ESAIM: PS 30 (2026) 192–241
https://doi.org/10.1051/ps/2026003

ESAIM: Probability and Statistics
www.esaim-ps.org

PERSISTENCE-BASED MODES INFERENCE

Hugo Henneuse*

Abstract. We address the problem of estimating multiple modes of a multivariate density using persistent homology, a central tool in Topological Data Analysis. We introduce a method based on the preliminary estimation of the H_0-persistence diagram to infer the number of modes, their locations, and the corresponding local maxima. For broad classes of piecewise-continuous functions with geometric control on discontinuity loci, we identify a critical separation threshold between modes, equivalently interpretable in our framework in terms of modes' prominence, below which modes inference is impossible and above which our procedure achieves minimax optimal rates.

Mathematics Subject Classification. 62R40, 62G05, 62C20.

Received May 23, 2025. Accepted January 18, 2026.

1. 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 [1]. Among them, modal approaches in clustering have gathered significant attention [2–8].

The problem of mode(s) inference dates back to Parzen [9] 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 [10–18]. Notably, in [11], the authors 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, i.e, there exists $\alpha, L, C > 0$ such that, for all y in a neighborhood of a mode x:

$$C\|x-y\|^\alpha \leq f(x) - f(y) \leq L\|x-y\|^\alpha. \tag{1.1}$$

Under this assumption, for $\alpha = 2$, they show that Parzen's method achieves minimax rates. In these works, a key question is how to choose the bandwidth for kernel estimation, which often involves knowing the regularity

Keywords and phrases: Modes inference, non-parametric statistics, topological data analysis, persistent homology.

Laboratoire de Mathématiques d'Orsay, 307 Rue Michel Magat, 91400 Orsay, France.

* Corresponding author: hugo.henneuse@universite-paris-saclay.fr

of the density. Interestingly, [19] proposes an adaptive estimator using Lepski's method [20] to overcome this issue.

An alternative, based on histogram estimation of the density, is proposed in [21]. Under (1.1), for all $1 \geq \alpha > 0$, 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 [22] with $\hat{f}$ a kernel estimator. It was proved in [23, 24] that this estimator essentially shares the same asymptotic behavior as Parzen's estimator, and convergence rates are not affected by the maximization over finite samples. In the same direction, in [25], the authors propose a procedure based on KNN-estimation of the density and show, under similar assumptions as [11], that their estimator achieves minimax rates.

For the inference of multiple modes, the available literature is comparatively scarce. Mean-shift and related procedures [4–6, 8, 26, 27] are widely used in practice to identify several modes simultaneously. Despite their empirical success, a rigorous theoretical understanding of their convergence properties remained elusive for a long time. Partial progress was made in [28], which established near-consistency results for mean-shift algorithms. Full consistency was obtained recently in [29], under the assumption that the underlying density is twice differentiable and has a non-degenerate Hessian.

Among the previously cited works on single-mode estimation, [25] also proposes a k-nearest neighbors-based procedure for estimating multiple modes. For this more general problem, their method is shown to be minimax under the assumption that the densities are twice differentiable around the modes, with a negative definite Hessian at each mode (which in particular implies (1.1), for $\alpha = 2$), and that the modes are "r-salient," essentially meaning they are separated by valleys whose depth and width are both controlled by a parameter $r > 0$. Furthermore, [7] extends this approach to capture modal sets, allowing for cases where modes are not single points but entire regions. This broader method is shown to be minimax optimal under the assumption that the densities are globally Hölder continuous, behave as power functions around the modes, and that the modal sets are r-salient. To our knowledge, no other procedure has been demonstrated to possess such broad convergence properties to date.

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 encoding, 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 the birth and death times of topological features. Under proper conditions, looking at H_0-persistence diagram (representing the evolution of connected components) permits identifying 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 [30] and [31]. In a slightly different context, we must also mention [3]. 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, which 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 deciding if a mode is significant or not based on its prominence. Outside the sphere of TDA, this question has also received significant attention, with a rich modal-testing literature [32–40]. Note that persistence diagrams alone do not permit to localize modes, only to estimate their numbers and associated local maxima. Additional information needs to be extracted to infer their positions.

We should also note that, although the formalism is different, the H_0-persistence diagram computed from the superlevel sets of a density is closely related to another fundamental concept in clustering: the cluster tree [41]. While the H_0-persistence diagram records the lifetimes of connected components as the level varies, the cluster tree captures the entire hierarchical organization of these components across all density levels in the form of a tree. The estimation of cluster trees, as well as their application to clustering, has been the focus of substantial

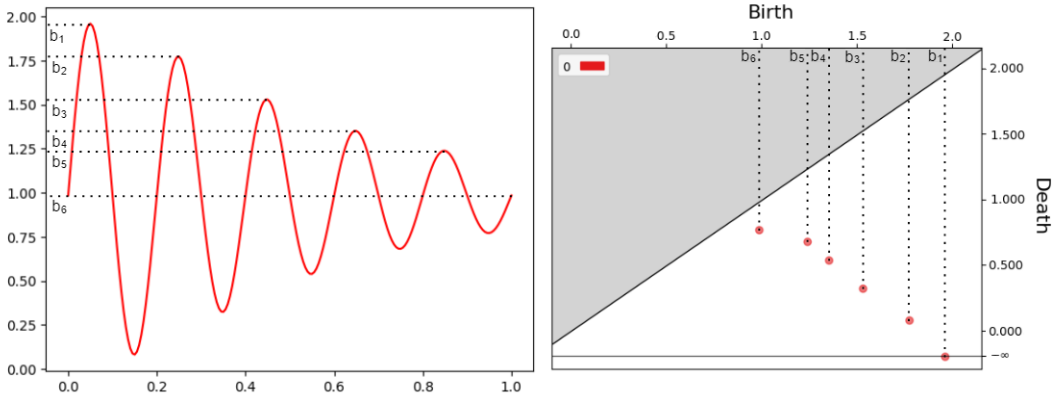


FIGURE 1. 1D illustration of the link between local maxima and H_0 persistence diagram. The birth times $b_1, ..., b_6$ correspond to the local maxima of the function.

research. In particular, several important consistency results have been established under a variety of metric (or pseudo-metric) and regularity assumptions [see *e.g.*, 41–44]. Another interesting method that is spiritually connected to persistence is SiZer [45]. SiZer aims to reveal which features (for instance, clusters or modes) of an estimated density are statistically significant across smoothing scales rather than across levels. By examining how the density estimate evolves as the bandwidth varies, it produces a multiscale map that highlights features that "persist" over a range of scales. In the context of mode detection, this approach can be used to assess the significance of estimated modes.

The estimation of persistence diagrams of a density is an interesting question by itself. A common approach [see *e.g.* 46] in persistence diagram estimation is to use sup-norm (or other popular functional norm) stability theorem [47] 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 for densities (or signals) that are not globally continuous. In [48], a different approach based on image persistence is proposed to move beyond traditional plug-in methods that rely on sup-norm stability. This framework allows for the estimation of persistence diagrams over a broad class of functions, specifically, q-tame bounded functions. However, the generality of this setting comes at a cost: the work does not provide convergence rates or establish formal consistency results, likely due to the inherent looseness of the underlying assumptions. Under the same motivation, for the Gaussian white noise model and the non-parametric regression, in [49] 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 set 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.

The estimation of (super)level sets of a density is an important subject in nonparametric statistics, with an extensive literature. Two popular metrics are commonly used: the volume of the symmetric set difference [50–56] and the Hausdorff distance [54, 57–61]. Both lines of work provide thorough statistical analyses with quantified convergence rates under various assumptions. For the purposes of this paper, we adopt a perspective aligned with studies based on the Hausdorff distance, since a key aspect of our analysis is, in some sense, controlling, in Hausdorff distance, the estimated level sets around the modes at the levels where the modes appear. From this viewpoint, it is interesting to note the similarity between the assumptions introduced for establishing consistency in both modes estimation and level set estimation in the Hausdorff distance. Notably, in [61], they assume (Asm. [A]) that around $\mathcal{F}_\lambda$, the λ-superlevel set of a density f (*i.e.*, the set they wish to estimate), f is α-regular.

That is, there exist constants $\varepsilon, \alpha, L, C > 0$, such that for all y satisfying $|\lambda - f(y)| \leq \varepsilon$:

$$C \min_{x \in \partial \mathcal{F}_\lambda} \|x - y\|_2^\alpha \leq |\lambda - f(y)| \leq L \min_{x \in \partial \mathcal{F}_\lambda} \|x - y\|_2^\alpha.$$

Assuming that f has a global maxima attained uniquely at some point x and that $\mathcal{F}_\lambda$ is the first non-empty superlevel set (as λ decreases), we have $\mathcal{F}_\lambda = \{x\}$, and the previous condition reduces to

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

for all y satisfying $|\lambda - f(y)| \leq \varepsilon$, which is very similar to (1.1). The framework we introduce in the present work also involves a similar condition (see Asms. **A1** and **A4**).

Contribution

We propose a new framework for mode inference that combines geometric conditions from the TDA literature with standard analytical assumptions by introducing a new class of piecewise-continuous functions, denoted as $S_d(L, \alpha, \mu, R_\mu, C, h_0)$. These classes contain (L, α)-piecewise-Hölder-continuous densities, with discontinuity loci having geometric complexity controlled by μ and R_μ, using the μ-reach, a geometric measure introduced in [62]. We furthermore require that the densities within these classes satisfy (1.1) on h_0-neighborhoods of the modes. This provides control over both mode separation and prominence, playing a role analogous to the r-saliency condition introduced in [25] and [7]. It is worth noting that the regularity conditions we assume are globally weaker than those imposed in [7] and [29], as we do not assume global continuity. They are also locally weaker than those in [25], as we allow non-differentiability and even discontinuities around the modes.

In this context, we propose a procedure based on H_0-persistence diagrams and "rough" super level sets estimation to infer the number of modes, their locations, and associated local maxima. The procedure can be outlined as:

- Estimate the superlevel sets of the density using histograms, and apply a thickening step (depending on the parameter μ, see Sect. 4). This yields a collection of "rough" superlevel set estimators $\hat{\mathcal{F}}_\lambda$, for $\lambda \in \mathbb{R}$. The thickening step ensures the correct recovery of the connectivity of the level sets.
- Compute the associated H_0-persistence diagram $\widehat{\mathrm{dgm}(f)}$.
- Construct $\overline{\mathrm{dgm}(f)}$ by removing the points in $\widehat{\mathrm{dgm}(f)}$ with lifetime shorter than δ, where δ is chosen depending on some parameters of the model we consider.
- The number of points $\hat{k}$ in $\overline{\mathrm{dgm}(f)}$ is our estimator of the number of modes.
- For each $(\hat{b}_i, \hat{d}_i) \in \overline{\mathrm{dgm}(f)}$, we can associate a connected component $\hat{C}_i$ of $\hat{\mathcal{F}}_{\hat{b}_i}$ (see Sect. 4.2). Take any $\hat{x}_i \in \hat{C}_i$. The collection $\{\hat{x}_i, i \in \{1, ..., \hat{k}\}\}$ is our estimator of the modes and $\{\hat{b}_i, i \in \{1, ..., \hat{k}\}\}$ our estimator of the associated local maxima.

We study the convergence properties of the proposed procedure over the classes $S_d(L, \alpha, \mu, R_\mu, C, h_0)$. More precisely, we identify thresholds on mode separation and prominence, both of which can be formalized in our framework as a threshold condition on the parameter h_0. Below this threshold, mode detection is impossible (Prop. 4.5); above it, our procedure recovers, with high probability, the exact number of modes and estimates their locations and associated maxima at minimax rates (Thms. 4.3 and 4.4).

These results are of significant interest for several reasons. First, they establish minimax statistical guarantees for our procedure within a broad framework. Notably, this framework accommodates densities with discontinuities, possibly near or at the modes, cases that typically challenge standard approaches such as mean-shift, as we demonstrate through numerical experiments. Secondly and more broadly, while the assumption that modes must be sufficiently prominent and sufficiently separated is central in the literature concerning modes estimation, for instance [7, 25], which formalized this as a r-saliency condition, the quantification of "how much is sufficient" has, to the best of our knowledge, not yet been addressed. Our results, in particular the identification of the

threshold on h_0, provide precise insights into the minimal separation and prominence needed for consistent modes estimation.

Furthermore, in the process of establishing our results on mode inference, we also derive convergence guarantees for our estimators of the H_0 persistence diagram (Props. 4.1, A.2, and Cor. A.1), thereby extending the results obtained in our earlier work [49]. These findings are of independent interest, as they demonstrate that consistent estimation of H_0 persistence diagrams is possible under significantly weaker geometric conditions than those previously required. In particular, the proposed method, based on thickened histograms, can recover consistently persistence diagrams in cases where the plug-in histogram estimator from our earlier work fails.

The paper is organized as follows. Section 2 reviews the geometric and topological background relevant to our work. Section 3 introduces the formal framework under study. Section 4 presents our procedure along with the main theoretical results. Section 5 contains the proofs of these results. Section 6 provides numerical illustrations. Additional results and the proofs of technical lemmas are deferred to the Appendix.

2. Background

This section provides the necessary background to follow this paper.

2.1. Distance function, generalized gradient and μ-reach

In this section, we recall some concepts from geometric measure theory involved in this work. For a set $K \subset [0,1]^d$, we denote $\overline{K}$ its adherence and ∂K its boundary. Let $K \subset [0,1]^d$ be 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 $\Theta_K(x)$ be the center of the unique smallest ball enclosing $\Gamma_K(x)$ (see Fig. 2) the **generalized gradient function** $\nabla_{d_K}(x)$ is then defined as,

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

We can now introduce the notion of μ-reach [62] that will be used to measure and control the geometric complexity of discontinuity sets (see Asm. **A3**, Sect. 3). Let $K \subset [0,1]^d$ be a compact set and $0 \leq \mu \leq 1$, its **μ-reach** is defined by:

$$\operatorname{reach}_\mu(K) = \inf \left\{ r \mid \inf_{d_K^{-1}(r)} \|\nabla_{d_K}\|_2 < \mu \right\}. \tag{2.1}$$

The μ-reach is positive for a large class of sets. For example, all piecewise linear compact sets have positive μ-reach (for some $\mu > 0$). For $\mu = 1$, this simply corresponds to the reach, a standard curvature measure, introduced in [63] and used in our earlier work [49] to control the geometry of the discontinuities. In our context, using the μ-reach instead of the reach to measure the complexity of the discontinuities allows considering significantly wider classes of irregular densities, tolerating corners and multiple points (*i.e.*, self-intersections) in their discontinuities set.

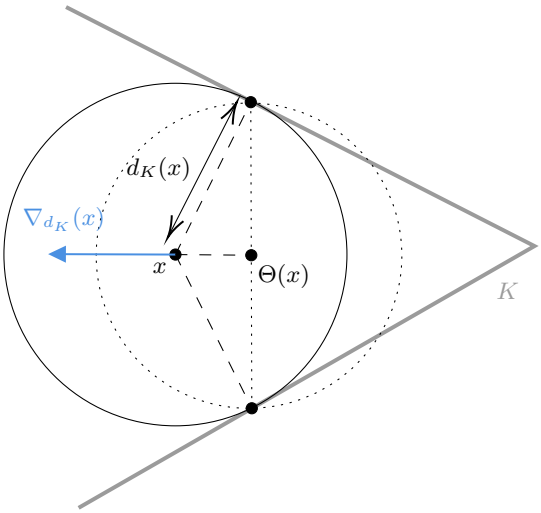


FIGURE 2. 2D example with 2 closest points.

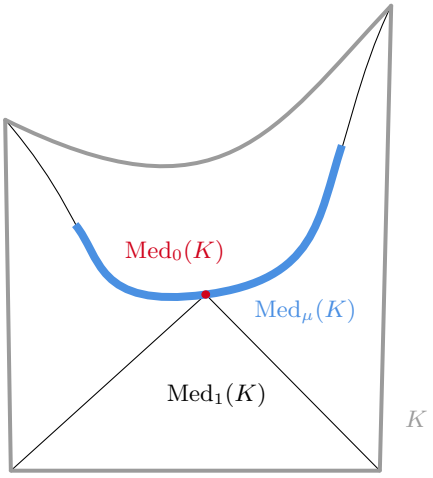


FIGURE 3. Illustration of μ-medial axis for a set K. In black is represented the 1-medial axis, in red the 0-medial axis and in blue the μ-medial axis for a small $0 < \mu < 1/2$.

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:

$$\mathrm{Med}_\mu(K) = \left\{ x \in [0,1]^d \quad | \quad \|\nabla_{d_K}(x)\|_2 < \mu \right\}.$$

The μ-reach of K is then equal to $d_2(K, \overline{\mathrm{Med}_\mu(K)})$ (see illustration in Fig. 3), where, for two sets $A, B \subset \mathbb{R}^d$, we write $d_2(A,B) = \inf_{a\in A} \inf_{b\in B} \|a-b\|_2$ for their Euclidean distance.

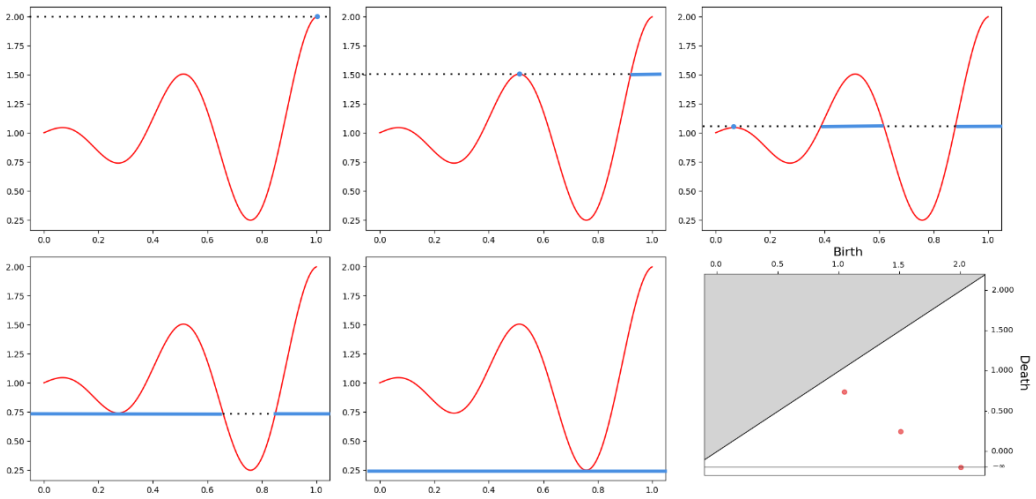


FIGURE 4. Super level sets filtration of $f(x) = x\cos(4\pi x)$ over $[0, 1]$ and the associated H_0-persistence diagram.

2.2. Filtration, persistence module and H_0-persistence diagram

In this section, we briefly present 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, which describes the evolution of connectivity. For a broader overview and visual illustrations of persistent homology, we recommend [64]. For detailed and rigorous constructions, see [65]. Additionally, since the construction discussed here involves (singular) homology, the reader can refer to [66].

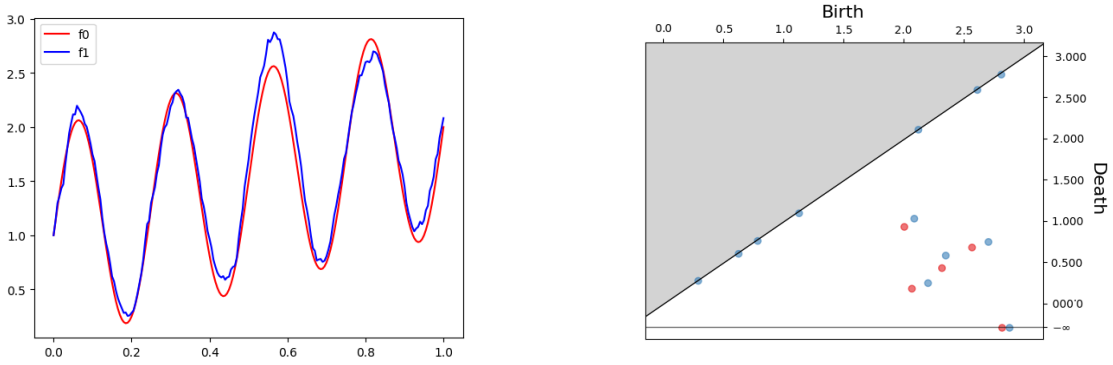
The typical filtration that we will consider in this paper is, for a function $f : \mathbb{R}^d \to \mathbb{R}$, the family of superlevel sets $(\mathcal{F}_\lambda)_{\lambda\in\mathbb{R}} = \left(f^{-1}([\lambda, +\infty[)\right)_{\lambda\in\mathbb{R}}$. The associated family of homology groups of degree 0, $\mathbb{V}_{f,0} = (H_0(\mathcal{F}_\lambda))_{\lambda\in\mathbb{R}}$, equipped with $v_\lambda^{\lambda'}$ the linear map 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, $H_0(.)$ is the singular homology functor in degree 0 with coefficients in the field $\mathbb{Z}/2\mathbb{Z}$. Hence, $H_0(\mathcal{F}_\lambda)$ is a vector space.

If, for all $\lambda > \lambda' \in \mathbb{R}$, $\text{rank}(v_\lambda^{\lambda'})$ 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. We encourage the reader to look at [65] for details. Furthermore, the algebraic structure of a such persistence module $\mathbb{V}$ can be summarized by a collection $\{(b_i, d_i), i \in I\} \subset \overline{\mathbb{R}}^2$, which defines the **persistence diagram**, denoted dgm($\mathbb{V}$). Following previous remarks, for H_0-persistent homology associated to the superlevel set filtration, b_i corresponds to the birth time of a connected component, d_i to its death time and $b_i - d_i$ to its lifetime (see Fig. 4).

