



HAL
open science

Persistence-based Modes Inference

Hugo Henneuse

► **To cite this version:**

| Hugo Henneuse. Persistence-based Modes Inference. 2024. hal-04655461v2

HAL Id: hal-04655461

<https://hal.science/hal-04655461v2>

Preprint submitted on 27 Aug 2024

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.

Persistence-based Modes Inference

Hugo Henneuse ^{1,2}

hugo.henneuse@universite-paris-saclay.fr

August 27, 2024

Abstract

We address the problem of estimating multiple modes of a multivariate density, using persistent homology, a key tool from Topological Data Analysis. We propose a procedure, based on a preliminary estimation of the H_0 -persistence diagram, to estimate the number of modes, their locations, and the associated local maxima. For large classes of piecewise-continuous functions, we show that these estimators are consistent and achieve nearly minimax rates. These classes involve geometric control over the set of discontinuities of the density, and differ from commonly considered function classes in mode(s) inference. Interestingly, we do not suppose regularity or even continuity in any neighborhood of the modes.

Introduction

Modes are simple measures to describe central tendencies of a density and one of the most used tool to process data. Modes inference finds applications in a wide variety of statistical tasks, as highlighted in the recent survey by [Chacón \(2020\)](#). Among them, modal approaches in clustering has gathered significant attention ([Fukunaga and Hostetler, 1975](#); [Cheng, 1995](#); [Comaniciu and Meer, 2002](#); [Li et al., 2007](#); [Chazal et al., 2011](#); [Chacón, 2012](#); [Jiang and Kpotufe, 2017](#)).

The problem of mode(s) inference dates back to [Parzen \(1962\)](#) and has received considerable attention since. The question of consistency and convergence rates of estimators has occupied a central place. The bulk of work on this question is large but has mainly been concentrated on single mode estimation. For this problem, the popular approach is to consider an estimator of the form $\hat{x} = \arg \max_{x \in [0,1]^d} \hat{f}(x)$ with \hat{f} an estimator of the density. Usually \hat{f} is a kernel estimator of the density, following Parzen's original work. This work already provides convergence rates but for univariate densities and under strong regularity assumptions (global regularity). Subsequent efforts have been made to weaken those assumptions and extend this work to multivariate densities ([Chernoff, 1964](#); [Samanta, 1973](#); [Konakov, 1974](#); [Eddy, 1980](#); [Romano, 1988](#); [Tsybakov, 1990](#); [Donoho and Liu, 1991](#); [Vieu, 1996](#); [Mokkadem and Pelletier, 2003](#)). Notably [Donoho and Liu \(1991\)](#) consider a general multivariate setting and weaken the assumptions to a local assumption around the mode. They suppose that the density essentially behaves around the mode as a power function, and show that Parzen's method then achieves minimax rates. In these works a key question is how to choose the bandwidth for kernel estimation, which often involves knowing the regularity of the density. Interestingly, [Klemelä \(2005\)](#) proposes an adaptive estimator using

¹Laboratoire de Mathématiques d'Orsay, Université Paris-Saclay, Orsay, France

²DataShape, Inria Saclay, Palaiseau, France

Lepski’s method [Lepski \(1992\)](#) to overcome this issue.

An alternative, based on histogram estimation of the density, is proposed in [Arias-Castro et al. \(2022\)](#). Under the same framework as [Donoho and Liu \(1991\)](#), they show that their estimator also achieves minimax rates. This approach seems rather isolated, but has great adaptivity and computational properties.

Another line of work, is to consider $\hat{x} = \arg \max_{i=1, \dots, n} \hat{f}(X_i)$ where X_1, \dots, X_n are the observations sampled from the density, which reduces considerably computational costs. This was initially proposed in [Devroye \(1979\)](#) with \hat{f} a kernel estimator. It was proved in [Abraham et al. \(2003, 2004\)](#) that this estimator essentially shares the same asymptotic behavior as Parzen’s estimator, convergence rates are not affected by the maximization over finite samples. In the same direction, [Dasgupta and Kpotufe \(2014\)](#) propose a procedure based on KNN-estimation of the density and show, under similar assumptions as [Donoho and Liu \(1991\)](#), that their estimator achieves minimax rates once again.

For multiples modes inference, the literature is scarcer. Mean-shift and related procedures ([Fukunaga and Hostetler, 1975](#); [Cheng, 1995](#); [Comaniciu and Meer, 2002](#); [Carreira-Perpinan and Williams, 2003](#); [Carreira-Perpinan, 2007](#); [Li et al., 2007](#)) are widely used to infer multiple modes. Although these approaches exhibit great practical performance, we lack theoretical understanding about their convergence properties. In this direction, some insights are given in [Arias-Castro et al. \(2016\)](#) proving near-consistency of the mean shift procedures.

Among the previously cited works for single mode estimation, [Dasgupta and Kpotufe \(2014\)](#) also propose a method based on KNN to estimate multiple modes. Their procedure is proved for this more general problem to be minimax under relatively weak conditions. Furthermore, it can even be adapted to capture modal sets (i.e. handle cases where modes are not limited to single points but can be sets of points), as shown in [Jiang and Kpotufe \(2017\)](#). Once again, under relatively weak conditions, the procedure is shown to be minimax. To our knowledge, no other procedures have been demonstrated to possess such broad convergence properties as of now.

Here we present an alternative relying on tools from Topological Data Analysis (TDA). TDA is a field that focuses on providing descriptors of the data, using tools from (algebraic) topology. One of these descriptors, persistent homology, permits to encode, in so-called "persistence diagrams", the evolution of the topology (in the homology sense) of the super-level sets of a density. Those diagrams compactly represent birth and death times of topological features. Under proper conditions, looking at H_0 -persistence diagram (representing the evolution of connected components) permits to identify local maxima (birth times in super-level sets persistence diagram). This is illustrated in [Figure 1](#).

The link between estimation of persistence diagrams and modes detection has already been highlighted several times, see for example [Bauer et al. \(2014\)](#) and [Genovese et al. \(2015\)](#) that propose methods to infer the number of modes from persistence diagrams. In a slightly different context, we must also mention [Chazal et al. \(2011\)](#). The authors propose an algorithm, *ToMATo*, combining ideas from mean-shift with persistent homology to perform modal-clustering. The two main questions in those works, that we will address here, are how to estimate persistence diagrams and how to determine if a point in these estimated diagrams is significant or not, i.e. at which distance from the diagonal points are not due to noise with high probability. Reformulated in terms of modes, it is equivalent to decide if a mode is significant or not based on its magnitude. Outside the sphere of TDA, this question has also received significant attention, with a rich modal-testing literature ([Silverman, 1981](#); [Hartigan and Hartigan, 1985](#); [Minnotte,](#)

1997; Cheng and Hall, 1998; Hall and York, 2001; Fisher and Marron, 2001; Duong et al., 2008; Burman and Polonik, 2009; Ameijeiras-Alonso et al., 2019). Note that persistence diagram alone does not permit to localize modes, only to estimates their numbers and associated local maxima. Additional information needs to be extracted in order to infer their positions.

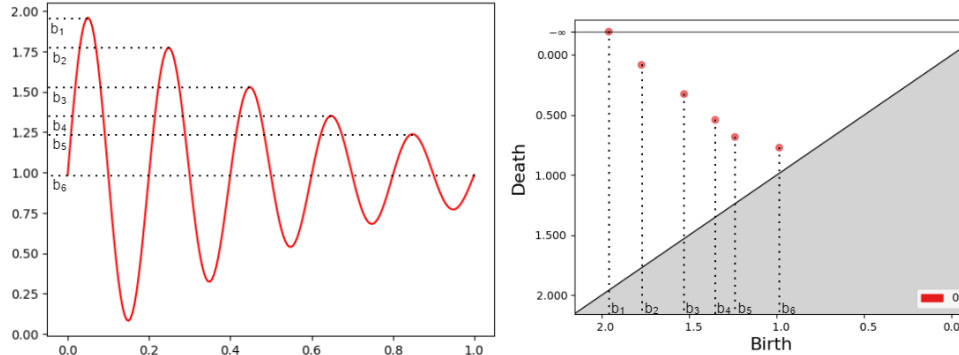


Figure 1: 1D illustration of the link between local maxima and H_0 persistence diagram. The birth times b_1, \dots, b_6 corresponds to the local maxima of the function.

The estimation of persistence diagrams of a density is an interesting question by itself. A common approach (see e.g. Bubenik et al., 2009) in persistence diagram estimation is to use sup-norm (or other popular functional norm) stability theorem (Chazal et al., 2009) to lift convergence results for density (or signal) estimation in sup norm. This approach is limited as it supposes to be able to consistently estimate the density (or the signal) in sup-norm, which will typically fail under the local assumptions designed for modes estimation cited earlier. Bobrowski et al. (2017) proposed to break free from this approach, allowing to consider estimation of persistence diagram for a wide class of functions (q -tame functions). Unfortunately, this work does not quantify convergence rates. Under the same motivation, for the Gaussian white noise model and the non-parametric regression, in Henneuse (2024) we studied classes of piecewise Hölder functions tolerating irregularities that are difficult to handle for signal estimation. But on which it is still possible to infer persistence diagrams with convergence rates following the minimax one classically known on Hölder spaces. The proposed procedure in this work is based on histogram estimation of the (sub or super) level sets. We believe that these level sets estimators contain significantly more information than the one contained in persistence diagrams. In particular, we illustrate it in this work, showing that they contain the missing information to move from persistence diagram estimation to modes inference.

We here precise and discuss the settings considered in this paper. Let $f : [0, 1]^d \rightarrow \mathbb{R}$ a probability density, we suppose that we observe $X = \{X_1, \dots, X_n\}$ points sampled from \mathbb{P}_f .

Regularity assumptions on the densities are inspired by the one consider in Henneuse (2024) (adapted for super level filtration). A notable difference is that we relax the positive reach assumption made there into a positive μ -reach assumption. This allows considerably wilder discontinuities sets, tolerating multiple points (i.e. self intersections of the discontinuities set) and corners. For a set $A \subset [0, 1]^d$, \bar{A} we denote its adherence and ∂A its boundary. We consider the following assumptions over f :

- **A1.** f is a piecewise (L, α) -Hölder-continuous probability density, i.e. there exists M_1, \dots, M_l

open sets of $[0, 1]^d$ such that,

$$\bigcup_{i=1}^l \overline{M_i} = [0, 1]^d$$

and $f|_{M_i}$ is in $H(L, \alpha)$, $\forall i \in \{1, \dots, l\}$, with,

$$H(L, \alpha) = \left\{ f : [0, 1]^d \rightarrow \mathbb{R} \text{ s.t. } |f(x) - f(y)| \leq L\|x - y\|_2^\alpha, \forall x, y \in [0, 1]^d \right\}.$$

Assumption **A1** control, on each M_i , how f is “spiky”. The smaller the value of α , the less mass there is around modes, making it more challenging to infer local maxima accurately.

- **A2.** f verifies, $\forall x_0 \in [0, 1]^d$,

$$\limsup_{x \in \bigcup_{i=1}^l M_i \rightarrow x_0} f(x) = f(x_0).$$

In this context, two signals, differing only on a null set, are statistically undistinguishable. Persistent homology is sensitive to point-wise irregularity, two signals differing only on a null set can have persistence diagrams that are arbitrarily far. Assumption **A2** prevents such scenario.

- **A3.** Let $\mu \in]0, 1]$ and $R_\mu > 0$, for all $I \subset \{1, \dots, l\}$

$$\text{reach}_\mu \left(\bigcup_{i \in I} \partial M_i \right) \geq R_\mu.$$

Here reach_μ denotes the μ -reach (defined by (1) in Section 1). μ -reach is a generalization of the reach (Federer, 1959), introduced by Chazal et al. (2006). This can be thought as geometric characterization of the regularity of the discontinuities set. The combination of Assumptions **A3** and **A1** ensures that at all point $x \in [0, 1]^d$ there is a small half cone of apex x and angles $2 \cos^{-1}(\mu)$ on which f is (L, α) -Hölder. Following the remark made on **A1**, it gives control over the mass of the density around modes. We invite the reader to see Section 1, where we recall the definition of the μ -reach and make additional remarks about Assumption **A3**.

The class of functions verifying **A1**, **A2** and **A3** is denoted $S_d(L, \alpha, \mu, R_\mu)$. As proved in Appendix D, densities in $S_d(L, \alpha, \mu, R_\mu)$ have well-defined persistence diagram. We prove in this work that these assumptions are sufficient to infer the H_0 -persistence diagram coming from the super level sets of f consistently. For modes estimation, we require an additional assumption :

- **A4.** Let $h_0 > 0$, $0 < C < L$ and $0 < \alpha \leq 1$, for any x local maxima of f and $y \in B_2(x, h_0)$,

$$C\|x - y\|_2^\alpha \leq \|f(x) - f(y)\|_2.$$

This assumption is common in the context of modes inference (Donoho and Liu, 1991; Dasgupta and Kpotufe, 2014; Jiang and Kpotufe, 2017; Arias-Castro et al., 2022). It has several implications. First, it ensures that local maxima are strict (modal sets are here singletons). Secondly, it ensures that the modes are well separated (at distance at least h_0) and that f is not too flat around modes. The larger the value of α , the more f flattens

around modes, making it more challenging to locate modes accurately. Moreover, denotes $l = Ch_0^\alpha$ and $\text{dgm}(f)$ the H_0 -persistence diagram of f , we have,

$$\inf_{(b,d) \in \text{dgm}(f)} b - d \geq l.$$

Thus **A4** ensures that $\text{dgm}(f)$ contains no point at distant less than l from the diagonal, i.e. the magnitude of the modes is lower bounded by l . It consequently avoids having arbitrarily small oscillations and undetectable modes.

The class of functions verifying **A1**, **A2**, **A3** and **A4** is denoted $S_d(L, \alpha, \mu, R_\mu, C, h_0)$.

Contribution

First, we propose a procedure based on histogram to estimate H_0 -persistence diagram. We show that the proposed estimator achieves minimax rates over $S_d(L, \alpha, \mu, R_\mu)$. The obtained convergence rates coincide with the known minimax ones over Hölder spaces. Secondly, taking advantage of this new result on persistence diagram estimation, we derive estimation procedures to infer the number of modes, the localization of the modes and the values of f at the modes. We show that these estimators are (up to a logarithmic factor) minimax over $S_d(L, \alpha, \mu, R_\mu, C, h_0)$. These results provide a formal framework and a rigorous study of convergence properties for inference of multiple modes based on persistence. It departs from existing statistical studies, considering new sets of assumptions. Typically, in comparison to the conditions imposed in [Dasgupta and Kpotufe \(2014\)](#), our conditions are stronger globally (the densities being piecewise Hölder) but weaker locally (tolerating non differentiability or even discontinuities around modes). To illustrate this point, we provide some numerical illustrations in dimension 1 and 2.

The paper is organized as follows. Section 1 provides background on the geometric and topological concepts employed in this work. Section 2 is dedicated to H_0 -persistence diagram estimation. Section 3 is dedicated to multiple modes estimation. Finally, Section 4 provides numerical illustrations. Proofs of technical lemmas and auxiliary results can be found in Appendix.

1 Background

This section provides the necessary background to follow this paper.

1.1 Distance function, generalized gradient and μ -reach

We here present some concepts from geometric measure theory, used in this paper to control the geometry of the discontinuities set. Let $K \subset [0, 1]^d$ a compact set, the **distance function** d_K is given by,

$$d_K : x \mapsto \min_{y \in K} \|x - y\|_2.$$

Generally the distance function is not differentiable everywhere, but we can define a generalized gradient function that matches the gradient at points where the distance function is differentiable. Consider the set of closest points to x in K ,

$$\Gamma_K(x) = \{y \in K \mid \|x - y\|_2 = d_K(x)\}$$

and for $x \in [0, 1]^d \setminus K$, let $\Theta_K(x)$ the center of the unique smallest ball enclosing $\Gamma_K(x)$, the **generalized gradient function** $\nabla_K(x)$ is then defined as,

$$\nabla_K(x) = \frac{x - \Theta_K(x)}{d_K(x)}.$$

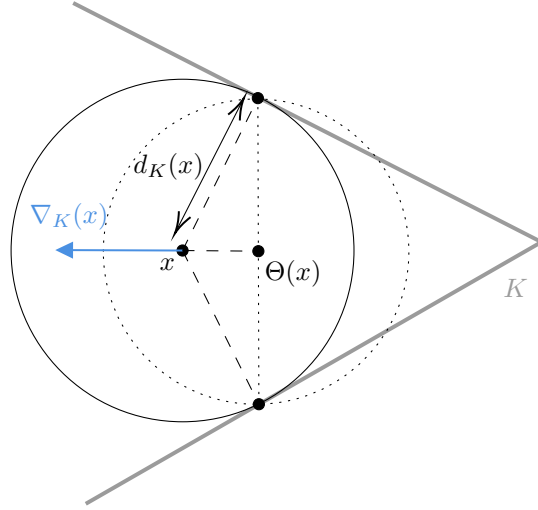


Figure 2: 2D example with 2 closest points

We can now introduce the notion of μ -reach (Chazal et al., 2006), involved in assumptions **A3**. Let $K \subset [0, 1]^d$ a compact set, its μ -reach $\text{reach}_\mu(K)$ is defined by,

$$\text{reach}_\mu(K) = \inf \left\{ r \mid \inf_{d_K^{-1}(r)} \|\nabla_K\|_2 < \mu \right\}. \quad (1)$$

For $\mu = 1$, this simply corresponds to the reach (Federer, 1959). The μ -reach is positive for a large class of sets. For example, all piecewise linear compact sets have positive μ -reach (for some μ). In our context, considering the μ -reach instead of the reach in assumption **A3** allows considering significantly wider classes of irregular densities, tolerating corners and multiple points.

