



**HAL**  
open science

# Local Convex Hull density and level set estimation

Catherine Aaron, Myriam Garrido

► **To cite this version:**

Catherine Aaron, Myriam Garrido. Local Convex Hull density and level set estimation. 2024. hal-04010316v2

**HAL Id: hal-04010316**

**<https://hal.science/hal-04010316v2>**

Preprint submitted on 23 Jan 2024

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

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

# Local Convex Hull density and level set estimation

C. Aaron and M. Garrido

January 23, 2024

## Abstract

Consider an i.i.d. sample  $\mathbb{X}_n = \{X_1, \dots, X_n\} \subset \mathbb{R}^D$  drawn from an unknown distribution  $\mathbb{P}_X$  associated to a density  $f$  and supported in  $S$ . We introduce and analyze a density estimator based on the Local Convex Hull (LoCoH) method, inspired by [20]. We prove that, if  $f$  is Lipschitz continuous and bounded below on a sufficiently regular support  $S$ , the proposed estimator achieves pointwise minimax rates at any point in  $S$ , even if  $S$  is unknown. If  $f$  is  $\mathcal{C}^2$ , when  $D \leq 7$ , under broad assumptions regarding the support set  $S$ , the proposed estimator can achieve minimax rates in  $\mathring{S}$ , with additional guarantees at the boundary of  $S$ . This offers guarantees for level set estimation (even for level sets intersecting the boundary). Additionally, we offer a generalization for density estimation when  $S$  is a  $d$ -dimensional sub-manifold of  $\mathbb{R}^D$  with similar convergence rates (depending on  $d$  instead of  $D$ ). We also present several numerical illustrations.

## 1 Introduction

The density estimation problem, got an important impulse after [25] and [28] fundamental pioneering works. It remains an active area of research. Numerous well-known books have been published on this topic. To name a few, refer to [31], [14], [15], and [30]. As evidenced by [10], density estimation remains a contemporary and actively researched topic. Presently, two focal points attract notable attention: mitigating bias near the support boundary and extending density estimation to encompass manifolds. The objective of this paper is to propose an asymptotically unbiased density estimator applicable across the entire support, while also offering adaptability for manifold scenarios.

Let us first focus on **the bias reduction problem in the full dimensional setting**. Consider an i.i.d. sample  $\mathbb{X}_n = \{X_1, \dots, X_n\} \subset \mathbb{R}^D$  drawn from an unknown distribution  $\mathbb{P}_X$  supported in  $S$  (which is also unknown). Recall the definition of  $S$  as the intersection of all closed sets  $E$  for which we have  $\mathbb{P}_X(E) = 1$ . Assuming that  $S$  is a regular set, specifically  $\bar{\mathring{S}} = S$ , and that the density restricted to  $S$ , denoted as  $f|_S$ , is continuous, we recall the Besicovitch condition through the Lebesgue differentiation theorem:

$$\text{Let } x \in \mathring{S}, \lim_{r \rightarrow 0} \frac{\int_{B(x,r)} f(t) dt}{|B(x,r)|_D} = f(x). \quad (1)$$

This forms the foundation of basic density estimation techniques. This is due to the fact that  $\int_{B(x,r)} f(t) dt$  can be readily estimated by the ratio of the number of observations within  $B(x,r)$  to the total number of observations. However, this approach introduces a bias for points  $x \in \partial S$  where  $f(x) > 0$ . Bias reduction techniques in density estimation have been extensively explored when  $D = 1$ , with notable works such as [29], [24], [21], and [23]. In higher dimensions, many

proposed bias correction methods rely on adjusting kernel density estimators with knowledge of the support. This is exemplified by approaches like those presented in [6], [11], [19], and [8]. To the best of our knowledge, the only method allowing for bias reduction without prior knowledge of the support can be found in [5]. A way to extend (1) on  $S$  is to use that, for all  $x \in S$ ,  $\lim_{r \rightarrow 0} \frac{\int_{B(x,r)} f(t) dt}{|B(x,r) \cap S|_D} = f(x)$ . By applying a plugin methodology, a straightforward class of density estimators emerges:

$$\check{f}(x) = \frac{\#\{\mathbb{X}_n \cap B(x,r)\}}{n|B(x,r) \cap S|_D}.$$

In this method, the crucial aspect lies in the choice of  $|B(x,r) \cap S|_D$ , an estimate for the  $D$ -dimensional volume of the intersection between  $B(x,r)$ , the close ball centered at  $x$  and of radius  $r$ , and the support  $S$ . A straightforward naive option for  $|B(x,r) \cap S|_D$ , utilizing the concept of local convex hull (LoCoH), is  $|H(B(X_i,r)) \cap \mathbb{X}_n|_D$  and thus  $\check{f}_{\text{naive}}(x) = \frac{\#\{\mathbb{X}_n \cap B(x,r)\}}{n|H(B(X_i,r)) \cap \mathbb{X}_n|_D}$ . Regrettably, this choice provides overestimation of the density and hampers the convergence rate for points  $x$  satisfying  $d(x, \partial S) \geq r$ , which asymptotically encompasses the most of the observations. It is important to notice that the situation is even worth when considering the density estimator implicitly proposed in [20], which is  $\check{f}_{\text{Getz}}(x) = \max_{X_i, x \in B(X_i,r)} \frac{\#\{\mathbb{X}_n \cap B(X_i,r)\}}{n|H(B(X_i,r)) \cap \mathbb{X}_n|_D}$ .

Let us introduce  $N_{x,r} = \#\{\mathbb{X}_n \cap B(x,r)\}$ ,  $C_{x,r} = H(B(x,r) \cap \mathbb{X}_n)$ ,  $N_{x,r}^\partial = \#\{\mathbb{X}_n \cap \partial C_{x,r}\}$  and  $N_{x,r}^o = N_{x,r} - N_{x,r}^\partial$ . Due to [3] we can propose the following LoCoH correction for the naive estimator:

$$\hat{f}_{r,A}(x) = \frac{N_{x,r}^o}{(n - N_{x,r}^\partial)|C_{x,r}|} \mathbb{I}_{|C_{x,r}|_D \geq A\omega_D r^D} \mathbb{I}_{N_{x,r}^\partial \leq n/2}. \quad (2)$$

#### Remarks:

1. The inclusion of the additional parameter  $A$  and the use of indicator functions is only intended to prevent division by zero.
2. The count of observations within  $\partial C_{x,r}$ , denoted as  $N_{x,r}^\partial$ , is non-zero. Indeed  $C_{x,r}$  is a convex polytope defined by vertices that are part of the observation sets, which also encompasses those within  $\partial C_{x,r}$ .
3. Although the initial presentation might appear intricate, the introduction of  $N_{x,r}^o$  and  $N_{x,r}^\partial$  is actually quite intuitive, as demonstrated by [3]. This is because  $N_{x,r}^o |C_{x,r} \sim \text{Binomial}(n - N_{x,r}^\partial, \int_{C_{x,r}} f(z) dz)$ .

Subject to relatively broad assumptions concerning  $f$  and  $S$ , the proposed estimator has nice theoretical properties, namely:

1. If  $f$  is Lipschitz continuous and bounded below on a sufficiently regular support  $S$ , the proposed estimator achieves pointwise minimax rates at any point in  $S$  even if  $S$  is unknown (see Theorem 1).
2. If  $f$  is  $\mathcal{C}^2$ , when  $D \leq 7$ , under broad assumptions regarding the support set  $S$ , the proposed estimator can achieve minimax rates in the interior of  $S$ , with additional guarantees at the boundary of  $S$  (see Theorem 2 and Corollary 1). This offers guarantees for level set estimation (even for level sets intersecting the boundary), see Theorem 4.

Assuming that  $S$  is a  $d$ -dimensional manifold with  $d < D$ , a common assumption when working with high-dimensional data-sets, our focus shifts towards **estimating a density on**

**a manifold.** Density estimation on a known manifold without a boundary was initially introduced in [26], where the conventional kernel density estimator was extended by substituting the Euclidean distance with the geodesic distance. In [22], a similar approach was employed, but kernels were applied to projections onto tangent spaces. The technique presented in [5] adapted the methodology from [22] to address potential boundary bias. Importantly, the prior knowledge of the boundary is not necessary in this context: the distance and direction from a point to the boundary can be estimated through a local barycenter shift. More recently, methods have been proposed for unknown manifolds without boundaries. The approach developed in [16] only requires knowledge of the dimension  $d$  of the support  $S$ , whereas in [4], the support  $S$  is entirely unknown. In section 3 we present an extension of the density estimator introduced in (2), designed to accommodate densities supported in a possibly unknown manifold that may have boundaries. We adopt the classical approach of using local projections onto estimated tangent spaces. When  $S$  represents a sufficiently regular compact manifold with either a regular boundary or no boundary, and  $f$  is Lipschitz continuous, we obtain  $L^1$  convergence (see Theorem 5).

The paper is organized as follows. Section 2 is devoted to the presentation of the main results in the full-dimensional context. Section 3 deals with the extension to the manifold setting. In Section 4 we present some numerical experiments. Finally, the proofs are given in Section 5.

## 2 Full dimensional context

### 2.1 Geometric assumptions and statistical model

Let  $\mathbb{X}_n = \{X_1, \dots, X_n\} \subset \mathbb{R}^D$ , be an i.i.d. sample drawn from a distribution supported in  $S$ , and characterized by an unknown density  $f$  that is uniformly continuous with respect to the Lebesgue measure. Additionally, we will assume certain regularity conditions on both  $S$  and  $f$ .

First, we define and discuss the geometric assumptions on  $S$ . We may either consider the ball standardness (Definition 1, illustrated in the left part of Figure 1) or the rolling ball condition (Definition 2, illustrated in the right part of Figure 1).

**Definition 1** (Ball-standardness). *A closed set  $E$  is  $(r_0, \delta)$ -ball standard if for all  $x \in E$  and all  $r \leq r_0$ ,  $B(x, r) \cap E$  is path connected and there exists  $y$  such that  $B(y, \delta r) \subset E \cap B(x, r)$ .*

Having a ball-standard support and a density bounded below by a positive constant imply the classical standardness of the distribution as introduced in [12]. This characteristic also accommodates the presence of corners (though not cusps). Opting for ball standardness rather than the traditional standardness separates the geometric aspects from the distributional ones, enabling a clearer distinction between the two components. Ball standardness guarantees consistency, but when aiming for convergence rates, more regular support yields significantly better results. Classically (as in [32]) we will characterize smooth support via the inside and outside rolling ball condition.

**Definition 2** (Inside and outside rolling ball condition). *A closed set  $E$  satisfies the  $r_0$ -inside and outside rolling ball condition if  $\overline{E} = E$  and, for all  $x \in \partial E$ , there exist two points  $O_x^{out}$  and  $O_x^{in}$  such that  $B(O_x^{out}, r_0) \cap E = \{x\}$  and  $B(O_x^{in}, r_0) \subset E$ .*

Such a condition has been extensively studied in [33]. It implies that  $\partial S$  is a  $\mathcal{C}^1$  manifold or, equivalently, that  $\partial S$  has a positive reach. It is easy to see that, if  $E$  satisfies the  $r_0$ -inside and outside rolling ball condition, then  $E$  and  $\overline{E}^c$  are  $(r_0, 1/2)$ -ball standard (see Proposition 4 and its proof in Appendix A).

Our LoCoH density estimator has minimax pointwise convergence rate at any point of  $S$  (even of the boundary) when considering the following model.

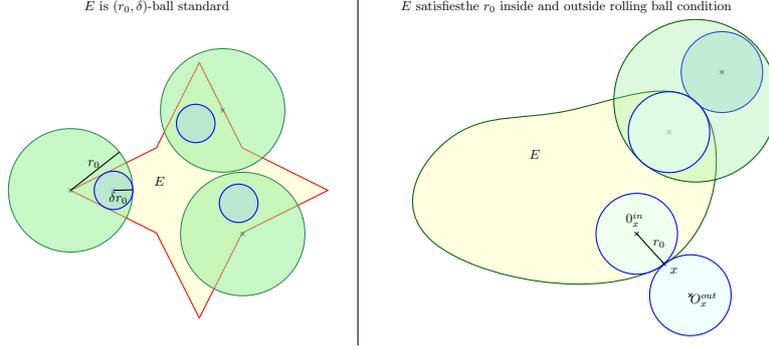


Figure 1:  $(r_0, \delta)$ -ball standardness (left) and  $r_0$ -inside an outside rolling ball condition (right).