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 $\Delta = \{(x, x) \text{ s.t. } x \in \overline{\mathbb{R}}\}$. The addition of Δ allows comparing diagrams with different cardinalities.



(A) graphs of f_0 and f_1 (B) persistence diagrams of f_0 and f_1

FIGURE 5. 1D illustration of stability theorems.

Now, we present the algebraic stability theorem for the bottleneck distance. This theorem was the key for proving upper bounds in [49], which will also be employed in this paper. This theorem relies on interleaving between modules. Let $\mathbb{V}$ and $\mathbb{W}$ be two persistence modules. $\mathbb{V}$ and $\mathbb{W}$ are said to be **ε-interleaved** if there exist two families of linear maps $\phi = (\phi_\lambda)_{\lambda\mathbb{R}}$ and $\psi = (\psi_\lambda)_{\lambda\mathbb{R}}$ (which we will refer to as morphisms between persistence modules) where $\phi_\lambda : \mathbb{V}_\lambda \to \mathbb{W}_{\lambda-\varepsilon}$, $\psi_\lambda : \mathbb{V}_\lambda \to \mathbb{W}_{\lambda-\varepsilon}$, and for all $\lambda > \lambda^{'}$ the following diagrams commute,

$$\begin{array}{ccc}
\mathbb{V}_\lambda & \xrightarrow{v_\lambda^{\lambda'}} & \mathbb{V}_{\lambda'} \\
{\scriptstyle\phi_\lambda}\downarrow & & \downarrow{\scriptstyle\phi_{\lambda'}} \\
\mathbb{W}_{\lambda-\varepsilon} & \xrightarrow[w_{\lambda-\varepsilon}^{\lambda'-\varepsilon}]{} & \mathbb{W}_{\lambda'-\varepsilon} \\
\mathbb{V}_\lambda & \xrightarrow{v_\lambda^{\lambda'-2\varepsilon}} & \mathbb{V}_{\lambda-2\varepsilon} \\
& {\scriptstyle\phi_\lambda}\searrow \quad \nearrow{\scriptstyle\psi_{\lambda-\varepsilon}} & \\
& \mathbb{W}_{\lambda-\varepsilon} &
\end{array}
\qquad
\begin{array}{ccc}
\mathbb{W}_\lambda & \xrightarrow{w_\lambda^{\lambda'}} & \mathbb{W}_{\lambda'} \\
{\scriptstyle\psi_\lambda}\downarrow & & \downarrow{\scriptstyle\psi_{\lambda'}} \\
\mathbb{V}_{\lambda-\varepsilon} & \xrightarrow[v_{\lambda-\varepsilon}^{\lambda'-\varepsilon}]{} & \mathbb{V}_{\lambda'-\varepsilon} \\
\mathbb{W}_\lambda & \xrightarrow{w_\lambda^{\lambda'-2\varepsilon}} & \mathbb{W}_{\lambda-2\varepsilon} \\
& {\scriptstyle\psi_\lambda}\searrow \quad \nearrow{\scriptstyle\phi_{\lambda-\varepsilon}} & \\
& \mathbb{V}_{\lambda-\varepsilon} &
\end{array}
\tag{2.2}$$

Theorem [47], "algebraic stability". *Let $\mathbb{V}$ and $\mathbb{W}$ be two q-tame persistence modules. If $\mathbb{V}$ and $\mathbb{W}$ are ε-interleaved then,*

$$d_b\left(\mathrm{dgm}(\mathbb{V}), \mathrm{dgm}(\mathbb{W})\right) \leq \varepsilon.$$

A direct corollary of this result, proved earlier in special cases [67, 68], is the sup-norm stability of persistence diagrams (illustrated in Fig. 5). We insist that this is a strictly weaker result than algebraic stability.

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

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

As highlighted in the introduction, 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.

3. Framework

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

Regularity assumptions on the densities are inspired by the one considered in [49] (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 discontinuity sets, tolerating multiple points and corners. We consider the following assumptions over f:

- **A1.** f is a piecewise (L, α)-Hölder-continuous probability density, *i.e.* there exists $M_1, ..., M_l$ disjoint open sets of $[0,1]^d$ such that:

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

and, for all $i \in \{1, ..., l\}$, $f|_{M_i}$ is (L, α)-Hölder-continuous, i.e,

$$|f(x) - f(y)| \leq L\|x - y\|_2^{\alpha}, \quad \forall x, y \in M_i.$$

- **A2**. f is upper semi-continuous, *i.e.*, for all $x_0 \in [0,1]^d$,

$$\limsup_{x \to x_0} f(x) \leq f(x_0),$$

and for all $x_0 \in [0,1]^d$ there exists $i \in \{1, ..., l\}$ and a sequence $(y_n)_{n\in\mathbb{N}}$ of M_i, converging to x_0 such that:

$$\lim_{n \to +\infty} f(y_n) = f(x_0).$$

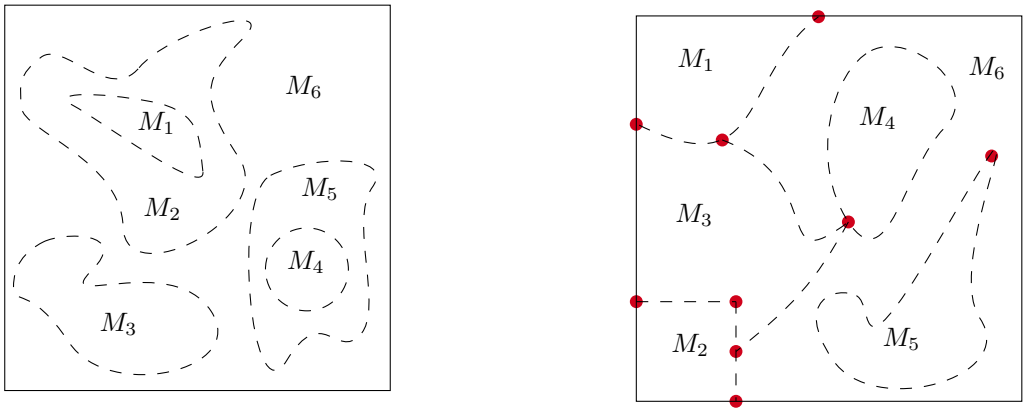
In this context, two signals that differ only on a null set are statistically indistinguishable. However, persistent homology is sensitive to pointwise irregularities: two such signals may have persistence diagrams that are arbitrarily far apart. Assumption **A2** prevents this pathology by requiring that, for every $x \in \bigcup_{i=1}^{l} \partial M_i$, the value $f(x)$ coincides with the supremum of $\lim_{n\to+\infty} f(x_n)$ over all sequences belonging to some M_j, $j \in \{1, ..., l\}$, and converging to x. Importantly, note that this condition does not enforce continuity. For example, consider $f : [0,1] \to \mathbb{R}$ defined by $f(x) = 2\,\mathbf{1}_{x\in[0,1/2]}$. The domain splits into two regular pieces, $[0, 1/2[$ and $]1/2, 1]$, with a single discontinuity at $x = 1/2$. For all $x \in [0,1]$, we have $f(x) \leq 2 = f(1/2)$ thus in particular f is upper semi-continuous and

$$\lim_{x \to 1/2^-} f(x) = 2 = f(1/2),$$

thus, Assumption **A2** holds. To illustrate the necessity of **A2**, define $g : [0,1] \to \mathbb{R}$ by $g(x) = 2\,\mathbf{1}_{x\in[0,1/2]} + \mathbf{1}_{x=1/2}$. The function g violates Assumption **A2**, since

$$\lim_{x \to 1/2^-} g(x) = 2 \neq 3 = g(1/2) \quad \text{and} \quad \lim_{x \to 1/2^+} g(x) = 0 \neq 3 = g(1/2).$$

Its persistence diagram (for the superlevel-set filtration) is $\mathrm{dgm}(g) = \{(3, -\infty)\}$, whereas f differs from g only at the single point $x = 1/2$, yet $\mathrm{dgm}(f) = \{(2, -\infty)\}$. Thus, without a boundary-value convention such as Assumption **A2**, two densities that differ only on a null set may yield different persistence diagrams, making consistent estimation impossible.



(A) Positive (1−)reach (B) Positive μ−reach (μ small)

FIGURE 6. (a) displays a partition M_1,..., M_6 such that $\text{reach}_1(\partial M_1 \cup ... \cup \partial M_6) > 0$. (b) displays a partition M_1,..., M_6 such that $\text{reach}_1(\partial M_1 \cup ... \cup \partial M_6) = 0$ (in red are highlighted problematic points) but for sufficiently small $\mu > 0$, $\text{reach}_\mu(\partial M_1 \cup ... \cup \partial M_6) > 0$.

- **A3.** There is $\mu \in]0, 1]$ and $R_\mu > 0$ such that for all $I \subset \{1, ..., l\}$

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

Here reach_μ denotes the μ-reach (defined by (2.1) in Sect. 2). This can be thought of as a geometric characterization of the regularity of the discontinuity set. In a similar spirit, in [49], we used the reach [63] to control the geometry of the discontinuities. As highlighted in Section 2, the reach coincides with the μ-reach considered here in the special case $\mu = 1$. As previously explained, this particular case imposes several important constraints on the discontinuity set: for instance, it cannot contain self-intersections (*i.e.*, multiple points) or corners. By considering $\mu < 1$, we can encompass a much broader class of discontinuity sets, including those that exhibit self-intersections and corners (as illustrated in Fig. 6). In deed, the combination of Assumptions **A3**, **A2** 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.

The class of functions satisfying **A1**, **A2** and **A3** is denoted $S_d(L, \alpha, \mu, R_\mu)$. As proved in Appendix F, densities in $S_d(L, \alpha, \mu, R_\mu)$ have well-defined persistence diagrams. We prove in this work (Prop. 4.1) 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 (local) mode x of f and $y \in [0, 1]^d$ such that $\|x - y\|_2 \leq h_0$,

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

This assumption is common in the context of modes inference [7, 11, 21, 25]. It has several implications.

First, it ensures that the modes are well separated, by a distance strictly greater than h_0 and thus, in particular, that local maxima are strict (*i.e.*, modal sets are singletons). If this were not the case, there would exist distinct modes x and y with $\|x - y\|_2 \leq h_0$. Assume without loss of generality that

$f(x) \leq f(y)$. Then Assumption **A4** at x implies

$$0 < C\|x-y\|_2^\alpha \leq f(x) - f(y) \leq 0,$$

a contradiction.

Secondly, note that it also implies that f is not too flat around modes and provides a lower bound on their "proeminences". More precisely, let $\delta = Ch_0^\alpha$ and let $\mathrm{dgm}(f)$ denote the H_0-persistence diagram of f. We claim that:

$$\inf_{(b,d)\in \mathrm{dgm}(f)} (b-d) \;\geq\; \delta.$$

To see this, recall that the lifetime $b-d$ of any point $(b,d) \in \mathrm{dgm}(f)$ is bounded below by the minimal gap between the time at which a connected component appears in the superlevel-set filtration and the time at which it subsequently merges with another connected component. Birth times coincide with the values of local maxima. Since modal sets are singletons, each connected component at its birth consists of exactly one mode and remains alive at least until this mode becomes connected to another one. Let x be a mode of height $b = f(x)$, and let C_λ denotes the connected component of $\mathcal{F}_\lambda := f^{-1}([\lambda, +\infty))$ that contains x, for $\lambda \leq b$. Modes are at a distance of at least h_0, and each connected component of the filtration contains at least a mode. Therefore, x cannot be connected with another mode in $\mathcal{F}_\lambda$ as long as C_λ remains contained in $\{u \in [0,1]^d, \|x-u\|_2 < h_0\}$. Let $\lambda \in]b-\delta,\, b]$ and suppose there exists $y \in C_\lambda \setminus \{u \in [0,1]^d, \|x-u\|_2 < h_0\}$. Then, there exists $z \in C_\lambda \cap \{u \in [0,1]^d, \|x-u\|_2 = h_0\}$, else we could separate $C_\lambda \cap \{u \in [0,1]^d, \|x-u\|_2 < h_0\}$ from $C_\lambda \setminus \{u \in [0,1]^d, \|x-u\|_2 < h_0\}$ by considering the two disjoint open sets $\{u \in [0,1]^d, \|x-u\|_2 < h_0\}$ and $\{u \in [0,1]^d, \|x-u\|_2 > h_0\}$, which would contradict the fact that C_λ is a connected component. Then, by Assumption **A4**, we have:

$$\delta = Ch_0^\alpha = C\|x-z\|_2^\alpha \leq f(x) - f(z) \leq b - \lambda < b - (b-\delta) = \delta,$$

which is a contradiction. Thus, for every $\lambda \in]\, b-\delta,\, b\,]$, $C_\lambda \subset \{u \in [0,1]^d, \|x-u\|_2 < h_0\}$. This shows that the connected component born at x remains isolated and does not merge with any other component for all $\lambda \in]b-\delta,\, b]$. As it holds for all modes, we have:

$$\inf_{(b,d)\in \mathrm{dgm}(f)} (b-d) \;\geq\; \inf_{(b,d)\in \mathrm{dgm}(f)} b - (b-\delta) = \delta,$$

which proves the claim. Thus, **A4** ensures that $\mathrm{dgm}(f)$ contains no point with lifetime less than δ. In other words, the prominence of each mode is lower bounded by δ. Consequently, it avoids having arbitrarily small oscillations and undetectable modes. To this extent, it is comparable to the "r-saliency" assumption from [25] and [7], in the sense that it ensures that modes are separated by "valleys" of controlled depth and width. Then, it also controls the number of modes that f can admit, as we are working on the compact set $[0,1]^d$, the maximal number of modes is of order $\asymp (1/h_0)^d$.

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

4. Procedures descriptions and results

In this section, we present our estimation strategy in detail and state several associated convergence results. Section 4.1 focuses on the preliminary step of estimating H_0 persistence diagrams, while Section 4.2 is devoted to the inference of multiple modes.

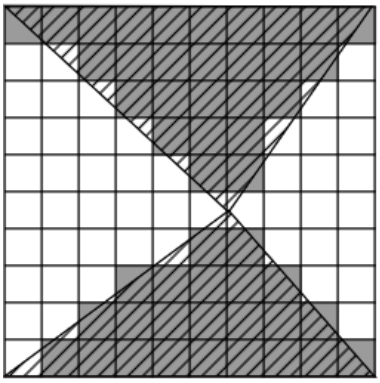


FIGURE 7. λ-Superlevel cubical approximation for f the function defined as 0 outside the hatched area and K inside (for arbitrarily large K). The histogram approximation fails to identify the connectivity of the two triangles for at least all $3K/4 < \lambda < K$.

4.1. H_0-persistence diagram estimation

This section presents our procedure for persistence diagram estimation, which serves as the foundation of our mode inference method. We prove its consistency and quantify its convergence rates over the classes $S_d(L, \alpha, \mu, R_\mu)$, which will be instrumental in later sections to establish our main results on mode inference. Unlike the setting of [49], where the positive reach assumption allows for direct plug-in estimation *via* histograms, the weaker assumptions considered here render this approach insufficient, as illustrated in Figure 7. The presence of self-intersections, sharp angles, or intersections with the boundary of $[0,1]^d$ in the piece boundaries may prevent histogram approximations from correctly recovering the connectivity of the superlevel set filtration, even in the absence of noise. To address these additional challenges, we introduce an extra thickening step, similarly to the methods proposed in [69] and Chapter 5 of [70].

Let us 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 $\mathfrak{C}_h$ the collection of all the closed hypercubes of side h composing $G(h)$. We define, for all $\lambda \in \mathbb{R}$, the λ-superlevel estimator,

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

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

$$\hat{v}_\lambda^{\lambda'} : H_0\left(\widehat{\mathcal{F}}_\lambda\right) \to H_0\left(\widehat{\mathcal{F}}_{\lambda'}\right)$$

the map induced by the inclusion $\widehat{\mathcal{F}}_\lambda \subset \widehat{\mathcal{F}}_{\lambda'}$. Now, we 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 $(\widehat{v}_\lambda^{\lambda'})_{\lambda>\lambda'}$ and $\widehat{\mathrm{dgm}(f)}$ the associated H_0-persistence diagram. This diagram is well-defined, as we prove in Appendix F 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 5.3, we choose h such that:

$$h^\alpha > \sqrt{\frac{\log(1/h^d)}{nh^d}} \tag{4.1}$$

In particular, we can choose,

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

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, it 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.

Consistency. We now present a key result that will play a central role in our proofs concerning mode inference: Proposition 4.1, which quantifies, in probabilistic terms, the convergence rates achieved over the class $S_d(L,\alpha,\mu,R_\mu)$.

Proposition 4.1. *Let $h \asymp (\log(n)/n)^{\frac{1}{d+2\alpha}}$. There exists $\tilde{c}_0$ and $\tilde{c}_1$ (depending on L, α, μ and R_μ) such that, for all $A \geq 0$,*

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

We believe that this result is also of independent interest, as it significantly extends the results of [49] to a much broader class of signals. To complement this, we provide in Appendix A a corollary establishing consistency guarantees in expectation (Cor. A.1), along with matching minimax lower bounds (Prop. A.2), proving that the obtained rates are optimal.

These rates match those of Theorem 1 in [49], derived in the slightly different context of nonparametric regression. They also coincide with those from Corollary 4.4 in [46], which were established for nonparametric regression over Hölder spaces. This demonstrates that even when the signal is only piecewise Hölder with discontinuity loci having a positive μ-reach, the H_0 persistence diagram can still be estimated at the same rates as in the fully Hölder-continuous setting.

The proof of Proposition 4.1, presented in Section 5.1, follows the same general strategy as the proof of Theorem 1 in [49], with several key refinements. In particular, under the weaker μ-reach assumption, we lose some of the geometric properties provided by the reach assumption, which were extensively leveraged in our previous work.

4.2. 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 $\widehat{\mathrm{dgm}(f)}$ gives a procedure that allows recovering consistently the number of modes and the associated values of f. Then, using this "regularized" persistence

diagram, the estimation of the modes can be derived from the filtration $\hat{\mathcal{F}}$. Let $\{x_1, ..., x_k\}$ the set of modes of f and $\{m_1, ..., m_k\}$ the associated maxima. In this setting, an estimator of the number of modes is given by:

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

where points in the diagram are counted with multiplicity, that is, if a pair (b,d) appears twice in $\widehat{\mathrm{dgm}(f)}$ and satisfies $b - d > \delta/2$, then it is counted twice. Furthermore, we denote

$$\left\{ \left(\hat{b}_i, \hat{d}_i \right) \right\}_{1 \leq i \leq \hat{k}} = \left\{ (b,d) \in \widehat{\mathrm{dgm}\,(f)}, b - d > \delta/2 \right\},$$

the collection of birt-death pairs of $\widehat{\mathrm{dgm}\,(f)}$ satisfying $b - d > \delta/2$ (also considered with multiplicity). We then define an estimator of the complete list of local maxima of f, by:

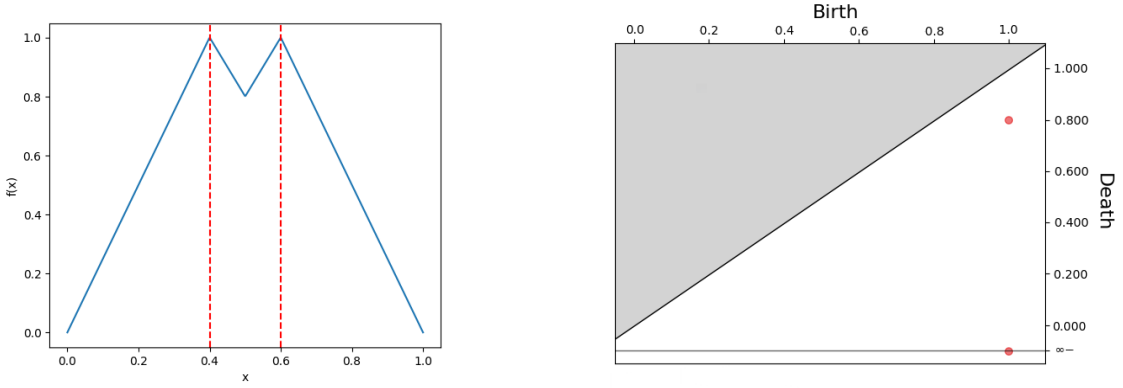
$$\hat{m} = \left\{ \hat{m}_1, \hat{m}_2, ..., \hat{m}_{\hat{k}} \right\} = \left\{ \hat{b}_1, ..., \hat{b}_{\hat{k}} \right\}.$$

Again, note that the $\hat{m}_1, ..., \hat{m}_{\hat{k}}$ are also counted with multiplicity, *i.e.*, it is possible that there are $1 \leq i \neq j \leq \hat{k}$ such that $\hat{m}_i = \hat{m}_j$.

Now, to define our estimator of the modes, we need to evaluate the lifetime (or prominence) of the connected components that appear in the filtration $\hat{\mathcal{F}}$, so that components with short lifetimes can be discarded. We know that each birth–death pair $(b,d) \in \widehat{\mathrm{dgm}(f)}$ corresponds to a connected component of $\hat{\mathcal{F}}$ that is created at level b and merges, at level d, with another component that was born no later than b. Our goal is therefore to associate each birth–death pair in $\widehat{\mathrm{dgm}(f)}$ with a corresponding connected component in the filtration $\hat{\mathcal{F}}$. However, when multiple components are born at the same level and then merge, this association is not necessarily unique. For example, in Figure 8, the signal produces two birth–death pairs in its persistence diagram for the superlevel-set filtration, namely $(1, -\infty)$ and $(1, 0.8)$. These correspond to the components $\{0.4\}$ and $\{0.6\}$, both born at $\lambda = 1$. A priori, either component could be matched to either birth–death pair because both components appear simultaneously in the filtration and then merge. Since they are born at the same time, neither is "older" than the other, so there is no natural way to decide which one "dies" when they merge.