Another way to understand the μ -reach is to see it as the distance from the μ -medial axis. The μ -**medial axis** of a set $K \subset [0, 1]^d$ is defined by,

$$\text{Med}_\mu(K) = \left\{ x \in [0, 1]^d \mid \|\nabla_K(x)\|_2 < \mu \right\}.$$

The μ -reach of K is then equal to $d_2(K, \overline{\text{Med}_\mu(K)})$.

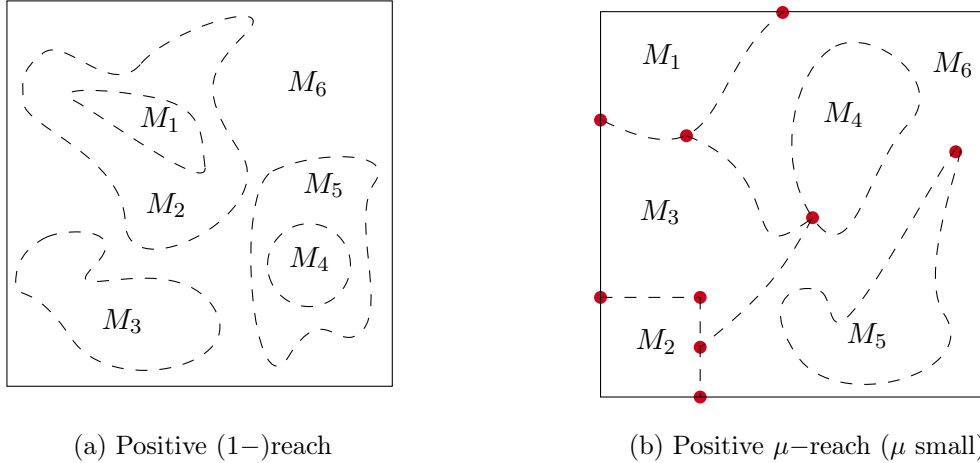


Figure 3: (a) displays a partition M_1, \dots, M_6 such that $\text{reach}_1(\partial M_1 \cup \dots \cup \partial M_6) > 0$. (b) displays a partition M_1, \dots, M_6 such that $\text{reach}_1(\partial M_1 \cup \dots \cup \partial M_6) = 0$ (in red are highlighted problematic points) but for sufficiently small $\mu > 0$, $\text{reach}_\mu(\partial M_1 \cup \dots \cup \partial M_6) > 0$.

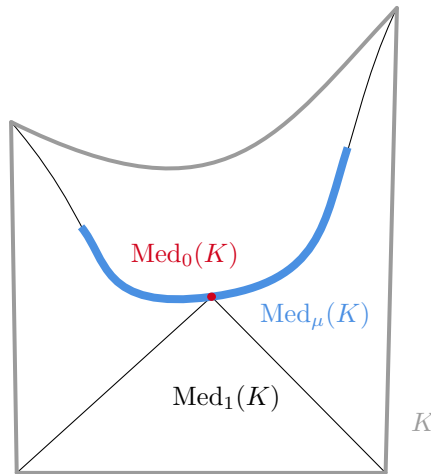


Figure 4: Illustration of μ -medial axis for a set K . In black is represented the 1-medial axis, in red the 0-medial axis and in blue the μ -medial axis for a small $0 < \mu < 1/2$.

1.2 Filtration, persistence module and H_0 -persistence diagram

We here present briefly some notions related to persistent homology. Persistent homology encodes the evolution of topological features (in the homology sense) along a family of nested spaces, called filtration. Moving along indices, topological features (connected components, cycles, cavities, ...) can appear or die (existing connected components merge, cycles or cavities are filled, ...). In this paper, we focus on H_0 -persistent homology, that describes the evolution of connectivity. For a broader overview and visual illustrations of persistent homology, we recommend [Chazal and Michel \(2021\)](#). For detailed and rigorous constructions, see [Chazal et al. \(2016\)](#). Additionally, since the construction discussed here involves (singular) homology, the reader can refer to [Hatcher \(2000\)](#).

The typical filtration that we will consider in this paper is, for a function $f : \mathbb{R}^d \rightarrow \mathbb{R}$, the

family of superlevel sets $(\mathcal{F}_\lambda)_{\lambda \in \mathbb{R}} = (f^{-1}([\lambda, +\infty[))_{\lambda \in \mathbb{R}}$. The associated family of homology groups of degree 0, $\mathbb{V}_{f,0} = (H_0(\mathcal{F}_\lambda))_{\lambda \in \mathbb{R}}$, equipped with v_λ^λ the linear application induced by the inclusion $\mathcal{F}_\lambda \subset \mathcal{F}_{\lambda'}$, for all $\lambda > \lambda'$ forms a **persistence module**. To be more precise, in this paper, $H_0(\cdot)$ is the singular homology functor in degree 0 with coefficients in a field (typically $\mathbb{Z}/2\mathbb{Z}$). Hence, $H_0(\mathcal{F}_\lambda)$ is a vector space.

if $\forall \lambda > \lambda' \in \mathbb{R}$, $\text{rank}(v_\lambda^{\lambda'})$ is finite, the module is said to be **q-tame**. It is then possible to show that the algebraic structure of the persistence module encodes exactly the evolution of the topological features along the filtration. For details, we encourage the reader to look at [Chazal et al. \(2016\)](#). Furthermore, the algebraic structure of such a persistence module can be summarized by a collection $\{(b_i, d_i), i \in I\} \subset \overline{\mathbb{R}}^2$, which defines the **persistence diagram**. Following previous remarks, for H_0 -persistent homology, b_i corresponds to the birth time of a connected component, d_i to its death time and $d_i - b_i$ to its lifetime.

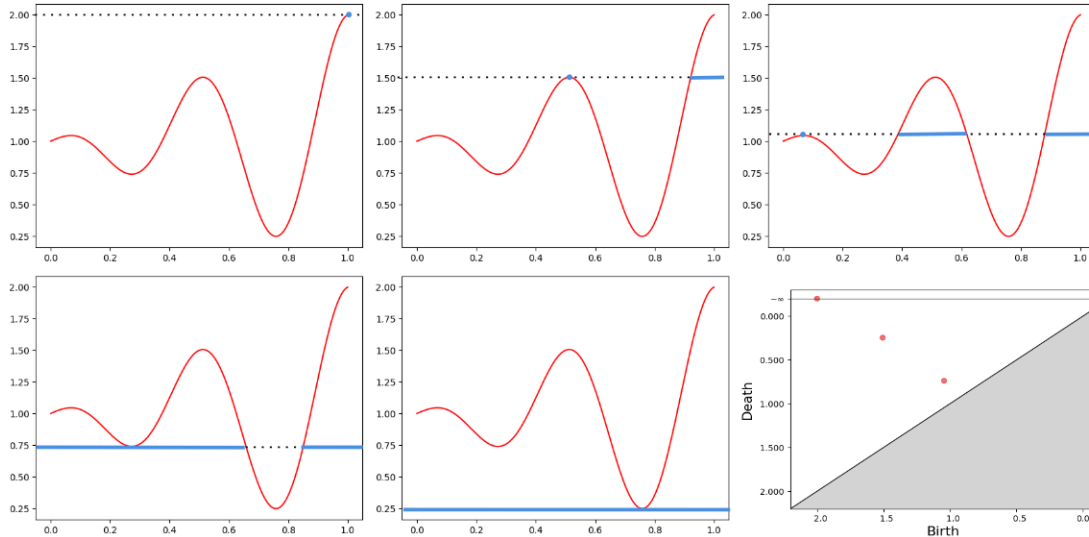


Figure 5: Super level sets filtration of $f(x) = x \cos(4\pi x)$ over $[0, 1]$ and the associated H_0 -persistence diagram.

To compare persistence diagrams, a popular distance, especially in statistical contexts, is the **bottleneck distance**, defined for two persistence diagrams D_1 and D_2 by,

$$d_b(D_1, D_2) = \inf_{\gamma \in \Gamma} \sup_{p \in D_1} \|p - \gamma(p)\|_\infty$$

with Γ the set of all bijections between D_1 and D_2 (both enriched with the diagonal).

We now present the algebraic stability theorem for the bottleneck distance. This theorem was the key for proving upper bounds in [Henneuse \(2024\)](#), it will be also employed in this paper. This theorem relies on interleaving between modules. Let $f : [0, 1]^d \rightarrow \mathbb{R}$ and $g : [0, 1]^d \rightarrow \mathbb{R}$, the associated H_0 -persistent modules for the superlevel sets filtrations, denoted respectively \mathbb{V} and \mathbb{W} , are said to be **ε -interleaved** if there exist two families of applications $\phi = (\phi_\lambda)_{\lambda \in \mathbb{R}}$ and $\psi = (\psi_\lambda)_{\lambda \in \mathbb{R}}$ where $\phi_\lambda : \mathbb{V}_\lambda \rightarrow \mathbb{W}_{\lambda-\varepsilon}$, $\psi_\lambda : \mathbb{V}_\lambda \rightarrow \mathbb{W}_{\lambda-\varepsilon}$, and for all $\lambda > \lambda'$ the following diagrams

commute,

$$\begin{array}{ccc}
 \mathbb{V}_\lambda & \xrightarrow{v_\lambda^{\lambda'}} & \mathbb{V}_{\lambda'} \\
 \phi_\lambda \downarrow & & \downarrow \phi_{\lambda'} \\
 \mathbb{W}_{\lambda-\varepsilon} & \xrightarrow{w_{\lambda-\varepsilon}^{\lambda'-\varepsilon}} & \mathbb{W}_{\lambda'-\varepsilon} \\
 & \xrightarrow{v_\lambda^{\lambda'-2\varepsilon}} & \\
 \mathbb{V}_\lambda & \xrightarrow{v_\lambda^{\lambda'-2\varepsilon}} & \mathbb{V}_{\lambda-2\varepsilon} \\
 \phi_\lambda \searrow & & \nearrow \psi_{\lambda-\varepsilon} \\
 & \mathbb{W}_{\lambda-\varepsilon} &
 \end{array}
 \qquad
 \begin{array}{ccc}
 \mathbb{W}_\lambda & \xrightarrow{w_\lambda^{\lambda'}} & \mathbb{W}_{\lambda'} \\
 \psi_\lambda \downarrow & & \downarrow \psi_{\lambda'} \\
 \mathbb{V}_{\lambda-\varepsilon} & \xrightarrow{v_{\lambda-\varepsilon}^{\lambda'-\varepsilon}} & \mathbb{V}_{\lambda'-\varepsilon} \\
 & \xrightarrow{w_\lambda^{\lambda'-2\varepsilon}} & \\
 \mathbb{W}_\lambda & \xrightarrow{w_\lambda^{\lambda'-2\varepsilon}} & \mathbb{W}_{\lambda-2\varepsilon} \\
 \psi_\lambda \searrow & & \nearrow \phi_{\lambda-\varepsilon} \\
 & \mathbb{V}_{\lambda-\varepsilon} &
 \end{array}
 \tag{2}$$

Theorem (Chazal et al. 2009, "algebraic stability"). *Let \mathbb{V} and \mathbb{W} two q -tame persistence modules. If \mathbb{V} and \mathbb{W} are ε -interleaved then,*

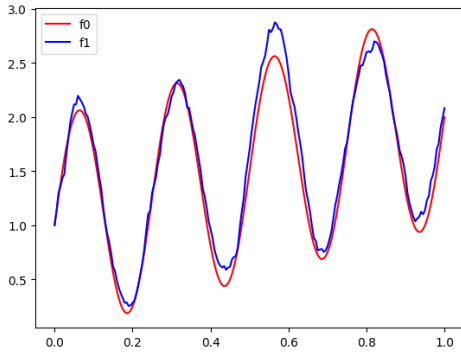
$$d_b(\text{dgm}(\mathbb{V}), \text{dgm}(\mathbb{W})) \leq \varepsilon.$$

A direct corollary of this result, proved earlier in special cases (Barannikov, 1994; Cohen-Steiner et al., 2005), is the sup-norm stability of persistence diagrams. We insist on the fact that this is a strictly weaker result than algebraic stability.

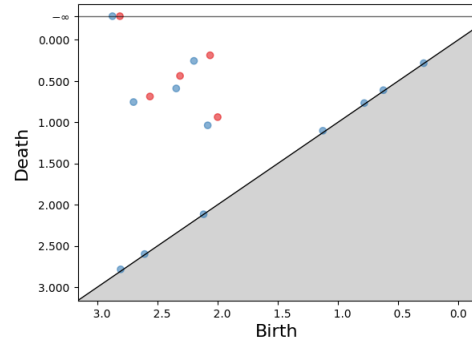
Theorem ("sup norm stability"). *Let f and g two real-valued function and $s \in \mathbb{N}$. If their persistent modules for the s -th order homology, denoted \mathbb{V} and \mathbb{W} , are q -tame, then,*

$$d_b(\text{dgm}(\mathbb{V}), \text{dgm}(\mathbb{W})) \leq \|f - g\|_\infty.$$

As highlighted in the introduction, the sup-norm stability is often used to upper bound the errors in bottleneck distance of "plug-in" estimators of persistence diagrams. It enables the direct translation of convergence rates in sup-norm to convergence rates in bottleneck distance, which for regular classes of signals provides minimax upper bounds. However, when the convergence in sup-norm of the preliminary estimator is not ensured, this approach falls short.



(a) graphs of f_0 and f_1



(b) persistence diagrams of f_0 and f_1

Figure 6: 1D Illustration of stability theorems.

2 H_0 -persistence diagram estimation

This section describes our procedure for persistence diagram estimation and provides associated convergence results.

2.1 Procedure description

We propose a procedure to estimate H_0 -persistence diagram over $S_d(L, \alpha, \mu, R_\mu, C, h_0)$. Contrary to the setting of [Henneuse \(2024\)](#), where the positive reach assumptions allows considering plug-in estimation directly through histogram, under the weaker assumptions made here, this approach falls short as illustrated in figure 7. To properly recover the connectivity of the sublevel sets filtration, we need an additional thickening.

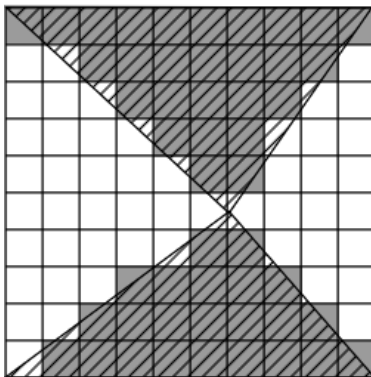


Figure 7: λ -superlevel cubical approximation for f the function defined as 0 outside the hatched area and K inside (for arbitrarily large K). The histogram approximation fails to identify the connectivity of the two triangles for at least all $3K/4 < \lambda < K$.