**Definition 3.** A distribution belongs to the family model  $\mathcal{M}_1$  if:

1. it is supported in a compact support  $S$  that either is convex and ball standard, or has the  $r_0$ -inside and outside rolling ball condition for some  $r_0 > 0$ ,
2.  $f|_S$  is  $K$ -Lipschitz continuous and bounded below by a positive constant  $f_{\min}^{(0)}$ .

Ensuring consistency at boundary points necessitates shape assumptions on  $S$ . Indeed it's required that, as  $r$  goes to 0,  $|B(x, r) \cap S| \sim |H(B(x, r) \cap S)|$  for any  $x \in \partial S$ . This entails considering a combination of convexity and rolling ball conditions, as illustrated in Figure 2.

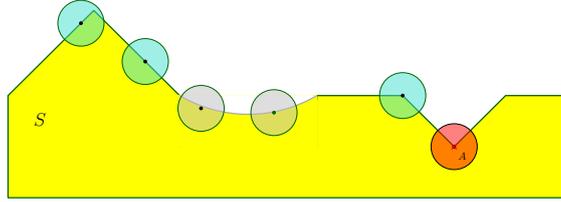


Figure 2: Convexity :  $H(B(x, r) \cap S) = B(x, r) \cap S$  for  $r$  small enough, when  $x$  belongs to blue part of  $\partial S$  (illustrated by the three blue balls). Rolling ball :  $|H(B(x, r) \cap S)| \sim |B(x, r) \cap S|$  when  $x$  belongs to the grey part of  $\partial S$  (illustrated with the two grey balls).  $|B(A, r) \cap S|$  and  $H(B(A, r) \cap S)$  are not similar when  $r \rightarrow 0$  (illustrated by the red ball).

The LoCoH estimator also achieves good convergence properties in  $\mathring{S}$  under the more classical hypotheses of  $\mathcal{C}^2$  regularity of the density and possibly unbounded support. Convergence is studied for distribution in the model  $\mathcal{M}_2$ , see Definition 5. Roughly speaking, the density must be  $\mathcal{C}^2$  with upper-bounded first derivatives and the support and "small level" sets must be ball-standard.

**Definition 4** ( $(r_0, \delta, \varepsilon_0)$ -standardness for a distribution). A distribution associated to a density  $f$  is  $(r_0, \delta, \varepsilon_0)$ -standard if, for all  $\varepsilon \in (0, \varepsilon_0)$ , the level set  $L_\varepsilon = \{x, f(x) > \varepsilon\}$  is  $(r_0, \delta)$ -ball standard.

The second model consists in  $(r_0, \delta, \varepsilon_0)$ -standard distributions with smooth enough densities.

**Definition 5.** A distribution belongs to the family model  $\mathcal{M}_2$  if:

1. The distribution is  $(r_0, \delta, \varepsilon_0)$ -standard,  $S$  is path connected and for all  $(x, y) \in S^2$  the geodesic distance (in  $S$ ) satisfies  $d_S(x, y) \leq K_{\max} \|x - y\|$ .
2. The associated density  $f$  is continuous on  $S$ , of class  $\mathcal{C}^2$  on  $\overset{\circ}{S}$  and:
  - (a) For all  $x \in S$ ,  $f(x) \leq f_{\max}^{(0)}$  and for all  $x \in \overset{\circ}{S}$ ,  $f(x) > 0$ .
  - (b) For all  $x \in S$ ,  $\|\nabla_f(x)\| \exp(3\|\nabla_{\ln(f)}(x)\|) \leq L_{\max}$ , where  $\nabla$  stands for the gradient. Such a condition implies that  $\|\nabla_f(x)\| \leq f_{\max}^{(1)}$  for all  $x$  in  $S$  (see Proposition 7 in Appendix C).
  - (c) For all  $x \in S$ ,  $\|\mathcal{H}_f(x)\|_{op} \leq f_{\max}^{(2)}$ , where  $\mathcal{H}_f(x)$  stands for the Hessian and  $\|\cdot\|_{op}$  for the operator norm.

Model  $\mathcal{M}_2$  inherently encompasses distributions supported in compact  $(r_0, \delta)$ -ball standard supports  $S$ , characterized by  $\mathcal{C}^2$  densities that exhibit a positive lower bound on  $S$ , and have a hessian matrix whose norm is bounded above by  $f_{\max}^{(2)}$ . This model further encompasses a diverse array of classical densities with non-compact supports, such as Gaussian or exponential distributions.

## 2.2 Main theoretical results for model $\mathcal{M}_1$

For bounded support, under regularity assumptions, and with a Lipschitz-continuous density bounded from below by a positive constant, the proposed estimator  $\hat{f}_{r_n, A}(x)$  exhibits  $L^2$  convergence. This is achieved since  $x$  is an element of  $S$ ,  $r_n$  tends to zero, and  $nr_n^D / \ln n$  tends to infinity. The general rate can be found in Equation (3), and by optimally choosing  $r_n = cn^{-\frac{1}{D+2}}$ , we achieve minimax  $L^2$  rates of order  $n^{-\frac{1}{D+2}}$  at any point  $x$  within  $S$ , as indicated in Equation (4). The advantage is that this correction doesn't require prior knowledge of the support. When the support is unknown, the only points that are guaranteed to belong to  $S$  are the observations, and Equation (5) claims that with high probability, the density estimation of every observation is close to the real density.

**Theorem 1.** *Let the distribution belong to the model  $\mathcal{M}_1$ . Let the window size  $r_n$  be a sequence such that  $r_n \rightarrow 0$  and  $nr_n^D / \ln n \rightarrow +\infty$ . Let  $A < \delta^D$  be a positive constant. There exist explicit constants depending only on the model such that,*

$$\text{for all } x \in S, \mathbb{E}((\hat{f}_{r_n, A}(x) - f(x))^2) \leq a_1(nr_n^D)^{-1} + a_2r_n^2 + a_3 \exp(-a_4nr_n^D) + a_5 \exp(-n/8). \quad (3)$$

$$\text{for all } x \in S, \text{ we have } \mathbb{E}((\hat{f}_{c_n^{-1/(D+2)}, A}(x) - f(x))^2) \leq a(c^2 + c^{-D})n^{-\frac{2}{D+2}}. \quad (4)$$

There exist explicit constants depending only on the model such that, with probability one, for  $n$  large enough,

$$\max_i \left\{ \left| \hat{f}_{r_n, A}(X_i) - f(X_i) \right| \right\} \leq b_1 \sqrt{\frac{\ln n}{nr_n^D}} + b_2 r_n. \quad (5)$$

## 2.3 Main theoretical results for Model $\mathcal{M}_2$

### 2.3.1 Pointwise $L^2$ convergence rates

We characterize the convergence with the use of the following notations.  $o_\alpha(1)$  (resp.  $O_\alpha(u_n)$ ) represents a sequence depending on  $\alpha$  and the constants of model  $\mathcal{M}_2$ , converging towards 0 (resp. depending on  $\alpha$  and the constants of model  $\mathcal{M}_2$  and being a  $O(u_n)$ ).

In Theorem 2 we establish that for points  $x \in \mathring{S}$ , the density estimator proposed in (2) exhibits a squared convergence rate of the order  $(nr_n^D)^{-1} + r_n^4 + r_n^2(nr_n^D)^{-\frac{4}{D+1}} + r_n^3(nr_n^D)^{-\frac{2}{D+1}}$ . This includes the "classical" density rate  $(nr_n^D)^{-1} + r_n^4$ , along with an additional term  $r_n^2(nr_n^D)^{-\frac{4}{D+1}} + r_n^3(nr_n^D)^{-\frac{2}{D+1}}$  coming from the estimation of LoCoH volume. Despite this added term, the advantage is that, under additional hypotheses on  $S$ , the bias is rectified at possible points of discontinuity located on the boundary of  $S$ . Moreover, as stated in Corollary 1, this supplementary term doesn't hinder the attainment of the minimax rate when  $D \leq 7$ .

**Theorem 2.** *Let the distribution belong to the model  $\mathcal{M}_2$ . Let the window size  $r_n$  be a sequence such that  $r_n \rightarrow 0$  and  $nr_n^D/l_n \rightarrow +\infty$ . Let  $A < \delta^D$  be a positive constant.*

1. For all  $x \in \mathring{S}$ ,  $\mathbb{E}((\hat{f}_{r_n, A}(x) - f(x))^2) = O_x\left((nr_n^D)^{-1} + r_n^4 + r_n^2(nr_n^D)^{-\frac{4}{D+1}} + r_n^3(nr_n^D)^{-\frac{2}{D+1}}\right)$ .
2. If  $S$  is either convex or has the  $R_0$  inside and outside rolling ball property, we have that, for all  $x \in S$ ,  $\mathbb{E}((\hat{f}_{r_n, A}(x) - f(x))^2) = O\left((nr_n^D)^{-1} + r_n^2\right)$ .

More explicit values of  $O$  can be found within proofs (see Equations (13) and (18)).

In Corollary 1, we present, for points in  $\mathring{S}$ , optimal radius size and associated convergence rates depending on the dimension.

**Corollary 1.** *Let the distribution belong to the model  $\mathcal{M}_2$ .*

1. if  $D \leq 7$ , the choice of  $r_n = cn^{-\frac{1}{D+4}}$  provide that, for all  $x \in \mathring{S}$ , the  $L^2$  error has the minimax convergence rate of order  $n^{-\frac{2}{D+4}}$ .
2. if  $D > 7$ , the choice of  $r_n = cn^{-\frac{2}{3D+1}}$  provide that, for all  $x \in \mathring{S}$ , the  $L^2$  error has a non-optimal convergence rate of order  $n^{-\frac{4}{3D+1}}$ .

It may be possible to develop a more sophisticated method that allows us to achieve the minimax rate of order  $n^{-2/(D+4)}$  everywhere, even at the boundary, by applying the following algorithm:

1. Compute  $\widehat{S}$  and  $\widehat{\partial S}$ , which are estimators of  $S$  and  $\partial S$  respectively.
2. If  $x \in \widehat{S}$  and  $d(x, \widehat{\partial S}) \geq h$  estimate the density with standard kernel estimator.
3. If  $x \in \widehat{S}$  and  $d(x, \widehat{\partial S}) < h$ , if possible, apply to our corrected estimator a high order bias correction near the boundary, similar to the proposition in section 3.3 in [5].

This approach is likely to yield methods with minimax optimal rates without any restrictions on the dimension. However, it comes with some computational and practical challenges. For instance:

1. Parameter tuning: this approach involves tuning more parameters. First, for the support estimation step, we need to choose an appropriate reference distance to the support's boundary to decide which density estimation method to apply. Furthermore, there are two density estimation parameters to be tuned.
2. Convex hull estimation: the convex hull estimation step may not be realistic when the dimension  $D$  is greater than 7.

Therefore, it might be more convenient to focus on the case where  $D \leq 7$  and use radius sequences of the form  $r_n = cn^{-\frac{1}{D+4}}$ . This choice is minimax for all points in the interior of  $S$  and remains consistent on all of  $S$ . Additionally, there are only two parameters to tune:  $A$  and  $r_n$ . Empirically, we have observed that the window size tuning method for the classical kernel works also very well for our modified version.

### 2.3.2 Uniform convergence rates and application to level set estimation

The previous section provided point-wise convergence rates for density estimation at a point  $x \in S$ . However, when  $S$  is unknown, these point-wise convergence rates do not allow us to achieve uniform convergence rates over the entire space  $\mathbb{R}^D$ . Nevertheless, because we know that all the observations are located on the set  $S$ , we can derive uniform convergence rates over the set of observations.