To remove this ambiguity, the idea is to fix an (arbitrary) ordering of the connected components that are born at the same filtration value. For instance, in the example illustrated by Figure 8, among the components born at $\lambda = 1$, we may declare that $\{0.4\}$ comes before $\{0.6\}$. Once such an ordering is chosen, we define the death time of any connected component as the first value of the filtration parameter at which it merges with another component that either was born strictly earlier, or has the same birth time but appears earlier in the chosen order. Furthermore, if no finite filtration value satisfies this condition, we say that the death time of the connected component is $-\infty$. In the example of Figure 8, both $\{0.4\}$ and $\{0.6\}$ are born at $\lambda = 1$. By fixing the order $\{0.4\} \prec \{0.6\}$, we treat $\{0.4\}$ as the "older" component. As the filtration parameter decreases, $\{0.6\}$ merges with $\{0.4\}$ and since $\{0.4\}$ has priority in the ordering, this merging event determines the death time of $\{0.6\}$, yielding the finite value 0.8 in the persistence diagram, thus associating $\{0.6\}$ to $(1, 0.8)$. In contrast, $\{0.4\}$ never merges with a component born earlier or with higher priority, so its lifetime is infinite, giving the pair $(1, -\infty)$. Observe that reversing the order to $\{0.6\} \prec \{0.4\}$ would result in $(1, -\infty)$ being associated with $\{0.6\}$ and $(1, 0.8)$ with $\{0.4\}$.

Let us formalize this idea. Let us denote by $\{(\check{b}_i, \check{d}_i)\}_{i \in I}$, the birth–death pairs in the diagram $\widehat{\mathrm{dgm}(f)}$; that is, the points of $\widehat{\mathrm{dgm}(f)}$. For each birth time $\check{b}_i$, let $\mathcal{C}(\check{b}_i)$ denote the collection of connected components that appear in the filtration $\hat{\mathcal{F}}$ at time $\check{b}_i$. More precisely, for all $i \in I$, an element $C \in \mathcal{C}(\check{b}_i)$ is a connected component of $\hat{\mathcal{F}}_{\check{b}_i}$ such that $C \cap \hat{\mathcal{F}}_\lambda = \emptyset$, for all $\lambda > \check{b}_i$. We can then define the notion of birth and death time of a connected component as follows.



(A) Graph of f

(B) Persistence diagram of f (superlevel set)

FIGURE 8. An example illustrating the ambiguity in associating birth–death pairs with connected components: there is no canonical matching.

Definition 4.2. For each $i \in I$, we say that the elements of $\mathcal{C}(\check{b}_i)$ have birth time $\check{b}_i$. We fix an arbitrary order on this set:

$$\mathcal{C}(\check{b}_i) = \{\tilde{C}_{i,1}, \dots, \tilde{C}_{i,p_i}\}.$$

For a component $\tilde{C}_{i,p} \in \mathcal{C}(\check{b}_i)$, we define its death time as the highest value $d_{i,p} < \check{b}_i$ of the filtration parameter such that, in $\hat{\mathcal{F}}_{d_{i,p}}$, the component $\tilde{C}_{i,p}$ becomes connected to either:

(i) a component born strictly earlier, that is, any element of $\mathcal{C}(\check{b}_j)$ with $\check{b}_j > \check{b}_i$, or
(ii) a component born at the same time but with higher priority in the fixed ordering, that is, some $\tilde{C}_{i,q} \in \mathcal{C}(\check{b}_i)$ with $q < p$.

and again, if no finite filtration value satisfies this condition, we say that the death time of $\tilde{C}_{i,p}$ is $-\infty$. The lifetime of $\tilde{C}_{i,p}$ is then simply defined as $\check{b}_i - d_{i,p}$.

With this definition in hand, observe that the collection of birth–death pairs produced by this construction coincides exactly with the set of points in $\widehat{\mathrm{dgm}(f)}$ (this is proved formally in Lem. G.1, in Appendix G). Now denote by $\hat{\mathcal{C}}$ the collection of connected components appearing in $\widehat{\mathcal{F}}$ whose lifetimes are strictly greater than $\delta/2$. The birth–death pairs associated with the components of $\hat{\mathcal{C}}$ are precisely those $(b, d) \in \widehat{\mathrm{dgm}(f)}$ satisfying $b - d > \delta/2$. Thus, $\hat{\mathcal{C}}$ contains $\hat{k}$ components, denoted $\hat{C}_1, \dots, \hat{C}_{\hat{k}}$, such that for each $1 \leq i \leq \hat{k}$, $\hat{C}_i$ has birth time $\hat{m}_i = \hat{b}_i$ and death time $\hat{d}_i$.

Then, we can estimate the modes by taking, for all $i \in \{1, ..., \hat{k}\}$, any $\hat{x}_i$ in $\hat{C}_i$ and define the collection of estimated modes by:

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

Note that $\hat{\mathcal{C}}$ (and thus $\hat{x}$) depends on the choice of ordering of the components in each $\mathcal{C}(\check{b}_i)$. A connected component associated to a birth–death pair with lifetime greater than $\delta/2$ for one ordering (and thus included in $\hat{\mathcal{C}}$ for that ordering) may be associated to a birth–death pair with lifetime smaller than $\delta/2$ for another ordering (and therefore not included in $\hat{\mathcal{C}}$ for that ordering). For instance, in the example of Figure 8, if we remove all birth–death pairs with lifetime smaller than 0.5, only $(1, -\infty)$ remains. As explained above, depending on the chosen ordering, the retained connected component can be either $\{0.4\}$ or $\{0.6\}$.

Nevertheless, we show that any such collection $\hat{\mathcal{C}}$, regardless of the choice of ordering, yields a consistent estimator $\hat{x}$ of the modes of f (*i.e.*, in any case, the estimated modes are close to the true ones). We show as well that $\hat{m}$ is a consistent estimator of the associated local maxima. The convergence rates achieved by these estimators are given by the following theorem.

Theorem 4.3. *Let $f \in S_d(L, \alpha, \mu, R_\mu, C, h_0)$, $h \asymp (\log(n)/n)^{\frac{1}{d+2\alpha}}$. There exist $\check{c}_0$, $\check{c}_1$, $\check{c}_2$ and $\check{c}_3$ (depending only on L ,α, μ, R_μ, and C) such that, for all $A \geq 0$ with probability at least $1 - \check{c}_0 \exp(-\check{c}_1 \min(A^{2\alpha}, A^2)) - \check{c}_2 \exp(-\check{c}_3 (h_0/h)^{2\alpha})$,*

$$k = \hat{k}$$

and for all $i \in \{1, ..., k\}$, 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

$$|\hat{m}_i - m_i| \leq Ah^\alpha.$$

Several conclusions can be drawn from this theorem. First, for fixed parameters L, α, μ, R_μ, C and h_0, Theorem 4.3 implies that, as $\lim_{n\to\infty} \frac{h_0}{h} = +\infty$, for sufficiently large n, we can infer with high probability the exact number of modes, locate them at a rate of $O\left((\log(n)/n)^{\frac{1}{2\alpha+d}}\right)$, and estimate the associated local maxima at a rate of $O\left((\log(n)/n)^{\frac{\alpha}{2\alpha+d}}\right)$. Secondly, this result also addresses a perhaps more interesting question: how far apart or how prominent must the modes be for our procedure to identify them (and their associated maxima) with high probability? Equivalently, how large should h_0 be, or how large should δ be, in comparison to n? In this regard, Theorem 4.3 indicates that the critical value of h_0 for our procedure to succeed is of the order $h \asymp (\log(n)/n)^{\frac{1}{2\alpha+d}}$. This critical h_0 can be interpreted as a minimal separation distance between modes that allows for their identification by our procedure. In particular, if $h_0 \geq Bh$, for some constant $B > 0$, we have, with probability at least $1 - \check{c}_0 \exp(-\check{c}_1 \min(A^{2\alpha}, A^2)) - \check{c}_2 \exp(-\check{c}_3 B^{2\alpha})$,

$$\hat{k} = k \text{ and, for all } 1 \leq i \leq k, \ \|\hat{x}_i - x_i\|_\infty \leq Ah \text{ and } |\hat{m}_i - m_i| \leq Ah^\alpha.$$

This also gives that, if $h = o(h_0)$ (and other parameters are fixed):

$$\lim_{n\to+\infty} \mathbb{P}\left(\left\{\hat{k} = k\right\} \cap \left(\bigcap_{i=1}^{k} (\{\|\hat{x}_i - x_i\|_\infty \leq Ah\} \cap \{|\hat{m}_i - m_i| \leq Ah\})\right)\right) \geq 1 - \check{c}_0 \exp(-\check{c}_1 \min(A^{2\alpha}, A^2)).$$

Note that, as previously emphasized, due to Assumption **A4** (recalling that $\delta = Ch_0^\alpha$), this condition can be equivalently reformulated in terms of a critical prominence δ. In this case, it implies a critical prominence of order $h^\alpha \asymp (\log(n)/n)^{\frac{\alpha}{2\alpha+d}}$. It can also be interpreted in terms of the maximal number of modes, which is then of order $\asymp (n/\log(n))^{\frac{d}{2\alpha+d}}$.

In addition to the above result, one may also wonder about the more realistic setting where δ is unknown. To address this, we propose in Appendix B a penalization procedure for selecting δ, and adapt our method accordingly. Under slightly stronger threshold assumption, requiring that $h_0 \gg (n/\log(n))^{\frac{1}{2(2\alpha+d)}}$ or equivalently $\delta \gg (n/\log(n))^{\frac{\alpha}{2(2\alpha+d)}}$, we show that, with this selected value, the number of modes can still be recovered with high probability, and both the locations of the modes and their associated local maxima can be estimated at the same rates as in Theorem 4.3.

In light of Theorem 4.3, several questions arise regarding its optimality: first, the optimality in terms of the convergence rates for modes locations and their associated local maxima; second, the optimality concerning the tightness of the critical h_0.

Regarding the convergence rates, it is noteworthy that, up to a logarithmic factor, the rates we have obtained align with the minimax rates for single-mode estimation established in [21]. By adapting their proof, we show in Theorem 4.4 that similar lower bounds also hold in our setting. Moreover, we refine this analysis further by recovering an additional $\log(1/h_0)$ factor.

Theorem 4.4. *For sufficiently small h_0, there are two strictly positive constants A and B (independent of n) and a family of densities of $S_d(L, \alpha, \mu, R_\mu, C, h_0)$ with (pairwise) modes separated by $A(\log(1/h_0)/n)^{\frac{1}{d+2\alpha}}$ and local maxima separated by $B(\log(1/h_0)/n)^{\frac{\alpha}{d+2\alpha}}$ that cannot be distinguished with more than probability $1/5$ for a sample size of n.*

In the case where the parameters L, α, μ, R_μ, C and h_0 are fixed and n tends to $+\infty$, our result does not provide any additional insight beyond Theorem 2 of [21]. This confirms that, in this setting, our proposed procedure is minimax (up to logarithmic factors). However, when considering h_0 that depend on n, our result establishes a slightly stronger lower bound, which permits to explain the $\log(n)$ factor appearing in the convergence rates of Theorem 4.3. More precisely, consider the class $S_d(L, \alpha, \mu, R_\mu, C, h_0)$ with $h_0 = n^{-\gamma}$ for some absolute constant $\gamma < 1/(d + 2\alpha)$. we have:

$$h \asymp (\log(n)/n)^{\frac{1}{d+2\alpha}} \ll h_0 = n^{-\gamma}.$$

Thus, Theorem 4.3 guarantees that, over $S_d(L, \alpha, \mu, R_\mu, C, h_0)$, our estimator is consistent and achieves the rates $O((\log n/n)^{\frac{1}{d+2\alpha}})$ for mode estimation and $O((\log n/n)^{\frac{\alpha}{d+2\alpha}})$ for the estimation of the associated local maxima. Complementarily, Theorem 4.4 states that, over the same class, no estimator can achieve faster rates than:

$$(\log(1/h_0)/n)^{\frac{1}{d+2\alpha}} = (\gamma \log(n)/n)^{\frac{1}{d+2\alpha}} \asymp (\log(n)/n)^{\frac{1}{d+2\alpha}},$$

for mode estimation, and

$$(\log(1/h_0)/n)^{\frac{\alpha}{d+2\alpha}} = (\gamma \log(n)/n)^{\frac{\alpha}{d+2\alpha}} \asymp (\log(n)/n)^{\frac{\alpha}{d+2\alpha}},$$

for the estimation of the associated local maxima. This establishes the minimaxity of our estimator over $S_d(L, \alpha, \mu, R_\mu, C, h_0)$ in this setting, fully accounting for the logarithmic factors appearing in Theorem 4.3.

Now, regarding the sharpness of the critical threshold on h_0, a natural question arises: can we still recover some information about the modes below this threshold, albeit potentially at slower rates? Proposition 4.5 shows that this is essentially impossible. More precisely, even if we only aim to recover the number of modes, disregarding their locations, Proposition 4.5 establishes that, below the threshold (up to logarithmic factors), this task is infeasible. More precisely, we show that we can construct, in this case, a bimodal density and an unimodal density that are statistically indistinguishable with high probability. Interestingly, we can even construct them such that they have sets of modes that are well separated (by some constant B, independent of n) in the sense of the Hausdorff distance, that is defined for two compact sets $A, B \subset \mathbb{R}^d$ by:

$$d_H(A, B) = \max\left\{ \sup_{x\in B} \inf_{y\in A} \|x - y\|_2, \sup_{x\in A} \inf_{y\in B} \|x - y\|_2 \right\}.$$

Proposition 4.5. *There exist strictly positive constants A and B (independent of n), a bimodal density $f_0 \in S_d(L, \alpha, \mu, R_\mu, C, h_0)$ and an unimodal density $f_1 \in S_d(L, \alpha, \mu, R_\mu, C, h_0)$, with $h_0 = n^{\frac{-1}{2\alpha+d}}/A$ such that the sets of modes of f_0 and f_1 are at Hausdorff distance B and f_0 cannot be distinguished from f_1 with more than probability 1/5 for a sample size n.*

Hence, this proposition tells us that, up to a logarithmic factor, the critical h_0 obtained in Theorem 4.3 is optimal, as if h_0 is smaller, even estimating the number of modes (without considering their locations or the associated maxima) becomes hopeless.

5. Proofs of main results

This section is devoted to the proofs of Proposition 4.1, Theorem 4.3, Theorem 4.4, and Proposition 4.5.

5.1. Proof of Proposition 4.1

This section is dedicated to the proof of Proposition 4.1. The strategy is to construct an interleaving between the persistence modules $\mathbb{V}_{f,0}$ and $\widehat{\mathbb{V}}_{f,0}$, allowing us to invoke the algebraic stability theorem [47]. To this end, we construct two morphisms, $\overline{\psi} : \mathbb{V}_{f,0} \to \widehat{\mathbb{V}}_{f,0}$ and $\overline{\phi} : \widehat{\mathbb{V}}_{f,0} \to \mathbb{V}_{f,0}$, satisfying the interleaving condition (2.2). The overall construction follows the blueprint used in the proof of Theorem 1 in [49], but it requires substantial modifications. In particular, the previous approach relied heavily on the positive reach assumption, which guaranteed the existence of a continuous projection map near the boundaries of the regular pieces. In the more general setting considered here, this assumption no longer holds, necessitating several non-trivial adaptations. To address this, we rely on key geometric properties of sets with positive μ-reach, as established in Lemma 3.1 of [71] and Theorem 12 of [72]. These results allow us to circumvent the lack of a projection map near the piece boundaries. The remainder of this section details the construction. Note that several lemmas we introduce here will also be reused later to prove Theorem 4.3.

5.1.1. 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 Lemma 5.2, a slightly weaker double inclusion.

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

Lemma 5.1. *Let $f \in S_d(L, \alpha, \mu, R_\mu)$, there exists a constant $\kappa = \kappa(d, L, \alpha, R_\mu)$ depending only on d, L, α and R_μ, such that $\|f\|_\infty \leq \kappa$.*

Denote N_h the random variable defined by:

$$N_h = \frac{\max_{H \in \mathfrak{C}_h} \left| \frac{|X \cap H|}{nh^d} - \mathbb{E}\left[\frac{|X \cap H|}{nh^d}\right] \right|}{\sqrt{3\kappa \frac{\log(1/h^d)}{nh^d}}}$$

Lemma 5.2. *Let $f \in S_d(L, \alpha, \mu, R_\mu)$ and $h \asymp (\log(n)/n)^{\frac{1}{d+2\alpha}}$. For n sufficiently large such that $h < \mu R_\mu/\sqrt{d}$, for all $\lambda \in \mathbb{R}$, we have:*

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

This proposition follows from Lemma 3.1 from [71] and the following lemma, whose proof can be found in Appendix D.

Lemma 5.3. *Let $f \in S_d(L, \alpha, \mu, R_\mu)$ and $h > 0$ satisfying (4.1). Let*

$$H \subset \mathcal{F}^c_{\lambda - 2\sqrt{3\kappa}N_h h^\alpha} \cap \mathfrak{C}_h$$

and

$$H^{'} \subset \mathcal{F}_{\lambda+2\sqrt{3\kappa}N_h h^\alpha} \cap \mathfrak{C}_h.$$

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 5.4. *[71], Lemma 3.1. Let $K \subset [0,1]^d$ be a compact set and let $\mu > 0, r > 0$ be such that $r < \text{reach}_\mu(K)$. For any $x \in \overline{B}_2(K,r) \backslash K$, we have:*

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

Proof of Lemma 5.2. We begin by proving the lower inclusion, let

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

Without loss of generality, let us suppose $x \in M_i$, or $x \in \partial M_i$ and $\limsup_{z \in M_i \to x} f(z) \geq f(x)$. If,

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

then, $H_{x,h}$, the hypercube of $\mathfrak{C}_{h,\lambda}$ containing x is included in $\overline{M}_i$. Assumption **A1** implies that for all $y \in B_2(x, \sqrt{d}h) \cap M_i$ we have:

$$|f(x) - f(y)| \leq L|x - y\|^\alpha \leq L(\sqrt{d}h)^\alpha,$$

thus, $y \in \mathcal{F}_{\lambda+2\sqrt{3\kappa}N_h h^\alpha}$. Now, for all $y \in B_2(x, \sqrt{d}h) \cap \overline{M}_i$, there exists a sequence $(y_n)_{n\in\mathbb{N}}$ of points of $B_2(x, \sqrt{d}h) \cap M_i$ such that $\lim_{n\to\infty} y_n = y$. As, for all $n \in \mathbb{N}$, we have shown that $y_n \in \mathcal{F}_{\lambda+2\sqrt{3\kappa}N_h h^\alpha}$, by Assumption **A2**, it follows that:

$$f(y) \geq \limsup_{n\to+\infty} f(y_n) \geq \lambda + 2\sqrt{3\kappa}N_h h^\alpha.$$

Hence $y \in \mathcal{F}_{\lambda+2\sqrt{3\kappa}N_h h^\alpha}$. Thus, as $H_{x,h} \subset B_2(x, \sqrt{d}h) \cap \overline{M}_i$, the combination of Assumption **A1** and **A2** ensures $H_{x,h} \subset \mathcal{F}_{\lambda+2\sqrt{3\kappa}N_h h^\alpha}$. Hence, it follows from Lemma 5.3 that $H_{x,h} \in \mathfrak{C}_{h,\lambda}$ and consequently $x \in \widehat{\mathcal{F}}_\lambda$. Else, as $\sqrt{d}h/\mu < R_\mu$, under Assumption **A3**, by Lemma 5.4, 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 $H_{y,h}$ be the closed hypercube of $\mathfrak{C}_h$ 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, again, Assumption **A1** and **A2** ensure that:

$$H_{y,h} \subset \mathcal{F}_{\lambda+2\sqrt{3\kappa}N_h h^\alpha}.$$

Then, Lemma 5.3 gives $H_{y,h} \in \mathfrak{C}_{h,\lambda}$ 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

$$x \in \left(\mathcal{F}^{\sqrt{d}h}_{\lambda-2\sqrt{3\kappa}N_h h^\alpha}\right)^c$$

and $H_{x,h}$ be the hypercube of $\mathfrak{C}_h$ containing x. We then have $H_{x,h} \subset \mathcal{F}^c_{\lambda-2\sqrt{3\kappa}N_h h^\alpha}$. Hence, Lemma 5.3 gives that:

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

and thus,

$$\bigcup_{H\in\mathfrak{C}_{h,\lambda}} H \subset \mathcal{F}^{\sqrt{d}h}_{\lambda-2\sqrt{3\kappa}N_h h^\alpha}.$$

Consequently,

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

and the proof is complete. □

5.1.2. Ingredient 2: geometry of the thickened level sets

The lower inclusion of Lemma 5.2 gives directly a morphism $\overline{\psi} : \mathbb{V}_{f,0} \to \widehat{\mathbb{V}}_{f,0}$. But the upper inclusion is not sufficient to provide similarly $\overline{\phi} : \widehat{\mathbb{V}}_{f,0} \to \mathbb{V}_{f,0}$. To overcome this issue, in Lemma 5.5, we exploit Assumption **A3** to construct morphisms from $(H_0(\mathcal{F}^r_\lambda))_{\lambda\in\mathbb{R}}$ 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 Lemma 5.2 will give us our morphism $\overline{\phi}$.