Let denote,

$$A^b = \left\{ x \in \mathbb{R}^d \text{ s.t. } d_\infty(x, A) \leq b \right\}$$

with

$$d_\infty(x, A) = \inf_{y \in A} \|x - y\|_\infty.$$

Let $h > 0$ such that $1/h$ is an integer, consider $G(h)$ the regular orthogonal grid over $[0, 1]^d$ of step h and C_h the collection of all the hypercubes of side h composing $G(h)$. We define, $\forall \lambda \in \mathbb{R}$, the λ -sublevel estimator,

$$\widehat{\mathcal{F}}_\lambda = \left(\bigcup_{H \in C_{h,\lambda}} H \right)^{\lceil \sqrt{d}/\mu \rceil h} \quad \text{with } C_{h,\lambda} = \left\{ H \in C_h \text{ such that } \frac{|X \cap H|}{nh^d} \geq \lambda \right\}$$

with $\lceil \cdot \rceil$ the ceiling function. Let $\lambda > \lambda'$, we denote,

$$\widehat{v}_\lambda^{\lambda'} : H_0(\widehat{\mathcal{F}}_\lambda) \rightarrow H_0(\widehat{\mathcal{F}}_{\lambda'})$$

the map induced by the inclusion $\widehat{\mathcal{F}}_\lambda \subset \widehat{\mathcal{F}}_{\lambda'}$. We now introduce $\widehat{\mathbb{V}}_{f,0}$ the persistence module associated to $(H_0(\widehat{\mathcal{F}}_\lambda))_{\lambda \in \mathbb{R}}$ equipped with the collection of maps $(\widehat{v}_{\lambda,h}^{\lambda'})_{\lambda > \lambda'}$ and $\widehat{\text{dgm}}(f)$ the associated H_0 -persistence diagram. These diagrams are well-defined, as we prove in Appendix D that $\widehat{\mathbb{V}}_{f,0}$ is q -tame.

Calibration. A natural question is how to choose the parameter h . Following the proof of Lemma 2, we choose h verifying,

$$h^\alpha > \sqrt{\frac{\log(1/h^d)}{nh^d}} \quad (3)$$

In particular, we can choose,

$$h \simeq \left(\frac{\log(n)}{n} \right)^{\frac{1}{d+2\alpha}}.$$

Computation. By construction, for all $\lambda \in \mathbb{R}$, $\widehat{\mathcal{F}}_\lambda$ is simply a union of cubes from the regular grid $G(h)$, thus can be thought as a (geometric realization of) a cubical complex or even a simplicial complex. Hence, it allows practical computation of its persistence diagram.

2.2 Convergence Rates

In this section, we state our main results concerning persistence diagram estimation, Theorem 1 and Theorem 2, establishing that the procedure proposed in the previous section achieves minimax rates over $S_d(L, \alpha, \mu, R_\mu)$.

Theorem 1. *Let $h \simeq \left(\frac{\log(n)}{n} \right)^{\frac{1}{d+2\alpha}}$. There exists \tilde{C}_0 and \tilde{C}_1 such that, for all $A \geq 0$,*

$$\mathbb{P} \left(\sup_{f \in S(L, R, \alpha)} d_\infty \left(\widehat{\text{dgm}}(f), \text{dgm}(f) \right) \geq A \left(\frac{\log(n)}{n} \right)^{\frac{\alpha}{d+2\alpha}} \right) \leq \tilde{C}_0 \exp \left(-\tilde{C}_1 A^2 \right).$$

From this, we can derive from this result a bound in expectation.

Corollary 1. *Let $h \simeq \left(\frac{\log(n)}{n} \right)^{\frac{1}{d+2\alpha}}$ and $p \geq 1$,*

$$\sup_{f \in S(L, R, \alpha)} \mathbb{E} \left(d_\infty \left(\widehat{\text{dgm}}(f), \text{dgm}(f) \right)^p \right) \lesssim \left(\frac{\log(n)}{n} \right)^{\frac{p\alpha}{d+2\alpha}}$$

Proof. The sub-Gaussian concentration provided by Proposition 1, gives that, for all $A \geq 0$,

$$\mathbb{P} \left(\frac{d_b \left(\widehat{\text{dgm}}(f), \text{dgm}(f) \right)}{h^\alpha} \geq A \right) \leq \tilde{C}_0 \exp \left(-\tilde{C}_1 A^2 \right).$$

Now, we have,

$$\begin{aligned} \mathbb{E} \left(\frac{d_b \left(\widehat{\text{dgm}}(f), \text{dgm}(f) \right)^p}{h^{p\alpha}} \right) &= \int_0^{+\infty} \mathbb{P} \left(\frac{d_b \left(\widehat{\text{dgm}}(f), \text{dgm}(f) \right)^p}{h^{p\alpha}} \geq A \right) dA \\ &\leq \int_0^{+\infty} \tilde{C}_0 \exp \left(-\tilde{C}_1 A^{2/p} \right) dA < +\infty. \end{aligned}$$

□

Theorem 2. *Let $p \geq 1$*

$$\inf_{\widehat{\text{dgm}}(f)} \sup_{f \in S_d(L, \alpha, \mu, R_\mu)} \mathbb{E} \left(d_\infty \left(\widehat{\text{dgm}}(f), \text{dgm}(f) \right)^p \right) \gtrsim \left(\frac{\log(n)}{n} \right)^{\frac{p\alpha}{d+2\alpha}}$$

Where the infimum is taken over all the estimators of $\text{dgm}(f)$.

2.3 Proof of Theorem 1

This section is devoted to the proofs of Theorem 1 and Corollary 1 stated in Section 2.2. The idea is to construct an interleaving between $\mathbb{V}_{f,0}$ and $\widehat{\mathbb{V}}_{f,0}$, to then apply the algebraic stability theorem Chazal et al. (2009). This requires constructing two morphisms $\overline{\psi} : \mathbb{V}_{f,0} \rightarrow \widehat{\mathbb{V}}_{f,0}$ and $\overline{\phi} : \widehat{\mathbb{V}}_{f,0} \rightarrow \mathbb{V}_{f,0}$ satisfying (2). The following sections give the necessary ingredients.

2.3.1 Ingredient 1 : inclusions between level sets

If there existed an $\varepsilon > 0$ such that for all $\lambda \in \mathbb{R}$, $\mathcal{F}_\lambda \subset \widehat{\mathcal{F}}_{\lambda-\varepsilon} \subset \mathcal{F}_{\lambda-2\varepsilon}$, an ε -interleaving would be directly given taking the morphisms induced by inclusions. Here we do not have such nice inclusions, but we show, in Proposition 1, a slightly weaker double inclusion.

First, remark that the $f \in S_d(L, \alpha, \mu, R_\mu)$ are uniformly bounded, as stated in the following lemma, which proof can be found in Appendix C.

Lemma 1. *Let $f \in S_d(L, \alpha, \mu, R_\mu)$, there exists a constant M_{d,L,α,R_μ} depending on d, L and R_μ , such that $\|f\|_\infty \leq M_{d,L,\alpha,R_\mu}$.*

And denotes, N_h the variable defined by,

$$N_h = \frac{\max_{H \in C_h} \left| \frac{|X \cap H|}{nh^d} - \mathbb{E} \left[\frac{|X \cap H|}{nh^d} \right] \right|}{\sqrt{3M_{d,L,\alpha,R_\mu} \frac{\log(1/h^d)}{nh^d}}}$$

Proposition 1. *Let $f \in S_d(L, \alpha, \mu, R_\mu)$ and $h \simeq (\log(n)/n)^{\frac{1}{d+2\alpha}}$. For n sufficiently large such that $h < \mu R_\mu / \sqrt{d}$, for all $\lambda \in \mathbb{R}$, we have,*

$$\mathcal{F}_{\lambda + (2\sqrt{3M_{d,L,\alpha,R_\mu}}N_h + L(\lceil \sqrt{d}/\mu \rceil + \sqrt{d})^\alpha)h^\alpha} \subset \widehat{\mathcal{F}}_\lambda \subset \mathcal{F}_{\lambda - 2\sqrt{3M_{d,L,\alpha,R_\mu}}N_h h^\alpha}.$$

This proposition then follows from Lemma 3.1 from Chazal et al. (2007) and the following lemma, which proof can be found in Appendix B.

Lemma 2. *Let $f \in S_d(L, \alpha, \mu, R_\mu)$ and $h > 0$ verifying (3). Let*

$$H \subset \mathcal{F}_{\lambda - 2\sqrt{3M_{d,L,\alpha,R_\mu}}N_h h^\alpha} \cap C_h$$

and

$$H' \subset \mathcal{F}_{\lambda + 2\sqrt{3M_{d,L,\alpha,R_\mu}}N_h h^\alpha} \cap C_h.$$

We then have that,

$$\frac{|X \cap H|}{nh^d} < \lambda \text{ and } \frac{|X \cap H'|}{nh^d} > \lambda.$$

For a set $K \subset [0, 1]^d$ and $r \geq 0$ we denote $B_2(K, r) = \{x \in [0, 1]^d \text{ s.t. } d_K(x) < r\}$ and $\overline{B}_2(K, r) = \{x \in [0, 1]^d \text{ s.t. } d_K(x) \leq r\}$.

Lemma 3. (Chazal et al., 2007, Lemma 3.1.) *Let $K \subset [0, 1]^d$ a compact set and let $\mu > 0, r > 0$ be such that $r < \text{reach}_\mu(K)$. For any $x \in \overline{B}_2(K, r) \setminus K$, we have,*

$$d_2(x, \partial \overline{B}_2(K, r)) \leq \frac{r - d_K(x)}{\mu} \leq \frac{r}{\mu}.$$

Proof of Proposition 1. We begin by proving the lower inclusion, let,

$$x \in \mathcal{F}_{\lambda + (2\sqrt{3M_{d,L,\alpha,R\mu}}N_h + L(\lceil\sqrt{d}/\mu\rceil + \sqrt{d}))^\alpha} h^\alpha.$$

Without loss of generality, let suppose $x \in M_i$, or $x \in \partial M_i$ and $\liminf_{z \in M_i \rightarrow x} f(z) \leq f(x)$. If,

$$B_2(x, \sqrt{d}h) \subset \left(\bigcup_{i=1}^l \partial M_i \right)^c$$

then, $H_{x,h}$, the hypercube of $C_{h,\lambda}$ containing x is included in \overline{M}_i . Assumption **A1** and **A2** then give $H_{x,h} \subset \mathcal{F}_{\lambda + 2\sqrt{3M_{d,L,\alpha,R\mu}}N_h} h^\alpha$. Hence, it follows from Lemma 2 that $H_{x,h} \in C_{h,\lambda}$ and consequently $x \in \widehat{\mathcal{F}}_\lambda$. Else, as $\sqrt{d}h/\mu < R_\mu$, under Assumption **A3**, by Lemma 3, there exists,

$$y \in \left(B_2 \left(\bigcup_{i=1}^l \partial M_i, \sqrt{d}h \right) \right)^c \cap M_i \text{ such that } \|x - y\|_2 \leq \sqrt{d}h/\mu.$$

Let $H_{y,h}$ the closed hypercube of C_h containing y . Hence, $H_{y,h} \subset \overline{M}_i$. and, for all $z \in H_{y,h}$ $\|z - x\|_2 \leq \sqrt{d}h(1 + 1/\mu)$. Then, Assumption **A1** and **A2** ensure that,

$$H_{y,h} \subset \mathcal{F}_{\lambda + 2\sqrt{3M_{d,L,\alpha,R\mu}}N_h} h^\alpha.$$

Then, Lemma 2 gives $H_{y,h} \in C_{h,\lambda}$ and thus, as $x \in H_{y,h}^{\sqrt{d}h/\mu}$, $x \in \widehat{\mathcal{F}}_\lambda$, which proves the lower inclusion.

For the upper inclusion, let $x \in \left(\mathcal{F}_{\lambda - 2\sqrt{3M_{d,L,\alpha,R\mu}}N_h} h^\alpha \right)^c$, and $H_{x,h}$ the hypercube of C_h containing x . We then have, $H_{x,h} \subset \mathcal{F}_{\lambda - 2\sqrt{3M_{d,L,\alpha,R\mu}}N_h}^c h^\alpha$. Hence, Lemma 2 gives that,

$$H_{x,h} \subset \left(\bigcup_{H \in C_{h,\lambda}} H \right)^c$$

and thus,

$$\bigcup_{H \in C_{h,\lambda}} H \subset \mathcal{F}_{\lambda - 2\sqrt{3M_{d,L,\alpha,R\mu}}N_h}^{\sqrt{d}h} h^\alpha.$$

Consequently,

$$\left(\bigcup_{H \in C_{h,\lambda}} H \right)^{\lceil\sqrt{d}/\mu\rceil h} = \widehat{\mathcal{F}}_\lambda \subset \mathcal{F}_{\lambda - 2\sqrt{3M_{d,L,\alpha,R\mu}}N_h}^{(\sqrt{d} + \lceil\sqrt{d}/\mu\rceil)h} h^\alpha$$

and the proof is complete. \square

2.3.2 Ingredient 2 : geometry of the thickened level sets

The lower inclusion of Proposition 1 gives directly a morphism $\overline{\psi} : \mathbb{V}_{f,0} \rightarrow \widehat{\mathbb{V}}_{f,0}$. But the upper inclusion is not sufficient to provide similarly $\overline{\phi} : \widehat{\mathbb{V}}_{f,0} \rightarrow \mathbb{V}_{f,0}$. To overcome this issue, in Proposition 2, we exploit assumption **A3** to construct morphisms from $(H_0(\mathcal{F}_\lambda^r))_{\lambda \in \mathbb{R}}$ into $V_{f,0}$, for all $r < R_\mu/\sqrt{d}$. For $r = (\sqrt{d} + \lceil\sqrt{d}/\mu\rceil)h$, composing this morphism with the one induced by the upper inclusion of Proposition 1 will give us our morphism $\overline{\phi}$.

Proposition 2. Let $f \in S_d(L, \alpha, \mu, R_\mu)$. $\forall s \in \mathbb{N}$, $\forall h < R_\mu/\sqrt{d}$, there exists a morphism ϕ such that, $\forall \lambda \in \mathbb{R}$,

$$\begin{array}{ccc} H_0(\mathcal{F}_\lambda) & \xrightarrow{\quad\quad\quad} & H_0\left(\mathcal{F}_{\lambda-L(\sqrt{d}(2+2/\mu^2))^\alpha h^\alpha}\right) \\ & \searrow & \nearrow \phi_\lambda \\ & H_0(\mathcal{F}_\lambda^h) & \end{array} \quad (4)$$

is a commutative diagram (unspecified map come from set inclusions).

To prove this proposition, we use the following lemma. This result involves the notion of deformation retract that we recall here.

Definition 1. A subspace A of X is called a **deformation retract** of X if there is a continuous $F : X \times [0, 1] \rightarrow X$ such that for all $x \in X$ and $a \in A$,

- $F(x, 0) = x$
- $F(x, 1) \in A$
- $F(a, 1) = a$.

The function $x \mapsto F(x, 1)$ is called a (deformation) **retraction** from X to A .

Lemma 4. (*Kim et al., 2020, Theorem 12*) Let $K \subset [0, 1]^d$, for all $0 \leq r < \text{reach}_\mu(K)$, $B_2(K, r)$ retracts by deformation onto K and the associated retraction $R : B_2(K, r) \rightarrow K$ verifies for all $x \in B_2(K, r)$, $R(x) \in B_2(x, 2r/\mu^2)$.

The first part of the claim is Theorem 12 from [Kim et al. \(2020\)](#) and the second part follows easily from their construction. Elements of proof can be found in [Appendix A](#).

Proof of Proposition 2. Any connected component B of \mathcal{F}_λ^h contains at least a connected component A of \mathcal{F}_λ . Suppose that B contains A and A' two disjoint connected components of \mathcal{F}_λ , then there exist $x \in A$ and $y \in A'$ such that $\|x - y\|_\infty \leq 2h$. Suppose that $x \in M_i$ and $y \in M_j$, $i, j \in \{1, \dots, l\}$.

If the segment $[x, y]$ is included in $\overline{M}_i \cup \overline{M}_j$, then by Assumptions **A1** and **A2**, directly we have $[x, y] \subset \mathcal{F}_{\lambda-L(2\sqrt{d}h)^\alpha}$ and thus A and A' are connected in $\mathcal{F}_{\lambda-L(2\sqrt{d}h)^\alpha}$.

Otherwise, there exists $\bar{x} \in \partial M_i$ and $\bar{y} \in \partial M_j$ such that $[\bar{x}, \bar{y}] \subset (M_i \cup M_j)^c$ and $[x, y] \setminus [\bar{x}, \bar{y}] \subset M_i \cup M_j$. It follows from Assumptions **A1** that $[x, y] \setminus [\bar{x}, \bar{y}] \subset \mathcal{F}_{\lambda-L(2\sqrt{d}h)^\alpha}$.

Now, let,

$$R : B_2(\partial M_i \cup \partial M_j, \sqrt{d}h) \rightarrow \partial M_i \cup \partial M_j$$

the deformation retraction of $B_2(\partial M_i \cup \partial M_j, \sqrt{d}h)$ onto $\partial M_i \cup \partial M_j$ given by Lemma 4 under Assumption **A3**. As $R(\bar{x}) = \bar{x}$, $R(\bar{y}) = \bar{y}$ and R continuous, $R([\bar{x}, \bar{y}])$ is a continuous path between \bar{x} and \bar{y} .