**Theorem 3.** *When  $D \leq 7$  and the distribution belongs to the model  $\mathcal{M}_2$ , let  $r_n = cn^{-1/(D+4)}$  and  $A < \delta^D$ .*

1. *For all  $t > 0$ , there exists  $C_t$  such that, with probability one, for  $n$  large enough:*

$$\max_i \left\{ \left| \hat{f}_{r_n, A}(X_i) - f(X_i) \right|, X_i \in L_t \text{ and } B(X_i, r_n) \subset S \right\} \leq C_t \sqrt{\ln n} n^{-2/(D+4)}.$$

2. *If  $f|_S \geq f_{\min}^{(0)} > 0$ , there exists  $C$  such that, with probability one, for  $n$  large enough:*

$$\max_i \left\{ \left| \hat{f}_{r_n, A}(X_i) - f(X_i) \right|, B(X_i, r_n) \subset S \right\} \leq C \sqrt{\ln n} n^{-2/(D+4)}.$$

3. *If  $S$  is either convex or satisfies the inside and outside  $R_0$  rolling ball condition, there exists  $C$  such that, with probability one, for  $n$  large enough:*

$$\max_i \left\{ \left| \hat{f}_{r_n, A}(X_i) - f(X_i) \right| \right\} \leq C n^{-1/(D+4)}.$$

This result allows us to employ the proposed density estimator for level set estimation using  $\hat{S}(\mathbb{X}_{r_n, A}^+(t))$  (where  $\hat{S}(\mathbb{Y})$  is a set estimator based on a sample  $\mathbb{Y}$  and  $\mathbb{X}_{r_n, A}^+(t) = \{X_i, \hat{f}_{r_n, A}(X_i) \geq t\}$ ). The chosen set estimator is the LoCoH estimator. This choice was made because it seems consistent to estimate both the density and the level sets with the same tool. It also pays homage to [20], who initially applied the LoCoH to create home range and core area maps. In theorem 4, we provide convergence rates for

$$\hat{L}_t = \bigcup_{X_i \in \mathbb{X}_{r_n, A}^+(t)} H(B(X_i, r_n) \cap \mathbb{X}_{r_n, A}^+(t)), \quad (6)$$

which differs from the original method proposed in [20]. The distinction between the two methods is discussed in the numerical study section. It's worth noting that the bias correction allows us to achieve convergence for level sets that may intersect the boundary of the support (as indicated in point 2 of Theorem 4). More precisely, we consider two kinds of levels. First, we deal with regular levels (see Definition 6) that satisfy the classical assumptions in the field of level set estimation, as in [27] for instance. In particular, regular levels do not intersect the boundary and, by application of Theorem 2 in [32], they have the inside and outside rolling ball property. Second, we consider standard levels (see Definition 7) that may intersect the boundary and are only ball standard.

**Definition 6** (Regular levels).  *$t > 0$  is a regular level if there exists  $\Delta_t > 0$  and  $m_t > 0$  such that*

$$L_{t-\Delta_t} \subset \overset{\circ}{S} \text{ and } \min_{z \in L_{t-\Delta_t} \setminus \overset{\circ}{L}_{t+\Delta_t}} \|\nabla f(z)\| = m_t > 0.$$

**Definition 7** (Standard levels).  $t > 0$  is a standard level if there exists  $\Delta_t, \beta_t, \delta_t, \gamma_t$  and  $\ell_t$  such that, for all  $r \leq \Delta_t$  and all  $x \in L_t$ ,  $|B(x, r) \cap L_t| \geq \delta_t \omega_D r^D$  and  $d_H(L_t, L_t \oplus rB) \leq \gamma_t r$ . Also, for all  $t'$  such that  $|t' - t| \leq \Delta_t$ , we have  $d_H(L_t, L_{t'}) \leq \beta_t |t - t'|$  and  $|\partial L_t|_{D-1} \leq \ell_t$ .

Roughly speaking, if  $t$  is a regular level, then all the  $t'$ -level sets,  $t'$  being close enough to  $t$ , have the rolling ball property with the same radius. The level  $t$  is considered "standard" if  $L_t$  is standard and partly expandable (as in [13]). Additionally, in a neighborhood of  $t$ , level sets are sufficiently close to  $L_t$ , and their perimeters are uniformly bounded. To illustrate this two kinds of level, consider for instance the density  $f(z) = c(2 - \|z\|^2)\mathbb{I}_{[-1,1]^2}(z)$ . For such a density, any levels  $t \in ]c, 2c[$  are regular and any levels  $t \in ]0, c[$  are standard.

**Theorem 4.** Suppose that  $D \leq 7$  and that the distribution belongs to the model  $\mathcal{M}_2$ , with  $S$  is either convex or have the inside and outside rolling ball property. Let  $r_n = cn^{-1/(D+4)}$  and  $A < \delta^D$ . Assume that  $\hat{S} = \{z, f(z) > 0\}$  and define  $\hat{L}_t$  as in Equation (6).

1. If  $t$  is a regular level, then, with probability 1 for  $n$  large enough,  $d_H(L_t, \hat{L}_t) = O(\ln nn^{-2/(D+4)})$ .
2. If  $t$  is a standard level and  $c \geq 4\beta_t C$  ( $C$  being the constant of Theorem 3 point 3.), with probability one, for  $n$  large enough,  $d_H(L_t, \hat{L}_t) = O(n^{-1/(D+4)})$ .

### 3 Density on manifolds

In this section, our aim is to generalize the results for the model  $\mathcal{M}_1$  to the density estimation of a distribution supported in a compact manifold. The associated model, see Definition 9, naturally focuses on lower-bounded and Lipschitz-continuous density (similar to model  $\mathcal{M}_1$ ). However, it requires additional regularity assumptions on the manifold  $S$  and its boundary  $\partial S$ , which are typically characterized by their reach.

**Definition 8.** Let  $S$  be a compact set. Its medial axis  $Med(S)$  and its reach  $\tau_S$  are defined by

$$Med(S) = \{z \in \mathbb{R}^D, \exists x \neq y \in S, \|x - z\| = \|y - z\| = d(z, S)\} \text{ and } \tau_S = \min_{x \in S} d(x, Med(S)).$$

Roughly speaking, manifolds with positive reach are regular enough to have "nice" tangent spaces (for detailed results on reach, see [17] and [18]). The rolling ball condition in the full dimensional context is generalized by manifolds whose boundary has a positive reach (see [1]). If  $\partial S = \emptyset$ , which is possible in the manifold case, we use the convention  $\tau_\emptyset = +\infty$ .

**Definition 9.** A distribution belongs to the family model  $\mathcal{M}'_d$  if

1.  $S$  is a compact  $d$ -dimensional manifold with positive reach  $\tau_S \geq \tau_{\min} > 0$ . If exists, its boundary  $\partial S$  is a  $(d - 1)$ -manifold without boundary and with reach  $\tau_{\partial S} \geq \tau_{\partial, \min} > 0$ .
2. Its density  $f$  satisfies that  $0 < f_{\min}^{(0)} \leq f(x) \leq f_{\max}^{(0)}$  for all  $x \in S$ . Also,  $f|_S$  is  $K$ -Lipshitz continuous.

Now, let's extend the density estimator by using local projections onto estimated tangent spaces. Suppose that, for all  $x \in S$ , we have  $\hat{T}_x$ , an estimation of  $T_x S$ . Let  $\pi_{\hat{T}_x}$  denote the orthogonal projection onto  $\hat{T}_x$ . Introduce the following notations for  $R \in \mathbb{R}$ :

1.  $\mathbb{X}_{n,R} = \pi_{\hat{T}_x}(B(x, R) \cap \mathbb{X}_n)$ , with as previously  $N_{x,R} = \#\{\mathbb{X}_n \cap B(x, R)\}$ ,
2.  $\hat{C}_{x,r,R} = H(\mathbb{X}_{n,R} \cap B(x, r))$ ,

$$3. \widehat{N}_{x,r,R} = \#\{\mathbb{X}_{n,R} \cap B(x,r)\}, \widehat{N}_{x,r,R}^\partial = \#\{\mathbb{X}_{n,R} \cap \partial\widehat{C}_{x,r,R}\} \text{ and } \widehat{N}_{x,r,R}^o = \widehat{N}_{x,r,R} - \widehat{N}_{x,r,R}^\partial.$$

The proposed LoCoH estimator is now:

$$\widehat{f}_{r,R,A}(x) = \frac{N_{x,R}}{n} \frac{\widehat{N}_{x,r,R}^o}{(N_{x,R} - \widehat{N}_{x,r,R}^\partial)|\widehat{C}_{x,r,R}|} \mathbb{I}_{|\widehat{C}_{x,r,R}| \geq A\omega_d r^d} \mathbb{I}_{\widehat{N}_{x,r,R}^\partial \leq N_{x,R}/2}. \quad (7)$$

The following theorem provides the convergence rates of the density estimator. In the manifold case, an additional error term is introduced, which depends on the quality of the tangent space approximation, denoted by  $\Theta_n$ . This quality is typically characterized by  $\angle\widehat{T}_x, T_x S$ , which represents the angle between the tangent space and its approximation. It can also be defined as the operator norm of  $\pi_{T_x S} - \pi_{\widehat{T}_x}$ .

**Theorem 5.** *Let the distribution belong to the model  $\mathcal{M}'_d$ . Let the window size  $r_n$  be a sequence such that  $r_n \rightarrow 0$  and  $nr_n^d/\ln n \rightarrow +\infty$ . Let  $A < 4^{-d}$  be a positive constant. Additionally suppose that there exists  $\Theta_n \rightarrow 0$  such that  $\angle T_x S, \widehat{T}_x \leq \Theta_n$ . Then there exists  $R_0$  such that, for all  $R < R_0$  there exists  $C$  such that  $\mathbb{E} \left( \left| \widehat{f}_{r_n,R,A}(x) - f(x) \right| \right) \leq C \left( \frac{1}{\sqrt{nr_n^d}} + r_n + \Theta_n \right)$ .*

When  $S \subset \mathbb{S}$ , where  $\mathbb{S}$  is a known  $d$ -manifold, we obtain a natural extension of Theorem 1 with  $\Theta_n = 0$ . This is the case, for instance, in the seismic example in the next section. When the support  $S$  is fully unknown, we can apply Theorem 5 at points  $x \in \mathbb{X}_n$ . Additionally, we need to estimate the tangent spaces with local PCA. This can be done with  $\Theta_n$  of the order  $(\ln n/n)^{1/d}$  (as described in [1]) therefore, the part of the convergence rates depending on  $\Theta_n$  can be considered negligible. Consequently, we obtain the minimax rate of order  $n^{-\frac{1}{d+2}}$ .

The additional parameter  $R$  is a macro parameter, distinct from the parameters used in local PCA tangent space estimation. Conceptually,  $R$  represents a radius such that, for all  $x \in S$ , the projection  $\pi_{T_x}|_{S \cap B(x,R)}$  is injective and possesses a boundary that is "sufficiently regular". As in [1], any radius  $R$  sufficiently small with regard to the reach of  $S$  and that of  $\partial S$  is suitable. In practice, it can be tuned as in [1] and approached in a similar manner (see Section 6.2.2 of [1] for details).

Extension to  $L^2$  convergence is still feasible but more technical. Also extending the results of model  $\mathcal{M}_2$  to the manifold setting is possible assuming  $\mathcal{C}^2$  manifolds and  $\mathcal{C}^2$  densities. The proofs, while a much more technical than those of Theorem 5, follow a similar structure. However, extending level set estimation to the manifold setting is considerably more challenging. This is because level sets on manifolds have boundaries, and only the method proposed in [1] could be directly applied to density estimation with little additional work. It is believed that the LoCoH still provides a viable estimator in this case, but rigorous proofs are required to establish its performance.

## 4 Numerical Experiments

### 4.1 Computation and complexity

In languages such as MATLAB, where the convex hull is already programmed, computation is particularly easy. In MATLAB, Let  $Y$  be the set of points at distance at most  $r$  from  $x$ , the function  $[F, \text{Vol}] = \text{convhulln}(Y)$  returns  $|C_{x,r}| = \text{Vol}$ ,  $N_{x,r}^\partial = \text{size}(\text{Unique}(F), 1)$  and  $N_{x,r}^o = \text{size}(Y, 1) - N_{x,r}^\partial$ . The complexity is upper-bounded by the computational complexity of the Delaunay triangulation of  $Y$  from which we can obtain the volume as the sum of the volumes of simplices (with

straightforward formulas) and the set of boundary observations. The computational time of  $\hat{f}_{r_n, A}(x)$  is then typically of the order  $2^D(N_{x,r} \log(N_{x,r}) + N_{x,r}^{D/2})$ . When computing the density for the entire observation set, the computational time is of order  $2^D(n^2 r^D \log(n) + n(nr^D)^{D/2})$  (notice that  $N_{x,r} = O(nr^d)$ ). For radius sequences  $r = cn^{-1/(D+2)}$  and  $r_n = cn^{-1/(D+4)}$ , respectively, the complexity is less than  $2^D n^2$  and less than  $2^D n^3$ , respectively.