Lemma 5.5. *Let $f \in S_d(L, \alpha, \mu, R_\mu)$. For all $h < R_\mu/\sqrt{d}$, there exists a morphism ϕ such that, for all $\lambda \in \mathbb{R}$,*

$$\begin{array}{ccc} H_0\left(\mathcal{F}_\lambda\right) & \longrightarrow & H_0\left(\mathcal{F}_{\lambda-L\left(\sqrt{d}(2+2/\mu^2)\right)^\alpha h^\alpha}\right) \\ & \searrow \quad \nearrow_{\phi_\lambda} & \\ & H_0\left(\mathcal{F}^h_\lambda\right) & \end{array} \tag{5.1}$$

is a commutative diagram (unspecified maps 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 5.6. A subspace A of X is called a **deformation retract** of X if there is a continuous $F : X \times [0,1] \to X$ 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 5.7. *[72], Theorem 12. Let $K \subset [0,1]^d$ be a compact set, for all $0 \leq r < \text{reach}_\mu(K)$, $\overline{B}_2(K,r)$ retracts by deformation onto K and the associated retraction $R : \overline{B}_2(K,r) \to K$ satisfies for all $x \in \overline{B}_2(K,r)$, $R(x) \in \overline{B}_2(x, 2r/\mu^2) \cap K$.*

The first part of the claim is Theorem 12 from [72], and the second part follows easily from their construction. Elements of proof can be found in Appendix C.

Proof of Lemma 5.5. Any connected component B of $\mathcal{F}_\lambda^h$ contains at least a connected component A of $\mathcal{F}_\lambda$. Suppose that B contains A and $A^{'}$ two disjoint connected components of $\mathcal{F}_\lambda$, then there exist $x \in A$ and $y \in A^{'}$ such that $\|x - y\|_\infty \leq 2h$. Suppose that $x \in \overline{M}_i$, $y \in \overline{M}_j$, $\lim_{z \in M_i \to x} f(z) = f(x)$ and $\lim_{z \in M_j \to y} f(z) = f(y)$, $i, j \in \{1, ..., l\}$.

Let

$$R : \overline{B}_2(\partial M_i \cup \partial M_j, \sqrt{d}h) \to \partial M_i \cup \partial M_j$$

the deformation retraction of $\overline{B}_2(\partial M_i \cup \partial M_j, \sqrt{d}h)$ onto $\partial M_i \cup \partial M_j$ given by Lemma 5.7 under Assumption **A3**. We use this deformation retraction to construct a continuous path from x to y included in $\overline{M}_i \cup \overline{M}_j$. To do so, consider, for all $z \in [x,y]$ the function:

$$\gamma(z) = \begin{cases} z & \text{if } z \in \overline{M}_i \cup \overline{M}_j \\ R(z) & \text{else} \end{cases}$$

Observe that $\gamma(x) = x$, $\gamma(y) = y$ and, as R is continuous and $R(z) = z$ for all $z \in \partial M_i \cup \partial M_j$, γ is a continuous path between x and y. Furthermore, as $R(z) \subset \partial M_i \cup \partial M_j$ for all $z \in \overline{B}_2(\partial M_i \cup \partial M_j, \sqrt{d}h)$, we have that $\gamma([x,y])$ is a continuous path between x and y included in $\overline{M}_i \cup \overline{M}_j$. Now, by Lemma 5.7 and as $\|x - y\|_\infty \leq 2h$, we know that, for all $z \in [x,y] \setminus (\overline{M}_i \cup \overline{M}_j)$,

$$R(z) \in \overline{B}_2(z, 2\sqrt{d}h/\mu^2) \subset \overline{B}_2(x, \sqrt{d}h(2 + 2/\mu^2)).$$

and, for all $z \in [x,y] \cap (\overline{M}_i \cup \overline{M}_j)$,

$$z \in \overline{B}_2(x, 2\sqrt{d}h).$$

Thus,

$$\gamma([x,y]) \subset \overline{B}_2(x, \sqrt{d}h(2 + 2/\mu^2)) \cap (\overline{M}_i \cup \overline{M}_j).$$

Assumptions **A1** and **A2** then ensure that $\gamma([x,y]) \subset \mathcal{F}_{\lambda - L((2+2/\mu^2)\sqrt{d}h)^\alpha}$. Consequently, there exists a continuous path between A and $A^{'}$ in $\mathcal{F}_{\lambda - L((2+2/\mu^2)\sqrt{d}h)^\alpha}$, and hence, A and $A^{'}$ are connected in $\mathcal{F}_{\lambda - L((2+2/\mu^2)\sqrt{d}h)^\alpha}$.

We can then properly define,

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

with y belonging to A any connected component of $\mathcal{F}_\lambda$ included in B, the connected component of $\mathcal{F}_\lambda^h$ containing x. By the foregoing, $[y]$ is independent of the choice of such A and such y in A, and Diagram 5.1 commutes. □

5.1.3. Ingredient 3 : concentration of N_h

The two previous ingredients will permit the establishment of 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.

Lemma 5.8. *For all $h > 0$,*

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

Proof of Lemma 5.8. Let $h>0$ and $H\in\mathfrak{C}_h$. We can write,

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

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

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

As $X_1,...,X_n$ are i.i.d., $Y_1^H,...,Y_n^H$ 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 distributions that:

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

Lemma 5.1 then gives that $\mathbb{P}(X_1\in H)\leq h^d\kappa$, hence,

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

By union bounds, it follows,

$$\begin{aligned}\mathbb{P}\left(\bigcup_{H\in\mathfrak{C}_h}\left\{\left|\frac{|X\cap H|}{nh^d}-\mathbb{E}\left[\frac{|X\cap H|}{nh^d}\right]\right|\geq t\right\}\right)&\leq 2\left|\mathfrak{C}_h\right|\exp\left(-\frac{t^2nh^d}{3\kappa}\right)\\&\leq\frac{2}{h^d}\exp\left(-\frac{t^2nh^d}{3\kappa}\right).\end{aligned}$$

□

Lemma 5.8 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 $h\leq 1$,

$$\mathbb{P}\left(N_h>t\right)\leq c_0\exp\left(-c_1t^2\right).$$

5.1.4. Assembling the ingredients

Equipped with Lemmas 5.2, 5.5, and 5.8, we can now proceed to the proof of Proposition 4.1.

Proof of Proposition 4.1. 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 Lemmas 5.2 and 5.5 used in this proof. We denote $k_1=\sqrt{d}+\lceil\sqrt{d}/\mu\rceil$ and $k_2=\sqrt{d}\left(2+2/\mu^2\right)\left(\sqrt{d}+\lceil\sqrt{d}/\mu\rceil\right)$.

First, we construct $\overline{\phi}$. Let $\lambda\in\mathbb{R}$, Lemma 5.3 imply that any connected component A of $\widehat{\mathcal{F}}_\lambda$ intersects a connected component B of $\mathcal{F}_{\lambda-2\sqrt{3\kappa}N_hh^\alpha}$. Now suppose that A intersects 2 such components B_1 and B_2, then, by Lemma 5.2, B_1 and B_2 are connected in $\mathcal{F}^{k_1h}_{\lambda-2\sqrt{3\kappa}N_hh^\alpha}$. Thus, as a consequence of Lemma 5.5 (applied for

$h^{'} = k_1 h$, which requires $h < R_\mu / k_1 \sqrt{d}$), B_1 and B_2 are connected in $\mathcal{F}_{\lambda - \left(2\sqrt{3\kappa} N_h + L k_2^\alpha\right) h^\alpha}$. We can then define the applications,

$$\begin{cases} \overline{\phi}_\lambda : H_0\left(\widehat{\mathcal{F}}_\lambda\right) & \longrightarrow H_0\left(\mathcal{F}_{\lambda - \left(2\sqrt{3\kappa} N_h + L k_2^\alpha\right) h^\alpha}\right) \\ [x] & \longmapsto [y] \end{cases}$$

with y belonging to B a connected component of $\mathcal{F}_{\lambda - 2\sqrt{3\kappa} N_h h^\alpha}$ 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 $\overline{\psi}$, by Lemma 5.2, we have:

$$\mathcal{F}_\lambda \subset \widehat{\mathcal{F}}_{\lambda - \left(2\sqrt{3\kappa} N_h + L k_1^\alpha\right) h^\alpha}$$

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

$$\begin{cases} \overline{\psi}_\lambda : H_0\left(\mathcal{F}_\lambda\right) & \longrightarrow H_0\left(\widehat{\mathcal{F}}_{\lambda - \left(2\sqrt{3\kappa} N_h + L k_1^\alpha\right) h^\alpha}\right) \\ [x] & \longmapsto [x] \end{cases}$$

Denote $K_1 = 2\sqrt{3\kappa} N_h + L k_1^\alpha$ and $K_2 = 2\sqrt{3\kappa} N_h + L k_2^\alpha$. The foregoing ensures the commutativity of the following diagrams:

$$\begin{array}{ccc} H_0\left(\mathcal{F}_\lambda\right) & \xrightarrow{v_\lambda^{\lambda^{'}}} & H_0\left(\mathcal{F}_{\lambda'}\right) \\ \downarrow{\scriptstyle \overline{\psi}_\lambda} & & \downarrow{\scriptstyle \overline{\psi}_{\lambda'}} \\ H_0\left(\widehat{\mathcal{F}}_{\lambda - K_1 h^\alpha}\right) & \xrightarrow{\hat{v}_{\lambda - K_1 h^\alpha}^{\lambda^{'} - K_1 h^\alpha}} & H_0\left(\widehat{\mathcal{F}}_{\lambda' - K_1 h^\alpha}\right) \end{array}$$

$$\begin{array}{ccc} H_0\left(\widehat{\mathcal{F}}_\lambda\right) & \xrightarrow{\hat{v}_\lambda^{\lambda^{'}}} & H_0\left(\widehat{\mathcal{F}}_{\lambda'}\right) \\ \downarrow{\scriptstyle \overline{\phi}_\lambda} & & \downarrow{\scriptstyle \overline{\phi}_{\lambda'}} \\ H_0\left(\mathcal{F}_{\lambda - K_2 h^\alpha}\right) & \xrightarrow{v_{\lambda - K_2 h^\alpha}^{\lambda^{'} - K_2 h^\alpha}} & H_0\left(\mathcal{F}_{\lambda' - K_2 h^\alpha}\right) \end{array}$$

$$\begin{array}{ccc} H_0\left(\mathcal{F}_\lambda\right) & \xrightarrow{v_\lambda^{\lambda + (K_1 + K_2) h^\alpha}} & H_0\left(\mathcal{F}_{\lambda - (K_1 + K_2) h^\alpha}\right) \\ & {\scriptstyle \overline{\psi}_\lambda} \searrow \quad \nearrow {\scriptstyle \overline{\phi}_{\lambda - K_1 h^\alpha}} & \\ & H_0\left(\widehat{\mathcal{F}}_{\lambda - K_1 h^\alpha}\right) & \end{array}$$

$$\begin{array}{ccc} H_0\left(\widehat{\mathcal{F}}_\lambda\right) & \xrightarrow{\hat{v}_\lambda^{\lambda + (K_1 + K_2) h^\alpha}} & H_0\left(\widehat{\mathcal{F}}_{\lambda - (K_1 + K_2) h^\alpha}\right) \\ & {\scriptstyle \overline{\phi}_\lambda} \searrow \quad \nearrow {\scriptstyle \overline{\psi}_{\lambda - K_2 h^\alpha}} & \\ & H_0\left(\mathcal{F}_{\lambda - K_2 h^\alpha}\right) & \end{array}$$

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 [47] that:

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

We then conclude, using the concentration of N_h given in Lemma 5.8.

$$\begin{aligned}
&\mathbb{P}\left(d_b\left(\widehat{\mathrm{dgm}(f)}, \mathrm{dgm}(f)\right) \geq th^\alpha\right) \\
&\leq \mathbb{P}\left(K_1+K_2 \geq t\right) \\
&= \mathbb{P}\left(N_h \geq \frac{t - L\left(k_1^\alpha + k_2^\alpha\right)}{4\sqrt{3\kappa}}\right) \\
&\leq c_0 \exp\left(-c_1\left(\frac{t - L\left(k_1^\alpha + k_2^\alpha\right)}{4\sqrt{3\kappa}}\right)^2\right) \\
&\leq c_0 \exp\left(-\frac{c_1}{48\kappa}t^2\right)\exp\left(c_1\frac{L\left(k_1^\alpha + k_2^\alpha\right)}{24\kappa}t\right)\exp\left(-c_1\left(\frac{L\left(k_1^\alpha + k_2^\alpha\right)}{4\sqrt{3\kappa}}\right)^2\right) \\
&= c_0 \exp\left(-\frac{c_1}{96\kappa}t^2\right)\exp\left(c_1\frac{L\left(k_1^\alpha + k_2^\alpha\right)}{24\kappa}t - \frac{c_1}{96\kappa}t^2\right)\exp\left(-c_1\left(\frac{L\left(k_1^\alpha + k_2^\alpha\right)}{4\sqrt{3\kappa}}\right)^2\right)
\end{aligned}$$

Thus, as for all $t \in \mathbb{R}$,

$$c_1\frac{L\left(k_1^\alpha + k_2^\alpha\right)}{24\kappa}t - \frac{c_1}{96\kappa}t^2 \leq \frac{c_1L^2\left(k_1^\alpha + k_2^\alpha\right)^2}{24\kappa},$$

we have:

$$\begin{aligned}
\mathbb{P}\left(d_b\left(\widehat{\mathrm{dgm}(f)}, \mathrm{dgm}(f)\right) \geq th^\alpha\right) &\leq \mathbb{P}\left(K_1+K_2 \geq t\right) \\
&\leq c_0 \exp\left(-c_1\left(\frac{L\left(k_1^\alpha + k_2^\alpha\right)}{4\sqrt{3\kappa}}\right)^2 + \frac{c_1L^2\left(k_1^\alpha + k_2^\alpha\right)^2}{24\kappa}\right)\exp\left(-\frac{c_1}{96\kappa}t^2\right) \\
&= c_0 \exp\left(c_1\frac{c_1L^2\left(k_1^\alpha + k_2^\alpha\right)^2}{48\kappa}\right)\exp\left(-\frac{c_1}{96\kappa}t^2\right) \qquad (5.2)
\end{aligned}$$

and the result follows. □

5.2. Proof of Theorem 4.3

We now proceed to prove Theorem 4.3. The high-level outline of the proof is the following:

- First, we prove that, with high probability, for all $i \in \{1, ..., k\}$, $\overline{C}_i$, the connected component of $\mathcal{F}^{k_1 h}_{m_i-(K_1+2\sqrt{3\kappa}N_h)h^\alpha}$ containing $C_i = \{x_i\}$, contains a $\hat{C}_i$, i.e $\overline{C}_i$ contains a connected component appearing in the filtration $\hat{\mathcal{F}}$ with associated lifetime exceeding $\delta/2$.
- We show, with high probability, that the birth time associated with $\hat{C}_i$ is close to m_i.
- We use Proposition 4.1, to show that, with high probability, $k = \hat{k}$.
- Finally, we use Assumption **A4** to bound, with high probability, the diameter of $\overline{C}_i$, and thus the distance between x_i and $\hat{x}_i$.

The last two steps are pretty straightforward, but the first two are more technical. The proofs of these first claims are then decomposed into the three following lemmas. Let us define, for all $i \in \{1, ..., k\}$, for all $\lambda < m_i$, $\overline{C}^i_\lambda$ the connected component of $\mathcal{F}^{k_1 h}_\lambda$ containing $C_i = \{x_i\}$ and denote,

$$\tilde{b}_i = \sup \left\{ \lambda \in \mathbb{R} \text{ s.t } \widehat{\mathcal{F}}_\lambda \cap \overline{C}^i_{m_i + (K_1+K_2)h^\alpha - \delta} \neq \emptyset \right\}$$

with k_1, K_1 and K_2 as defined in the proof of Proposition 4.1. We start by establishing upper and lower bounds on $\tilde{b}_i$ through the following lemma.

Lemma 5.9. *Consider the event E_1:*

$$(K_1 + K_2)\, h^\alpha < \delta/8.$$

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

$$m_i - K_1 h^\alpha \leq \tilde{b}_i \leq m_i + 2\sqrt{3\kappa} N_h h^\alpha.$$

Proof. We suppose along the proof that we are under E_1. The lower inclusion of Lemma 5.2 ensures that $x_i \in \widehat{\mathcal{F}}_{m_i - K_1 h^\alpha}$ and thus,

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

Let $\check{C}_i$ the connected component of $\mathcal{F}_{m_i + (K_1+K_2)h^\alpha - \delta}$ containing x_i. By Assumption **A4**, we know that C_i is included in $B_2(x_i, h_0)$. Moreover, $\check{C}_i$ is the unique connected component of $\mathcal{F}_{m_i + (K_1+K_2)h^\alpha - \delta}$ intersecting $B_2(x_i, h_0)$. Indeed, since $B_2(x_i, h_0)$ contains only one mode and since each connected component contains at least one mode, if this were not the case, there would exist another connected component C of $\mathcal{F}_{m_i + (K_1+K_2)h^\alpha - \delta}$ intersecting both $B_2(x_i, h_0)$ and its complement $B_2(x_i, h_0)^c$. Consequently, there would exist a point $z \in C \cap \partial B_2(x_i, h_0)$, which would imply:

$$\delta = C\|x_i - z\|_2^\alpha \leq f(x_i) - f(z) \leq \delta - (K_1 + K_2)h^\alpha < \delta,$$

yielding a contradiction. Furthermore, Assumption **A4** also ensures that for all $x \in \check{C}_i$:

$$C\|x_i - x\|_2^\alpha \leq \delta - (K_1 + K_2)h^\alpha = Ch_0^\alpha - (K_1 + K_2)h^\alpha.$$

Thus,

$$h_0 - \|x_i - x\|_2 \geq (h_0^\alpha - \|x_i - x\|_2^\alpha)^{1/\alpha} \geq \left(\frac{K_1 + K_2}{C}\right)^{1/\alpha} h \geq \left(\frac{2Lk_1^\alpha}{C}\right)^{1/\alpha} h > 2k_1 h.$$

Hence, $d_\infty(\check{C}_i, C) > 2k_1 h$ for any other connected components C of $\mathcal{F}_{m_i + (K_1+K_2)h^\alpha - \delta}$, and thus, we have:

$$\overline{C}^i_{m_i + (K_1+K_2)h^\alpha - \delta} = \check{C}_i^{k_1 h} \subset B_2(x_i, h_0) \tag{5.3}$$

Now, if $\tilde{b}_i > m_i + 2\sqrt{3\kappa} N_h h^\alpha$ then by Assumption **A4**, $\mathcal{F}_{\tilde{b}_i - 2\sqrt{3\kappa} N_h h^\alpha} \cap B_2(x_i, h_0) = \emptyset$. Thus, $\mathcal{F}_{\tilde{b}_i - 2\sqrt{3\kappa} N_h h^\alpha} \cap \overline{C}^i_{m_i + (K_1+K_2)h^\alpha - \delta} = \emptyset$. But, by Lemma 5.3, any connected component of $\widehat{\mathcal{F}}_{\tilde{b}_i}$ intersects $\mathcal{F}_{\tilde{b}_i - 2\sqrt{3\kappa} N_h h^\alpha}$. Hence, we have:

$$\tilde{b}_i \leq m_i + 2\sqrt{3\kappa} N_h h^\alpha.$$

□

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

Lemma 5.10. *Under the event E_1, for all $i \in \{1, ..., k\}$, any connected component $\tilde{C}^i$ of $\widehat{\mathcal{F}}_{\tilde{b}_i}$ intersecting $\overline{C}^i_{m_i+(K_1+K_2)h^\alpha-\delta}$ is included in $\overline{C}_i$.*

Proof. Let $\tilde{C}^i$ a connected component of $\widehat{\mathcal{F}}_{\tilde{b}_i}$ intersecting $\overline{C}^i_{m_i+(K_1+K_2)h^\alpha-\delta}$. Under E_1, Lemma 5.9 and Lemma 5.2 yields:

$$\hat{\mathcal{F}}_{\tilde{b}_i} \subset \mathcal{F}^{k_1 h}_{\tilde{b}_i - 2\sqrt{3\kappa}N_h h^\alpha} \subset \mathcal{F}^{k_1 h}_{m_i-(K_1+2\sqrt{3\kappa}N_h)h^\alpha} \subset \mathcal{F}^{k_1 h}_{m_i+(K_1+K_2)h^\alpha-\delta}. \tag{5.4}$$

Consequently, $\tilde{C}^i$ is included in $\overline{C}^i_{m_i+(K_1+K_2)h^\alpha-\delta}$. Additionally, by Lemma 5.3, $\tilde{C}^i$ intersects $\mathcal{F}_{\tilde{b}_i-2\sqrt{3\kappa}N_h h^\alpha}$. Thus,

$$\tilde{C}^i \cap \left(\mathcal{F}_{\tilde{b}_i-2\sqrt{3\kappa}N_h h^\alpha} \cap \overline{C}^i_{m_i+(K_1+K_2)h^\alpha-\delta}\right) \neq \emptyset. \tag{5.5}$$

Now, note that, under E_1, by Assumption **A4**, there exists a unique connected component of $\mathcal{F}_{m_i-(K_1+2\sqrt{3\kappa}N_h)h^\alpha}$ intersecting $B(x_i, h_0)$ (by the same argument as in the proof of Lemma 5.9, where we showed that there is a unique connected component of $\mathcal{F}_{m_i+(K_1+K_2)h^\alpha-\delta}$ intersecting $B_2(x_i, h_0)$), that is $\overline{C}_i$. Thus,

$$\mathcal{F}_{m_i-(K_1+2\sqrt{3\kappa}N_h)h^\alpha} \cap B(x_i, h_0) \subset \overline{C}_i.$$

It then follows from Lemma 5.9 and (5.3) that:

$$\mathcal{F}_{\tilde{b}_i-2\sqrt{3\kappa}N_h h^\alpha} \cap \overline{C}^i_{m_i+(K_1+K_2)h^\alpha-\delta} \subset \mathcal{F}_{m_i-(K_1+2\sqrt{3\kappa}N_h)h^\alpha} \cap B(x_i, h_0) \subset \overline{C}_i. \tag{5.6}$$

Hence, combining (5.6) and (5.5), $\tilde{C}^i$ intersects $\overline{C}_i$. Consequently, by (5.4), $\tilde{C}^i$ is included in $\overline{C}_i$. □

Note that, under E_1, by (5.4), we have:

$$\overline{C}_i \subset \overline{C}^i_{m_i+(K_1+K_2)h^\alpha-\delta}. \tag{5.7}$$

Then, by definition of $\tilde{b}_i$, a consequence of Lemma 5.10 is that, under E_1, for all $i \in \{1, ..., k\}$, any connected component $\tilde{C}^i$ of $\widehat{\mathcal{F}}_{\tilde{b}_i}$ intersecting $\overline{C}^i_{m_i+(K_1+K_2)h^\alpha-\delta}$ has associated birth time $\tilde{b}_i$. It now suffices to prove that, for all $i \in \{1, ..., k\}$, there exists such a $\tilde{C}^i$ with associated death time $\tilde{d}_i$ satisfying $\tilde{b}_i - \tilde{d}_i > \delta/2$.