Furthermore, also by Lemma 4, we have,

$$R([\bar{x}, \bar{y}]) \subset (\partial M_i \cup \partial M_j) \cap \overline{B}_2([\bar{x}, \bar{y}], 2\sqrt{d}h/\mu^2).$$

Let $z \in R([\bar{x}, \bar{y}]) \cap \partial M_i$ then, $z \in \overline{B}_2(x, (2 + 2/\mu^2)\sqrt{d}h)$ and thus by Assumptions **A1** and **A2**, $z \in \mathcal{F}_{\lambda-L((2+2/\mu^2)\sqrt{d}h)^\alpha}$.

Similarly, if $z \in R([\bar{x}, \bar{y}]) \cap \partial M_j$, then, $z \in \bar{B}_2(y, (2 + 2/\mu^2)\sqrt{dh})$ and $z \in \mathcal{F}_{\lambda-L((2+2/\mu^2)\sqrt{dh})^\alpha}$. Hence,

$$R([\bar{x}, \bar{y}]) \subset \mathcal{F}_{\lambda-L((2+2/\mu^2)\sqrt{dh})^\alpha}.$$

Thus, $[x, \bar{x}] \cup R([\bar{x}, \bar{y}]) \cup [\bar{y}, \bar{x}]$ is a continuous path between x and y , included in $\mathcal{F}_{\lambda-L((2+2/\mu^2)\sqrt{dh})^\alpha}$. And consequently, in this case also A and A' are connected in $\mathcal{F}_{\lambda-L((2+2/\mu^2)\sqrt{dh})^\alpha}$.

We can then properly define,

$$\begin{cases} \phi_\lambda : H_0(\mathcal{F}_\lambda^h) & \rightarrow H_0(\mathcal{F}_{\lambda-L((2+2/\mu^2)\sqrt{dh})^\alpha h^\alpha}) \\ [x] & \mapsto [y] \end{cases}$$

with y belonging to A any connected component of \mathcal{F}_λ included in B , the connected component of \mathcal{F}_λ^h containing x . By the foregoing, $[y]$ is independent of the choice of such A (and such y in A) and Diagram 4 commutes. \square

2.3.3 Ingredient 3 : concentration of N_h

The two previous ingredients will permit to establish an interleaving between $\widehat{\mathbb{V}}_{f,0}$ and $\mathbb{V}_{f,0}$. This interleaving will depend on the variable N_h . Consequently, to derive convergence rates from this interleaving, we need concentration inequalities over N_h .

Proposition 3. *For all $h > 0$,*

$$\mathbb{P} \left(\bigcup_{H \in C_h} \left| \frac{|X \cap H|}{nh^d} - \mathbb{E} \left[\frac{|X \cap H|}{nh^d} \right] \right| \geq t \right) \leq \frac{2}{h^d} \exp \left(-\frac{t^2 nh^d}{3M_{d,L,\alpha,R_\mu}} \right).$$

Proof of Proposition 3. Let $h > 0$ and $H \in C_h$. We can write,

$$|X \cap H| = \sum_{i=1}^n Y_i^H$$

with, for all $i \in \{1, \dots, n\}$,

$$Y_i^H = \mathbb{1}_{X_i \in H}$$

As X_1, \dots, X_n are i.i.d., Y_1^H, \dots, Y_n^H are i.i.d. Bernoulli variable, and $|X \cap H|$ is then a binomial variable of parameters $(n, \mathbb{P}(X_1 \in H))$. It follows from the Chernoff bound for binomial distribution that,

$$\mathbb{P} \left(\left| \frac{|X \cap H|}{nh^d} - \mathbb{E} \left[\frac{|X \cap H|}{nh^d} \right] \right| \geq t \right) \leq 2 \exp \left(-\frac{t^2 nh^d}{3\mathbb{P}(X_1 \in H)} \right).$$

Lemma 1 then gives that $\mathbb{P}(X_1 \in H) \leq h^d M_{d,L,\alpha,R_\mu}$, hence,

$$\mathbb{P} \left(\left| \frac{|X \cap H|}{nh^d} - \mathbb{E} \left[\frac{|X \cap H|}{nh^d} \right] \right| \geq t \right) \leq 2 \exp \left(-\frac{t^2 nh^d}{3M_{d,L,\alpha,R_\mu}} \right).$$

By union bounds, it follows,

$$\mathbb{P} \left(\bigcup_{H \in C_h} \left\{ \left| \frac{|X \cap H|}{nh^d} - \mathbb{E} \left[\frac{|X \cap H|}{nh^d} \right] \right| \geq t \right\} \right) \leq 2|C_h| \exp \left(-\frac{t^2 nh^d}{3M_{d,L,\alpha,R_\mu}} \right) \leq \frac{2}{h^d} \exp \left(-\frac{t^2 nh^d}{3M_{d,L,\alpha,R_\mu}} \right).$$

\square

Proposition 3 implies that N_h is sub Gaussian. More precisely, there exist two constants C_0 and C_1 , depending only on d, L, α and R_μ , such that, for all h ,

$$\mathbb{P}(N_h > t) \leq C_0 \exp(-C_1 t^2).$$

2.3.4 Main proof

Equipped with Propositions 1, 2 and 3 we can now proceed to the proof of Theorem 1.

Proof of Theorem 1. It suffices to show the result for (arbitrarily) large n (up to rescaling \tilde{C}_0), thus suppose that n is sufficiently large for the application of Propositions 1 and 2 used in this proof. We denote $k_1 = \sqrt{d} + \lceil \sqrt{d}/\mu \rceil$ and $k_2 = \sqrt{d}(2 + 2/\mu^2)(\sqrt{d} + \lceil \sqrt{d}/\mu \rceil)$.

First, we construct $\bar{\phi}$. Let $\lambda \in \mathbb{R}$, Lemma 2 imply that any connected component A of $\widehat{\mathcal{F}}_\lambda$ intersects a connected component B of $\mathcal{F}_{\lambda - 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h}h^\alpha}$. Now suppose that A intersects 2 such components B_1 and B_2 , then, by Proposition 1, B_1 and B_2 are connected in $\mathcal{F}_{\lambda - 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h}h^\alpha}^{k_1 h}$. Thus, as a consequence of Proposition 2, B_1 and B_2 are connected in $\mathcal{F}_{\lambda - (2\sqrt{3M_{d,L,\alpha,R_\mu}N_h} + Lk_2^\alpha)h^\alpha}$.

We can then define properly the applications,

$$\begin{cases} \bar{\phi}_\lambda : H_0(\widehat{\mathcal{F}}_\lambda) & \longrightarrow H_0\left(\mathcal{F}_{\lambda - (2\sqrt{3M_{d,L,\alpha,R_\mu}N_h} + Lk_2^\alpha)h^\alpha}\right) \\ [x] & \longmapsto [y] \end{cases}$$

with y belonging to B a connected component of $\mathcal{F}_{\lambda - 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h}h^\alpha}$ intersecting A , the connected component of $\widehat{\mathcal{F}}_\lambda$ containing x . The previous remark ensures that $[y]$ is independent of the choice of B (and y in B).

Now, for the construction of $\bar{\psi}$, by Proposition 1, we have,

$$\mathcal{F}_\lambda \subset \widehat{\mathcal{F}}_{\lambda - (2\sqrt{3M_{d,L,\alpha,R_\mu}N_h} + Lk_1^\alpha)h^\alpha}$$

and we simply take $\bar{\phi}$ the morphism induced by this inclusion, i.e.,

$$\begin{cases} \bar{\psi}_\lambda : H_0(\mathcal{F}_\lambda) & \longrightarrow H_0\left(\widehat{\mathcal{F}}_{\lambda - (2\sqrt{3M_{d,L,\alpha,R_\mu}N_h} + Lk_1^\alpha)h^\alpha}\right) \\ [x] & \longmapsto [x] \end{cases}$$

Denote $K_1 = 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h} + Lk_1^\alpha$ and $K_2 = 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h} + Lk_2^\alpha$. The foregoing ensures the commutativity of following diagrams (unspecified maps are the one induced by set inclusion),

$$\begin{array}{ccc} H_0(\mathcal{F}_\lambda) & \longrightarrow & H_0(\mathcal{F}_{\lambda'}) \\ \downarrow \bar{\psi}_\lambda & & \downarrow \bar{\psi}_{\lambda'} \\ H_0(\widehat{\mathcal{F}}_{\lambda - K_1 h^\alpha}) & \longrightarrow & H_0(\widehat{\mathcal{F}}_{\lambda' - K_1 h^\alpha}) \end{array}$$

$$\begin{array}{ccc}
H_0(\widehat{\mathcal{F}}_\lambda) & \longrightarrow & H_0(\widehat{\mathcal{F}}_{\lambda'}) \\
\downarrow \bar{\phi}_\lambda & & \downarrow \bar{\phi}_{\lambda'} \\
H_0(\mathcal{F}_{\lambda-K_2h^\alpha}) & \longrightarrow & H_0(\mathcal{F}_{\lambda'-K_2h^\alpha}) \\
\\
H_0(\mathcal{F}_\lambda) & \xrightarrow{\quad \bar{\psi}_\lambda \quad} & H_0(\mathcal{F}_{\lambda-(K_1+K_2)h^\alpha}) \\
& \searrow & \nearrow \\
& H_0(\widehat{\mathcal{F}}_{\lambda-K_1h^\alpha}) & \\
\\
H_0(\widehat{\mathcal{F}}_\lambda) & \xrightarrow{\quad \bar{\psi}_{\lambda-K_2h^\alpha} \quad} & H_0(\widehat{\mathcal{F}}_{\lambda-(K_1+K_2)h^\alpha}) \\
& \searrow & \nearrow \\
& H_0(\mathcal{F}_{\lambda-K_2h^\alpha}) &
\end{array}$$

Hence $\widehat{\mathbb{V}}_{f,0}$ and $\mathbb{V}_{f,0}$ are $(K_1 + K_2)h^\alpha$ -interleaved, and thus we get from the algebraic stability theorem (Chazal et al., 2009) that,

$$d_b\left(\mathrm{dgm}\left(\widehat{\mathbb{V}}_{f,0}\right), \mathrm{dgm}\left(\mathbb{V}_{f,0}\right)\right) \leq (K_1 + K_2)h^\alpha$$

and as it holds for all $f \in S_d(L, \alpha, \mu, R_\mu)$,

$$\sup_{f \in S_d(L, \alpha, \mu, R_\mu)} d_b\left(\widehat{\mathrm{dgm}}(f), \mathrm{dgm}(f)\right) \leq (K_1 + K_2)h^\alpha.$$

We then conclude, using the concentration of N_h given in Proposition 3.

$$\begin{aligned}
& \mathbb{P}\left(\sup_{f \in S_d(L, \alpha, \mu, R_\mu)} d_b\left(\widehat{\mathrm{dgm}}(f), \mathrm{dgm}(f)\right) \geq th\right) \\
& \leq \mathbb{P}(K_1 + K_2 \geq t) \\
& = \mathbb{P}\left(N_h \geq \frac{t - L(k_1^\alpha + k_2^\alpha)}{4\sqrt{3M_{d,L,\alpha,R_\mu}}}\right) \\
& \leq C_0 \exp\left(-C_1 \left(\frac{t - L(k_1^\alpha + k_2^\alpha)}{4\sqrt{3M_{d,L,\alpha,R_\mu}}}\right)^2\right) \\
& \leq C_0 \exp\left(-\frac{C_1}{48M_{d,L,\alpha,R_\mu}}t^2\right) \exp\left(C_1 \frac{L(k_1^\alpha + k_2^\alpha)}{2\sqrt{3M_{d,L,\alpha,R_\mu}}}t\right) \exp\left(-C_1 \left(\frac{L(k_1^\alpha + k_2^\alpha)}{4\sqrt{3M_{d,L,\alpha,R_\mu}}}\right)^2\right)
\end{aligned}$$

and the result follows. \square

2.4 Proof of Theorem 2

The proof follows as the proof of Theorem 2 from [Henneuse \(2024\)](#). It uses a standard technique to provide minimax lower bounds, as presented in section 2 of [Tsybakov \(2008\)](#). The idea is, for all,

$$r_n = o\left(\left(\frac{\log(n)}{n}\right)^{\frac{\alpha}{d+2\alpha}}\right)$$

to exhibit a finite collection of functions in $S_d(L, \alpha, \mu, R_\mu)$ such that their persistence diagrams are two by two at distance $2r_n$ but indistinguishable, with high certainty.

Proof of Theorem 2. For m integer in $[0, \lfloor 1/h \rfloor]$, we define,

$$f_{h,m}(x_1, \dots, x_d) = \frac{L}{2} x_1^\alpha + L(h^\alpha - \|(x_1, \dots, x_d) - m/\lfloor 1/h \rfloor(1, \dots, 1)\|_\infty)^\alpha_+$$

and

$$\tilde{f}_{m,h} = \frac{1 + f_{h,m}}{\|1 + f_{h,m}\|_1}$$

The $f_{h,m}$ are (L, α) -Hölder. As for all m , $f_{h,m}$ are positive functions,

$$\|1 + f_{h,m}\|_1 \geq 1$$

consequently, $\tilde{f}_{m,h}$ is (L, α) -Hölder. As by construction $\tilde{f}_{m,h}$ are probability densities, they belong to $S_d(L, \alpha, \mu, R_\mu)$ for all $\mu \in]0, 1]$ and $R_\mu > 0$.

We have, for sufficiently small h and all $\frac{1}{4}\lfloor 1/h \rfloor < m < \frac{3}{4}\lfloor 1/h \rfloor$,

$$\text{dgm}(f_{h,m}) = \left\{ (L/2, +\infty), \left(\frac{L}{2} \left(\frac{m}{\lfloor 1/h \rfloor} \right)^\alpha - \frac{L}{2} h^\alpha, \frac{L}{2} \left(\frac{m}{\lfloor 1/h \rfloor} \right)^\alpha - Lh^\alpha \right) \right\}$$

and consequently,

$$\text{dgm}(\tilde{f}_{h,m}) = \left\{ \left(\frac{L}{2\|1 + f_{h,m}\|_1}, -\infty \right), \left(\frac{1 + \frac{L}{2} \left(\frac{m}{\lfloor 1/h \rfloor} \right)^\alpha - \frac{L}{2} h^\alpha}{\|1 + f_{h,m}\|_1}, \frac{1 + \frac{L}{2} \left(\frac{m}{\lfloor 1/h \rfloor} \right)^\alpha - Lh^\alpha}{\|1 + f_{h,m}\|_1} \right) \right\}$$

As for all $\frac{1}{4}\lfloor 1/h \rfloor < m, m' < \frac{3}{4}\lfloor 1/h \rfloor$, $\|1 + f_{h,m}\|_1 = \|1 + f_{h,m'}\|_1$, we then have, for all $\frac{1}{4}\lfloor 1/h \rfloor < m \neq m' < \frac{3}{4}\lfloor 1/h \rfloor$,

$$d_\infty \left(\text{dgm}(f_{h,m}), \text{dgm}(f_{h,m'}) \right) \geq \frac{Lh^\alpha}{2\|1 + f_{h,m}\|_1} \geq \frac{Lh^\alpha}{2(1+L)} \quad (5)$$