## 4.2 Experiments on simulated data

First, we propose to test the LoCoH density estimator on a sample drawn from a Gaussian distribution restricted to the square  $[0, 1]^2$ . Specifically, in this first example, the density is  $f(x, y) = c \exp(-2(x^2 + y^2))$ . In Figure 3, we present a comparison of the LoCoH density estimator  $\hat{f}$  and  $f_{r_n}^*(x) = N_{x, r_n} / (n \omega_D r_n^D)$ , the classical kernel estimator, on a sample of size  $n = 5000$ . In panels (a) and (b) on the left, we present the results associated with the LoCoH density estimator ( $r = 0.18$  and  $A = 0.1$ ). The top figure (a) presents the sample cloud with colors corresponding to the rank of the estimated density value (isopleth presentation). The bottom figure (b) presents the associated level sets estimated with the LoCoH method (also with  $r = 0.18$ ) for levels in  $\{0, 0.5, 1, 1.5, 2, 2.5\}$ . In the same figure, panels (c) and (d) illustrate the results with the kernel density estimator  $f^*$  (using the same window size). In Figure 3, we can observe that, as expected, the LoCoH correction improves the density estimation near the boundary of the support.

To investigate this improvement further, we also tested the proposed method on data drawn from a uniform distribution within the unit disk. The results are presented in Figure 4. We experimented with different sample sizes, specifically  $n \in \{1000, 2000, 3000, 5000, 7000, 10000\}$ . The tested parameters were set to  $r_n = 0.5n^{-1/6}$  and  $A = 1/8$ . For each sample size, we conducted 100 replications. For each replication  $k$ , we computed  $Q_\alpha(k)$ , the  $\alpha$ -quantile (on  $i$ ) of  $|\hat{f}_{r_n, A}(X_i) - f(X_i)|$ . We then plotted the results as follows: the median (on  $k$ ) of  $Q_\alpha(k)$  is shown in solid blue lines, the 25th and 75th percentiles (on  $k$ ) of  $Q_\alpha(k)$  are presented as blue dashed lines, and blue points represent the maximum and minimum (on  $k$ ) of  $Q_\alpha(k)$ . The red curves display the same quantities calculated for  $f_{r_n}^*$ , which is the classical uniform kernel method with the same radius. Since the LoCoH correction primarily affects data points near the boundary, we can observe differences between the blue and red curves for high values of  $\alpha$  (e.g., 100%, 99%, and 95% quantiles). Moreover, as the sample size  $n$  increases, the corrected estimator outperforms the classical kernel estimator. For "small" quantiles (e.g., 75% and 50%), we can observe the similarity between the two estimators, reflecting their similarity at points far from the boundary.

## 4.3 Examples on real data

### 4.4 Comparison with standard KDE : Camping and seasonal rental

An example based on camping and seasonal rentals in France (see Figures 5 and 6) illustrates how LoCoH density estimation can correct density estimations at highly touristic boundary locations. The two density maps (LoCoH vs classical KDE) are presented in Figure 5, with the LoCoH method on the left showing much higher contrast. To be more precise on difference between the two density estimations, in Figure 6 we present the relative difference between  $\hat{f}$  and  $f^*$ . As expected, we observe more differences in the density estimation, along the boundary of France, especially at its corners, and in specific cities with significant tourist or economic interest, where seasonal lodging of the whole surrounding is concentrated (such as Toulouse, Bordeaux, or Rennes, for instance).

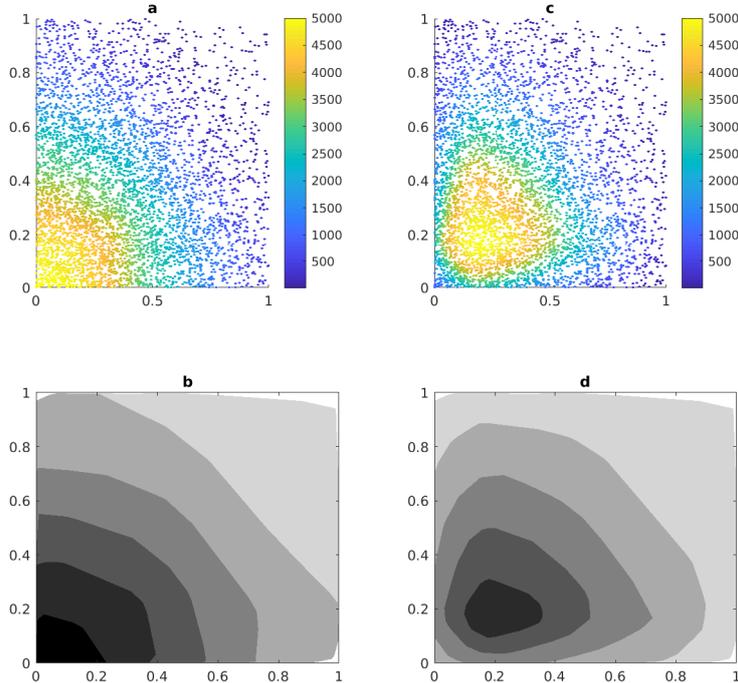


Figure 3: Comparison of Density Estimation Methods for a Gaussian Restricted to a Square. Panels (a) and (b) show the LoCoH estimator, while panels (c) and (d) depict the results with the classical kernel method. In the top graphs, the sample cloud is colored based on the estimated density rank value. In the bottom graphs, we present the level sets estimated using the LoCoH method for levels  $\{0, 0.5, 1, 1.5, 2, 2.5\}$ .

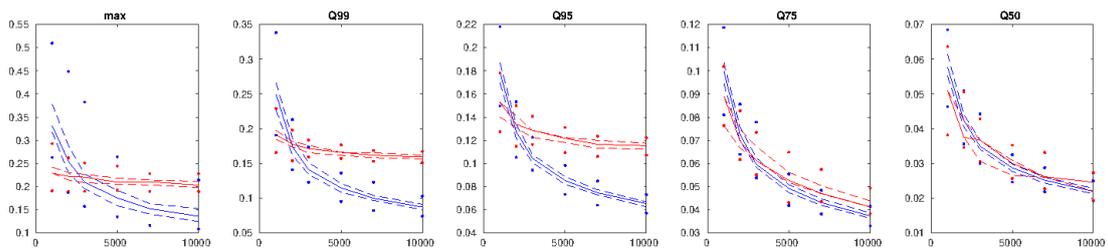


Figure 4: Comparison of density estimation methods for samples uniformly drawn on a disk. Computation of different quantiles of  $|\hat{f}_{T_n, A}(X_i) - f(X_i)|$  with the LoCoH estimator we propose in blue and the usual kernel one in red.

#### 4.4.1 Comparison with standard LoCoH : Panther Jitter

In this section, we aim to compare the original LoCoH method proposed by [20] with our proposed LoCoH correction. As a tribute to Getz and Wilmer, we test both approaches for core-area esti-

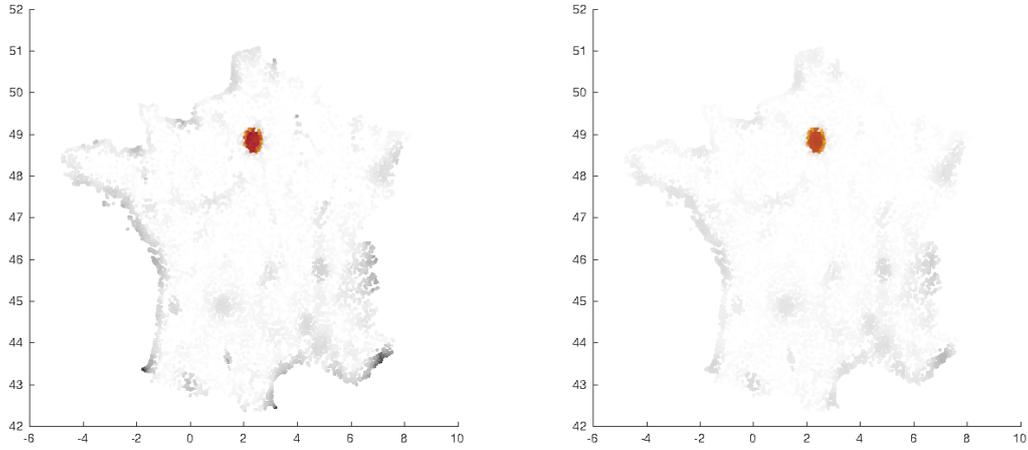


Figure 5: Density Estimation for Camping and Seasonal Rental. On the left, the LoCoH density estimator; on the right, the classical KDE. The two densities are presented with the same scale of grey or brown (for Paris and its surrounding).

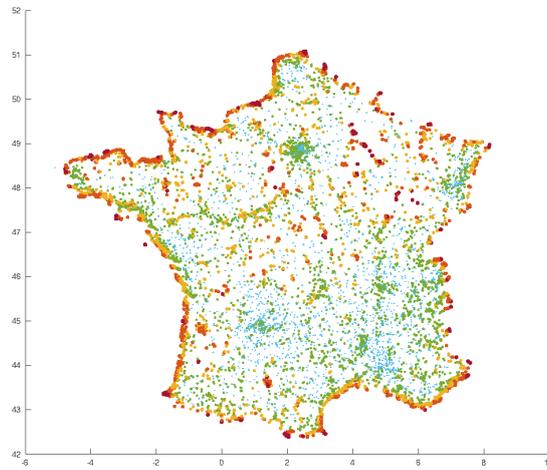


Figure 6:  $(\hat{f} - f^*)/f^*$  presented in five color level : less than 20% blue, between 20% and 50% green, between 50% and 100% mustard, between 100% and 200% orange and more than 200% brown

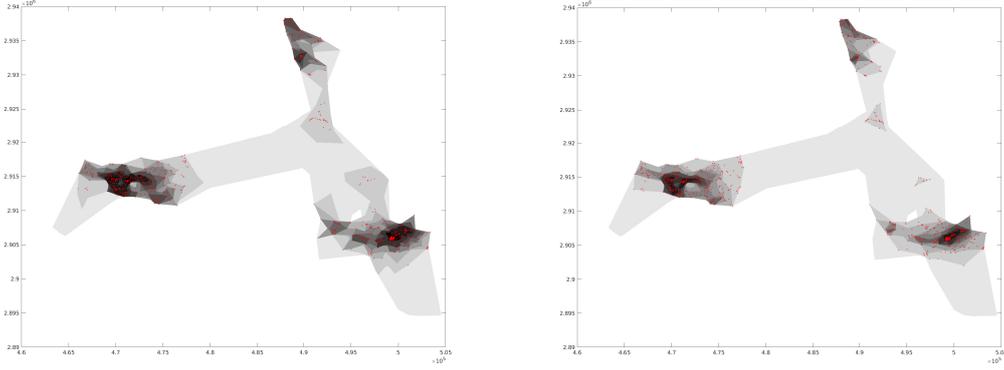


Figure 7: Isopleth for Panther Jitter, left original method from [20] and right the proposed corrected method,

mation on the Panther Jitter dataset (downloaded from the Penn State Department of Ecosystem Science and Management). In Figure 7, we present the  $x\%$  isopleths, where  $x$  is chosen from  $\{0, 10, 20, 30, 40, 50, 60, 70, 80, 90\}$ . On the left, we apply the original LoCoH method and, on the right, we apply our proposed LoCoH correction.

Let recall that the level sets in [20] are estimated via

$$\check{L}_t^{\text{Getz}} = \bigcup_{i, \frac{\#\{X_n \cap B(X_i, r)\}}{|H(B(X_i, r) \cap X_n)|} \geq nt} H(B(X_i, r) \cap X_n).$$

Note that with the estimator we propose, the estimated levels are more focused on the data points, whereas the original estimator extends the different levels quite away from the data points (see *e.g.* mid-grey levels).

#### 4.4.2 A manifold example

In Figure 8, we illustrate the manifold method. In this case, the data consists of earthquake localizations on Earth, specified by latitude and longitude (a known 2-dimensional manifold). We only kept earthquakes with a magnitude greater than 3. We conducted computations using the LoCoH method proposed in Section 3, assuming a known manifold and real tangent space.

The two graphs depict a projection on a planisphere of the data sample, with colors indicating the rank of the density. The top graph illustrates the LoCoH method, while the second one employs the constant kernel for manifolds. Although the two densities appear quite similar, some differences still exist. For instance, our correction allows for more mass to be placed on the Mid-Atlantic Ridge, where the density is underestimated in the classical kernel method due to its proximity to a lower-dimensional structure. Many other estimation differences can be observed all along the different ridges. This highlights the accuracy of our new LoCoH estimator in the manifold case.

Additionally, in the circled area, we can observe a density maximum located in a corner of the support. This would typically be impossible without bias correction.