Lemma 5.11. *Under E_1, for all $i \in \{1, ..., k\}$, there exists a connected component $\tilde{C}^i$ of $\widehat{\mathcal{F}}_{\tilde{b}_i}$ intersecting $\overline{C}^i_{m_i+(K_1+K_2)h^\alpha-\delta}$ with an associated death time $\tilde{d}_i$ satisfying:*

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

Proof. Let $\tilde{C}^i$ be the oldest connected component of $\hat{\mathcal{F}}_{\tilde{b}_i}$ intersecting $\overline{C}^i_{m_i+(K_1+K_2)h^\alpha-\delta}$, according to definition 4.2, for a fixed (but arbitrary) ordering on $\mathcal{C}(\tilde{b}_i)$. For all $\lambda \leq \tilde{b}_i$, we denote $\tilde{C}^i_\lambda$ the connected component of $\widehat{\mathcal{F}}_\lambda$ containing $\tilde{C}^i$. Combining Lemma 5.2 and Lemma 5.9, we have that, for all $\tilde{b}_i - 3\delta/4 \leq \lambda \leq \tilde{b}_i$, under E_1,

$$\widehat{\mathcal{F}}_\lambda \subset \mathcal{F}^{k_1 h}_{\lambda-2\sqrt{3\kappa}N_h h^\alpha} \subset \mathcal{F}^{k_1 h}_{\tilde{b}_i-3\delta/4-2\sqrt{3\kappa}N_h h^\alpha} \subset \mathcal{F}^{k_1 h}_{m_i-3\delta/4-(K_1+K_2)h^\alpha} \subset \mathcal{F}^{k_1 h}_{m_i+(K_1+K_2)h^\alpha-\delta}.$$

Thus, under E_1, for all $\tilde{b}_i - 3\delta/4 \leq \lambda \leq \tilde{b}_i$,

$$\tilde{C}^i_\lambda \subset \overline{C}^i_{m_i+(K_1+K_2)h^\alpha-\delta}.$$

Now, by definition of $\tilde{C}_i$, this implies that under E_1, for all $\tilde{b}_i - 3\delta/4 \leq \lambda \leq \tilde{b}_i$, in $\tilde{C}^i_\lambda$, $\tilde{C}_i$ is not connected with any older connected components appearing in $\widehat{\mathcal{F}}$. Hence, the lifetime of $\tilde{C}_i$ is lower bounded by $\tilde{b}_i - (\tilde{b}_i - 3\delta/4) = 3\delta/4 > \delta/2$, which proves Lemma 5.11. □

Equipped with Lemma 5.9, 5.10 and 5.11, we can now prove Theorem 4.3.

Proof of Theorem 4.3. Let $f \in S_d(L, \alpha, \mu, R_\mu, C, h_0)$. From the proof of Proposition 4.1, we have that:

$$d_\infty\left(\widehat{\mathrm{dgm}(f)}, \mathrm{dgm}(f)\right) \leq (K_1 + K_2)\, h^\alpha. \tag{5.8}$$

Then, under E_1, it follows that $\hat{k} = k$. Recalling that $\delta = Ch_0^\alpha$, it follows from (5.2), that there exist A_1 and B_1 such that E_1 occurs with probability at least $1 - A_1 \exp(-B_1(h_0/h)^{2\alpha})$.

The combination of Lemma 5.10 and 5.11 ensures that, under E_1, for all $i \in \{1, ..., k\}$ there exists a connected component $\tilde{C}^i$ appearing in $\widehat{\mathcal{F}}$ at time $\tilde{b}_i$, included in $\overline{C}^i$, such that its associated lifetime exceeds $\delta/2$. Thus, by definition of $\hat{\mathcal{C}}$, for all $i \in \{1, ..., k\}$, $\tilde{C}^i \in \hat{\mathcal{C}}$. Combining (5.3) and (5.7), we have that, under E_1, $\overline{C}_1 \subset B(x_1, h_0)$, ..., $\overline{C}_k \subset B(x_k, h_0)$. Thus, as by Assumption **A4** the modes are at least at a distance h_0 from each other, $\overline{C}_1, ..., \overline{C}_k$ are disjoint. Consequently $\tilde{C}^1, ..., \tilde{C}^k$ are also disjoint. Hence, as $k = \hat{k}$, we have:

$$\hat{\mathcal{C}} = \left\{\hat{C}_1, ..., \hat{C}_k\right\} = \left\{\tilde{C}^1, ..., \tilde{C}^k\right\}$$

and

$$\hat{m} = \{\hat{m}_1, ..., \hat{m}_k\} = \left\{\tilde{b}_1, ..., \tilde{b}_k\right\}$$

As, for all $i \in \{1, ..., k\}$, $x_i \in C_i \subset \overline{C}_i$ and (by Lem. 5.10), $\hat{x}_i \in \hat{C}_i = \tilde{C}^i \subset \overline{C}_i$, to bound the distance between $\hat{x}_i$ and x_i, it suffices to bound the diameter of $\overline{C}_i$, defined by:

$$\mathrm{diam}\left(\overline{C}_i\right) = \sup_{x,y \in \overline{C}_i} \|x - y\|_\infty.$$

Recall that, by (5.3) and (5.7), we have that $\overline{C}_i \subset B_2(x_i, h_0)$, thus $\overline{C}_i \subset \left(\mathcal{F}_{m_i-(K_1+2\sqrt{3\kappa}N_h)h^\alpha} \cap B_2(x_i, h_0)\right)^{k_1 h}$. Otherwise, there would exist some $y \in \mathcal{F}_{m_i-(K_1+2\sqrt{3\kappa}N_h)h^\alpha} \setminus B_2(x_i, h_0)$ such that $B_\infty(y, k_1 h) \cap \overline{C}_i \neq \varnothing$, which would imply that y and $\overline{C}_i$ are connected in $\mathcal{F}^{k_1 h}_{m_i-(K_1+2\sqrt{3\kappa}N_h)h^\alpha}$ and hence that $y \in \overline{C}_i$, contradicting the inclusion $\overline{C}_i \subset B(x_i, h_0)$. Now, by Assumption **A4**, we have for all $y \in \mathcal{F}_{m_i-(K_1+2\sqrt{3\kappa}N_h)h^\alpha} \cap B(x_i, h_0)$:

$$\|x_i - y\|_\infty \leq \|x_i - y\|_2 \leq \left(\frac{f(x_i) - f(y)}{C}\right)^{1/\alpha} \leq \frac{(K_1 + 2\sqrt{3\kappa}N_h)^{1/\alpha}}{C^{1/\alpha}} h.$$

Thus,

$$\begin{aligned}\operatorname{diam}\left(\mathcal{F}_{m_i-(K_1+2\sqrt{3\kappa}N_h)h^\alpha}\cap B(x_i,h_0)\right) &= \sup_{z,y\in\mathcal{F}_{m_i-(K_1+2\sqrt{3\kappa}N_h)h^\alpha}\cap B(x_i,h_0)}\|z-y\|_\infty\\ &\le 2\sup_{y\in\mathcal{F}_{m_i-(K_1+2\sqrt{3\kappa}N_h)h^\alpha}\cap B(x_i,h_0)}\|x_i-y\|_\infty\\ &\le \frac{2(K_1+2\sqrt{3\kappa}N_h)^{1/\alpha}}{C^{1/\alpha}}h.\end{aligned}$$

It then follows that:

$$\operatorname{diam}(\overline{C}_i)\le \operatorname{diam}\left(\mathcal{F}_{m_i-(K_1+2\sqrt{3\kappa}N_h)h^\alpha}\cap B(x_i,h_0)\right)+2k_1h\le\left(\frac{2(K_1+2\sqrt{3\kappa}N_h)^{1/\alpha}}{C^{1/\alpha}}+2(\sqrt{d}+\lceil\sqrt{d}/\mu\rceil)\right)h.$$

And thus,

$$\|x_i-\hat{x}_i\|_\infty\le\left(\frac{2(K_1+2\sqrt{3\kappa}N_h)^{1/\alpha}}{C^{1/\alpha}}+2(\sqrt{d}+\lceil\sqrt{d}/\mu\rceil)\right)h. \tag{5.9}$$

Consider the event E_2:

$$\left(\frac{2(K_1+2\sqrt{3\kappa}N_h)^{1/\alpha}}{C^{1/\alpha}}+2(\sqrt{d}+\lceil\sqrt{d}/\mu\rceil)\right)\le A$$

and the event $E_3: K_1\le A$.

As $2\sqrt{3\kappa}N_h\le K_2$ and $2(\sqrt{d}+\lceil\sqrt{d}/\mu\rceil)=2k_1\le\frac{2}{L^{1/\alpha}}(K_1+K_2)^{1/\alpha}$, by (5.2), there exist A_2, B_2, such that E_2 occurs with probability $1-A_2\exp(-B_2A^{2\alpha})$. And as $K_1\le K_1+K_2$, by (5.2) there also exist A_3, B_3, such that E_3 occurs with probability $1-A_3\exp(-B_3A^2)$. Hence, it follows from Lemma 5.9 and (5.9) that, with probability at least

$$\begin{aligned}1-\mathbb{P}\left(E_1^c\right)-\mathbb{P}\left(E_2^c\right)-\mathbb{P}\left(E_3^c\right)&\ge 1-A_1\exp(-B_1(h_0/h)^{2\alpha})-A_2\exp(-B_2A^{2\alpha})-A_3\exp(-B_3A^2)\\ &\ge 1-\check{c}_2\exp(-\check{c}_3(h_0/h)^{2\alpha})-\check{c}_0\exp(-\check{c}_1\min(A^{2\alpha},A^2))\end{aligned}$$

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

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

and

$$|\hat{m}_i-m_i|\le Ah^\alpha.$$

□

5.3. Proof of Theorem 4.4 and Proposition 4.5

This section is devoted to the proofs of Theorem 4.4 and Proposition 4.5, both of which follow standard arguments for constructing minimax lower bounds.

Proof of Theorem 4.4. The proof essentially follows the same idea as the proof of Theorem 2 of [21]. For simplicity, we fix $L = 3$ and $C = 1/\sqrt{d}$ and suppose that $h_0 < 1/(2\sqrt{d})^\alpha$ and such that $1/(4\sqrt{d}h_0)$ is an even integer, other cases can be treated similarly. Let

$$\beta(x) = \left((2\sqrt{d})^\alpha h_0^\alpha - \|x\|_\infty^\alpha\right)_+.$$

Let $x_{i_1,...,i_d} = 4\sqrt{d}h_0(i_1 - 1/2, ..., i_d - 1/2)$, $i_1, ..., i_d \in \left[1, 1/(4\sqrt{d}h_0)\right] \cap \mathbb{N}$, be the centers of the hypercubes of the regular grid $G_{4\sqrt{d}h_0}$. We distinguish the sets:

$$P_1 = \left\{ i_1, ..., i_d \in \left[1, 1/(4\sqrt{d}h_0)\right] \cap \mathbb{N} \text{ s.t. } \sum_{j=1}^d i_j \text{ is even} \right\}$$

and

$$P_2 = \left\{ i_1, ..., i_d \in \left[1, 1/(4\sqrt{d}h_0)\right] \cap \mathbb{N} \text{ s.t. } \sum_{j=1}^d i_j \text{ is odd} \right\}.$$

We define:

$$f_0(x) = 1 + \sum_{I \in P_1} \beta(x - x_I) - \sum_{I \in P_2} \beta(x - x_I).$$

Note that, as $|P_1| = |P_2|$ and for all $I, J \in P_1 \cup P_2$, $\int_{[0,1]^d} \beta(x - x_I)\mathrm{d}x = \int_{[0,1]^d} \beta(x - x_J)\mathrm{d}x$, we have:

$$\int_{[0,1]^d} f_0(x)\mathrm{d}x = 1 + \sum_{I \in P_1} \int_{[0,1]^d} \beta(x - x_I)\mathrm{d}x - \sum_{I \in P_2} \int_{[0,1]^d} \beta(x - x_I)\mathrm{d}x = 1$$

Now, as $h_0 < 1/(2\sqrt{d})^\alpha$, we have $f_0 > 0$ and thus f_0 is a probability density. As β is $(1, \alpha)$-Hölder continuous and the maps $x \mapsto \beta(x - x_I)$, $I \in P_1 \cup P_2$ have disjoint support, f_0 is $(1, \alpha)$-Hölder continuous. It also follows that f_0 admits $1/(2(4\sqrt{d}h_0)^d)$ modes : the $(x_I)_{I \in P_1}$ all associated with the global maximum: $1 + (2\sqrt{d})^\alpha h_0^\alpha$. Furthermore, as, for any $I \in P_1$, and any $x \in \{\|x - x_I\|_2 \leq h_0\} \subset \{\|x - x_I\|_\infty \leq h_0\}$, $f_0(x) = \beta(x - x_I)$ and $x \mapsto \beta(x - x_I)$ satisfies **A4** for $C = 1/\sqrt{d}$, it follows that f_0 satisfies **A4** for $C = 1/\sqrt{d}$. Thus, f_0 belongs to $S_d(L, \alpha, \mu, R_\mu, C, h_0)$ for $C = 1/\sqrt{d}$ and all $L > C$, $0 < \mu \leq 1$, $R_\mu > 0$. Now, let $0 < h < \min(((1 - (2\sqrt{d})^\alpha h_0^\alpha)/2)^{1/\alpha}, h_0)$ and define:

$$g(x) = 2(h^\alpha - \|x\|_\infty^\alpha)_+.$$

We fix $J_1 \in P_2$ and, for all $I \in P_1$, we then define:

$$f_I(x) = f_0(x) + g(x - x_I - (h, 0, ..., 0)) - g(x - x_{J_1} - (h, 0, ..., 0))$$

Now, as $h < ((1 - (2\sqrt{d})^\alpha h_0^\alpha)/2)^{1/\alpha}$, $f_I \geq 1 - (2\sqrt{d})^\alpha h_0^\alpha - 2h^\alpha \geq 0$ and as $\int_{[0,1]^d} g(x - x_I)\mathrm{d}x = \int_{[0,1]^d} g(x - x_{J_1})\mathrm{d}x$, we have $\int_{[0,1]^d} f_I(x)\mathrm{d}x = \int_{[0,1]^d} f_0(x)\mathrm{d}x = 1$. Thus, f_I is a probability density. As f_0 and g are $(1, \alpha)$ and $(2, \alpha)$-Hölder continuous, f_I is $(3, \alpha)$-Hölder continuous. Hence, it satisfies **A1**, **A2** and **A3** for $L = 3$, and

all $\mu \in]0,1]$ and $R_\mu > 0$. Observe that, as $h \leq h_0$ and $\|x_I - \tilde{x}_I\|_\infty = h$, then:

$$\{\|x - \tilde{x}_i\|_\infty \leq h\} \subset \{\|x - \tilde{x}_i\|_\infty \leq h_0\} \subset \{\|x - x_i\|_\infty \leq 2h_0\}$$

and thus f_I admits $1/(2(4\sqrt{d}h_0)^d)$ modes : the $(x_J)_{J\in P_1\setminus\{I\}}$ and $\tilde{x}_I = x_I + (h, 0, ..., 0)$ respectively associated to the local maxima:

$$1 + (2\sqrt{d})^\alpha h_0^\alpha \text{ and } 1 + (2\sqrt{d})^\alpha h_0^\alpha + h^\alpha.$$

For all $J \in P_1\setminus\{I\}$, on $\{\|x - x_J\|_2 \leq h_0\} \subset \{\|x - x_J\|_\infty \leq 2\sqrt{d}h_0\}$, $f_I(x) = f_0(x)$ and thus f_I satisfies **A4** around all $(x_J)_{J\in P_1\setminus\{I\}}$ for $C = 1/\sqrt{d}$. Now, if $x \in \{\|x - \tilde{x}_I\|_2 \leq h_0\} \subset \{\|x - \tilde{x}_I\|_\infty \leq \sqrt{d}h_0\} \subset \{\|x - x_i\|_\infty \leq 2\sqrt{d}h_0\}$, as $\alpha \in]0,1]$ we have:

$$\begin{aligned}
f_I(\tilde{x}_I) - f_I(x) &= h^\alpha + \|x - x_I\|_\infty^\alpha - 2(h^\alpha - \|x - \tilde{x}_I\|_\infty^\alpha)_+ \\
&\geq -h^\alpha + 2\|x - \tilde{x}_I\|_\infty^\alpha + \|x - x_I\|_\infty^\alpha \\
&\geq -h^\alpha + \|x - \tilde{x}_I\|_\infty^\alpha + (\|x - \tilde{x}_I\|_\infty + \|x - x_I\|_\infty)^\alpha \\
&\geq -h^\alpha + \|x - \tilde{x}_I\|_\infty^\alpha + \|x_I - \tilde{x}_I\|_\infty^\alpha \\
&= \|x - \tilde{x}_I\|_\infty^\alpha \geq \frac{1}{\sqrt{d}}\|x - \tilde{x}_I\|_2^\alpha
\end{aligned}$$

which proves that f_I satisfies **A4** around $\tilde{x}_I$ for $C = 1/\sqrt{d}$. Then, $\{f_0\} \cup \{f_I, I \in P_1\}$ is a collection of probability densities of $S_d(L, \alpha, \mu, R_\mu, C, h_0)$, for $C = 1/\sqrt{d}$, $L = 3$ and all $\mu \in]0,1]$ and $R_{\mu>0}$, that have, pairwise, modes separated by h and associated local maxima separated by h^α. More precisely for any distinct $f, \tilde{f} \in \{f_0\} \cup \{f_I, I \in P_1\}$, let $\mathcal{M}_f$ (resp. $\mathcal{M}_{\tilde{f}}$) be the set of modes of f (resp. $\tilde{f}$), $\Gamma(f, \tilde{f})$ the set of one to one maps from $\mathcal{M}_f$ to $\mathcal{M}_{\tilde{f}}$ and

$$\gamma^* \in \operatorname{argmin}_{\gamma\in\Gamma(f,\tilde{f})} \max_{x\in\mathcal{M}_f} \|x - \gamma(x)\|_\infty$$

we have:

$$\max_{x\in\mathcal{M}_f} \|x - \gamma^*(x)\|_\infty \geq h \text{ and } \max_{x\in\mathcal{M}_f} |f(x) - \tilde{f}(\gamma^*(x))| \geq h^\alpha.$$

Now, we prove that if $h \lesssim (\log(1/h_0)/n)^{\frac{1}{2\alpha+d}}$ then $\{f_I, I \in P_1\}$ are not all distinguishable from f_0 with high probability. To do so we use Theorem 2.6 from [73], which reduces the problem to studying the χ^2 divergence between $\mathbb{P}_{f_0}^n$ and the $\mathbb{P}_{f_I}^n$, $I \in P_1$. More precisely, it suffices to prove that:

$$\frac{1}{|P_1|^2}\sum_{I\in P_1} \chi^2(\mathbb{P}_{f_0}^n, \mathbb{P}_{f_I}^n) = \frac{1}{|P_1|^2}\sum_{I\in P_1}\left(\int_{[0,1]^d}\frac{f_I^2}{f_0^2}\right)^n - 1 \underset{n\to+\infty}{\longrightarrow} 0 \tag{5.10}$$

whenever $h = o\left((\log(1/h_0)/n)^{\frac{1}{2\alpha+d}}\right)$. Let $I \in P_1$, we have:

$$\begin{aligned}
\int_{[0,1]^d}\frac{f_I^2}{f_0} &= \int_{[0,1]^d}\frac{(f_0 + g(x - \tilde{x}_I) - g(x - \tilde{x}_{J_1}))^2}{f_0} \\
&\leq 1 + \frac{1}{1 - (2\sqrt{d})^\alpha h_0^\alpha}\int_{[0,1]^d} g(x - \tilde{x}_I)^2 \mathrm{d}x + \frac{1}{1 - (2\sqrt{d})^\alpha h_0^\alpha}\int_{[0,1]^d} g(x - \tilde{x}_{J_1})^2 \mathrm{d}x
\end{aligned}$$

$$\begin{aligned}
&\leq 1 + \frac{2}{1-(2\sqrt{d})^{\alpha}h_0^{\alpha}} \int_{\|x\|_\infty \leq h} g^2(x)\mathrm{d}x \\
&\leq 1 + \frac{2}{1-(2\sqrt{d})^{\alpha}h_0^{\alpha}} \int_{\|x\|_\infty \leq h} h^{2\alpha}\mathrm{d}x \\
&= 1 + O\left(h^{2\alpha+d}\right)
\end{aligned}$$