We set $r_n = \frac{Lh^\alpha}{4(1+L)}$, then, by (5), for all $\frac{1}{4}\lfloor 1/h \rfloor < m \neq m' < \frac{3}{4}\lfloor 1/h \rfloor$,

$$d_\infty \left(\text{dgm}(f_{h,m}), \text{dgm}(f_{h,m'}) \right) \geq 2r_n$$

For a fixed signal f , denote \mathbb{P}_f^n the joint probability distribution of X_1, \dots, X_n . From section 2 of [Tsybakov \(2008\)](#), it now suffices to show that if $r_n = o\left(\left(\frac{\log(n)}{n}\right)^{\frac{p\alpha}{d+2\alpha}}\right)$, then,

$$\frac{2}{\lfloor \frac{1}{h} \rfloor} \sum_{\frac{1}{4}\lfloor 1/h \rfloor < m' \neq m < \frac{3}{4}\lfloor 1/h \rfloor} \chi^2 \left(\mathbb{P}_{\tilde{f}_{h,m'}}^n, \mathbb{P}_{\tilde{f}_{h,m}}^n \right) \quad (6)$$

$$= \frac{2}{\lfloor \frac{1}{h} \rfloor - 2} \sum_{\frac{1}{4} \lfloor 1/h \rfloor < m' \neq m < \frac{3}{4} \lfloor 1/h \rfloor} \mathbb{E}_{\mathbb{P}_{\tilde{f}_{h,m}}^n} \left[\left(\frac{d\mathbb{P}_{\tilde{f}_{h,m'}}^n}{d\mathbb{P}_{\tilde{f}_{h,m}}^n} \right)^2 \right] - 1 \quad (7)$$

converges to zero when n converges to infinity. Note that,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}_{\tilde{f}_{h,m}}^n} \left(\left(\frac{d\mathbb{P}_{\tilde{f}_{h,m'}}^n}{d\mathbb{P}_{\tilde{f}_{h,m}}^n} \right)^2 \right) &= \mathbb{E}_{\mathbb{P}_{\tilde{f}_{h,m}}^n} \left(\frac{\prod_{i=1}^d \tilde{f}_{h,m'}^2(X_i)}{\prod_{i=1}^d \tilde{f}_{h,m}^2(X_i)} \right) \\ &= \prod_{i=1}^d \mathbb{E}_{\mathbb{P}_{\tilde{f}_{h,m}}} \left(\frac{\tilde{f}_{h,m'}^2(X_i)}{f_{d,h,m^1,L,\alpha}^2(X_i)} \right) \\ &= \left(\mathbb{E}_{\mathbb{P}_{\tilde{f}_{h,m}}} \left(\frac{\tilde{f}_{h,m'}^2(X_1)}{\tilde{f}_{h,m}^2(X_1)} \right) \right)^n \end{aligned} \quad (8)$$

we denote H_m the hypercube defined by $\|(x_1, \dots, x_d) - m/\lfloor 1/h \rfloor(1, \dots, 1)\| \leq h$. Remarking that,

$$\begin{aligned} \tilde{f}_{h,m'} &= \tilde{f}_{h,m} + \frac{L}{\|1 + f_{h,m}\|_1} (h^\alpha - \|(x_1, \dots, x_d) - m/\lfloor 1/h \rfloor(1, \dots, 1)\|_\infty^\alpha)_+ \\ &\quad - \frac{L}{\|1 + f_{h,m}\|_1} (h^\alpha - \|(x_1, \dots, x_d) - m'/\lfloor 1/h \rfloor(1, \dots, 1)\|_\infty^\alpha)_+ \end{aligned}$$

and

$$\tilde{f}_{h,m}(x) \geq \frac{1}{1+L}$$

We then have,

$$\begin{aligned} &\mathbb{E}_{\mathbb{P}_{\tilde{f}_{h,m}}} \left(\frac{\tilde{f}_{h,m'}^2(X_1)}{\tilde{f}_{h,m}^2(X_1)} \right) \\ &= 1 + 2 \frac{L}{\|1 + f_{h,m}\|_1} \int_{H_m} h^\alpha - \|(x_1, \dots, x_d) - m/\lfloor 1/h \rfloor(1, \dots, 1)\|_\infty^\alpha dt_1 \dots dt_d \\ &\quad - 2 \frac{L}{\|1 + f_{h,m}\|_1} \int_{H_{m'}} h^\alpha - \|(x_1, \dots, x_d) - m'/\lfloor 1/h \rfloor(1, \dots, 1)\|_\infty^\alpha dt_1 \dots dt_d \\ &\quad + \frac{L^2}{\|1 + f_{h,m}\|_1^2} \int_{H_m} \frac{(h^\alpha - \|(x_1, \dots, x_d) - m/\lfloor 1/h \rfloor(1, \dots, 1)\|_\infty^\alpha)^2}{\tilde{f}_{h,m}} dt_1 \dots dt_d \\ &\quad + \frac{L^2}{\|1 + f_{h,m}\|_1^2} \int_{H_{m'}} \frac{(h^\alpha - \|(x_1, \dots, x_d) - m'/\lfloor 1/h \rfloor(1, \dots, 1)\|_\infty^\alpha)^2}{\tilde{f}_{h,m}} dt_1 \dots dt_d \\ &= 1 + 2 \frac{L^2}{\|1 + f_{h,m}\|_1^2} \int_{H_m} \frac{(h^\alpha - \|(x_1, \dots, x_d) - m/\lfloor 1/h \rfloor(1, \dots, 1)\|_\infty^\alpha)^2}{\tilde{f}_{h,m}} dt_1 \dots dt_d \\ &\leq 1 + 2L^2(1+L) \int_{H_m} (h^\alpha - \|(x_1, \dots, x_d) - m/\lfloor 1/h \rfloor(1, \dots, 1)\|_\infty^\alpha)^2 dt_1 \dots dt_d \end{aligned}$$

$$\begin{aligned}
&\leq 1 + 2L^2(1 + L) \left(\int_{H_m} h^{2\alpha} dt_1 \dots dt_d + \int_{H_m} \|(x_1, \dots, x_d) - m/[1/h](1, \dots, 1)\|_\infty^{2\alpha} dt_1 \dots dt_d \right) \\
&\leq 1 + 4L^2(1 + L)h^{2\alpha+d} = 1 + O(h^{2\alpha+d})
\end{aligned}$$

Hence, if $h = o\left(\left(\frac{\log(n)}{n}\right)^{\frac{1}{d+2\alpha}}\right)$, we have that (6) converges to zero when n converges to infinity.

Consequently, if $r_n = o\left(\left(\frac{\log(n)}{n}\right)^{\frac{\alpha}{d+2\alpha}}\right)$, then $h = o\left(\left(\frac{\log(n)}{n}\right)^{\frac{1}{d+2\alpha}}\right)$ and we get the conclusion. \square

3 Multiple modes estimation

In this section, we aim to derive, from the previous procedure, estimators for the number of modes of f , their locations, and the value of f at the modes, i.e. infer the number of local maxima, their values and their locations. These local maxima simply correspond to the birth times in the H_0 -persistence diagram of f . We show that removing points "too close" to the diagonal in $\widehat{\text{dgm}}(f)$ gives a procedure that permits to recover consistently the number of modes and the associated values of f for these modes. Then, using this "regularized" persistence diagram, the estimation of the modes can then be derived from the filtration $\hat{\mathcal{F}}$.

3.1 Case where l is known

We first consider the easier setting where we suppose knowing l (or a lower bound on l), defined in Assumption **A4**. Even if this knowledge is unlikely in practice, it offers a first basic framework to work with. In this setting, an estimator of the number of modes is given by,

$$\hat{k} = \left| \left\{ (b, d) \in \widehat{\text{dgm}}(f), b - d > l/2 \right\} \right|$$

and an estimator of the complete list of local maxima of f by,

$$\hat{M} = \left\{ \hat{M}_1, \hat{M}_2, \dots, \hat{M}_{\hat{k}} \right\} = \left\{ b | (b, d) \in \widehat{\text{dgm}}(f), b - d > l/2 \right\}.$$

Now, for the modes, let denote $\{M_1, \dots, M_k\}$ the local maximum of f . Thanks to Condition **A4**, maximum are strict, thus, the collection of modes is a finite collection of singletons. For all birth times M_i , $i \in \{1, \dots, k\}$, and corresponding point (M_i, d_i) in $\text{dgm}(f)$ we denote $C_i = \{x_i\}$ the associated connected component of \mathcal{F}_{M_i} which corresponds to the mode associated to M_i (x_i being a mode of f). Similarly, for all birth times \hat{M}_i , $i \in \{1, \dots, \hat{k}\}$, and corresponding point (\hat{M}_i, \hat{d}_i) in the estimated persistence diagram, we denote \hat{C}_i the associated connected component of $\hat{\mathcal{F}}_{\hat{M}_i}$ and,

$$\hat{C} = \left\{ \hat{C}_1, \dots, \hat{C}_{\hat{k}} \right\}.$$

We can then estimate the modes by taking, for all $i \in \{1, \dots, \hat{k}\}$, any \hat{x}_i in \hat{C}_i and define the collection of estimated modes by,

$$\hat{x} = \left\{ \hat{x}_1, \dots, \hat{x}_{\hat{k}} \right\}.$$

The convergence rates achieved by this estimator are given by the following theorem.

Theorem 3. *Let $f \in S_d(L, \alpha, \mu, R_\mu, C, h_0)$ and $h \simeq \left(\frac{\log(n)}{n}\right)^{\frac{1}{d+2\alpha}}$. There exist $\check{C}_0, \check{C}_1, \check{c}_0$ and*

\check{c}_1 depending only on L, α, μ, R_μ, C , and h_0 such that, for all $A \geq 0$ with probability at least $1 - \check{C}_0 \exp(-\check{C}_1 A^2) - \check{c}_0 \exp(-\check{c}_1 h^{-2\alpha})$,

$$k = \hat{k}$$

and for all $i \in \{1, \dots, k\}$, there exists distinct $(\hat{x}_i, \hat{M}_i) \in \hat{x} \times \hat{M}$ such that,

$$\|\hat{x}_i - x_i\|_\infty \leq Ah$$

and

$$|\hat{M}_i - M_i| \leq Ah^\alpha.$$

The idea to prove Theorem 3 is the following :

- First, we prove that, with high probability, for all $i \in \{1, \dots, k\}$, \bar{C}_i , the connected component of $\mathcal{F}_{M_i - (K_1 + 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h})h^\alpha}^{k_1 h}$ containing $C_i = \{x_i\}$, contains a \hat{C}_i , i.e \bar{C}_i contains a connected component appearing in the filtration $\hat{\mathcal{F}}$ with associated lifetime exceeding $l/2$ (and furthermore we show that the birth times associated \hat{C}_i is close to M_i).
- Then, we use Theorem 1, to show that, with high probability, $k = \hat{k}$.
- Finally, we use assumption **A4** to bound, with high probability, the diameter of \bar{C}_i , and thus the Hausdorff distance between x_i and \hat{x}_i .

The last two steps are pretty straightforward, but the first one is more technical. The proof of this first claim is then decomposed into the three following lemmas. Let define, for all $i \in \{1, \dots, l\}$, for all $\lambda < M_i$, \bar{C}_λ^i the connected component of $\mathcal{F}_\lambda^{k_1 h}$ containing C_i . Denote,

$$\tilde{b}_i = \sup \left\{ \lambda \in \mathbb{R} \text{ s.t. } \hat{\mathcal{F}}_\lambda \cap \bar{C}_{M_i + (K_1 + K_2)h^\alpha - l}^i \neq \emptyset \right\}.$$

We start by establishing upper and lower bounds on \tilde{b}_i through the following lemma.

Lemma 5. Consider the event E_1 :

$$(K_1 + K_2) \left(\frac{\log(n)}{n} \right)^{\frac{\alpha}{d+2\alpha}} < l/8.$$

Under E_1 , we have, for all $i \in \{1, \dots, k\}$,

$$M_i - K_1 h^\alpha \leq \tilde{b}_i \leq M_i + 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h}h^\alpha.$$

Proof. The lower inclusion of Proposition 1 ensures that $x_i \in \hat{\mathcal{F}}_{M_i - K_1 h^\alpha}$ and thus, under E_1 ,

$$\tilde{b}_i \geq M_i - K_1 h^\alpha.$$

The upper inclusion of Proposition 1 then implies that, under E_1 , any connected component $C_{\tilde{b}_i}^i$ of $\hat{\mathcal{F}}_{\tilde{b}_i}$ intersecting $\bar{C}_{M_i + (K_1 + K_2)h^\alpha - l}^i$ is included in $\bar{C}_{M_i + (K_1 + K_2)h^\alpha - l}^i$. By Lemma 2, any connected component of $\hat{\mathcal{F}}_{\tilde{b}_i}$ intersects $\mathcal{F}_{\tilde{b}_i - 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h}h^\alpha}$. By assumption **A4**, $\bar{C}_{M_i + (K_1 + K_2)h^\alpha - l}^i \subset B_2(x, h_0)$ and hence, if,

$$\tilde{b}_i > M_i + 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h}h^\alpha$$

then,

$$\mathcal{F}_{\tilde{b}_i - 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h h^\alpha}} \cap \overline{C}_{M_i + (K_1 + K_2)h^\alpha - l}^i = \emptyset.$$

Thus, we have,

$$\tilde{b}_i \leq M_i + 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h h^\alpha}.$$

□

We now prove that for all $i \in \{1, \dots, k\}$, \overline{C}_i contains a connected component of $\widehat{\mathcal{F}}_{\tilde{b}_i}$ with associated lifetime in $\widehat{\text{dgm}}(f)$ exceeding $l/2$. The proof is divided into the two following lemmas.

Lemma 6. *Under E_1 , for all $i \in \{1, \dots, k\}$, any connected component $\tilde{C}_{\tilde{b}_i}^i$ of $\widehat{\mathcal{F}}_{\tilde{b}_i}$ intersecting $\overline{C}_{M_i + (K_1 + K_2)h^\alpha - l}^i$ is included in \overline{C}_i .*

Proof. First note that, under E_1 , we have,

$$\overline{C}_i \subset \overline{C}_{M_i + (K_1 + K_2)h^\alpha - l}^i$$

and by Assumption **A4** and Lemma 5,

$$\mathcal{F}_{\tilde{b}_i - 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h h^\alpha}} \cap \overline{C}_{M_i + (K_1 + K_2)h^\alpha - l}^i \subset \overline{C}_i.$$

It then follows, by Lemma 2, that any connected component $C_{\tilde{b}_i}^i$ of $\widehat{\mathcal{F}}_{\tilde{b}_i}$ intersecting $\overline{C}_{M_i + (K_1 + K_2)h^\alpha - l}^i$ intersects $\mathcal{F}_{\tilde{b}_i - 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h h^\alpha}}$ and thus intersects \overline{C}_i . As $\tilde{b}_i > M_i - K_1 h^\alpha$ by Lemma 5, Proposition 1 ensures that any such $C_{\tilde{b}_i}^i$ is then included in \overline{C}_i . □

By definition of \tilde{b}_i , a consequence of Lemma 6 is that, under E_1 , any connected component $\tilde{C}_{\tilde{b}_i}^i$ of $\widehat{\mathcal{F}}_{\tilde{b}_i}$ intersecting $\overline{C}_{M_i + (K_1 + K_2)h^\alpha - l}^i$ has birth time \tilde{b}_i in $\widehat{\text{dgm}}(f)$, for all $i \in \{1, \dots, k\}$. It now suffices to find such a $\tilde{C}_{\tilde{b}_i}^i$ with associated death time \tilde{d}_i verifying $\tilde{b}_i - \tilde{d}_i > l/2$.

Lemma 7. *Under E_1 , for all $i \in \{1, \dots, k\}$, there exists a connected component $\tilde{C}_{\tilde{b}_i}^i$ of $\widehat{\mathcal{F}}_{\tilde{b}_i}$ intersecting $\overline{C}_{M_i + (K_1 + K_2)h^\alpha - l}^i$ with an associated death time \tilde{d}_i in $\widehat{\text{dgm}}(f)$, verifying,*

$$\tilde{b}_i - \tilde{d}_i > l/2.$$

Proof. To prove Lemma 7, it suffices to show that, if any such $\tilde{C}_{\tilde{b}_i}^i$ merges with another connected component in $\widehat{\mathcal{F}}_\lambda$, for all $\tilde{b}_i - l/2 \geq \lambda \geq \tilde{b}_i$, this connected component is born “after” \tilde{b}_i . More formally, for all such $\tilde{C}_{\tilde{b}_i}^i$, $\tilde{b}_i - l/2 \leq \lambda \leq \tilde{b}_i$, if $x \in \tilde{C}_\lambda^i \cap \left(\tilde{C}_{\tilde{b}_i}^i\right)^c$ then $x \notin \widehat{\mathcal{F}}_{\lambda'}$, for all $\lambda' > \tilde{b}_i$.

Let a connected component $\tilde{C}_{\tilde{b}_i}^i$ of $\widehat{\mathcal{F}}_{\tilde{b}_i}$ intersecting $\overline{C}_{M_i + (K_1 + K_2)h^\alpha - l}^i$. For all $\lambda \geq \tilde{b}_i$, we denote \tilde{C}_λ^i the connected component of $\widehat{\mathcal{F}}_\lambda$ containing $\tilde{C}_{\tilde{b}_i}^i$. Applying Proposition 1 and Lemma 5, we have that for all $\tilde{b}_i - 3l/4 < \lambda \leq \tilde{b}_i$, under E_1 , any connected components of $\widehat{\mathcal{F}}_\lambda$ intersecting $\overline{C}_{M_i + (K_1 + K_2)h^\alpha - l}^i$ is included in $\overline{C}_{M_i + (K_1 + K_2)h^\alpha - l}^i$. By Lemma 6, under E_1 , \tilde{C}_λ^i intersects \overline{C}_i and thus intersects $\overline{C}_{M_i + (K_1 + K_2)h^\alpha - l}^i$. Hence, we have, under E_1 , for all $\tilde{b}_i - 3l/4 < \lambda \leq \tilde{b}_i$,

$$\tilde{C}_\lambda^i \subset \overline{C}_{M_i + (K_1 + K_2)h^\alpha - l}^i$$

Then, by definition of \tilde{b}_i , under E_1 , for all $\tilde{b}_i - 3l/4 < \lambda \leq \tilde{b}_i$, if $x \in \tilde{C}_\lambda^i \cap \left(\tilde{C}_{\tilde{b}_i}^i\right)^c$ then $x \notin \widehat{\mathcal{F}}_{\lambda'}$, for all $\lambda' > \tilde{b}_i$ and Lemma 7 follows. □

Equipped with Lemma 5, 6 and 7, we can now prove Theorem 3.