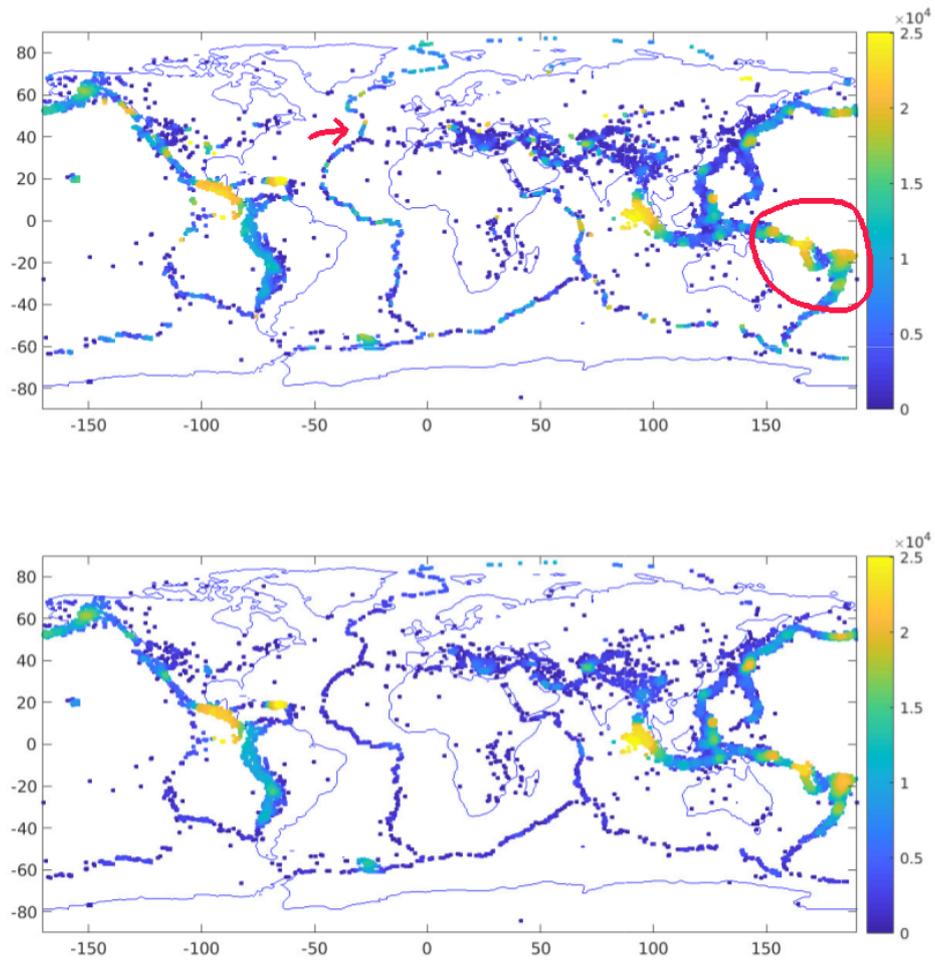


Figure 8: Earthquake density with two methods : the LoCoH for known manifold (top) and classical kernel for known manifold (down).

## 5 Proofs

In the proofs, we make use of the following notations. For a set  $A$  and a real number  $\rho$ , the  $\rho$  offset of  $A$ , denoted by  $A \oplus \rho B$ , is defined as  $A \oplus \rho B = \{x \in \mathbb{R}^D, d(x, A) \leq \rho\}$ . For a compact set  $A$ , when well defined, we denote by  $\pi_A$  the projection onto  $A$  which is defined as follows:  $\pi_A(x)$  is the point in  $A$  such that  $\|\pi_A(x) - x\| = d(x, A)$ . When  $A$  is convex,  $\pi_A$  is defined everywhere ; when  $A$  has the  $r_0$  inside and outside rolling ball property,  $\pi_A$  is defined on  $A \oplus r_0 B$  and  $\pi_{\partial A}$  on  $\partial A \oplus r_0 B$ .

### 5.1 Notations and definitions

First, for the sake of clarity, we will use  $r$  instead of  $r_n$  in this section. The probability content is  $\Gamma_{x,r} = \int_{B(x,r)} f(z) dz$  and its approximation with the help of LoCoH is  $\tilde{\Gamma}_{x,r} = \int_{C_{x,r}} f(z) dz$ .

Due to the indicator function in the definition of the density estimation we introduce the event  $E_{x,r} = \{|C_{x,r}| \geq A\omega_D r^D, \tilde{\Gamma}_{x,r} \geq f(x)A\omega_D r^D \text{ and } N_{x,r}^\partial \leq n/2\}$ . Conditionally to  $E_{x,r}$ , the error decomposition in the density estimation is  $\hat{f}_{r,A}(x) - f(x)|E_{x,r} = \varepsilon_1(x) + \varepsilon_2(x)$ , where:

$$\varepsilon_1(x) = \frac{N_{x,r}^o - \tilde{\Gamma}_{x,r}(n - N_{x,r}^\partial)}{|C_{x,r}|(n - N_{x,r}^\partial)} \mathbb{I}_{E_{x,r}} \quad \text{and} \quad \varepsilon_2(x) = \left( \frac{\tilde{\Gamma}_{x,r}}{|C_{x,r}|} - f(x) \right) \mathbb{I}_{E_{x,r}}. \quad (8)$$

Roughly speaking,  $\varepsilon_1$  is the variance term of the error and  $\varepsilon_2$  the bias one. We will make use of the following error decompositions:

$$\mathbb{E}((\hat{f}_{r,A}(x) - f(x))^2) \leq \mathbb{E}(\varepsilon_1^2(x)|E_{x,r}) + \mathbb{E}(\varepsilon_2^2(x)|E_{x,r}) + 2\mathbb{E}(\varepsilon_1(x)\varepsilon_2(x)|E_{x,r}) + f(x)^2\mathbb{P}(E_{x,r}^c), \quad (9)$$

$$\mathbb{P}(|\hat{f}_{r,A}(x) - f(x)| \geq t_1 + t_2) \leq \mathbb{P}(|\varepsilon_1(x)| \geq t_1|E_{x,r}) + \mathbb{P}(|\varepsilon_2(x)| \geq t_2|E_{x,r}) + \mathbb{P}(E_{x,r}^c). \quad (10)$$

### 5.2 Bound on $\mathbb{P}(E_{x,r}^c)$ and rough bounds on $\varepsilon_1$ and $\varepsilon_2$

**Lemma 1** (Bound on  $\mathbb{P}(E_{x,r}^c)$ ). *Suppose that the distribution is  $(r_0, \delta, \varepsilon_0)$ -standard and the density  $f$  is upper-bounded by  $f_{\max}^{(0)}$ . For all  $A < \delta^D$  and  $r < \min(r_0, (4\omega_D)f_{\max}^{(0)})^{-1/D}$ , there exist  $\beta_1$  and  $\beta_2$  positive constants such that, for all  $x$  with  $f(x) > 0$ , we have*

$$\mathbb{P}(E_{x,r}^c) \leq \beta_1 \exp(-nr^D \min(f(x), \varepsilon_0)\beta_2) + \exp(-n/8).$$

*Proof.* First, tackle the condition on  $N_{x,r}^\partial$ . Remark that  $\mathbb{P}(N_{x,r}^\partial > n/2) \leq \mathbb{P}(N_{x,r} > n/2)$ . Since  $N_{x,r} \sim \text{Binom}(n, \Gamma_{x,r})$  with  $\Gamma_{x,r} < 1/4$ , by condition  $r \leq (4\omega_D f_{\max}^{(0)})^{-1/D}$ , we have that

$$\mathbb{P}(N_{x,r} > n/2) = \mathbb{P}\left(N_{x,r} - n\Gamma_{x,r} > n\left(\frac{1}{2} - \Gamma_{x,r}\right)\right) \leq \mathbb{P}(N_{x,r} - n\Gamma_{x,r} > n/4).$$

Then, by use of Hoeffding, we obtain  $\mathbb{P}(N_{x,r}^\partial > n/2) \leq \mathbb{P}(N_{x,r} > n/2) \leq \exp(-n/8)$ .

Let now  $L_t$  (with  $t > 0$ ) be a level set that has the  $(\delta, r_1)$ -inside rolling ball property. We aim at proving that, for all  $x \in L_t$ , we have that

$$\mathbb{P}\left(\frac{|C_{x,r}|}{r^D} \leq A\omega_D \text{ or } \tilde{\Gamma}_{x,r} \leq \omega_D r^D A t\right) \leq C_D 4^D \frac{\exp\left(-n\omega_D t \left(\left(1 - \frac{A^{1/D}}{\delta}\right) \frac{\delta r}{4}\right)^D\right)}{\left(1 - \frac{A^{1/D}}{\delta}\right)^D}, \quad (11)$$

where  $C_D$  is the "covering constant" of a  $D$ -Dimensional Ball.

Notice that  $C_{y,\delta r} \subset C_{x,r}$ , where  $C_{y,\delta r}$  denotes the convex hull of the observations in  $B(y, \delta r)$  the ball of the "ball-standardness" assumption ( $B(y, \delta r) \subset B(x, r) \cap L_t$ ). Set  $\alpha \in ]0, 1[$ . Suppose that  $B(y, \alpha\delta r)$  is not included in  $C_{y,\delta r}$ . Thus, there exists  $y_1$  in  $B(y, \delta\alpha r)$  which is not in  $C_{y,\delta r}$ . Because  $y_1$  is not in  $C_{y,\delta r}$ , there exists  $u_1$  a unit vector such that, for all observation  $X_i$  in  $B(y, \delta r)$ , we have  $\langle X_i - y_1, u_1 \rangle < 0$ . Introduce  $z = y + \frac{\langle y_1 - y, u_1 \rangle + \delta r}{2} u_1$ . We have that  $z$  belongs to  $B(y, \delta r)$  and  $B(z, (1-\alpha)\delta r/2)$  contains no observation. We have then proved that, if  $B(y, \alpha\delta r)$  is not included in  $C_{y,\delta r}$ , there exists  $z \in B(y, \delta r)$  with  $B(z, (1-\alpha)\delta r/2) \cap \mathbb{X}_n = \emptyset$  (see Figure 9).

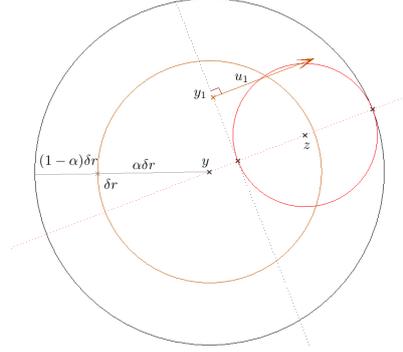


Figure 9: Construction of  $z$

Now, deterministically cover  $B(y, \delta r)$  with  $\nu \leq C_D 4^D (1-\alpha)^{-D}$  balls of radius  $(1-\alpha)\delta r/4$  centered at points  $z'_1, \dots, z'_\nu$ . It clearly comes that, if  $B(y, \alpha\delta r)$  is not included in  $C_{y,\delta r}$ , there exists  $z'_i \in B(y, \delta r)$  with  $B(z'_i, (1-\alpha)\delta r/4) \cap \mathbb{X}_n = \emptyset$ . Thus, we have:

$$\mathbb{P}(B(y, \alpha\delta r) \not\subset C_{y,\delta r}) \leq \sum_{i=1}^{\nu} \mathbb{P}(B(z'_i, (1-\alpha)\delta r/4) \cap \mathbb{X}_n = \emptyset),$$

$$\mathbb{P}(B(y, \alpha\delta r) \not\subset C_{y,\delta r}) \leq C_D 4^D \frac{(1 - \omega_D t ((1-\alpha)\delta r/4)^D)^n}{(1-\alpha)^D} \leq C_D 4^D \frac{\exp(-n\omega_D t ((1-\alpha)\delta r/4)^D)}{(1-\alpha)^D}.$$

Because  $B(y, \alpha\delta r) \subset C_{y,\delta r} \subset C_{x,r}$  implying both that  $|C_{x,r}| \geq \omega_D (\alpha\delta r)^d$  and  $\tilde{\Gamma}_{x,r} \geq \omega_D (\alpha\delta r)^d t$ , we get

$$\mathbb{P}\left(\frac{|C_{x,r}|}{r^D} \leq \omega_D (\delta\alpha)^D \text{ or } \tilde{\Gamma}_{x,r} \leq \omega_D (\delta\alpha)^D t\right) \leq C_D 4^D (1-\alpha)^{-D} \exp(-n\omega_D t ((1-\alpha)\delta r/4)^D).$$

And so, if  $A < \delta^D$ , with  $\alpha = A^{1/D}/\delta$ , we obtain (11).