It follows that if $h = o\left((\log(|P_1|)/n)^{\frac{1}{2\alpha+d}}\right)$, then we have (5.10). Finally, observing that $|P_1| \asymp 1/h_0$, Theorem 4.4 is proved. □

Proof of Proposition 4.5. The proof is a standard Le Cam's two-point argument, using again χ^2 divergence. Once again, for simplicity, we will consider specific parameters L and C, but other cases can be treated similarly. Let $f_1 = 1 + \|x\|_\infty^\alpha$ and $\tilde{f}_1 = f_1/Z_1$ with $Z_1 = \int_{[0,1]^d} f_1$. The function $\tilde{f}_1$ is probability density satisfying **A1**-**A4**, for $C = 1/(\sqrt{d}Z_1)$ and all $L > C$, $\mu \in]0,1]$, $R_\mu > 0$ and $h_0 \leq 1$. It admits a unique mode: $(1,...,1)$. Now, let $h < 1/4$ and consider:

$$\tilde{f}_0(x) = \tilde{f}_1(x) + \frac{1}{Z_1}\left[(h^\alpha - \|x - (1/2,...,1/2)\|_\infty^\alpha)_+ - 2^d\,(h^\alpha - \|x\|_\infty^\alpha)_+\right].$$

Observing that:

$$\int_{[0,1]^d} (h^\alpha - \|x - (1/2,...,1/2)\|_\infty^\alpha)_+\,\mathrm{d}x = 2^d \int_{[0,1]^d} (h^\alpha - \|x\|_\infty^\alpha)_+$$

we have that $\int_{[0,1]^d} \tilde{f}_0 = 1$. For sufficiently small h, we have $\tilde{f}_0 \geq 0$. Thus, $\tilde{f}_0$ is a probability density. It admits two modes $(1,...,1)$ and $(1/2,...,1/2)$ and by construction satisfies **A1**–**A4**, for $C = 1/(Z_1\sqrt{d})$, $h_0 \asymp h$ and all $L > 2^d C$, $\mu \in]0,1]$ and $R_\mu > 0$. Observe that, if we denote $\mathcal{M}_0 = \{(1,...,1),(1/2,...,1/2)\}$ and $\mathcal{M}_1 = \{(1,...,1))\}$ the sets of modes of f_0 and f_1, we have:

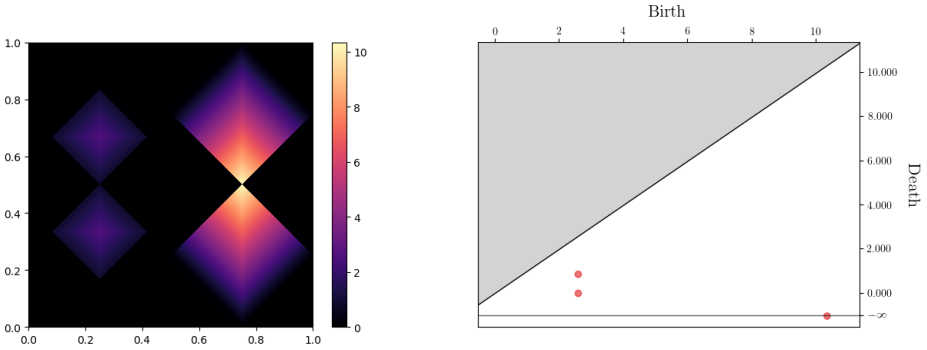
$$d_H\left(\mathcal{M}_1, \mathcal{M}_0\right) \geq 1/2.$$

Now, it suffices to show that:

$$n\chi^2(\tilde{f}_0, \tilde{f}_1) \underset{n\to+\infty}{\longrightarrow} 0$$

whenever $h = o(n^{-\frac{1}{2\alpha+d}})$.

$$\begin{aligned}
\chi^2\left(\tilde{f}_0, \tilde{f}_1\right) &= \int_{[0,1]^d} \frac{\tilde{f}_0^2}{\tilde{f}_1} - 1 \\
&= \int_{[0,1]^d} \frac{\left(\tilde{f}_1(x) + \frac{1}{Z_1}\left((h^\alpha - \|x - (1/2,...,1/2)\|_\infty^\alpha)_+ - 2^d\,(h^\alpha - \|x\|_\infty^\alpha)_+\right)\right)^2}{\tilde{f}_1(x)}\mathrm{d}x - 1 \\
&\leq \frac{1}{Z_1}\int_{\|x-(1/2,...,1/2)\|_\infty \leq h} (h^\alpha - \|x - (1/2,...,1/2)\|_\infty^\alpha)^2\,\mathrm{d}x \\
&\quad + \frac{2^d}{Z_1}\int_{\{\|x\|_\infty \leq h\}\cap[0,1]^d} (h^\alpha - \|x\|_\infty^\alpha)^2\,\mathrm{d}x
\end{aligned}$$



(A) Graph (B) H_0-persistence diagram

FIGURE 9. Graph and H_0-persistence diagram of f (normalized).

$$
\begin{aligned}
&= \frac{2^{d+1}}{Z_1} \int_{\{\|x\|_\infty \le h\} \cap [0,1]^d} \left(h^\alpha - \|x\|_\infty^\alpha\right)^2 \mathrm{d}x \\
&\le \frac{2^{d+1}}{Z_1} \int_{\{\|x\|_\infty \le h\} \cap [0,1]^d} h^{2\alpha} \mathrm{d}x \\
&= O\left(h^{2\alpha+d}\right).
\end{aligned}
$$

Thus, if $h = o(n^{-\frac{1}{2\alpha+d}})$ then $n\chi^2(\tilde{f}_0, \tilde{f}_1)$ converges to zero and the result is proved. □

6. A NUMERICAL ILLUSTRATION

From a more practical perspective, we now present a 2D numerical illustration of our results. As previously emphasized, an interesting aspect of our work is that we do not assume continuity of the density at the modes; for instance, modes may lie on the boundary of multiple distinct regular pieces. We consider a simple toy example of a density in the class $S_d(L, \alpha, \mu, R_\mu, C, h_0)$, which exhibits such typical discontinuities around some of its modes. In this example, we compare the performance of our persistence-based procedure with that of two algorithms: the mean-shift [5] and ToMATo [3].

More precisely, we consider the 2D density given by the normalization of f defined by:

$$
f(x,y) = \begin{cases}
\frac{1}{2}\max\left(0, \frac{1}{4} - \left|x - \frac{1}{4}\right| - \left|y - \frac{1}{3}\right|\right) & \text{if } |x - \frac{1}{4}| + |y - \frac{1}{3}| \le \frac{1}{6}, \\
\frac{1}{2}\max\left(0, \frac{1}{4} - \left|x - \frac{1}{4}\right| - \left|y - \frac{2}{3}\right|\right) & \text{if } |x - \frac{1}{4}| + |y - \frac{2}{3}| \le \frac{1}{6}, \\
\max\left(0, \frac{1}{2} - \left|x - \frac{3}{4}\right| - \left|y - \frac{1}{2}\right|\right) & \text{if } |x - \frac{3}{4}| + |y - \frac{3}{4}| \le \frac{1}{4} \\
\max\left(0, \frac{1}{2} - \left|x - \frac{3}{4}\right| - \left|y - \frac{1}{2}\right|\right) & \text{if } |x - \frac{3}{4}| + |y - \frac{1}{4}| \le \frac{1}{4}, \\
0 & \text{otherwise.}
\end{cases}
$$

This function admits three modes: two located within regular pieces, $x_1 = (1/4, 1/3)$ and $x_2 = (1/4, 2/3)$, and one at the boundary of three regular pieces, $x_3 = (3/4, 1/2)$. Around x_3, the function is not continuous in any full neighborhood, but only within half-cones whose boundaries intersect at x_3. One can check that, this function belongs to the class $S_d(L, 1, \mu, R_\mu, C, h_0)$ for any $\mu \le 1/\sqrt{2}$, and for some constants $L > 0$, $R_\mu > 0$, $C > 0$, and $h_0 > 0$. Its graph and persistence diagram are provided in Figure 9.

We sample $n = 40{,}000$ points according to the normalized version of f using the rejection sampling method. We then compare the performance of the different estimators, both qualitatively, by examining their behavior on the same sample, and quantitatively, by evaluating, over 100 independent samples, how often each approach

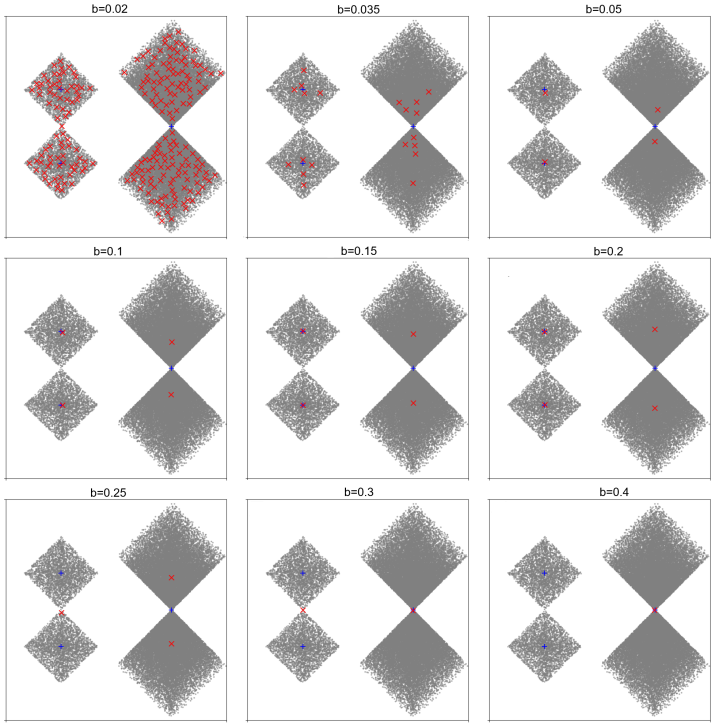


FIGURE 10. Behaviors of the mean-shift algorithm for various bandwidths $b \in \{0.02, 0.035, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4\}$. In gray is represented the sample, in blue the true modes, and in red the estimated modes. The algorithm fails to properly recover the modes for all bandwidth values.

correctly recovers the true number of modes and the mean Hausdorff distance between the estimated and true modes. The results, along with the chosen construction and calibration parameters, are presented below.

Mean-Shift. Mean-shift operates by first estimating the density using a kernel estimator and then performing a hill-climbing procedure to identify the modes. It has the advantages of being easy to use (essentially, the only parameter to tune is the bandwidth of the kernel) and computationally efficient. As previously highlighted, it is also proven to be consistent under the assumption that the underlying density is twice differentiable and has a non-degenerate Hessian [29]. There exist plenty of variant of the mean shift algorithm, here we use `MeanShift` implementation from the `sklearn.cluster` module, which follows the algorithm proposed in [5].

The example we consider lies outside the regularity classes for which mean-shift is known to be consistent, and our experiments suggest that it is in fact a typical situation where mean-shift fails to reliably recover the modes, regardless of the bandwidth choice. As illustrated in Figure 10, when the bandwidth is too small, the mode at x_3 is split into two (or more) distinct modes, whereas when the bandwidth is too large, the modes x_1 and x_2 are merged. For intermediate bandwidth values (*e.g.*, $b = 0.25$ in Fig. 10), the correct number of modes is recovered, but the result remains unsatisfactory: x_1 and x_2 still appear merged, while x_3 is again split into two. Overall, there is no clear "sweet spot" where x_1 and x_2 are correctly separated while x_3 is identified as a single mode. This observation is further supported by the quantitative results presented in Table 1, which show that although the bandwidth $b = 0.25$ often leads to the correct number of inferred modes, it also yields a large mean Hausdorff error. This confirms that the pathological configuration displayed in Figure 10 is not an exception but rather the typical outcome for this example.

TABLE 1. Comparison of the methods over 100 independent samples. For each method, we report: the number of samples for which the correct set of modes is recovered, the average number of estimated modes, the Hausdorff error between the estimated and true modes, and the Hausdorff error restricted to the samples where the correct number of modes is recovered.

	Number of time # modes=3	Avg. number of modes	Avg. Hausdorff	Avg. Hausdorff (if # modes=3)
Our estimator	98	2.98	0.0770	0.0717
Mean Shift				
bdwd = 0.02	0	225.71	0.4505	–
bdwd = 0.035	0	19.96	0.2827	–
bdwd = 0.05	0	4.20	0.0860	–
bdwd = 0.1	0	4.03	0.1191	–
bdwd = 0.15	0	4.00	0.1561	–
bdwd = 0.2	0	4.00	0.1784	–
bdwd = 0.25	86	2.86	0.2106	0.2115
bdwd = 0.3	0	2.00	0.1689	–
bdwd = 0.4	0	1.00	0.5265	–
ToMATo (DTM)				
$r = 0.01,\ \tau = 0.6$	0	161.81	0.4821	–
$r = 0.01,\ \tau = 1$	0	43.10	0.4818	–
$r = 0.01,\ \tau = 1.4$	0	21.65	0.4818	–
$r = 0.05,\ \tau = 0.6$	59	3.60	0.1448	0.1004
$r = 0.05,\ \tau = 1$	99	2.99	0.1010	0.0997
$r = 0.05,\ \tau = 1.4$	77	2.77	0.1510	0.0972
$r = 0.1,\ \tau = 0.6$	98	3.00	0.1019	0.0992
$r = 0.1,\ \tau = 1$	56	2.56	0.2014	0.0946
$r = 0.1,\ \tau = 1.4$	7	2.07	0.3199	0.0947

ToMATo. ToMATo is a persistence-based clustering algorithm that first constructs a graph from the sample points (typically a Rips graph or a k-NN graph), estimates the probability density, and then performs a hill-climbing procedure on the constructed graph to follow the gradient of the estimated density. By grouping points according to the "hill" they converge to, an initial clustering is obtained. These initial clusters are then merged based on their prominence (persistence-wise), using a given threshold. The hill-climbing step is conceptually very close to the mean-shift approach, while the merging step is similar to the "denoising" step in our method, where we remove connected components associated with birth–death pairs that have lifetime smaller than $\delta/2$. In [3], several theoretical guarantees are provided. In particular, it is shown that, with an appropriate choice of graph construction, density estimator, and merging threshold, the resulting clusters, with high probability, essentially match the basins of attraction of the underlying density, assuming it is Lipschitz-continuous. Although ToMATo is primarily a clustering algorithm, it can also be used to estimate modes. A straightforward approach is to take the output clusters and the corresponding estimated density, and then, within each cluster, identify the point where the estimated density attains its maximum. These points are then taken as the estimated modes.

For our numerical experiments, we use the implementation of ToMATo available in the `GUDHI` package [74, 75]. Several choices are possible for constructing the underlying graph. `GUDHI` provides two options: k-NN graphs (connecting each point to its k-nearest neighbors) and r-Rips graphs (connecting points whose pairwise distance is at most r). We choose to work with r-Rips graphs, as this is the setting used in the theoretical analysis of ToMATo in [3], Theorems 9.2 and 10.1. For the density estimation step, several choices are available. The package includes a KDE–based estimator, but also offers an alternative that uses the DTM (Distance-To-Measure) density estimator introduced in [76]. By default, when ToMATo is used with an r-Rips graph, the KDE bandwidth is set to r, and the number of neighbors used to compute the DTM density estimator is set to

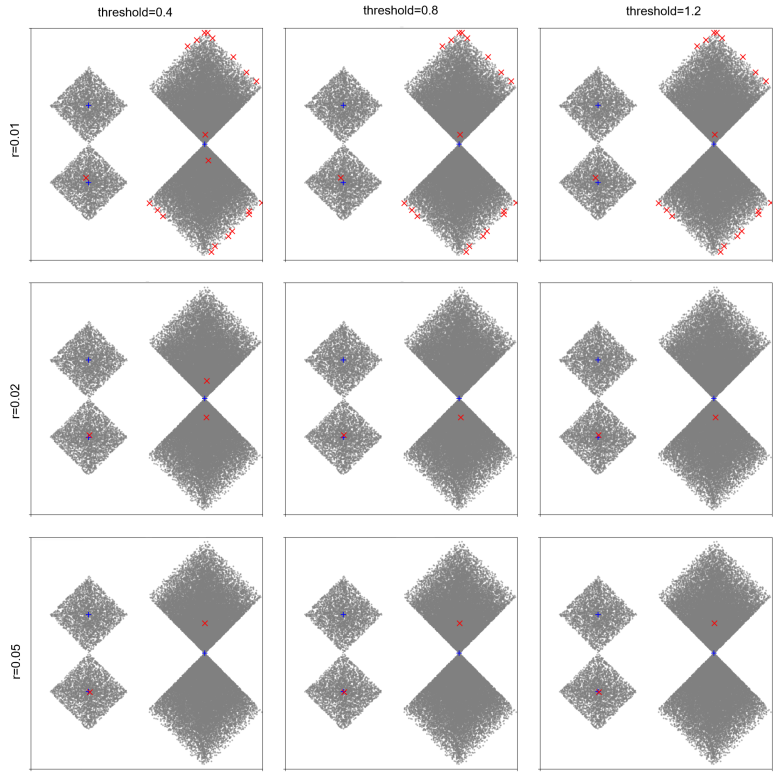
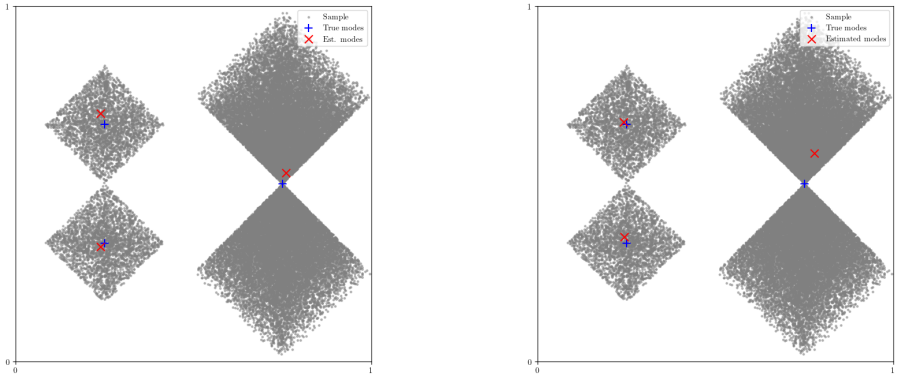


FIGURE 11. Behaviors of ToMATo with KDE for $r \in [0.01, 0.02, 0.05]$ and $\tau \in [0.4, 0.8, 1.2]$. For very small bandwidths ($r = 0.01$), the number of estimated modes remains too high, even for large threshold values. For large bandwidths ($r = 0.05$), the number of modes is too low, even for small thresholds. An intermediate bandwidth ($r = 0.02$) can recover the correct total number of modes, but this is achieved through the merging of modes x_1 and x_2 and the splitting of mode x_3.

10. We follow these default choices in our simulations, which reduces the calibration to only two parameters: r (the radius for the rips graph) and τ (the merging threshold). All other parameters are chosen by default.

From our experiments, when using a KDE, the modes returned by ToMATo exhibit the same difficulties as those observed with mean-shift. The theoretical results of [3] suggest choosing τ in the interval $[O(r), d_{\min} - O(r)]$, where $d_{\min}$ denotes the minimal distance between points in the persistence diagram of the normalized function f. In our case, $d_{\min} \simeq 1.6$. Consequently, we considered the grids $r \in [0.01 : 0.1 : 0.01]$ and $\tau \in [0.2 : 1.2 : 0.2]$, where the notation $[a : b : c]$ denotes the grid from a to b with step c. Over this grid of parameters, no pair (r, τ) yields satisfactory results. For small values of r, the number of estimated modes is systematically too large, regardless of τ; for large values of r, it is systematically too small. In the intermediate regime, some pairs do return the correct number of modes, but these again correspond to a splitting of x_3 and a merging of x_1 and x_2. Partial results from this experiment are shown in Figure 11, which illustrate these effects. As no calibration yielded satisfactory results in our experiments, and because the KDE-based variant of ToMATo is computationally intensive for the sample size we consider, we decided to exclude it from our quantitative evaluation.

In contrast, using the DTM density estimator within ToMATo yields far more satisfactory results. A choice that performs well on our example is $(r, \tau) = (0.05, 1)$, as illustrated in Figure 12b and Table 1. Moreover, DTM–ToMATo enjoys substantially lower computation times on large samples compared with KDE–ToMATo, as highlighted in the GUDHI documentation [74].



(A) Our approach (B) ToMATo (DTM)

FIGURE 12. Behaviors of our approach and the ToMATo estimator using Rips graph and DTM density (with parameters $r = 0.05$ and $\tau = 1$).In both case, the number of modes is exactly recovered, and the modes are well located.

Our approach. We implement our estimator using the package `GUDHI` [75], and in particular its class `CubicalComplex` [77] to compute persistence diagrams. Following our theoretical results, we calibrate our estimator by choosing the parameters $\mu = 1/\sqrt{2}$ and $h = (\log(n)/n)^{1/4}/6$. We choose δ using the selection procedure of Appendix B.

As illustrated in Figure 12a and Table 1, our method effectively estimates the modes, correctly recovering both their number and locations. Compared to ToMATo (with DTM), it performs on par in terms of the number of modes (both correctly identifying the number of modes in nearly all of the 100 samples) and achieves slightly lower mean Hausdorff distances, indicating more precise mode localization.