Proof of Theorem 3. Let $f \in S_d(L, \alpha, \mu, R_\mu, C, h_0)$. From the proof of Theorem 1, we have that,

$$d_\infty \left(\widehat{\text{dgm}}(f), \text{dgm}(f) \right) \leq (K_1 + K_2) \left(\frac{\log(n)}{n} \right)^{\frac{\alpha}{d+2\alpha}}. \quad (9)$$

Then, under E_1 it follows that $\hat{k} = k$. By applying Proposition 3, one can check that there exist A_1 and B_1 such that E_1 occurs with probability at least $1 - A_1 \exp(-B_1 h^{-2\alpha})$.

The combination of Lemma 6 and 7 ensures that, under E_1 , for all $i \in \{1, \dots, k\}$ there exists a connected component $\tilde{C}_{\tilde{b}_i}^i$ of $\hat{\mathcal{F}}_{\tilde{b}_i}^i$ included in \bar{C}^i , such that its lifetime exceeds $l/2$. Thus, by definition of \hat{C} , for all $i \in \{1, \dots, k\}$, there exists $\hat{C}_i \in \hat{C}$ such that, $\hat{C}_i \subset \tilde{C}_{\tilde{b}_i}^i$ and $(\tilde{b}_i, \tilde{d}_i) = (\hat{M}_i, \hat{d}_i)$. Proposition 2 and Assumption A4 imply that, under E_1 , $\bar{C}_1, \dots, \bar{C}_k$ are disjoint, thus $\tilde{C}_{\tilde{b}_1}^1, \dots, \tilde{C}_{\tilde{b}_k}^k$ are disjoint. Hence, there exists at least $\hat{C}_1, \dots, \hat{C}_k$ in \hat{C} , all disjoint two by two, such that $\hat{C}_1 \subset \tilde{C}_{\tilde{b}_1}^1, \dots, \hat{C}_k \subset \tilde{C}_{\tilde{b}_k}^k$. Thus, as $k = \hat{k}$, we have,

$$\hat{C} = \left\{ \hat{C}_1, \dots, \hat{C}_k \right\}.$$

As, for all $i \in \{1, \dots, k\}$, $C_i \subset \bar{C}_i$ and $\hat{C}_i \subset \bar{C}_i$, to bound the Hausdorff distance between C_i and \hat{C}_i (and consequently the distance between \hat{x}_i and x_i), it suffices to bound the diameter of \bar{C}_i , defined by,

$$\text{diam}(\bar{C}_i) = \max_{x, y \in \bar{C}_i} \|x - y\|_\infty.$$

Consider the event E_2 :

$$\left(\frac{(K_1 + 2\sqrt{3M_{d,L,\alpha,R_\mu} N_h})^{1/\alpha}}{C} + (\sqrt{d} + \lceil \sqrt{d}/\mu \rceil) \right) h < h_0.$$

Again, one can check that, by Proposition 3, there exist A_2 and B_2 such that E_2 occurs with probability at least $1 - A_2 \exp(-B_2 h^{-2\alpha})$.

As f verifies Assumption A4, under E_2 , for all $i \in \{1, \dots, k\}$,

$$\text{diam}(\bar{C}_i) \leq \left(\frac{(K_1 + 2\sqrt{3M_{d,L,\alpha,R_\mu} N_h})^{1/\alpha}}{C} + (\sqrt{d} + \lceil \sqrt{d}/\mu \rceil) \right) h.$$

Hence,

$$d_H(C_i, \hat{C}_i) \leq \text{diam}(\bar{C}_i) \leq \left(\frac{(K_1 + 2\sqrt{3M_{d,L,\alpha,R_\mu} N_h})^{1/\alpha}}{C} + (\sqrt{d} + \lceil \sqrt{d}/\mu \rceil) \right) h$$

and thus,

$$\|x_i - \hat{x}_i\|_\infty \leq \left(\frac{(K_1 + 2\sqrt{3M_{d,L,\alpha,R_\mu} N_h})^{1/\alpha}}{C} + (\sqrt{d} + \lceil \sqrt{d}/\mu \rceil) \right) h. \quad (10)$$

Consider the event E_3 :

$$\left(\frac{(K_1 + 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h})^{1/\alpha}}{C} + (\sqrt{d} + \lceil \sqrt{d}/\mu \rceil) \right) \leq A$$

and the event E_4 : $K_1 \leq A$.

Using Proposition 3, one can check that there exist A_3, B_3 , such that E_3 occurs with probability $1 - A_3 \exp(-B_3 A^2)$ and A_4, B_4 , such that E_4 occurs with probability $1 - A_4 \exp(-B_4 A^2)$. Hence, it follows from Lemma 5 and (10) that, with probability at least,

$$\begin{aligned} & 1 - \mathbb{P}(E_1) - \mathbb{P}(E_2) - \mathbb{P}(E_3) - \mathbb{P}(E_4) \\ & \geq 1 - A_1 \exp(-B_1 h^{-2\alpha}) - A_2 \exp(-B_2 h^{-2\alpha}) - A_3 \exp(-B_3 A^2) - A_4 \exp(-B_4 A^2) \end{aligned}$$

$\hat{k} = k$ and for all $i \in \{1, \dots, k\}$ there exists distinct $(\hat{x}_i, \hat{M}_i) \in \hat{x} \times \hat{M}$ such that,

$$\|x_i - \hat{x}_i\|_\infty \leq Ah$$

and

$$|\hat{M}_i - M_i| \leq Ah^\alpha.$$

□

3.2 Case where l is unknown

Let's now consider the more realistic case where l is unknown. The idea is selecting l through penalization. The procedure remains the same, except for the construction of \hat{k} . Let,

$$R(l) = d_\infty \left(\widehat{\text{dgm}}(f), \overline{\text{dgm}}(f)^l \right) + \frac{h^\alpha}{l}$$

with

$$\overline{\text{dgm}}(f)^l = \left\{ (b, d) \in \widehat{\text{dgm}}(f) \text{ s.t } b - d > l \right\}.$$

We consider,

$$\hat{l} = \operatorname{argmin} R(l)$$

the existence of this minimum is proved in Lemma 8 and \hat{k} is then defined as,

$$\hat{k} = \left| \left\{ (b, d) \in \widehat{\text{dgm}}(f), b - d > \hat{l} \right\} \right|.$$

With this choice for \hat{l} we describe the obtained convergence rates achieved in the following theorem.

Theorem 4. *Let $f \in S_d(L, \alpha, \mu, R_\mu, C, h_0)$ and $h \simeq \left(\frac{\log(n)}{n} \right)^{\frac{1}{d+2\alpha}}$. There exist $\check{C}_0, \check{C}_1, \check{c}_0, \check{c}_1$ depending only on L, α, μ, R_μ, C and h_0 such that, for all $A \geq 0$ with probability at least $1 - \check{C}_0 \exp(-\check{C}_1 A^2) - \check{c}_0 \exp(-\check{c}_1 h^{-\alpha})$,*

$$k = \hat{k}$$

and, for all $i \in \{1, \dots, k\}$, there exists distinct (\hat{x}_i, \hat{M}_i) such that,

$$\|\hat{x}_i - x_i\|_\infty \leq Ah$$

and

$$|\hat{M}_i - M_i| \leq Ah^\alpha.$$

The key idea to prove this result is contained in the following lemma. Let l^* , the distance between $\text{dgm}(f)$ and the diagonal and \tilde{l} the distance between the set,

$$\left\{ (b, d) \in \widehat{\text{dgm}}(f) \text{ s.t. } b - d > (K_1 + K_2)h^\alpha \right\}$$

and the diagonal. Note that l^* and \tilde{l} are possibly infinite.

Lemma 8. *Let $h \simeq \left(\frac{\log(n)}{n}\right)^{\frac{1}{d+2\alpha}}$, and E_5 the event :*

$$\max\left(\left((K_1 + K_2)^2, K_1 + K_2\right) h^\alpha < 1/2.\right.$$

For n sufficiently large, under $E_5 \cap E_1$, R admits a minimum over $]0, 1]$ attained only at $l = \min(\tilde{l}, 1)$.

Proof. Under E_1 , by Theorem 1,

$$7l^*/8 \leq l^* - (K_1 + K_2)h^\alpha \leq \tilde{l} \leq l^* + (K_1 + K_2)h^\alpha \leq 9l^*/8$$

Let's first tackle the case where $\tilde{l} < 1$, we have, under E_1

$$R(\tilde{l}) \leq (K_1 + K_2)h^\alpha + \frac{h^\alpha}{\tilde{l}} \leq (K_1 + K_2)h^\alpha + \frac{8h^\alpha}{7l^*} < \frac{l^*}{8} + \frac{8h^\alpha}{7l^*}$$

- If $0 < l \leq (K_1 + K_2)h^\alpha$, under E_1 and E_5 , we have,

$$R(l) \geq \frac{h^\alpha}{(K_1 + K_2)h^\alpha} = \frac{1}{(K_1 + K_2)} > (K_1 + K_2)h^\alpha + \frac{8h^\alpha}{7l^*} \geq R(\tilde{l}).$$

- If $(K_1 + K_2)h^\alpha < l < \tilde{l}$, then by definition of \tilde{l} ,

$$d_\infty\left(\widehat{\text{dgm}}(f), \overline{\text{dgm}}(f)^l\right) = d_\infty\left(\widehat{\text{dgm}}(f), \overline{\text{dgm}}(f)^{\tilde{l}}\right)$$

and $h^\alpha/l > h^\alpha/\tilde{l}$. Thus, $R(l) > R(\tilde{l})$.

- If $l > \tilde{l}$, under E_1 and for n sufficiently large

$$R(l) \geq \tilde{l} \geq 7l^*/8 > \frac{l^*}{8} + \frac{8h^\alpha}{7l^*} \geq R(\tilde{l}).$$

Hence, in these cases, if $l \neq \tilde{l}$, $R(l) > R(\tilde{l})$. Now for the case where $\tilde{l} \geq 1$,

$$R(1) \leq (K_1 + K_2)h^\alpha + h^\alpha \leq \frac{l^*}{8} + h^\alpha$$

- If $0 < l \leq (K_1 + K_2)h^\alpha$, under E_5 and for n sufficiently large,

$$R(l) \geq \frac{h^\alpha}{(K_1 + K_2)h^\alpha} = \frac{1}{(K_1 + K_2)} > (K_1 + K_2)h^\alpha + h^\alpha \geq R(1).$$

- If $1 \geq l > (K_1 + K_2)h^\alpha$,

$$d_\infty \left(\widehat{\text{dgm}}(f), \overline{\text{dgm}}(f)^l \right) = d_\infty \left(\widehat{\text{dgm}}(f), \overline{\text{dgm}}(f)^1 \right)$$

and $h^\alpha/l > h^\alpha$. Thus, $R(l) > R(1)$.

Hence, in this case, if $l \neq 1$, $R(l) > R(1)$. Combining both cases, the result is proved. \square

Proof of Theorem 4. Lemma 8, imply that, under $E_1 \cap E_5$, for sufficiently large n ,

$$\hat{l} = \min(\tilde{l}, 1) \leq \tilde{l}$$

Thus, by definition of \tilde{l} ,

$$\begin{aligned} d_\infty \left(\text{dgm}(f), \overline{\text{dgm}}(f)^{\hat{l}} \right) &\leq d_\infty \left(\widehat{\text{dgm}}(f), \text{dgm}(f) \right) + d_\infty \left(\widehat{\text{dgm}}(f), \overline{\text{dgm}}(f)^{\tilde{l}} \right) \\ &\leq 2(K_1 + K_2)h^\alpha. \end{aligned}$$

One can check that by Proposition 3 that there exist A_5 and B_5 such that E_5 occurs with probability $1 - B_5 \exp(-A_5 h^{-\alpha})$. The proof then follows as in the proof of Theorem 3. \square

3.3 Lower bounds

Here, we state that, up to a log factor, the upper bounds attained by our estimation procedure are optimal in the minimax sense.

Theorem 5. *There are two constants A and B and two densities in $S_d(L, \alpha, \mu, R_\mu, C, h_0)$ with modes separated by $n^{\frac{-1}{d+2\alpha}}/A$ and local maxima separated by $n^{\frac{-\alpha}{d+2\alpha}}/B$ that cannot be distinguished with more than probability $1/5$ for a sample size of n .*

Proof of Theorem 5. We consider the subproblem of modes estimation for unimodal densities, adapting the proof of Theorem 2 of Arias-Castro et al. (2022). The idea here follows Le Cam's two-points reasoning. For simplicity, we fix $C = 1$ (other cases can be treated similarly), let $L > 1$,

$$\begin{aligned} f_0(x_1, \dots, x_d) &= (1 - \|x - (1/2, \dots, 1/2)\|_\infty^\alpha)_+ \\ \tilde{f}_0 &= 1 - \|f_0\|_1 + f_0 \end{aligned}$$

and denoted Δ_d the set of vertices of the cube $[0, 1]^d$,

$$\tilde{f}_1 = \tilde{f}_0 + L \left(h^\alpha - \left\| x - \left(\frac{1}{2} + h \right) (1, \dots, 1) \right\|_\infty^\alpha \right)_+ - \sum_{y \in \Delta_d} L (h^\alpha - \|x - y\|_\infty^\alpha)_+$$

As $\|f_0\|_1 < 1$, \tilde{f}_0 is a density. As $\|\tilde{f}_1\|_1 = \|\tilde{f}_0\|_1$ and for sufficiently small h , \tilde{f}_1 is positive, thus \tilde{f}_1 is also a probability distribution. Since f_0 and \tilde{f}_1 are continuous, they verify **A2** and **A3** for all $1 \geq \mu > 0$ and R_μ . As \tilde{f}_0 and \tilde{f}_1 are also (L, α) -Hölder thus satisfy **A1** And by construction verify **A4** for $C = 1$ and all $h_0 \leq 1$.