Let now  $x$  be such that  $f(x) > 0$ . If  $f(x) < \varepsilon_0$ , by application of (11) with  $t = f(x)$ , we have the existence of positive constants  $\beta_1$  and  $\beta_2$  such that  $\mathbb{P}\left(\frac{|C_{y,\delta r}|}{r^D} \leq A\omega_D \text{ or } \tilde{\Gamma}_{x,r} \leq \omega_D r^D A f(x)\right) \leq \beta_1 \exp(-nr^D f(x)\beta_2)$ . If  $f(x) \geq \varepsilon_0$ , also by application of (11) with  $t = \varepsilon_0$ , to have

$$\mathbb{P}\left(\frac{|C_{y,\delta r}|}{r^D} \leq A\omega_D \text{ or } \tilde{\Gamma}_{x,r} \leq \omega_D r^D \varepsilon_0 A\right) \leq \beta_1 \exp(-nr^D \varepsilon_0 \beta_2)$$

That concludes the proof under model  $\mathcal{M}_2$ .  $\square$

**Lemma 2** (Bounds on  $\varepsilon_1$ ). *Suppose that the density is upper-bounded by  $f_{\max}^{(0)}$ . Then we have that*

$$\mathbb{E}(\varepsilon_1(x)\varepsilon_2(x)|E_{x,r}) = 0, \quad (12)$$

$$\mathbb{E}(\varepsilon_1^2(x)|E_{x,r}) \leq \frac{2 \max_{z \in C_{x,r}}(f(z))}{A\omega_D r^D n} \leq \frac{2f_{\max}^{(0)}}{A\omega_D r^D n}, \quad (13)$$

$$\mathbb{P}\left(|\varepsilon_1(x)| > \sqrt{\frac{32 \ln n f_{\max}^{(0)}}{A^2 \omega_D n r^D}} |E_{x,r}\right) \leq 2n^{-4}(1 + o(1)). \quad (14)$$

*Proof.* Recall that

$$\varepsilon_1(x) = \frac{N_{x,r}^o - \tilde{\Gamma}_{x,r}(n - N_{x,r}^\partial)}{|C_{x,r}|(n - N_{x,r}^\partial)} \mathbb{I}_{E_{x,r}}.$$

See [3] for the possibility to define  $N_{x,r}^o|C_{x,r}$ . Given  $C_{x,r}$ , the quantities  $|C_{x,r}|$ ,  $\tilde{\Gamma}_{x,r}$  and  $\varepsilon_2(x)$  are known. As a first conclusion  $\mathbb{E}(\varepsilon_1(x)\varepsilon_2(x)|C_{x,r}) = \varepsilon_2(x)\mathbb{E}(\varepsilon_1(x)|C_{x,r})$ . Also, with probability one,  $N_{x,r}^\partial$  is the number of vertices of the convex polygon  $C_{x,r}$  and thus is also known. So, almost surely,  $N_{x,r}^o|C_{x,r} \sim \text{Binom}(n - N_{x,r}^\partial, \tilde{\Gamma}_{x,r})$ . We thus also have  $\mathbb{E}(\varepsilon_1(x)|E_{x,r}) = \frac{\mathbb{E}(N_{x,r}^o|E_{x,r}) - \tilde{\Gamma}_{x,r}(n - N_{x,r}^\partial)}{|C_{x,r}|(n - N_{x,r}^\partial)} = 0$ . We then have  $\mathbb{E}(\varepsilon_1(x)\varepsilon_2(x)|C_{x,r}) = \varepsilon_2(x)\mathbb{E}(\varepsilon_1(x)|C_{x,r}) = 0$ , which proves (12). We also have  $\mathbb{E}(\varepsilon_1^2(x)|C_{x,r}) = \frac{\mathbb{V}(N_{x,r}^o|E_{x,r})}{|C_{x,r}|^2(n - N_{x,r}^\partial)^2} = \frac{\tilde{\Gamma}_{x,r}(1 - \tilde{\Gamma}_{x,r})}{|C_{x,r}|^2(n - N_{x,r}^\partial)^2}$ . Then, due to  $\frac{\tilde{\Gamma}_{x,r}(1 - \tilde{\Gamma}_{x,r})}{|C_{x,r}|} \leq \frac{\tilde{\Gamma}_{x,r}}{|C_{x,r}|} \leq \max_{C_{x,r}}(f(z))$ , it comes that  $\mathbb{E}(\varepsilon_1^2(x)|C_{x,r}) \leq \frac{\max_{z \in B(x,r)} f(z)}{n - N_{x,r}^\partial} \frac{1}{|C_{x,r}|}$ . Thus we clearly roughly obtain

$$\mathbb{E}(\varepsilon_1^2(x)|E_{x,r}) \leq \frac{2 \max_{z \in B(x,r)} f(z)}{A n \omega_D r^D} \leq \frac{2f_{\max}^{(0)}}{A n \omega_D r^D},$$

which proves (13). By Bennett's inequality, it comes that:

$$\mathbb{P}\left(|N_{x,r}^o - \tilde{\Gamma}_{x,r}(n - N_{x,r}^\partial)| > \sqrt{16 \ln n (n - N_{x,r}^\partial) \tilde{\Gamma}_{x,r}} |C_{x,r}\right) \leq 2 \exp\left(- (n - N_{x,r}^\partial) \tilde{\Gamma}_{x,r} h\left(\sqrt{\frac{16 \ln n}{(n - N_{x,r}^\partial) \tilde{\Gamma}_{x,r}}}\right)\right),$$

with  $h(u) = (1 + u) \ln(1 + u) - u$ . Because  $h(u) \geq u^2/4$  when  $u \leq 4$ , we get

$$\mathbb{P}\left(|\varepsilon_1(x)| > \sqrt{\frac{32 \ln n f_{\max}^{(0)}}{A^2 \omega_D n r^D}} |E_{x,r} \text{ and } \tilde{\Gamma}_{x,r} \geq \frac{8 \ln n}{n}\right) \leq 2n^{-4}. \quad (15)$$

To conclude the proof we aim at bounding the probability of large  $\varepsilon_1$  given  $E_{x,r}$  and  $\tilde{\Gamma}_{x,r} \leq \frac{8 \ln n}{n}$  (that could happen for points  $x$  with  $f(x) \leq 8 \ln n / (A\omega_D n r^D)$ ). Recall that, given  $C_{x,r}$ ,  $N_{x,r}^\partial$  has a binomial law. Then, by use of first Bernstein inequality, we have:

$$\mathbb{P}\left(|N_{x,r}^o - \tilde{\Gamma}_{x,r}(n - N_{x,r}^\partial)| > t |C_{x,r}\right) \leq 2 \exp\left(- \frac{t^2}{2\tilde{\Gamma}_{x,r}(1 - \tilde{\Gamma}_{x,r})(n - N_{x,r}^\partial) + 2t/3}\right).$$

Introduce  $n^\partial = n - N_{x,r}^\partial$  and choose  $t = \sqrt{a'n^\partial \ln n r^D}$  to obtain

$$\mathbb{P}\left(|N_{x,r}^o - n^\partial \tilde{\Gamma}_{x,r}| > \sqrt{a'n^\partial r^D \ln n} |C_{x,r}\right) \leq 2 \exp\left(- \frac{a'n^\partial r^D \ln n}{2n^\partial \tilde{\Gamma}_{x,r}(1 - \tilde{\Gamma}_{x,r}) + 2\sqrt{a'n^\partial r^D \ln n}/3}\right).$$

Thus (division by  $n^\partial |C_{x,r}|$  in the left term and trivial bound in the right one) we have

$$\mathbb{P} \left( |\varepsilon_1(x)| > \sqrt{\frac{a' r^D \ln n}{n^\partial |C_{x,r}|^2}} |C_{x,r}| \right) \leq 2 \exp \left( -\frac{a' n^\partial r^D \ln n}{2n^\partial \tilde{\Gamma}_{x,r} + 2\sqrt{a' n^\partial r^D \ln n}/3} \right).$$

Now, choose  $a' = 16\omega_D f_{\max}^{(0)}$  and notice that, given  $E_{x,r}$ ,  $n^\partial \geq n/2$ . It comes

$$\mathbb{P} \left( |\varepsilon_1(x)| > \sqrt{\frac{32f_{\max}^{(0)} \ln n}{A^2 \omega_D n r^D}} \Big| E_{x,r} \text{ and } \tilde{\Gamma}_{x,r} \leq \frac{8 \ln n}{n} \right) \leq 2 \exp \left( -\frac{a' n r^D \ln n}{16 \ln n + \frac{2}{3} \sqrt{2a' r^D n \ln n}} \right).$$

And, because  $n r^D / \ln n \rightarrow +\infty$ , we obtain

$$\mathbb{P} \left( |\varepsilon_1(x)| > \sqrt{\frac{32f_{\max}^{(0)} \ln n}{A^2 \omega_D n r^D}} \Big| E_{x,r} \text{ and } \tilde{\Gamma}_{x,r} \leq \frac{8 \ln n}{n} \right) \leq 2 \exp \left( -\frac{3}{2} \sqrt{a' \ln n n r^D} \right) (1 + o(1)).$$

Apply again  $\frac{n r^D}{\ln n} \rightarrow +\infty$  to have that  $\mathbb{P} \left( |\varepsilon_1(x)| > \sqrt{\frac{32f_{\max}^{(0)} \ln n}{A^2 \omega_D n r^D}} \Big| E_{x,r} \text{ and } \tilde{\Gamma}_{x,r} \leq \frac{8 \ln n}{n} \right) = o(n^{-4})$ .

That, together with (15), concludes the proof of (14).  $\square$

**Lemma 3** (Rough bound on  $\varepsilon_2$ ). *Suppose that the support  $S$  of the distribution is either convex or has the  $R_0$ -inside and outside rolling ball condition. Also suppose that the density  $f$  is  $K$ -Lipschitz continuous on  $S$  and upper-bounded by  $f_{\max}^{(0)}$ . Then, if  $S$  is convex, we have*

$$\mathbb{P} (|\varepsilon_2(x)| > rK |E_{x,r}) = 0. \quad (16)$$

And, if  $S$  has the  $R_0$ -inside and outside rolling ball property, we have

$$\mathbb{P} \left( |\varepsilon_2(x)| > r \left( K + \frac{4f_{\max}^{(0)} \omega_{D-1}}{3AR_0 \omega_D} \right) \Big| E_{x,r} \right) = 0. \quad (17)$$

*Proof.* Recall that

$$\varepsilon_2(x) = \left( \frac{\tilde{\Gamma}_{x,r}}{|C_{x,r}|} - f(x) \right) \mathbb{I}_{E_{x,r}},$$

$$|\varepsilon_2(x)| = \frac{\left| \int_{C_{x,r}} (f(y) - f(x)) dy \right|}{|C_{x,r}|} \leq \frac{\left| \int_{C_{x,r} \cap S} (f(y) - f(x)) dy \right|}{|C_{x,r}|} + \frac{\left| \int_{C_{x,r} \setminus S} (f(y) - f(x)) dy \right|}{|C_{x,r}|}.$$

Thus, we have  $|\varepsilon_2(x)| \leq rK + f(x) \frac{|C_{x,r} \setminus S|}{|C_{x,r}|}$ . Then, roughly bounding the second term given  $E_{x,r}$ , we get  $|\varepsilon_2(x)| \leq rK + f(x) \frac{|H(B(x,r) \cap S) \setminus S|}{A \omega_D r^D}$ . Obviously, if  $S$  is convex, we obtain  $|\varepsilon_2(x)| \leq r f_{\max}^{(1)}$ . When  $S$  has the rolling ball condition, by application of proposition 5 (see Appendix B) it comes that  $|\varepsilon_2(x)| \leq r(f_{\max}^{(1)} + f(x) \frac{4\omega_{D-1}}{3AR_0 \omega_D})$ .  $\square$

### 5.3 Proof of Theorem for model $\mathcal{M}_1$

*Proof of Theorem 1.* By Equation (9), we have

$$\mathbb{E}((\hat{f}_{r,A}(x) - f(x))^2) \leq \mathbb{E}(\varepsilon_1^2(x) | E_{x,r}) + 2\mathbb{E}(\varepsilon_1 \varepsilon_2 | E_{x,r}) + \mathbb{E}(\varepsilon_2^2(x) | E_{x,r}) + f(x)^2 \mathbb{P}(E_{x,r}^c).$$