7. Discussion

Exploiting the link between modes and persistence, we propose an estimator that consistently infers the number of modes, their locations, and the associated local maxima under mild assumptions. The central contribution of our work is the identification of a threshold for mode separation (and, equivalently, prominence): below this threshold, mode detection is impossible, while above it, our procedure achieves optimal (or near-optimal) performance in the minimax sense.

Along the way, we extend the approach and results of [49] on H_0-persistence diagram estimation in the density model. These findings are of independent interest, as they broaden the applicability of persistent homology inference for densities, which has so far been largely restricted to regular settings. Our results demonstrate strong robustness to discontinuities in H_0-persistence diagram estimation. We are currently working on generalizing these methods to higher-order homology.

Despite its theoretical appeal and the encouraging simulations we ran, the proposed method has some practical limitations. In particular, it lacks adaptivity, as it relies on prior knowledge of the parameters α, μ, and δ, which are typically unknown in real-world scenarios. The dependence on α can likely be mitigated using standard adaptive estimation techniques, such as Lepski's method [78]. To reduce the dependence on δ, we provide a selection procedure in Appendix B. We show that, with this additional selection step and under a slightly stronger threshold condition h_0, it is still possible to estimate the modes and associated local maxima at minimax rates. This procedure is also used in the simulations presented in Section 6, where it yields good empirical results. The dependence on μ, however, appears more problematic, and we currently have no clear approach to address it. Nevertheless, it is worth noting that this lack of adaptivity is not specific to our method but is a common limitation of existing multiple modes estimators. Developing adaptive strategies for multimodal

inference, possibly through adaptive variants of our procedure, thus remains a promising direction for future research.

Data availability statement

Data/code are available on request from the authors.

References

[1] J. Chacón, The modal age of statistics. *Int. Statist. Rev.* **88** (2020) 122–141.

[2] J. Chacón, Clusters and water flows: a novel approach to modal clustering through morse theory (2013). arXiv preprint: arXiv:math.ST/1212.1384.

[3] F. Chazal, L. Guibas, S. Oudot and P. Skraba, Persistence-based clustering in Riemannian manifolds. *J. ACM* **60** (2011) 1–38.

[4] Y. Cheng, Mean shift, mode seeking, and clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **17** (1995) 790–799.

[5] D. Comaniciu and P. Meer, Mean shift: a robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **24** (2002) 603–619.

[6] K. Fukunaga and L. Hostetler, The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans. Inform. Theory* **21** (1975) 32–40.

[7] H. Jiang and S. Kpotufe, Modal-set estimation with an application to clustering, in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, vol. 54 of *Proceedings of Machine Learning Research*. PMLR (2017) 1197–1206.

[8] J. Li, S. Ray and B. Lindsay, A nonparametric statistical approach to clustering via mode identification. *J. Mach. Learn. Res.* **8** (2007) 1687–1723.

[9] E. Parzen, On estimation of a probability density function and mode. *Ann. Math. Statist.* **33** (1962) 1065–1076.

[10] H. Chernoff, Estimation of the mode. *Ann. Inst. Statist. Math.* **16** (1964) 31–41.

[11] D. Donoho and R. Liu, Geometrizing rates of convergence. III. *Ann. Statist.* **19** (1991) 668–701.

[12] W. Eddy, Optimum kernel estimators of the mode. *Ann. Statist.* **8** (1980) 870–882.

[13] V. Konakov, On the asymptotic normality of the mode of multidimensional distributions. *Theory Probab. Appl.* **18** (1974) 794–799.

[14] A. Mokkadem and M. Pelletier. The law of the iterated logarithm for the multivariate kernel mode estimator. *ESAIM: Probab. Statist.* **7** (2003) 1–21.

[15] J. Romano, On weak convergence and optimality of kernel density estimates of the mode. *Ann. Statist.* **16** (1988) 629–647.

[16] M. Samanta, Nonparametric estimation of the mode of a multivariate density. *South Afr. Statist. J.* **7** (1973) 109–117.

[17] A. Tsybakov, Recursive estimation of the mode of a multivariate distribution. *Probl. Pered. Inform.* **26** (1990) 38–45.

[18] P. Vieu, A note on density mode estimation. *Statist. Probab. Lett.* **26** (1996) 297–307.

[19] J. Klemelä, Adaptive estimation of the mode of a multivariate density. *J. Nonparametric Statist.* **17** (2005) 83–105.

[20] O. Lepski, Asymptotically minimax adaptive estimation. i: Upper bounds. optimally adaptive estimates. *Theory Probab. Appl.* **36** (1992) 682–697.

[21] E. Arias-Castro, W. Qiao and L. Zheng, Estimation of the global mode of a density: minimaxity, adaptation, and computational complexity. *Electron. J. Statist.* **16** (2022) 2774–2795.

[22] L. Devroye, Recursive estimation of the mode of a multivariate density. *Can. J. Statist.* **7** (1979) 159–167.

[23] C. Abraham, G. Biau and B. Cadre, Simple estimation of the mode of a multivariate density. *Can. J. Statist.* **31** (2003) 23–34.

[24] C. Abraham, G. Biau and B. Cadre, On the asymptotic properties of a simple estimate of the mode. *ESAIM: Probab. Statist.* **8** (2004) 1–11.

[25] S. Dasgupta and S. Kpotufe, Optimal rates for k-nn density and mode estimation, in *Advances in Neural Information Processing Systems*, Vol. 27. Curran Associates, Inc. (2014).

[26] M. Carreira-Perpinan, Gaussian mean-shift is an EM algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.* **29** (2007) 767–776.
[27] M. Carreira-Perpinan and C. Williams, On the number of modes of a gaussian mixture, in *Scale Space Methods in Computer Vision*. Springer Berlin Heidelberg (2003) 625–640.
[28] E. Arias-Castro, D. Mason and B. Pelletier, On the estimation of the gradient lines of a density and the consistency of the mean-shift algorithm. *J. Mach. Learn. Res.* **17** (2016) 1–28.
[29] E. Arias-Castro and W. Qiao, Clustering by hill-climbing: consistency results. *Ann. Statist.* **53** (2025) 2536–2562.
[30] U. Bauer, A. Munk, H. Sieling and M. Wardetzky, Persistence barcodes versus kolmogorov signatures: detecting modes of one-dimensional signals. *Found. Computat. Math.* **17** (2014) 33.
[31] C. Genovese, M. Perone-Pacifico, I. Verdinelli and L. Wasserman, Non-parametric inference for density modes. *J. Roy. Statist. Soc. Ser. B: Statist. Methodol.* **78** (2015) 99–126.
[32] J. Ameijeiras-Alonso, R.M. Crujeiras and A. Rodríguez-Casal, Mode testing, critical bandwidth and excess mass. *TEST* **28** (2019) 900–919.
[33] P. Burman and W. Polonik, Multivariate mode hunting: data analytic tools with measures of significance. *J. Multivariate Anal.* **100** (2009) 1198–1218.
[34] M.-Y. Cheng and P. Hall, Calibrating the excess mass and dip tests of modality. *J. Roy. Statist. Soc. Ser. B (Statist. Methodol.)* **60** (1998) 579–589.
[35] T. Duong, A. Cowling, I. Koch and M. Wand, Feature significance for multivariate kernel density estimation. *Computat. Statist. Data Anal.* **52** (2008) 4225–4242.
[36] N. Fisher and S. Marron, Mode testing via the excess mass estimate. *Biometrika* **88** (2001) 499–517.
[37] P. Hall and M. York, On the calibration of silverman's test for multimodality. *Statist. Sinica* **11** (2001) 515–536.
[38] J. Hartigan and P. Hartigan, The dip test of unimodality. *Ann. Statist.* **13** (1985) 70–84.
[39] M. Minnotte, Nonparametric testing of the existence of modes. *Ann. Statist.* **25** (1997) 1646–1660.
[40] B. Silverman, Using kernel density estimates to investigate multimodality. *J. Roy. Statist. Soc.: Ser. B (Methodol.)* **43** (1981) 97–99.
[41] J.A. Hartigan, Consistency of single linkage for high-density clusters. *J. Am. Statist. Assoc.* **76** (1981) 388–394.
[42] K. Chaudhuri, S. Dasgupta, S. Kpotufe and U. von Luxburg, Consistent procedures for cluster tree estimation and pruning. *IEEE Trans. Inform. Theory* **60** (2014) 7900–7912.
[43] M. Penrose, Single linkage clustering and continuum percolation. *J. Multivariate Anal.* **53** (1995) 94–109.
[44] D. Wang, X. Lu and A. Rinaldo, Dbscan: optimal rates for density-based cluster estimation. *J. Mach. Learn. Res.* **20** (2019) 1–50.
[45] P. Chaudhuri and J. Marron, Sizer for exploration of structures in curves. *J. Am. Statist. Assoc.* **94** (1999) 807–823.
[46] P. Bubenik, G. Carlsson, P. Kim and Z. Luo, Statistical topology via Morse theory persistence and nonparametric estimation. *Contemp. Math.* **516** (2009) 75–92.
[47] F. Chazal, D. Cohen-Steiner, M. Glisse, L.J. Guibas and S.Y. Oudot, Proximity of persistence modules and their diagrams, in *Proceedings of the Twenty-Fifth Annual Symposium on Computational Geometry*, SCG'09 (2009) 237–246.
[48] O. Bobrowski, S. Mukherjee and J. Taylor, Topological consistency via kernel estimation. *Bernoulli* **23** (2017) 288–328.
[49] H. Henneuse, Persistence diagram estimation of multivariate piecewise Hölder-continuous signals (2025). arXiv preprint: arXiv:math.ST/2403.19396.
[50] W. Polonik, Measuring mass concentrations and estimating density contour Clusters – an excess mass approach. *Ann. Statist.* **23** (1995) 855–881.
[51] P. Rigollet and R. Vert, Optimal rates for plug-in estimators of density level sets. *Bernoulli* **15** (2009) 1154–1178.
[52] C. Scott and M. Davenport, Regression level set estimation via cost-sensitive classification. *IEEE Trans. Signal Process.* **55** (2007) 2752–2757.
[53] C. Scott and R. Nowak, Learning minimum volume sets, in *Advances in Neural Information Processing Systems*, Vol. 18. MIT Press (2005).
[54] A. Tsybakov, On nonparametric estimation of density level sets. *Ann. Statist.* **25** (1997) 948–969.
[55] R. Willett and R. Nowak, Level set estimation via trees, in *Proceedings (ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005*, vol. 5 (2005) v/1089–v/1092.

[56] R. Willett and R. Nowak, Minimax optimal level-set estimation. *IEEE Trans. Image Process.* **16** (2007) 2965–2979.
[57] A. Casal and P. Saavedra-Nieves, A data-adaptive method for estimating density level sets under shape conditions. *Ann. Statist.* **50** (2022) 1653–1668.
[58] L. Cavalier, Nonparametric estimation of regression level sets. *Statistics* **29** (1997) 131–160.
[59] A. Cuevas, W. González-Manteiga and A. Rodríguez-Casal, Plug-in estimation of general level-sets. *Austr. N. Z. J. Statist.* **48** (2006) 7–19.
[60] A. Rodríguez-Casal and P. Saavedra-Nieves, Minimax Hausdorff estimation of density level sets (2019). arXiv preprint: arXiv:math.ST/1905.02897.
[61] A. Singh, C. Scott and R. Nowak, Adaptive Hausdorff estimation of density level sets. *Ann. Statist.* **37** (2009) 2760–2782.
[62] F. Chazal, D. Cohen-Steiner and A. Lieutier, A sampling theory for compact sets in euclidean space. *Discrete Computat. Geom.* **41** (2006) 461–479.
[63] H. Federer, Curvature measures. *Trans. Am. Math. Soc.* **93** (1959) 418–491.
[64] F. Chazal and B. Michel, An introduction to topological data analysis: fundamental and practical aspects for data scientists (2021). arXiv preprint: arXiv:math.ST/1710.04019.
[65] F. Chazal, S. Oudot, M. Glisse and V. de Silva, *The structure and stability of persistence modules.* SpringerBriefs in Mathematics Springer Verlag (2016).
[66] A. Hatcher, Algebraic Topology. Cambridge University Press, Cambridge (2000).
[67] S. Barannikov, The framed morse complex and its invariants. *Adv. Soviet Math.* **21** (1994) 93–115.
[68] D. Cohen-Steiner, H. Edelsbrunner and J. Harer, Stability of persistence diagrams. *Discrete Computat. Geom.* **37** (2005) 263–271.
[69] J. Shin, J. Kim, A. Rinaldo and L. Wasserman, Confidence sets for persistent homology of the KDE filtration (2017). Preprint.
[70] J. Kim, *Statistical inference for geometric data.* PhD thesis, Carnegie Mellon University (2018).
[71] F. Chazal, D. Cohen-Steiner, A. Lieutier and B. Thibert, Shape smoothing using double offset. *Proceedings – SPM 2007: ACM Symposium on Solid and Physical Modeling* (2007).
[72] J. Kim, J. Shin, F. Chazal, A. Rinaldo and L.A. Wasserman, Homotopy reconstruction via the cech complex and the vietoris-rips Complex, in *International Symposium on Computational Geometry* (2019).
[73] A. Tsybakov, Introduction to Nonparametric Estimation. Springer Publishing Company, Incorporated (2008).
[74] M. Glisse, Persistence-based clustering, in *GUDHI User and Reference Manual*, 3.11.0 edn. GUDHI Editorial Board (2025). URL https://gudhi.inria.fr/python/3.11.0/clustering.html.
[75] GUDHI. *GUDHI user and reference manual*, 3.11.0 edn. GUDHI Editorial Board (2025). URL https://gudhi.inria.fr/doc/3.11.0/.
[76] G. Biau, F. Chazal, D. Cohen-Steiner, L. Devroye and C. Rodriguez, A weighted k-nearest neighbor density estimate for geometric inference. *Electron. J. Statist.* **5** (2011) 204–237.
[77] P. Dlotko, Cubical complex, in *GUDHI User and Reference Manual*, 3.11.0 edn. GUDHI Editorial Board (2025). URL https://gudhi.inria.fr/doc/3.11.0/group__cubical__complex.html.
[78] O. Lepski, On a problem of adaptive estimation in gaussian white noise. *Theory Probab. Appl.* **35** (1991) 454–466.
[79] W. Crawley-Boevey, Decomposition of pointwise finite-dimensional persistence modules. *J. Algebra Appl.* **14** (2015) 1550066.

Appendix A. Auxiliary results on persistence diagram estimation

In this section, we provide complementary results to Proposition 4.1. Namely, we establish Corollary A.1, which provides consistency guarantees in expectation, and Proposition A.2, in which we derive minimax lower bounds matching the rates obtained in Proposition 4.1, thereby proving the optimality of our procedures.

Corollary A.1. *Let* $h \asymp \left(\frac{\log(n)}{n}\right)^{\frac{1}{d+2\alpha}}$*, we have:*

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

Proof. The sub-Gaussian concentration provided by Proposition 4.1, gives that, for all $A \geq 0$,

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

Now, we have:

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

□

Proposition A.2.

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

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

The proof of Proposition A.2 follows essentially as the proof of Theorem 3 from [49]. This follows from the minimax lower bound technique presented in Theorem 2.6 of [73]. 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 pairwise at distance $2r_n$ but indistinguishable, with high certainty.

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

$$f_{h,m}(x_1, ..., x_d) = \frac{L}{4} x_1^\alpha + \frac{L}{2}\left(h^\alpha - \|(x_1, ..., x_d) - m/\lfloor 1/h \rfloor (1, ..., 1)\|_\infty^\alpha\right)_+$$

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

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

$$\mathrm{dgm}(f_{h,m}) = \left\{(L/4,-\infty), \left(\frac{L}{4}\left(\frac{m}{\lfloor 1/h\rfloor}\right)^\alpha - \frac{L}{4}h^\alpha, \frac{L}{4}\left(\frac{m}{\lfloor 1/h\rfloor}\right)^\alpha - \frac{L}{2}h^\alpha\right)\right\}$$

and consequently,

$$\mathrm{dgm}(\tilde{f}_{h,m}) = \left\{\left(\frac{L}{4\|1+f_{h,m}\|_1},-\infty\right), \left(\frac{1+\frac{L}{4}\left(\frac{m}{\lfloor 1/h\rfloor}\right)^\alpha - \frac{L}{4}h^\alpha}{\|1+f_{h,m}\|_1}, \frac{1+\frac{L}{4}\left(\frac{m}{\lfloor 1/h\rfloor}\right)^\alpha - \frac{L}{2}h^\alpha}{\|1+f_{h,m}\|_1}\right)\right\}.$$

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 $\frac{1}{4}\lfloor 1/h\rfloor < m \neq m' < \frac{3}{4}\lfloor 1/h\rfloor$,

$$d_\infty\left(\mathrm{dgm}(f_{h,m}), \mathrm{dgm}(f_{h,m'})\right) \geq \frac{Lh^\alpha}{4\|1+f_{h,m}\|_1} \geq \frac{Lh^\alpha}{4(1+L)} \tag{A.1}$$

We set $r_n = \frac{Lh^\alpha}{8(1+L)}$, then, by (A.1), for all $\frac{1}{4}\lfloor 1/h\rfloor < m \neq m' < \frac{3}{4}\lfloor 1/h\rfloor$,

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

For a fixed signal f, denote $\mathbb{P}_f^n$ the joint probability distribution of $X_1,...,X_n$. From Theorem 2.6 of [73], it now suffices to show that if $r_n = o\left(\left(\log(n)/n\right)^{\frac{p\alpha}{d+2\alpha}}\right)$, then, for a fixed m,

$$\left(\frac{2}{\lfloor \frac{1}{h}\rfloor - 2}\right)^2 \sum_{\frac{1}{4}\lfloor 1/h\rfloor < m' < \frac{3}{4}\lfloor 1/h\rfloor, m' \neq m} \chi^2\left(\mathbb{P}^n_{\tilde{f}_{h,m'}}, \mathbb{P}^n_{\tilde{f}_{h,m}}\right) \tag{A.2}$$

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

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

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

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

$$\tilde{f}_{h,m'} = \tilde{f}_{h,m} + \frac{L}{2\|1+f_{h,m}\|_1}\left(h^\alpha - \|(x_1,...,x_d) - m/\lfloor 1/h \rfloor (1,...,1)\|_\infty^\alpha\right)_+ - \frac{L}{2\|1+f_{h,m}\|_1}\left(h^\alpha - \|(x_1,...,x_d) - m'/\lfloor 1/h \rfloor (1,...,1)\|_\infty^\alpha\right)_+$$

and

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

We then have:

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}_{\tilde{f}_{h,m}}}\left(\frac{\tilde{f}^2_{h,m'}(X_1)}{\tilde{f}^2_{h,m}(X_1)}\right)\\
&= 1 + \frac{L}{\|1+f_{h,m}\|_1}\int_{H_m} h^\alpha - \|(x_1,...,x_d) - m/\lfloor 1/h \rfloor (1,...,1)\|_\infty^\alpha \mathrm{d}t_1...\mathrm{d}t_d\\
&\quad - \frac{L}{\|1+f_{h,m}\|_1}\int_{H_{m'}} h^\alpha - \|(x_1,...,x_d) m'/\lfloor 1/h \rfloor (1,...,1)\|_\infty^\alpha \mathrm{d}t_1...\mathrm{d}t_d\\
&\quad + \frac{L^2}{4\|1+f_{h,m}\|_1^2}\int_{H_m} \frac{\left(h^\alpha - \|(x_1,...,x_d) - m/\lfloor 1/h \rfloor (1,...,1)\|_\infty^\alpha\right)^2}{\tilde{f}_{h,m}} \mathrm{d}t_1...\mathrm{d}t_d\\
&\quad + \frac{L^2}{4\|1+f_{h,m}\|_1^2}\int_{H_{m'}} \frac{\left(h^\alpha - \|(x_1,...,x_d) - m'/\lfloor 1/h \rfloor (1,...,1)\|_\infty^\alpha\right)^2}{\tilde{f}_{h,m}} \mathrm{d}t_1...\mathrm{d}t_d\\
&= 1 + \frac{L^2}{2\|1+f_{h,m}\|_1^2}\int_{H_m} \frac{\left(h^\alpha - \|(x_1,...,x_d) - m/\lfloor 1/h \rfloor (1,...,1)\|_\infty^\alpha\right)^2}{\tilde{f}_{h,m}} \mathrm{d}t_1...\mathrm{d}t_d\\
&\leq 1 + L^2(1+L)\int_{H_m} \left(h^\alpha - \|(x_1,...,x_d) - m/\lfloor 1/h \rfloor (1,...,1)\|_\infty^\alpha\right)^2 \mathrm{d}t_1...\mathrm{d}t_d\\
&\leq 1 + L^2(1+L)\int_{H_m} h^{2\alpha} \mathrm{d}t_1...\mathrm{d}t_d\\
&\quad + L^2(1+L)\int_{H_m} \|(x_1,...,x_d) - m/\lfloor 1/h \rfloor (1,...,1)\|_\infty^{2\alpha} \mathrm{d}t_1...\mathrm{d}t_d\\
&\leq 1 + 2L^2(1+L)h^{2\alpha+d} = 1 + O(h^{2\alpha+d})
\end{aligned}$$

Hence, if $h = o\left(\left(\frac{\log(n)}{n}\right)^{\frac{1}{d+2\alpha}}\right)$, we have that (A.2) 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. □

Appendix B. Complement on modes inference: handling unknown δ

In this section, we adapt the procedure for modes inference presented in 4 to the more realistic case where δ is unknown. The idea is to select δ through penalization. The procedure remains the same, except for the construction of $\hat{k}$. Let

$$R(\delta) = d_\infty\left(\widehat{\mathrm{dgm}(f)}, \overline{\mathrm{dgm}(f)}^\delta\right) + \frac{h^\alpha}{\delta} \tag{B.1}$$

with

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

We consider,

$$\hat{\delta} = \text{argmin}_{\delta \in]0,+\infty[} R(\delta) = \text{argmin}_{\delta = |b-d|, (b,d) \in \widehat{\text{dgm}(f)}} R(\delta)$$

the existence of this minimum is proved in Lemma B.2 and $\hat{k}$ is then defined as,

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

With this choice of $\hat{\delta}$, we now show that, under the assumption :

$$\delta = Ch_0^\alpha > \sqrt{\frac{12}{7} h^\alpha},\ h \asymp \left({}^{\log(n)}/_{n}\right)^{\frac{1}{d+2\alpha}}$$

our procedure allows us to infer, with high probability, the exact number of modes, and to estimate their locations and associated maxima at the minimax rates (up to a logarithmic factor), as in Theorem 4.3. This additional assumption requires that h_0 is of order at least $\asymp \left({}^{\log(n)}/_{n}\right)^{\frac{1}{2(d+2\alpha)}}$ or equivalently that δ is of order at least $\left({}^{\log(n)}/_{n}\right)^{\frac{\alpha}{2(d+2\alpha)}}$. In particular, this covers the case where the parameters L, α, μ, R_μ, C and h_0 are fixed and n is sufficiently large.