For sufficiently small h , they admit a unique local maximum, attained at $x_0 = (1/2, \dots, 1/2)$ for \tilde{f}_0 and at $x_1 = (\frac{1}{2} + h)(1, \dots, 1)$ for \tilde{f}_1 . Consequently, their H_0 -persistence diagrams admit

only one (infinite) point and \tilde{f}_0 and \tilde{f}_1 belongs to $S_d(L, \alpha, \mu, R_\mu, C, h_0)$ for all $0 < \mu \leq 1$, $R_\mu > 0$, $C = 1$, $L > C$ and $0 < h_0 \leq 1$. Additionally, remark that,

$$\|x_0 - x_1\|_\infty = h.$$

Let $r_n = h/2$, we then have,

$$\|x_0 - x_1\|_\infty \geq 2r_n$$

From [Tsybakov \(2008\)](#), it now suffices to prove that if $r_n = o\left(n^{-\frac{1}{2\alpha+d}}\right)$ then $n\chi^2\left(\tilde{f}_0, \tilde{f}_1\right)$ converges to zero when n tends to infinity. Let then look at the chi-squared divergence between \tilde{f}_0 and \tilde{f}_1 ,

$$\begin{aligned} & \chi^2\left(\tilde{f}_1, \tilde{f}_0\right) \\ &= \int_{[0,1]^d} \frac{\tilde{f}_1^2}{\tilde{f}_0} - 1 \\ &= \int_{[0,1]^d} \frac{\left(\tilde{f}_0 + L\left(h^\alpha - \left\|t - \left(\frac{1}{2} + h\right)(1, \dots, 1)\right\|_\infty\right)_+ - \sum_{y \in \Delta_d} L\left(h^\alpha - \|t - y\|_\infty\right)_+\right)^2}{\tilde{f}_0} - 1 \\ &\leq 2L \left(\int_{\left\|t - \left(\frac{1}{2} + h\right)(1, \dots, 1)\right\|_\infty \leq h} h^\alpha - \left\|t - \left(\frac{1}{2} + h\right)(1, \dots, 1)\right\|_\infty^\alpha dt - 2^d \int_{\|t\|_\infty \leq h} h^\alpha - \|t\|_\infty^\alpha dt \right) \\ &\quad + \frac{L^2}{1 - \|f_0\|_1} \int_{\left\|t - \left(\frac{1}{2} + h\right)(1, \dots, 1)\right\|_\infty \leq h} \left(h^\alpha - \left\|t - \left(\frac{1}{2} + h\right)(1, \dots, 1)\right\|_\infty^\alpha\right)^2 dt \\ &\quad + \frac{2^d L^2}{1 - \|f_0\|_1} \int_{\|t\|_\infty \leq h} (h^\alpha - \|t\|_\infty^\alpha)^2 dt \\ &= \frac{L^2}{1 - \|f_0\|_1} \int_{\left\|t - \left(\frac{1}{2} + h\right)(1, \dots, 1)\right\|_\infty \leq h} \left(h^\alpha - \left\|t - \left(\frac{1}{2} + h\right)(1, \dots, 1)\right\|_\infty^\alpha\right)^2 dt \\ &\quad + \frac{2^d L^2}{1 - \|f_0\|_1} \int_{\|t\|_\infty \leq h} (h^\alpha - \|t\|_\infty^\alpha)^2 dt \\ &\leq \frac{L^2}{1 - \|f_0\|_1} \left(\int_{\left\|t - \left(\frac{1}{2} + h\right)(1, \dots, 1)\right\|_\infty \leq h} h^{2\alpha} dt + \int_{\left\|t - \left(\frac{1}{2} + h\right)(1, \dots, 1)\right\|_\infty \leq h} \left\|t - \left(\frac{1}{2} + h\right)(1, \dots, 1)\right\|_\infty^{2\alpha} dt \right) \\ &\quad + \frac{2^d L^2}{1 - \|f_0\|_1} \left(\int_{\|t\|_\infty \leq h} h^{2\alpha} dt + \int_{\|t\|_\infty \leq h} \|t\|_\infty^{2\alpha} dt \right) \\ &\leq \frac{2^{d+1} L^2}{1 - \|f_0\|_1} h^{2\alpha+d} \\ &= O\left(h^{2\alpha+d}\right) \end{aligned}$$

Now, if $r_n = o\left(n^{-\frac{1}{2\alpha+d}}\right)$ then $h = o\left(n^{-\frac{1}{2\alpha+d}}\right)$ and thus $n\chi^2\left(\tilde{f}_0, \tilde{f}_1\right)$ converges to zero when n goes to infinity, which gives the first assertion. For the second one, remark that, M_0 the (only) local maximum of \tilde{f}_0 (and of f_0) is $2 - \|f_0\|_1$ and M_1 the (only) local maximum of \tilde{f}_1 (and of f_1) is $2 - \|f_0\|_1 + Lh^\alpha$. Thus,

$$|M_0 - M_1| = Lh^\alpha$$

Taking $r_n = Lh^\alpha/2$, we then have,

$$|M_0 - M_1| > 2r_n.$$

As we proved that $\chi^2(\tilde{f}_0, \tilde{f}_1) = O(h^{2\alpha+d})$, if $r_n = o(n^{-\frac{\alpha}{2\alpha+d}})$ then $h = o(n^{-\frac{1}{2\alpha+d}})$ and $n\chi^2(\tilde{f}_0, \tilde{f}_1)$ converges to zero when n goes to infinity, which complete the proof. \square

4 Numerical illustrations

This section provides some numerical 1D and 2D illustrations for both persistence diagram and modes inference. These examples all belong to some $S_d(L, \alpha, \mu, R_\mu, C, h_0)$, but do not verify usual local regularity assumptions around modes. In dimension 1, all piecewise density have discontinuity sets with positive reach. Thus, in this case, the estimators are simply derived from a simple histogram estimator of the density. In dimension 2, this is no longer the case, the estimators are then derived from thickened histograms. The choice of the parameters h is made as suggested by the theory, i.e $h \simeq (\log(n)/n)^{1/(d+2\alpha)}$.

Example 1. We start by giving a simple 1D illustration. $f(x) = \cos(2\pi x) + 2\mathbb{1}_{x \geq 1/2}$ normalized to be a density through standard numerical scheme. We sample $n = 1000$ points in $[0, 1]$ according to f . We estimate the persistence diagram and the modes of f with the following choice of parameters : $\mu = 1$ and $h = (\log(n)/n)^{1/2}/4$ (as f belongs to $S(1, 1/2, 1, 1/2, 1, 1/4)$). This example is of particular interest as f is not continuous in all neighborhoods of the modes.

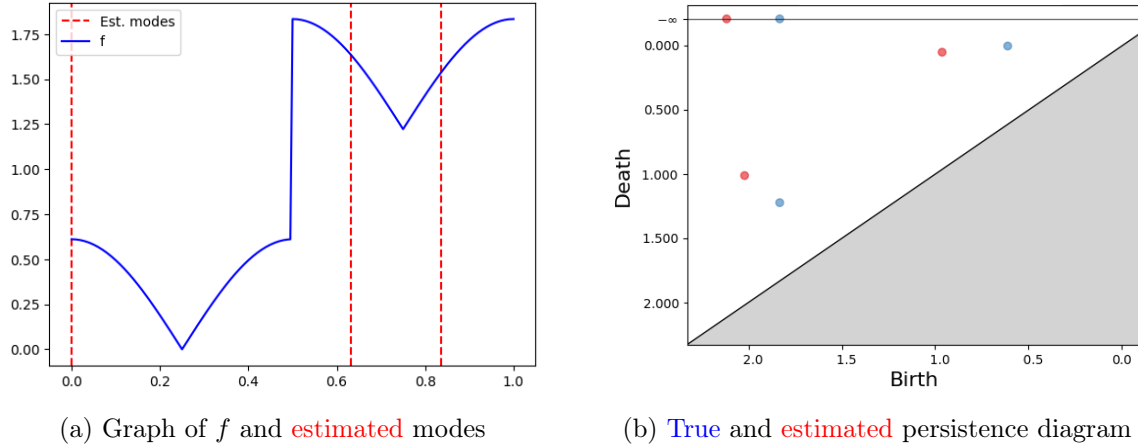


Figure 8: 1D numerical illustration, example 1.

Example 2.1. Here, we consider the 2D density given by the normalization of f define by,

$$f(x, y) = 2 \left(\frac{1}{20} - \|(x, y) - (1/4, 1/4)\|_2^2 \right)_+ + 4 \left(\frac{1}{20} - \|(x, y) - (3/4, 3/4)\|_2^2 \right)_+ \\ + \left(\frac{1}{16} - \|(x, y) - (3/4, 1/4)\|_2^2 \right)_+ + \left(\frac{1}{16} - \|(x, y) - (1/4, 3/4)\|_2^2 \right)_+$$

if $1/4 \leq y \leq 3/4$ and $1/4 - (y - 1/4)(y - 3/4) \leq x \leq 3/4 + (y - 1/4)(y - 3/4)$ and else $f(x, y) = 0$.

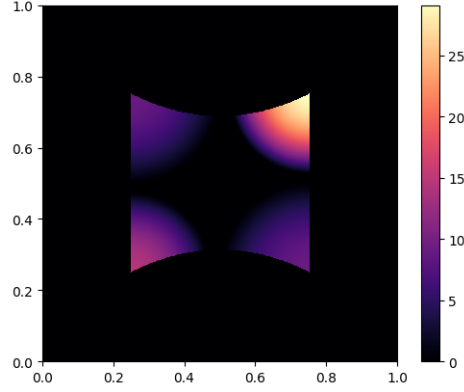
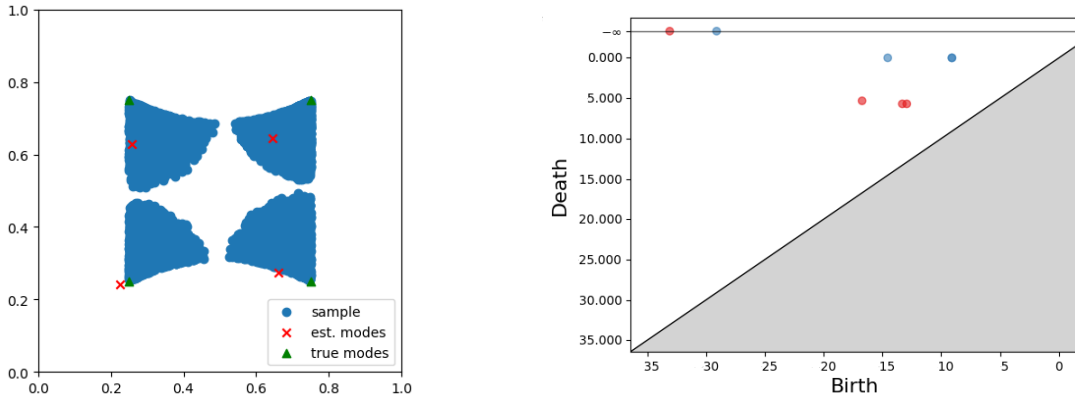


Figure 9: Graph of f

Again, this function is interesting as it is not continuous in any neighborhood of the modes, but only in a small half cone containing the modes. Additionally, one can check that this function belongs to $S_d(L, 1/2, \mu, R_\mu, C, h_0, l)$ for $\mu \leq 1/2$ and some $L > 0$, $R_\mu > 0$, $C > 0$, $h_0 > 0$. We sample $n = 10000$ points according to f , and choose the parameters $\mu = 1/2$ and $h = (\log(n)/n)^{1/3}/6$.



(a) Sampled points, true and estimated modes (b) True and estimated persistence diagrams

Figure 10: 2D numerical illustration, example 2.1.

Example 2.2. We propose another 2D example by considering the normalization of f defined by,

$$f(x, y) = \frac{1}{2} (1/8 - \|(x, y) - (1/4, 1/2)\|_2)_+$$

if $|x - 1/4| + |y - 1/4| \leq 1/4$ or $|x - 1/4| + |y - 3/4| \leq 1/4$,

$$f(x, y) = (1/8 - \|(x, y) - (3/4, 1/2)\|_2)_+$$

if $|x - 3/4| + |y - 3/4| \leq 1/4$ or $|x - 3/4| + |y - 1/4| \leq 1/4$, and $f(x, y) = 0$ else.

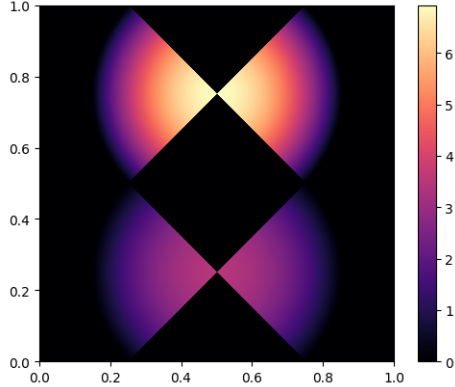
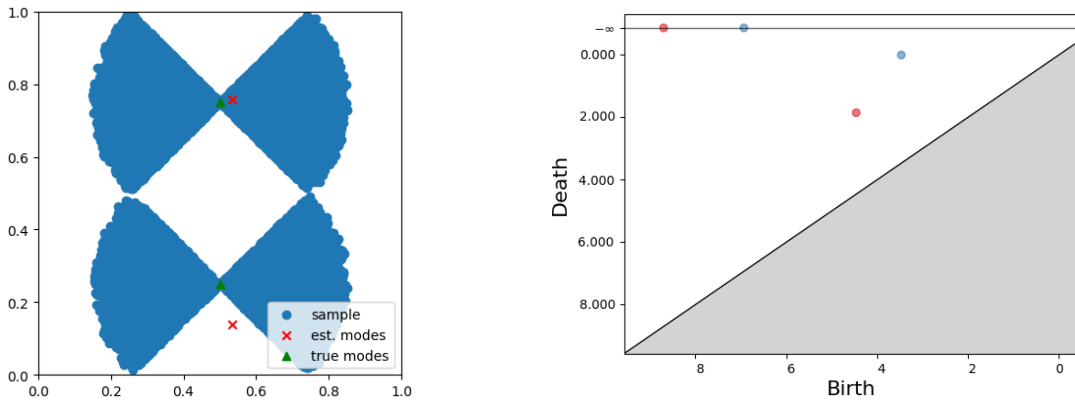


Figure 11: Graph of f

In all neighborhoods of the modes f is discontinuous. But again, one can check that this function belongs to $S_d(L, 1/2, \mu, R_\mu, C, h_0)$ for $\mu < \sqrt{2}/2$ and some $L > 0, R_\mu > 0, C > 0, h_0 > 0$. This example is interesting as, for each mode, capturing the connectivity of the two "half cones" intersecting at the mode can be difficult, as illustrated in Figure 7. We sample $n = 30000$ points according to f , and choose the parameters $\mu = 1/2$ and $h = (\log(n)/n)^{1/3}/4$.



(a) Sampled points, true and estimated modes (b) True and estimated persistence diagrams

Figure 12: 2D numerical illustration, example 2.2.

5 Discussion

Exploiting the link between modes and persistence, we propose an estimator that permits to infer consistently the number of modes, their location and associated local maxima. We lead a rigorous study of the convergence properties of this estimator, showing that it achieves minimax rates over large classes of densities. In contrast to existing work, our approach involves distinct density assumptions. While we impose a stronger global regularity assumption, we concurrently relax requirements around modes. Notably, we do not presume continuity in the neighborhood of modes, but only in small half cones containing modes. This provides new insights and perspectives on multiple modes inference.

Along the way, we also extend the approach and results of Henneuse (2024) for H_0 -persistence diagram estimation, in the context of the density model. We believe that these results are interesting by themselves as they extend the current scope on persistent homology inference for density, that has mainly remained limited to regular density. These results show some strong robustness to discontinuity of H_0 -persistence diagram estimation. We currently work on generalization for higher order homology.

Another point of interest is the adaptivity of such methods. Here, the proposed estimators depend on the regularity parameters α and the curvature measure of the discontinuity set μ . While the dependence in α can be handled via standard methods, such as Lepski's method (as done in Henneuse (2024)), overcoming the dependence in μ is less straightforward. Exploring methodologies to develop adaptive procedures to μ requires further dedicated research.

Acknowledgements

The author would like to thank Frédéric Chazal and Pascal Massart for our (many) helpful discussions. The author acknowledges the support of the ANR TopAI chair (ANR-19-CHIA-0001).

References

- Christophe Abraham, Gérard Biau, and Benoît Cadre. Simple estimation of the mode of a multivariate density. *Canadian Journal of Statistics*, 31(1):23–34, 2003. doi: 10.2307/3315901.
- Christophe Abraham, Gérard Biau, and Benoît Cadre. On the asymptotic properties of a simple estimate of the mode. *ESAIM: PS*, 8:1–11, 2004. doi: 10.1051/ps:2003015.
- Jose Ameijeiras-Alonso, Rosa M. Crujeiras, and Alberto Rodríguez-Casal. Mode testing, critical bandwidth and excess mass. *TEST*, 28:900–919, 2019. doi: 10.1007/s11749-018-0611-5.
- Ery Arias-Castro, David Mason, and Bruno Pelletier. On the estimation of the gradient lines of a density and the consistency of the mean-shift algorithm. *Journal of Machine Learning Research*, 17(43):1–28, 2016.
- Ery Arias-Castro, Wanli Qiao, and Lin Zheng. Estimation of the global mode of a density: Maximality, adaptation, and computational complexity. *Electronic Journal of Statistics*, 16(1):2774 – 2795, 2022. doi: 10.1214/21-EJS1972.
- Serguei Barannikov. The framed morse complex and its invariants. *Adv. Soviet Math.*, 21:93–115, 1994.
- Ulrich Bauer, Axel Munk, Hannes Sieling, and Max Wardetzky. Persistent homology meets statistical inference - a case study: Detecting modes of one-dimensional signals. 2014.
- Omer Bobrowski, Sayan Mukherjee, and Jonathan E. Taylor. Topological consistency via kernel estimation. *Bernoulli*, 23(1):288 – 328, 2017. doi: 10.3150/15-BEJ744.
- Peter Bubenik, Gunnar Carlsson, Peter Kim, and Zhiming Luo. Statistical topology via morse theory persistence and nonparametric estimation. *Contemporary Mathematics*, 516:75–92, 2009. doi: 10.1090/conm/516/10167.