Compactness of  $S$  and Lipschitz continuity of  $f$  imply that the density is upper bounded by  $f_{\max}^{(0)} \leq f_{\min}^{(0)} + K \frac{\text{diam}(S)}{2}$  so we can apply Lemmas 2 and 3. We then have  $\mathbb{E}(\varepsilon_1^2(x)|E_{x,r}) \leq \frac{2f_{\max}^{(0)}}{A\omega_D nr^D}$ ,  $\mathbb{E}(\varepsilon_1(x)\varepsilon_2(x)|E_{x,r}) = 0$  and  $\mathbb{E}(\varepsilon_2(x)^2|E_{x,r}) \leq r^2 \left( K + \frac{4f_{\max}^{(0)}\omega_{D-1}}{3AR_0\omega_D} \right)^2$  (take  $R_0 = +\infty$  in case of convex support). We also can apply Lemma 1 (the density is upper bounded and the distribution is  $(r, \delta, f_{\min}^{(0)})$ -standard) to achieve the proof of the first equation, for  $n$  large enough:

$$\mathbb{E}((\hat{f}_{r,A}(x) - f(x))^2) \leq \frac{2f_{\max}^{(0)}}{A\omega_D nr^D} + r^2 \left( K + \frac{4f_{\max}^{(0)}\omega_{D-1}}{3AR_0\omega_D} \right)^2 + (f_{\max}^{(0)})^2 \left( \beta_1 \exp(-nr^D f_{\min}^{(0)} \beta_2) + \exp(-n/8) \right).$$

To prove the second equation we can proceed the same way, by use of equation (10), and obtain

$$\mathbb{P}(|\hat{f}_{r,A}(x) - f(x)| \geq t_1 + t_2) \leq \mathbb{P}(|\varepsilon_1(x)| \geq t_1|E_{x,r}) + \mathbb{P}(|\varepsilon_2(x)| \geq t_2|E_{x,r}) + \mathbb{P}(E_{x,r}^c).$$

Choose then  $t_1 = \sqrt{\frac{32f_{\max}^{(0)} \ln n}{A^2\omega_D nr^D}}$  and  $t_2 = r \left( K + \frac{4f_{\max}^{(0)}\omega_{D-1}}{3AR_0\omega_D} \right)$ . By Lemma 3,  $\mathbb{P}(|\varepsilon_2(x)| \geq t_2|E_{x,r}) = 0$ ; and by Lemmas 2, 3 and 1  $\sum n\mathbb{P}(|\varepsilon_1(x)| \geq t_1|E_{x,r}) < +\infty$  and  $\sum n\mathbb{P}(E_{x,r}^c) < +\infty$ . Thus we can conclude by use of Borrel-Cantelli Lemma.  $\square$

## 5.4 Proofs for model $\mathcal{M}_2$

**Lemma 4** (Bounds on  $\varepsilon_2$  for model  $\mathcal{M}_2$ ). *Suppose that the distribution belongs to the model  $\mathcal{M}_2$ . For  $x \in \dot{S}$  and  $r$  small enough to have  $B(x,r) \subset \{z, f(z) > 0\}$ , introduce  $H_{x,r} = \max_{z \in B(x,r)} \|\mathcal{H}_f(x)\|_{\text{op}}$ . There exists an explicit constant  $C$  such that*

$$\mathbb{E}(\varepsilon_2(x)^2|E_{x,r}) \leq \left( H_{x,r}^2 r^4 + Cr^2 \left( \left( \frac{n\Gamma_{x,r}}{2} \right)^{-\frac{4}{D+1}} + r \left( \frac{n\Gamma_{x,r}}{2} \right)^{-\frac{2}{D+1}} \right) \right) (1 + o(1)). \quad (18)$$

And there exist explicit constants  $C, C'$  and  $C_2$  such that

$$\mathbb{P} \left( |\varepsilon_2| \geq \frac{Cr \ln n}{nr^D \gamma(x)} + r \left( \frac{C'}{nr^D \gamma(x)} \right)^{\frac{2}{D+1}} + r^2 f_{\max}^{(2)} |E_{x,r} \right) \leq C_2 n^{-4} + \exp(-n(\delta r)^D \gamma(x)/10), \quad (19)$$

where  $\gamma(x) = \min(f(x), \varepsilon_0)$ ,  $\varepsilon_0$  being the constant in the  $(r_0, \delta, \varepsilon_0)$ -standardness of the distribution.

*Proof.* As  $C_{x,r} \subset B(x,r) \subset S$ , by the use of a second order Taylor expansion (see Proposition 2 in Appendix A), we get:

$$\left| \tilde{\Gamma}_{x,r} - f(x)|C_{x,r}| - \int_{C_{x,r}} (y-x)\nabla f(x)' dy \right| \leq \int_{C_{x,r}} \|y-x\|^2 \max_{z \in C_{x,r}} \|\mathcal{H}_f(z)\|_{\text{op}} dy.$$

Because  $\int_{B(x,r)} (y-x)\nabla f(x)' dy = 0$ , we have that:

$$|\tilde{\Gamma}_{x,r} - f(x)|C_{x,r}| \leq \int_{B(x,r) \setminus C_{x,r}} r \|\nabla f(x)\| dy + |C_{x,r}| \max_{z \in C_{x,r}} \|\mathcal{H}_f(z)\|_{\text{op}} r^2,$$

from which we derive  $\frac{|\tilde{\Gamma}_{x,r} - f(x)|_{C_{x,r}}}{|C_{x,r}|} \leq r \|\nabla f(x)\| \frac{|B(x,r) \setminus C_{x,r}|}{|C_{x,r}|} + \max_{z \in C_{x,r}} \|\mathcal{H}_f(z)\|_{\text{op}} r^2$ . Thus, given  $|C_{x,r}| \geq A \omega_D r^D$ , we obtain

$$|\varepsilon_2(x)| \leq \frac{\|\nabla f(x)\|}{A} r \frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} + \max_{z \in C_{x,r}} \|\mathcal{H}_f(z)\|_{\text{op}} r^2. \quad (20)$$

To control the convergence rate of  $|B(x,r) \setminus C_{x,r}|$  we will make use of results from [9] recalled in Appendix C (see Proposition 6). To apply the aforementioned proposition let us introduce  $M_{x,r} = \max_{y \in B(0,1)} \frac{r^D f(x+ry)}{\int_{B(x,r)} f(z) dz}$  (defined because  $\int_{B(x,r)} f(z) dz > 0$ ).

Let us first consider the  $L^2$  convergence. Because  $\frac{\min_{B(x,r)} f(y)}{\max_{B(x,r)} f(y)} \frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} \leq \frac{\int_{B(x,r) \setminus C_{x,r}} f(z) dz}{\int_{B(x,r)} f(z) dz}$  and by application of equation (30) (Appendix C), it comes that, for  $k > 0$ ,

$$\mathbb{E} \left( \left( \frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} \right)^q |N_{x,r} = k \right) \leq A_q (M_{x,r} + 1)^q \left( \frac{\max_{B(x,r)} f(y)}{\min_{B(x,r)} f(y)} \right)^q k^{-\frac{2q}{D+1}}.$$

Notice that  $\mathbb{E} \left( \left( \frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} \right)^q |N_{x,r} = 0 \right) = 1$ . Then, we have:

$$\begin{aligned} \mathbb{E} \left( \left( \frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} \right)^q \right) &\leq A_q (M_{x,r} + 1)^q \left( \frac{\max_{B(x,r)} f(y)}{\min_{B(x,r)} f(y)} \right)^q \sum_{k=1}^n \binom{n}{k} k^{-\frac{2q}{D+1}} \Gamma_{x,r}^k (1 - \Gamma_{x,r})^{n-k} \\ &\quad + (1 - \Gamma_{x,r})^n. \end{aligned}$$

A rough bound gives  $M_{x,r} \leq \frac{1}{\omega_D} \frac{\max_{B(x,r)} f(y)}{\min_{B(x,r)} f(y)}$ .

Introduce  $B_{x,r,q} = A_q \left( \frac{1}{\omega_D} \frac{\max_{B(x,r)} f(y)}{\min_{B(x,r)} f(y)} + 1 \right)^q \left( \frac{\max_{B(x,r)} f(y)}{\min_{B(x,r)} f(y)} \right)^q$ . By application of Proposition 1 (Appendix A), we have that

$$\mathbb{E} \left( \left( \frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} \right)^q \right) \leq B_{x,r,q} \left( \left( \frac{n\Gamma_{x,r}}{2} \right)^{-\frac{2q}{D+1}} + \exp \left( -\frac{n\Gamma_{x,r}}{10} \right) \right) + (1 - \Gamma_{x,r})^n. \quad (21)$$

And thus, because  $\mathbb{E}(\varepsilon_2(x)^2 | E_{x,r}) \leq \mathbb{E}(\varepsilon_2(x)^2)$  ( $\varepsilon_2(x)^2$  is a positive random variable), by use of (20), we obtain

$$\begin{aligned} \mathbb{E}(\varepsilon_2(x)^2 | E_{x,r}) &\leq \frac{\|\nabla f(x)\|^2}{A^2} r^2 \left( B_{x,r,2} \left( \left( \frac{n\Gamma_{x,r}}{2} \right)^{-\frac{4}{D+1}} + \exp \left( -\frac{n\Gamma_{x,r}}{10} \right) \right) + (1 - \Gamma_{x,r})^n \right) \\ &\quad + 2 \frac{\|\nabla f(x)\| H_{x,r}}{A} r^3 \left( B_{x,r,1} \left( \left( \frac{n\Gamma_{x,r}}{2} \right)^{-\frac{2}{D+1}} + \exp \left( -\frac{n\Gamma_{x,r}}{10} \right) \right) + (1 - \Gamma_{x,r})^n \right) + H_{x,r}^2 r^4. \end{aligned}$$

Because  $(1 - \Gamma_{x,r})^n \leq \exp(-n\Gamma_{x,r})$  and  $nr^D / \ln n \rightarrow +\infty$ , and by application of Proposition 7 (Appendix C) we obtain the existence of an explicit constant  $C$  such that

$$\mathbb{E}(\varepsilon_2(x)^2 | E_{x,r}) \leq \left( H_{x,r}^2 r^4 + Cr^2 \left( \left( \frac{n\Gamma_{x,r}}{2} \right)^{-\frac{4}{D+1}} + r \left( \frac{n\Gamma_{x,r}}{2} \right)^{-\frac{2}{D+1}} \right) \right) (1 + o(1)).$$

Let us second consider the convergence in probability.

Introduce  $B'_{x,r} = C_1 \left( \frac{1}{\omega_D} \frac{\max_{B(x,r)} f(y)}{\min_{B(x,r)} f(y)} + 1 \right) \frac{\max_{B(x,r)} f(y)}{\min_{B(x,r)} f(y)}$ . By use of Equation (29) (Appendix C), we obtain

$$\mathbb{P} \left[ k \left( \frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} - B'_{x,r} k^{-2/(D+1)} \right) > t \frac{\max_{B(x,r)} f(z)}{\min_{B(x,r)} f(z)} |N_{x,r} = k \right] \leq C_2 e^{-t}.$$

From this, it comes, by Proposition 7 (Appendix C), that there exist  $C$  and  $C'$  such that

$$\mathbb{P} \left[ \left( \|\nabla_f(x)\| \frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} \geq C \frac{t}{k} + C' k^{-\frac{2}{D+1}} |N_{x,r} = k \right) | E_{x,r} \right] \leq C_2 e^{-t}.$$

Thus, for all  $K$ , we obtain

$$\mathbb{P} \left( \|\nabla_f(x)\| \frac{|B(x,r) \setminus C_{x,r}|}{|B(x,r)|} \geq C \frac{t}{K} + C' K^{-\frac{2}{D+1}} | E_{x,r} \right) \leq C_2 e^{-t} + \mathbb{P}(N_{x,r} < K),$$

which, with (20), gives

$$\mathbb{P} \left( |\varepsilon_2| \geq C \frac{tr}{K} + C' r K^{-\frac{2}{D+1}} + r^2 f_{\max}^{(2)} | E_{x,r} \right) \leq C_2 e^{-t} + \mathbb{P}(N_{x,r} < K).$$

Finally, to deal with  $\mathbb{P}(N_{x,r} < K)$ , recall that  $\gamma(x) = \min(f(x), \varepsilon_0)$  and choose  $K = \frac{n\omega_D \delta^D r^D \gamma(x)}{2}$ . Also recall that we have  $\Gamma_{x,r} \geq \omega_D (\delta r)^D \gamma(x)$  by the ball standardness assumption. Thus, by application of Bennets inequality ( $h(0.5) \geq 0.1$ ) with the choice of  $t = 4 \ln n$ , it comes the existence of (new) constants  $C$  and  $C'$  such that:

$$\mathbb{P} \left( |\varepsilon_2| \geq r \frac{C \ln n}{nr^D \gamma(x)} + r \left( \frac{C'}{nr^D \gamma(x)} \right)^{\frac{2}{D+1}} + r^2 f_{\max}^{(2)} | E_{x,r} \right) \leq C_2 n^{-4} + \exp(-n(\delta r)^D \gamma(x)/10).$$