Theorem B.1. *Let $f \in S_d(L, \alpha, \mu, R_\mu, C, h_0)$, $h \simeq \left({}^{\log(n)}/_{n}\right)^{\frac{1}{d+2\alpha}}$ and suppose $\delta = Ch_0^\alpha > \sqrt{\frac{12}{7} h^\alpha}$. There exist C'_0,..., C'_3 depending only on L ,α, μ, R_μ, C such that, for all $A \geq 0$ with probability at least $1 - C'_0 \exp(-C'_1 A^{2\alpha}) - C'_2 \exp(-C'_3 (h_0/h)^{-\alpha})$,*

$$k = \hat{k}$$

and, for all $i \in \{1, ..., k\}$, 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 δ^* be the minimal lifetime from points of dgm(f) and $\tilde{\delta}$ the minimal lifetime from the points of

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

Note that δ^* and $\tilde{\delta}$ are possibly infinite.

Lemma B.2. *Let $h \simeq \left({}^{\log(n)}/_{n}\right)^{\frac{1}{d+2\alpha}}$, and E_4 the event:*

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

If $\delta^* \geq \delta > \sqrt{\frac{12}{7} h^\alpha}$, under $E_4 \cap E_1$, R admits a minimum over $]0,1]$ attained only at $\delta = \tilde{\delta}$.

Proof. Under E_1, by Theorem 4.1,

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

Then, we have, under E_1,

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

- If $0 < \delta \leq (K_1+K_2)h^\alpha$, under E_1 and E_4,

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

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

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

  and $h^\alpha/\delta > h^\alpha/\tilde{\delta}$. Thus, $R(\delta) > R\left(\tilde{\delta}\right)$.
- If $\delta > \tilde{\delta}$, under E_1 and as $\delta^* \geq \delta > \sqrt{\frac{12}{7}h^\alpha}$,

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

Hence, in all cases, if $\delta \neq \tilde{\delta}$, $R(\delta) > R(\tilde{\delta})$. □

Proof of Theorem B.1. Lemma B.2, imply that, under $E_1 \cap E_4$, for sufficiently large n,

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

Thus, by definition of $\tilde{\delta}$,

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

By (5.2) there exist A_4 and B_4 such that E_4 occurs with probability at least:

$$1 - B_4\exp(-A_5 h^{-2\alpha}) \geq 1 - B_4\exp(-A_4(h_0/h)^{2\alpha}).$$

The proof then follows as in the proof of Theorem 4.3. □

Appendix C. Proof of Lemma 5.7

This section is dedicated to the proof of Lemma 5.7.

Proof of Lemma 5.7. The first part of the claim is Theorem 12 of [72]. A standard technique to construct deformation retraction in differential topology is to exploit the flow coming from a smooth underlying vector field. In [72], their idea is to use the vector field defined on $\overline{B}_2(K,r) \setminus K$, by $W(x) = -\nabla_{d_K}(x)$. 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 $\overline{B}_2(K,r) \setminus K$ 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 $\overline{B}_2(K,r) \times [0,+\infty[$ such that for all $x \in \overline{B}_2(K,r)$, for all $t \geq 2r/\mu^2$, $C(x,t) = C(x, 2r/\mu^2) \in K$. We make an additional remark, denote d_C the arc length distance along C, as $\|W\| \leq 1$, we have:

$$\begin{aligned} d_C(x, C(x, 2r/\mu^2)) &= \int_0^{2r/\mu^2} \left| \frac{\partial}{\partial t} C(x,t) \right| \mathrm{d}t \\ &\leq \int_0^{2r/\mu^2} \|W(C(x,t)\|_2 \mathrm{d}t \\ &\leq 2r/\mu^2 \end{aligned} \tag{C.1}$$

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 $\overline{B}_2(K,r)$ onto K and the associated retraction $R : x \mapsto F(x,1)$ satisfies, $R(x) \in \overline{B}_2(x, 2r/\mu^2)$ by (C.1). □

Appendix D. Proof of Lemma 5.3

This Appendix is dedicated to the proof of Lemma 5.3.

Proof of Lemma 5.3. Let $f \in S_d(L, \alpha, \mu, R_\mu)$. Let $H \subset \mathcal{F}_\lambda \cap C_h$, by definition of N_h, we have that:

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

As $|X \cap H|$ is a $(n, \mathbb{P}(X_1 \in H))$-binomial and $\mathbb{P}(X_1 \in H) \leq \lambda h^d$, we have:

$$\mathbb{E}\left[\frac{|X \cap H|}{nh^d}\right] = \frac{\mathbb{P}(X_1 \in H)}{h^d} \leq \lambda.$$

Thus,

$$\frac{|X \cap H|}{nh^d} \leq \lambda + 2\sqrt{3M_{d,L,\alpha,R_\mu}} N_h \sqrt{\frac{\log(1/h^d)}{nh^d}}.$$

As this holds for all $\lambda \in \mathbb{R}$, we have, for all $H \subset \mathcal{F}_{\lambda - 2\sqrt{3M_{d,L,\alpha,R_\mu}} N_h h^\alpha} \cap C_h$:

$$\frac{|X \cap H|}{nh^d} \leq \lambda + 2\sqrt{3M_{d,L,\alpha,R_\mu}} N_h \sqrt{\frac{\log(1/h^d)}{nh^d}} - 2\sqrt{3M_{d,L,\alpha,R_\mu}} N_h h^\alpha.$$

Now, the choice of h ensures that:

$$\sqrt{\frac{\log(1/h^d)}{nh^d}} - h^\alpha < 0,$$

which yields:

$$\frac{|X \cap H|}{nh^d} < \lambda.$$

Similarly, we can show that for all $H' \subset \mathcal{F}_{\lambda + 2\sqrt{3M_{d,L,\alpha,R_\mu}} N_h h^\alpha} \cap C_h$, we have:

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

and the result is proved. □

Appendix E. Proof of Lemma 5.1

This Appendix is dedicated to the proof of Lemma 5.1.

Proof of Lemma 5.1. Let $i \in \{1, ..., l\}$. 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**,

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

Hence, as f is a density,

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

and

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

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 [72], 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:

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

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

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

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:

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

To conclude, it suffices to note 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 $x \in [0,1]^d$,

$$f(x) \leq \kappa(d, L, \alpha, R_\mu).$$

□

Appendix F. Proof for the q-tameness

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

Proposition F.1. *Let $f \in S_d(L, \alpha, \mu, R_\mu)$ then f is q-tame.*

Proof. Let $s \in \mathbb{N}$ and $\mathbb{V}_{0,f}$ be the persistence module (for the 0-th homology) associated to the superlevel filtration $\mathcal{F}$, and for fixed levels $\lambda > \lambda'$ let us denote $v_\lambda^{\lambda'}$ the associated map. Let $\lambda \in \mathbb{R}$ and $h < \frac{R_\mu}{2}$. By Lemma 5.5,

$$v_\lambda^{\lambda - L\left(\sqrt{d}(2+2/\mu^2)\right)^\alpha h^\alpha} = \phi_\lambda \circ \tilde{i}_\lambda$$

with,

$$\tilde{i}_\lambda : H_0\left(\mathcal{F}_\lambda\right) \to H_0\left(\mathcal{F}_\lambda^h\right) \text{ induced by inclusion.}$$

By assumption **A1** and **A2**, $\mathcal{F}_\lambda$ is compact. As $[0,1]^d$ is triangulable, $\mathcal{F}_\lambda$ 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, $\tilde{i}_\lambda$ factors through the finite-dimensional space $H_0(K)$ and is then of finite rank.

Thus, $v_\lambda^{\lambda - L\left(\sqrt{d}(2+2/\mu^2)\right)^\alpha h^\alpha}$ 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. □

Proposition F.2. *Let $f : [0,1]^d \to \mathbb{R}$ 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 [79]. □

Appendix G. Equivalence of birth-death pairs

In this section, we rigorously prove in Lemma G.1 that the collection of all birth–death pairs (b, d), where b is the birth time of a connected component appearing in $\hat{\mathcal{F}}$ and d is its death time (according to Definition 4.2, for an arbitrary but fixed ordering), coincides with the birth–death pairs of $\widehat{\mathrm{dgm}}(f)$. Denote by $\mathcal{C}$ the collection of connected components that appear (*i.e.*, are born) at some level of the filtration:

$$\mathcal{C} := \left\{C \subset \widehat{\mathcal{F}}_{b(C)} \;\middle|\; C \text{ is a connected component born at time } b(C)\right\} = \bigcup_{i \in I} \mathcal{C}(\check{b}_i).$$

Furthermore, consider the collection of birth-death pairs defined by:

$$\mathcal{P} = \{(b(C), d(C)), C \in \mathcal{C}\},$$

where $b(C)$ and $d(C)$ denote respectively the birth and death time of $C \in \mathcal{C}$ (birth-death pairs are considered with multiplicity).

Lemma G.1. *We have:*

$$\mathcal{P} = \widehat{\mathrm{dgm}}(f).$$

Before turning to the proof of Lemma G.1, we introduce the following technical lemma, which will play a key role in that proof. Since $\hat{\mathcal{F}}_\lambda$ is a union of cubes from the grid $G(h)$, it necessarily has a finite number of distinct connected components, which we denote by $A_1, \ldots, A_q$. Recall that the homology group $H_0(\hat{\mathcal{F}}_\lambda)$ is generated by the equivalence classes $[x_1], \ldots, [x_q]$, where, for all $1 \leq i \leq q$, x_i is a point of A_i. As representatives can be chosen arbitrarily within each

connected component, we take $x_i \in C_i$, where C_i is the oldest connected component of $\mathcal{C}$ contained in A_i. This choice allows us to clearly link these equivalence classes with elements of $\mathcal{C}$. Furthermore, up to swapping indices, we can assume that $C_1, \dots, C_q$ are ordered from the youngest to the oldest (according to the rules of Definition 4.2). In addition, for each $1 \le i \le q-1$, we denote by $B(i) \in \{1, \dots, q\}$ the index of the older connected component among $\{C_1, \dots, C_q\}$ that become connected to C_i at its death time. With these conventions in place, we obtain the following result.

Lemma G.2. *Define*

$$b_i \;:=\; [x_i] - [x_{B(i)}], \quad \textit{for } 1 \le i \le q-1, \;\textit{and } b_q = [x_q]$$

and set

$$V_\lambda \;:=\; \{\, b_i :\; 1 \le i \le q \,\}.$$

Then V_λ is a basis of $H_0(\widehat{\mathcal{F}}_\lambda)$.

Proof. The classes $[x_1], \dots, [x_q]$ form the canonical basis of $H_0(\widehat{\mathcal{F}}_\lambda)$, we will show that each one of this class of equivalence can be expressed as a linear combination of elements of V_λ.

Fix $1 \le i \le q-1$ and consider the sequence $x_i, x_{B(i)}, x_{B^{(2)}(i)}, \dots$ obtained by iterating the map B. Since there are only q index and that $B(i) > i$, for all $1 \le i \le q-1$, there exists an integer $1 \le m \le q$ such that $B^{(m)}(i) = q$. Using the definition of the b_i, the chain $x_i, x_{B(i)}, \dots, x_{B^{(m)}(i)} = x_q$ yields the identities:

$$\begin{aligned}
[x_i] &= b_i + [x_{B(i)}], \\
[x_{B(i)}] &= b_{B(i)} + [x_{B^{(2)}(i)}], \\
&\vdots \\
[x_{B^{(m-1)}(i)}] &= b_{B^{(m-1)}(i)} + [x_{B^{(m)}(i)}] = b_{B^{(m-1)}(i)} + [x_q].
\end{aligned}$$

Substituting these identities successively produces the telescoping expansion:

$$[x_i] \;=\; b_i + b_{B(i)} + \cdots + b_{B^{(m-1)}(i)} + [x_q] = b_i + b_{B(i)} + \cdots + b_{B^{(m-1)}(i)} + b_q.$$

Hence every class $[x_i]$ with $1 \le i \le q-1$ is a linear combination of elements of V_λ, and $[x_q] \in V_\lambda$ itself. Since the classes $[x_1], \dots, [x_q]$ span $H_0(\widehat{\mathcal{F}}_\lambda)$, this shows that V_λ spans $H_0(\widehat{\mathcal{F}}_\lambda)$. As V_λ contains exactly q elements, it is a basis of $H_0(\widehat{\mathcal{F}}_\lambda)$. □

Proof of Lemma G.1. The main idea of the proof is to show that the persistence module obtain by the direct sum of the interval modules associated to birth-pairs of connected component appearing in $\hat{\mathcal{F}}$ is isomorphic to (*i.e.* 0-interleaved with) $\widehat{\mathbb{V}}_{f,0}$ (hereafter referred to as $\hat{\mathbb{V}}$ to simplify notation), which implies that they yield the same persistence diagram by the algebraic stability theorem [47].

For each $C \in \mathcal{C}$, denote by $b(C)$ its birth time and by $d(C)$ its death time. Define the interval module $I(C)$, equipped with the linear maps $(u(C)_\lambda^{\lambda'})_{\lambda' \le \lambda}$, by:

$$I(C)_\lambda = \begin{cases} \mathbb{Z}/2\mathbb{Z} & \text{if } d(C) < \lambda \le b(C), \\ 0 & \text{otherwise} \end{cases}, \qquad u(C)_\lambda^{\lambda'} = \begin{cases} \mathrm{id} & \text{if } d(C) < \lambda' \le \lambda \le b(C), \\ 0 & \text{otherwise} \end{cases}.$$

Consider the persistence module given by the direct sums:

$$\mathbb{W} \;:=\; \bigoplus_{C \in \mathcal{C}} I(C), \qquad w_\lambda^{\lambda'} = \bigoplus_{C \in \mathcal{C}} u(C)_\lambda^{\lambda'}.$$

Note that each interval module $I(C)$, $C \in \mathcal{C}$, in the interval decomposition of $\mathbb{W}$ contributes a single generator e_C, which exists exactly at the parameter values inside the interval $[b(C), d(C))$. In particular, the family W_λ of all e_C such that

$d(C) < \lambda \leq B(C)$ forms a basis of $\mathbb{W}_\lambda$. Note that the $C \in \mathcal{C}$ satisfying $d(C) < \lambda \leq B(C)$ are exactly the $C_1, ..., C_q \in \mathcal{C}$ associated to the connected component $A_1, ..., A_q$ of $\hat{\mathcal{F}}_\lambda$. Thus, denoting $e_i = e_{C_i}$ for all $1 \leq i \leq q$, the basis W_λ is the family $(e_i)_{1 \leq i \leq q}$. Note that, by definition of w, for all $1 \leq i \leq q$, for all $\lambda' \leq \lambda$, we have:

$$w_\lambda^{\lambda'}(e_i) = \begin{cases} e_i & \text{if } \lambda' > d(C_i) \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, by the definition of the persistence diagram of a decomposable module (see section 1.6 in [65]), *i.e.*, a module that is a direct sum of interval modules, we have:

$$\mathrm{dgm}(\mathbb{W}) = \{(b(C),\, d(C)) \,:\, C \in \mathcal{C}\} = \mathcal{P}.$$

Hence, to prove Lemma G.1, it is enough to construct an 0-interleaving between $\mathbb{W}$ and $\widehat{\mathbb{V}}$, that is, constructing two collections of linear maps $(\phi_\lambda : \mathbb{W}_\lambda \to \hat{\mathbb{V}}_\lambda)_{\lambda \in \mathbb{R}}$ and $(\psi_\lambda : \hat{\mathbb{V}}_\lambda \to \mathbb{W}_\lambda)_{\lambda \in \mathbb{R}}$ such that, for all $\lambda' \leq \lambda$,

$$\phi_{\lambda'} \circ w_\lambda^{\lambda'} = v_\lambda^{\lambda'} \circ \phi_\lambda, \text{ and } \psi_{\lambda'} \circ v_\lambda^{\lambda'} = w_\lambda^{\lambda'} \circ \psi_\lambda, \tag{G.1}$$

and moreover,

$$\psi_\lambda \circ \phi_\lambda = \mathrm{id}_{\mathbb{W}_\lambda} \text{ and } \phi_\lambda \circ \psi_\lambda = \mathrm{id}_{\hat{\mathbb{V}}_\lambda}. \tag{G.2}$$

To do so, it suffices to define, for all $\lambda \in \mathbb{R}$, ψ_λ and ϕ_λ on bases of $\mathbb{V}_\lambda$ and $\mathbb{W}_\lambda$. Using the basis V_λ from Lemma G.2 and the basis W_λ defined above, we set:

- ϕ_λ to be the linear map sending $e_i \in W_\lambda$ to b_i, for all $1 \leq i \leq q$;
- ψ_λ to be the linear map sending $b_i \in V_\lambda$ to e_i, for all $1 \leq i \leq q$.

With these definitions, for all $1 \leq i \leq q$, we have:

$$\phi_\lambda \circ \psi_\lambda(b_i) = \phi_\lambda(e_i) = b_i \text{ and } \psi_\lambda \circ \phi_\lambda(e_i) = \psi_\lambda(b_i) = e_i.$$

Since two linear maps that agree on a basis agree everywhere, (G.2) is satisfied. Now, let us show that (G.1) is also satisfied. let $\lambda' \leq \lambda$, $1 \leq i \leq q$ and denotes B'_i the connected component of $\widehat{\mathcal{F}}_{\lambda'}$ containing C_i. There is two possible cases:

- If $\lambda' \leq d(C_i)$: then C_i and $C_{B(i)}$ are connected in $\hat{\mathcal{F}}_{\lambda'}$, so they are represented by the same equivalence class in $\hat{\mathbb{V}}_{\lambda'}$:

  $$v_\lambda^{\lambda'}([x_i]) = v_\lambda^{\lambda'}([x_{B(i)}]) \iff v_\lambda^{\lambda'}(b_i) = 0.$$

  Thus,

  $$\psi_{\lambda'} \circ v_\lambda^{\lambda'}(b_i) = \psi_{\lambda'}(0) = 0 \quad \text{and} \quad w_\lambda^{\lambda'} \circ \psi_\lambda(b_i) = w_\lambda^{\lambda'}(e_i) = 0.$$

  and

  $$v_\lambda^{\lambda'} \circ \phi_\lambda(e_i) = v_\lambda^{\lambda'}(b_i) = 0 \quad \text{and} \quad \phi_{\lambda'} \circ w_\lambda^{\lambda'}(e_i) = \phi_{\lambda'}(0) = 0.$$

- If $\lambda' > d(C_i)$: then the older connected component of $\mathcal{C}$ contained in B'_i is also C_i. Otherwise, this would imply the existence of some $C'_i \in \mathcal{C}$, older than C_i, such that $C'_i \subset B'_i$. In turn, C'_i and C_i would be connected in $\widehat{\mathcal{F}}_{\lambda'}$,

which would imply $d(C_i) \geq \lambda'$, a contradiction. Thus, we have $b_i \in V_{\lambda'}$ and $e_i \in W_{\lambda'}$. Consequently, $\psi_{\lambda'}$ maps b_i to e_i and $\phi_{\lambda'}$ maps e_i to b_i. Hence:

$$\psi_{\lambda'} \circ v_\lambda^{\lambda'}(b_i) = \psi_{\lambda'}(b_i) = e_i \text{ and } w_\lambda^{\lambda'} \circ \psi_\lambda(b_i) = w_\lambda^{\lambda'}(e_i) = e_i.$$

and

$$v_\lambda^{\lambda'} \circ \phi_\lambda(e_i) = v_\lambda^{\lambda'}(b_i) = b_i \quad \text{and} \quad \phi_{\lambda'} \circ w_\lambda^{\lambda'}(e_i) = \phi_{\lambda'}(e_i) = b_i.$$

In both case, we have $\psi_{\lambda'} \circ v_\lambda^{\lambda'}(b_i) = w_\lambda^{\lambda'} \circ \psi_\lambda(b_i)$ and $v_\lambda^{\lambda'} \circ \phi_\lambda(e_i) = \phi_{\lambda'} \circ w_\lambda^{\lambda'}(e_i)$. As it holds for all $1 \leq i \leq q$, and since two linear maps that agree on a basis agree everywhere, (G.1) follows.

We have thus shown that the collections $(\phi_\lambda)_{\lambda \in \mathbb{R}}$ and $(\psi_\lambda)_{\lambda \in \mathbb{R}}$ provide a 0-interleaving between $\mathbb{W}$ and $\hat{\mathbb{V}}$ (*i.e.*, they are isomorphic). By the algebraic stability theorem, this implies that they yield the same persistence diagram, which completes the proof.

□