- Prabir Burman and Wolfgang Polonik. Multivariate mode hunting: Data analytic tools with measures of significance. *Journal of Multivariate Analysis*, 100(6):1198–1218, 2009. doi: 10.1016/j.jmva.2008.10.015.
- Miguel Carreira-Perpinan. Gaussian mean-shift is an em algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(5):767–776, 2007. doi: 10.1109/TPAMI.2007.1057.
- Miguel Carreira-Perpinan and Christopher Williams. On the number of modes of a gaussian mixture. In *Scale Space Methods in Computer Vision*, pages 625–640. Springer Berlin Heidelberg, 2003. doi: 10.1007/3-540-44935-3_44.
- José Chacón. Clusters and water flows: a novel approach to modal clustering through morse theory. 2012.
- José Chacón. The modal age of statistics. *International Statistical Review*, 88, 2020. doi: 10.1111/insr.12340.
- Frédéric Chazal, David Cohen-Steiner, Marc Glisse, Leonidas J. Guibas, and Steve Y. Oudot. Proximity of persistence modules and their diagrams. In *Proceedings of the Twenty-Fifth Annual Symposium on Computational Geometry*, SCG '09, page 237–246, 2009. doi: 10.1145/1542362.1542407.
- Frédéric Chazal, Steve Oudot, Marc Glisse, and Vin de Silva. *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics. Springer Verlag, 2016. doi: 10.1007/978-3-319-42545-0.
- Frédéric Chazal and Bertrand Michel. An introduction to topological data analysis: fundamental and practical aspects for data scientists, 2021.
- Frédéric Chazal, David Cohen-Steiner, and André Lieutier. A sampling theory for compact sets in euclidean space. *Discrete & Computational Geometry*, 41:461–479, 2006. doi: 10.1007/s00454-009-9144-8.
- Frédéric Chazal, David Cohen-Steiner, André Lieutier, and Boris Thibert. Shape smoothing using double offset. *Proceedings - SPM 2007: ACM Symposium on Solid and Physical Modeling*, 2007. doi: 10.1145/1236246.1236273.
- Frédéric Chazal, Leonidas Guibas, Steve Oudot, and Primoz Skraba. Persistence-based clustering in riemannian manifolds. *Journal of the ACM*, 60, 2011. doi: 10.1145/2535927.
- Ming-Yen Cheng and Peter Hall. Calibrating the excess mass and dip tests of modality. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60(3):579–589, 1998.
- Yizong Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799, 1995. doi: 10.1109/34.400568.
- Herbert Chernoff. Estimation of the mode. *Annals of the Institute of Statistical Mathematics*, 16, 1964.
- David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete and Computational Geometry - DCG*, 37:263–271, 2005. doi: 10.1007/s00454-006-1276-5.

- Dorin Comaniciu and Peter Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002. doi: 10.1109/34.1000236.
- William Crawley-Boevey. Decomposition of pointwise finite-dimensional persistence modules, 2012.
- Sanjoy Dasgupta and Samory Kpotufe. Optimal rates for k-nn density and mode estimation. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.
- Luc Devroye. Recursive estimation of the mode of a multivariate density. *Canadian Journal of Statistics*, 7(2):159–167, 1979. doi: 10.2307/3315115.
- David Donoho and Richard Liu. Geometrizing Rates of Convergence, III. *The Annals of Statistics*, 19(2):668 – 701, 1991. doi: 10.1214/aos/1176348115.
- Tarn Duong, Arianna Cowling, Inge Koch, and Matt Wand. Feature significance for multivariate kernel density estimation. *Computational Statistics & Data Analysis*, 52:4225–4242, 2008. doi: 10.1016/j.csda.2008.02.035.
- William Eddy. Optimum Kernel Estimators of the Mode. *The Annals of Statistics*, 8(4):870 – 882, 1980. doi: 10.1214/aos/1176345080.
- Herbert Federer. Curvature measures. *Trans. Amer. Math. Soc.*, 1959.
- Nick Fisher and Steve Marron. Mode testing via the excess mass estimate. *Biometrika*, 88(2): 499–517, 2001.
- Keinosuke Fukunaga and Larry Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1): 32–40, 1975. doi: 10.1109/TIT.1975.1055330.
- Christopher Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Non-Parametric Inference for Density Modes. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(1):99–126, 2015.
- Peter Hall and Matthew York. On the calibration of silverman’s test for multimodality. *Statistica Sinica*, 11(2):515–536, 2001.
- John Hartigan and P. M. Hartigan. The Dip Test of Unimodality. *The Annals of Statistics*, 13 (1):70 – 84, 1985. doi: 10.1214/aos/1176346577.
- Allen Hatcher. *Algebraic topology*. Cambridge Univ. Press, Cambridge, 2000.
- Hugo Henneuse. Persistence diagram estimation of multivariate piecewise hölder-continuous signals. 2024.
- Heinrich Jiang and Samory Kpotufe. Modal-set estimation with an application to clustering. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 1197–1206. PMLR, 2017.
- Jisu Kim, Jaehyeok Shin, Frédéric Chazal, Alessandro Rinaldo, and Larry Wasserman. Homotopy reconstruction via the cech complex and the vietoris-rips complex, 2020.

- Jussi Klemelä. Adaptive estimation of the mode of a multivariate density. *Journal of Nonparametric Statistics*, 17(1):83–105, 2005. doi: 10.1080/10485250410001723151.
- Valentin Konakov. On the asymptotic normality of the mode of multidimensional distributions. *Theory of Probability & Its Applications*, 18(4):794–799, 1974. doi: 10.1137/1118104.
- Oleg Lepski. Asymptotically minimax adaptive estimation. i: Upper bounds. optimally adaptive estimates. *Theory of Probability & Its Applications*, 36(4):682–697, 1992. doi: 10.1137/113608.
- Jia Li, Surajit Ray, and Bruce Lindsay. A nonparametric statistical approach to clustering via mode identification. *Journal of Machine Learning Research*, 8:1687–1723, 2007.
- Michael Minnotte. Nonparametric testing of the existence of modes. *The Annals of Statistics*, 25(4):1646 – 1660, 1997. doi: 10.1214/aos/1031594735.
- Abdelkader Mokkadem and Mariane Pelletier. The law of the iterated logarithm for the multivariate kernel mode estimator. *Esaim: Probability and Statistics*, 7:1–21, 2003.
- Emanuel Parzen. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, 33(3):1065 – 1076, 1962.
- Joseph Romano. On Weak Convergence and Optimality of Kernel Density Estimates of the Mode. *The Annals of Statistics*, 16(2):629 – 647, 1988. doi: 10.1214/aos/1176350824.
- Mrityunjay Samanta. Nonparametric estimation of the mode of a multivariate density. *South African Statistical Journal*, 7:109–117, 1973.
- Bernard Silverman. Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society: Series B (Methodological)*, 43(1):97–99, 1981.
- Alexandre Tsybakov. Recursive estimation of the mode of a multivariate distribution. *Problemy Peredachi Informatsii*, 26:38–45, 1990.
- Alexandre Tsybakov. *Introduction to Nonparametric Estimation*. Springer Publishing Company, Incorporated, 2008.
- Phillipe Vieu. A note on density mode estimation. *Statist. Probab. Lett.*, 26:297–307, 1996. doi: 10.1016/0167-7152(95)00024-0.

A Proof of Lemma 4

This section is dedicated to the proof of Lemma 4.

Proof of Lemma 4. The first part of the claim is Theorem 12 of [Kim et al. \(2020\)](#). A standard technique to construct deformation retraction in differential topology is to exploit the flow coming from a smooth underlying vector field. In [Kim et al. \(2020\)](#), their idea is to use the vector field defined on $B_2(K, r) \setminus K$, by $W(x) = -\nabla_K(x)$. But this vector field is not continuous. To overcome this issue, they construct a locally finite covering $(U_{x_i})_{i \in \mathbb{N}}$ of $B_2(K, r) \setminus K$ and an associated partition of the unity $(\rho_i)_{i \in \mathbb{N}}$, such that $\overline{W}(x) = \sum_{i \in \mathbb{N}} \rho_i(x) w(x_i)$ shares the same dynamic as W . More precisely, they show that \overline{W} induces a smooth flow C that can be extended on $B_2(K, r) \times [0, +\infty[$ such that for all $x \in B_2(K, r)$, for all $t \geq 2r/\mu^2$, $C(x, t) = C(x, 2r/\mu^2) \in K$.

We make an additional remark, denotes d_C the arc length distance along C , as $\|W\| \leq 1$, we have,

$$\begin{aligned} d_C(x, C(x, 2r/\mu^2)) &= \int_0^{2r/\mu^2} \left| \frac{\partial}{\partial t} C(x, t) \right| dt \\ &\leq \int_0^{2r/\mu^2} \|W(C(x, t))\|_2 dt \\ &\leq 2r/\mu^2 \end{aligned} \tag{11}$$

Thus for all $t \in [0, +\infty[$, $\|x - C(x, t)\|_2 \leq 2r/\mu^2$. Now taking, $F(x, s) = C(x, 2rs/\mu^2)$, provide a deformation retract of $B_2(K, r)$ onto K and the associated retraction $R : x \mapsto F(x, 1)$ verifies, $R(x) \in \overline{B}_2(x, 2r/\mu^2)$ by (11). \square

B Proof of Lemma 2

This Appendix is dedicated to the proof of Lemma 2.

Proof of Lemma 2. Let $f \in S_d(L, \alpha, \mu, R_\mu)$. Let $H \subset \mathcal{F}_{\lambda - 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h h^\alpha}}^c \cap C_h$ and $H' \subset \mathcal{F}_{\lambda + 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h h^\alpha}} \cap C_h$. Remark that $f|_{H'} < \lambda - 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h h^\alpha}$ and $f|_H \geq \lambda + 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h h^\alpha}$. Thus,

$$\mathbb{P}(X_1 \in H | N_h) < (\lambda - 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h h^\alpha})h^d$$

and

$$\mathbb{P}(X_1 \in H' | N_h) \geq (\lambda + 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h h^\alpha})h^d.$$

As, supposing N_h known, $|X \cap H|$ is $(n, \mathbb{P}(X_1 \in H | N_h))$ -binomial and $|X \cap H'|$ is $(n, \mathbb{P}(X_1 \in H' | N_h))$ -binomial we have that,

$$\mathbb{E} \left[\frac{|X \cap H|}{nh^d} | N_h \right] = \frac{\mathbb{P}(X_1 \in H | N_h)}{h^d} < \lambda - 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h h^\alpha}$$

and

$$\mathbb{E} \left[\frac{|X \cap H'|}{nh^d} | N_h \right] = \frac{\mathbb{P}(X_1 \in H' | N_h)}{h^d} \geq \lambda + 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h h^\alpha}.$$

Then, by the choice made for h ,

$$\begin{aligned} \frac{|X \cap H|}{nh^d} &= \mathbb{E} \left[\frac{|X \cap H|}{nh^d} | N_h \right] + \frac{|X \cap H|}{nh^d} - \mathbb{E} \left[\frac{|X \cap H|}{nh^d} | N_h \right] \\ &= \mathbb{E} \left[\frac{|X \cap H|}{nh^d} | N_h \right] + \frac{|X \cap H|}{nh^d} - \mathbb{E} \left[\frac{|X \cap H|}{nh^d} \right] - \mathbb{E} \left[\frac{|X \cap H|}{nh^d} - \mathbb{E} \left[\frac{|X \cap H|}{nh^d} \right] | N_h \right] \\ &\leq \mathbb{E} \left[\frac{|X \cap H|}{nh^d} | N_h \right] + 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h} \sqrt{\frac{\log(1/h^d)}{nh^d}} < \lambda \end{aligned}$$

and similarly,

$$\frac{|X \cap H'|}{nh^d} \geq \mathbb{E} \left[\frac{|X \cap H'|}{nh^d} | N_h \right] - 2\sqrt{3M_{d,L,\alpha,R_\mu}N_h} \sqrt{\frac{\log(1/h^d)}{nh^d}} > \lambda$$

and the proof is complete. \square

C Proof of Lemma 1

This Appendix is dedicated to the proof of Lemma 1.

Proof of Lemma 1. Let $i \in \{1, \dots, l\}$. First suppose that $d = 1$, in this case, ∂M_i is the union of two singletons $\{x_1\} \cup \{x_2\}$. The μ -medial axis is equal here to the set of critical points, which is simply given by $\{(x_1 + x_2)/2\}$. Thus, by Assumption **A3**, we have $1 \geq |x_1 - x_2| \geq 2R_\mu$. Let $x \in M_i$, by Assumption **A1**,

$$2R_\mu (f(x) - L) \leq \int_{x_1}^{x_2} f(u) du.$$

Hence, as f is a density,

$$\int_{x_1}^{x_2} f(u) du \leq 1$$

and

$$f(x) \leq \frac{1}{2R_\mu} + L := M_{1,L,R_\mu}.$$

Now, for $d > 1$, if the μ -medial axis of ∂M_i intersected with M_i is empty, then M_i^c has infinite μ -reach. By Theorem 12 of [Kim et al. \(2020\)](#), this implies that for any $r > 0$, $(M_i^c)^r$ is homotopy equivalent to M_i^c . For $r > \sqrt{d}$, $(M_i^c)^r = [0, 1]^d$, and thus has trivial s -homotopy groups for $s > 0$ but M_i^c contains a non-trivial $d-1$ cycles represented by ∂M_i , which gives a contradiction. Consequently, M_i contains a point x_1 belonging to the μ -medial axis of ∂M_i . By Assumption **A3**, $B_2(x_1, R_\mu/2)$ is contained into M_i . Thus, Assumption **A1** ensures that,

$$V_d(R_\mu/2)(f(x_1) - L(2R_\mu)^\alpha) \leq \int_{B_2(x_1, R_\mu/2)} f(u) du$$

with $V_d(R_\mu/2)$ the volume of the d -dimensional Euclidean ball of radius $R_\mu/2$. As f is a density,

$$\int_{B_2(x_1, R_\mu/2)} f(u) du \leq 1$$

and hence,

$$f(x_1) \leq \frac{1}{V_d(R_\mu/2)} + LR_\mu^\alpha.$$

Now, let $x \in M_i$, we have $\|x - x_1\| \leq \sqrt{d}$, using Assumption **A1** again, it follows that,

$$f(x) \leq f(x_1) + Ld^{\alpha/2} \leq \frac{1}{V_d(R_\mu)} + L(R_\mu^\alpha + d^{\alpha/2}) := M_{d,L,\alpha,R_\mu}.$$

To conclude, it suffices to remark that Assumption **A2** ensures that,

$$\sup_{x \in \bigcup_{i=1}^l M_i} f(x) = \sup_{x \in [0,1]^d} f(x)$$

and thus for all $x \in [0, 1]^d$,

$$f(x) \leq M_{d,L,\alpha,R_\mu}.$$

□

D Proof for the q -tameness

We here prove the claim that the persistence modules we introduced are q -tame and consequently their persistence diagrams are well-defined.

Proposition 4. *Let $f \in S_d(L, \alpha, \mu, R_\mu)$ then f is q -tame.*

Proof. Let $s \in \mathbb{N}$ and $\mathbb{V}_{0,f}$ the persistence module (for the 0-th homology) associated to the sublevel filtration \mathcal{F} , and for fixed levels $\lambda > \lambda'$ let denote $v_\lambda^{\lambda'}$ the associated map. Let $\lambda \in \mathbb{R}$ and $h < \frac{R_\mu}{2}$. By Proposition 2,

$$v_\lambda^{\lambda-L(\sqrt{d}(2+2/\mu^2))^\alpha h^\alpha} = \phi_\lambda \circ \tilde{i}_\lambda, \text{ with } \tilde{i}_\lambda : H_0(\mathcal{F}_\lambda) \rightarrow H_0(\mathcal{F}_\lambda^h) \text{ induced by inclusion.}$$

By assumption **A1** and **A2**, \mathcal{F}_λ is compact. As $[0, 1]^d$ is triangulable, \mathcal{F}_λ is covered by finitely many cells of the triangulation, and so there is a finite simplicial complex K such that $\overline{\mathcal{F}_\lambda} \subset K \subset \mathcal{F}_\lambda^h$. Consequently, \tilde{i}_λ factors through the finite dimensional space $H_0(K)$ and is then of finite rank by Theorem 1.1 of [Crawley-Boevey \(2012\)](#).

Thus, $v_\lambda^{\lambda-L(\sqrt{d}(2+2/\mu^2))^\alpha h^\alpha}$ is of finite rank for all $0 < h < \frac{R_\mu}{2}$. As for any $\lambda > \lambda' > \lambda''$, $v_\lambda^{\lambda''} = v_\lambda^{\lambda''} \circ v_\lambda^{\lambda'}$ we then have that $v_\lambda^{\lambda'}$ is of finite rank for all $\lambda > \lambda'$. Hence, f is q -tame. \square

Proposition 5. *Let $f : [0, 1]^d \rightarrow \mathbb{R}$ then, for all $s \in \mathbb{N}$ $\widehat{\mathbb{V}}_{0,f}$ is q -tame.*

Proof. Let $h > 0$ and $\lambda \in \mathbb{R}$. By construction $\widehat{\mathcal{F}}_\lambda$ is a union of cubes of the regular grid $G(h)$, thus $H_0(\widehat{\mathcal{F}}_\lambda)$ is finite dimensional. Thus $\widehat{\mathbb{V}}_{0,f}$ is q -tame by Theorem 1.1 of [Crawley-Boevey \(2012\)](#). \square