This achieves the proof of (19).  $\square$

*Proof of Theorem 2.* By Equation (9), we have

$$\mathbb{E}((\hat{f}_{r,A}(x) - f(x))^2) \leq \mathbb{E}(\varepsilon_1^2(x) | E_{x,r}) + 2\mathbb{E}(\varepsilon_1 \varepsilon_2 | E_{x,r}) + \mathbb{E}(\varepsilon_2^2(x) | E_{x,r}) + f^2(x) \mathbb{P}(E_{x,r}^c).$$

By Equation (13),  $\mathbb{E}(\varepsilon_1^2(x) | E_{x,r}) \leq \frac{2f_{\max}^{(0)}}{A\omega_D r^D n}$ . By Equation (12),  $\mathbb{E}(\varepsilon_1(x)\varepsilon_2(x) | E_{x,r}) = 0$ . By Equation (18), for  $n$  large enough to have  $B(x,r) \subset S$ ,

$$\mathbb{E}(\varepsilon_2(x)^2 | E_{x,r}) \leq \left( H_{x,r}^2 r^4 + Cr^2 \left( \left( \frac{n\Gamma_{x,r}}{2} \right)^{-\frac{4}{D+1}} + r \left( \frac{n\Gamma_{x,r}}{2} \right)^{-\frac{2}{D+1}} \right) \right) (1 + o(1)).$$

This gives

$$\mathbb{E}(\varepsilon_2(x)^2 | E_{x,r}) \leq \left( H_{x,r}^2 r^4 + Cr^2 \left( \left( \frac{n\omega_D f(x) r^D}{2} \right)^{-\frac{4}{D+1}} + r \left( \frac{n\omega_D f(x) r^D}{2} \right)^{-\frac{2}{D+1}} \right) \right) (1 + o_x(1)).$$

Finally, by Lemma 1, we have

$$f^2(x) \mathbb{P}(E_{x,r}^c) \leq \beta_1 f^2(x) \exp(-nr^D \min(f(x), \varepsilon_0) \beta_2) + (f_{\max}^{(0)})^2 \exp(-n/8).$$

Thus, we obtain

$$f^2(x) \mathbb{P}(E_{x,r}^c) \leq \max \left( \frac{4\beta_1 e^{-2}}{(\beta_2 nr^D)^2}, \beta_1 (f_{\max}^{(0)})^2 \exp(-nr^D \beta^2 \varepsilon_0) \right) + (f_{\max}^{(0)})^2 \exp(-n/8).$$

And because  $nr^D \rightarrow +\infty$ , we have

$$f^2(x) \mathbb{P}(E_{x,r}^c) = O((nr^D)^{-2}).$$

This achieves the proof of the first point of Theorem 2.

Second point of Theorem 2 is proved the same by use of Lemma 3 instead of equation(18).  $\square$

As is common and classical, the proof of Corollary 1 is left for the reader.

*Proof of Theorem 3.* Firstly focus on the first case. Let us fix  $t > 0$ . For all  $x \in L_t$  such that  $B(x, r) \subset S$ , due to Lemma 1, we have:

$$\mathbb{P}(E_{x,r}^c) \leq \beta_1 \exp(-\beta_2 n r^D \min(t, \varepsilon_0)) + \exp(-n/8) = o_t(n^{-4}).$$

Also due to Equation (14), we have  $\mathbb{P}\left(|\varepsilon_1(x)| > \sqrt{\frac{32 \ln n f_{\max}^{(0)}}{A^2 \omega_D n r^D}} |E_{x,r}\right) \leq 2n^{-4}(1 + o(1))$ . Finally, Equation (19) for  $r = cn^{-\frac{1}{D+4}}$  and  $D \leq 7$  gives

$$\mathbb{P}\left(|\varepsilon_2| \geq O_t(r^2) |E_{x,r}\right) \leq C_2 n^{-4} + \exp(-n r^D \min(t, \varepsilon_0)/10).$$

Thus, due to  $\mathbb{P}(|\hat{f}_{r,A}(x) - f(x)| \geq t_1 + t_2) \leq \mathbb{P}(|\varepsilon_1(x)| \geq t_1 |E_{x,r}) + \mathbb{P}(|\varepsilon_2| \geq t_2 |E_{x,r}) + \mathbb{P}(E_{x,r}^c)$ ,

$$\mathbb{P}\left(|\hat{f}_{r,A}(x) - f(x)| \geq \sqrt{\frac{32 \ln n f_{\max}^{(0)}}{A^2 \omega_D n r^D}} + O_t(r^2)\right) \leq C n^{-4}(1 + o_t(1)).$$

Introduce  $\mathbb{X}(t, r) = \{X_i \in \mathbb{X}_n, f(X_i) > t, B(X_i, r) \subset S\}$ . Recall that  $r = cn^{-\frac{1}{D+4}}$ . All the above equations provide the existence of a constant  $C_t$  such that, for  $n$  large enough,

$$\mathbb{P}\left(\exists X_i \in \mathbb{X}(t, r) \text{ s.t. } |\hat{f}_{r,A}(X_i) - f(X_i)| \geq C_t \sqrt{\ln n n^{-\frac{2}{D+4}}}\right) \leq C n^{-3}(1 + o(1)).$$

That, with Borrel Cantelli Lemma, concludes the proof of the first point of Theorem 3.

The second point of Theorem 3 is the same using the lower bound of the density to obtain uniform rates and is left to the reader.

We now prove the third point of Theorem 3. First suppose that  $f(x) \geq (\ln n / (n r^D))^{1/2}$ . Then, by Lemma 1 and because  $\frac{n r^D}{\ln n} \rightarrow +\infty$ , we have  $\mathbb{P}(E_{x,r}^c) \leq \beta_1 \exp\left(-\beta_2 \ln n \sqrt{\frac{n r^D}{\ln n}}\right) + \exp(-n/8) = o(n^{-4})$ . Due to Equation (14), we also have  $\mathbb{P}\left(|\varepsilon_1(x)| > \sqrt{\frac{32 \ln n f_{\max}^{(0)}}{A^2 \omega_D n r^D}} |E_{x,r}\right) \leq 2n^{-4}(1 + o(1))$ . And, under the rolling ball condition or the convexity condition, due to Lemma 3, there exists  $C$  such that  $\mathbb{P}(|\varepsilon_2| > Cr |E_{x,r}) = 0$ . Thus, due to  $\mathbb{P}(|\hat{f}_{r,A}(x) - f(x)| \geq t_1 + t_2) \leq \mathbb{P}(|\varepsilon_1(x)| \geq t_1 |E_{x,r}) + \mathbb{P}(|\varepsilon_2| \geq t_2 |E_{x,r}) + \mathbb{P}(E_{x,r}^c)$ , we obtain

$$\mathbb{P}\left(|\hat{f}_{r,A}(x) - f(x)| \geq \sqrt{\frac{32 \ln n f_{\max}^{(0)}}{A^2 \omega_D n r^D}} + Cr\right) \leq 2n^{-4}(1 + o(1)).$$

Second, suppose that  $0 < f(x) \leq (\ln n / (n r^D))^{1/2}$ . Because, conditionally to  $E_{x,r}^c$ ,  $\hat{f}_{r,A}(x) = 0$ , we have  $\mathbb{P}(|\hat{f}_{r,A}(x) - f(x)| > (\ln n / (n r^D))^{1/2} |E_{x,r}^c) = 0$ . It comes that, when  $t \geq 2(\ln n / (n r^D))^{1/2}$ ,

$$\mathbb{P}(|\hat{f}_{r,A}(x) - f(x)| \geq t) \leq \mathbb{P}(|\hat{f}_{r,A}(x) - f(x)| \geq t |E_{x,r}).$$

Then, applying the same calculus as in the first case, we obtain

$$\mathbb{P}\left(|\hat{f}_{r,A}(x) - f(x)| \geq \sqrt{\frac{\ln n}{n r^D}} \max\left(\sqrt{\frac{32 f_{\max}^{(0)}}{A^2 \omega_D}}, 2\right) + Cr\right) \leq 2n^{-4}(1 + o(1)).$$

Thus, for all  $x$  in  $S$ , we have

$$\mathbb{P} \left( |\hat{f}_{r,A}(x) - f(x)| \geq \sqrt{\frac{\ln n}{nr^D}} \max \left( \sqrt{\frac{32f_{\max}^{(0)}}{A^2\omega_D}}, 2 \right) + Cr \right) \leq 2n^{-4}(1+o(1)).$$

Thus, when  $r = cn^{-\frac{1}{D+4}}$ ,  $\mathbb{P}(\exists X_i \text{ s.t. } |\hat{f}_{r,A}(X_i) - f(X_i)| \geq Cr(1+o(1))) \leq 2n^{-3}(1+o(1))$ . That, with Borrel Cantelli Lemma, concludes the proof of last point of Theorem 3.  $\square$

*Proof of Theorem 4.* Recall the LoCoH definition: for a set  $E$ ,

$$H_r(E) := \bigcup_{x \in E} H(B(x, r) \cap E).$$

**Suppose first that  $t$  is a regular level.** The  $\mathcal{C}^2$  condition on  $f$  implies that for all  $(x, y) \in S \times S$   $\|\nabla_f(x) - \nabla_f(y)\| \leq f_{\max}^{(2)}\|x - y\|$ . Thus, by application of Theorem 2 in [32], for all  $t' \in [t - \Delta_t, t + \Delta_t]$ ,  $L_{t'}$  has the  $\frac{m_t}{f_{\max}^{(2)}}$ -inside and outside rolling ball condition.

We now prove that, for all  $0 < \varepsilon < \min\left(\Delta_t, \frac{m_t d(L_{t-\Delta_t}, \partial S)}{2}, \frac{m_t^2}{2f_{\max}^{(2)}}\right)$ , we have that  $L_{t-\varepsilon} \subset L_t \oplus \frac{2\varepsilon}{m_t}B$ . Indeed, consider  $x \in L_{t-\varepsilon}$  and  $y = x + 2\frac{\varepsilon}{m_t} \frac{\nabla_f(x)}{\|\nabla_f(x)\|}$ . We have that  $x \in L_{t-\Delta_t}$  and  $\|y - x\| < d(L_{t-\Delta_t}, \partial S)$ , thus  $y \in \mathring{S}$ . That allows to write, according to Proposition 2, that

$$f(y) \geq f(x) + 2\frac{\varepsilon}{m_t} \|\nabla_f(x)\| - \frac{1}{2} \left(2\frac{\varepsilon}{m_t}\right)^2 f_{\max}^{(2)}.$$

We have  $f(y) \geq t - \varepsilon + 2\varepsilon - \varepsilon \frac{2\varepsilon f_{\max}^{(2)}}{m_t^2}$ . Thus, since  $\varepsilon < m_t^2/(2f_{\max}^{(2)})$ , we also have  $f(y) \geq t$  and thus  $y \in L_t$ . That, together with  $\|y - x\| = \frac{2\varepsilon}{m_t}$ , concludes the proof of the inclusion:

$$L_{t-\varepsilon} \subset L_t \oplus \frac{2\varepsilon}{m_t}B. \quad (22)$$

By last point of Theorem 3, with probability one for  $n$  large enough, for all  $X_i$  such that  $\hat{f}(X_i) \geq t$ , we have  $f(X_i) \geq t - Cn^{-\frac{1}{D+4}}$ . Thus, when  $n$  is large enough to have  $Cn^{-\frac{1}{D+4}} \leq \Delta_t$ , we have  $X_i \in L_{t-\Delta_t}$  and thus  $d(X_i, \partial S) \geq d(L_{t-\Delta_t}, \partial S) > 0$ . When  $r \leq d(L_{t-\Delta_t}, \partial S)$  (that is when  $n$  is large enough), from Theorem 3 (first point), with probability one for  $n$  large enough, for all  $X_i$  such that  $\hat{f}(X_i) \geq t$ , we have  $|f(X_i) - \hat{f}(X_i)| \leq C_{t-\Delta_t} \sqrt{\ln nn^{-\frac{2}{D+4}}}$ . Thus, we get that  $X_i \in L_{t-\varepsilon_n}$  with  $\varepsilon_n = C_{t-\Delta_t} \sqrt{\ln nn^{-\frac{2}{D+4}}}$ . We then have, with probability one for  $n$  large enough:

$$\hat{L}_t \subset H_r(\mathbb{X}_n \cap L_{t-\varepsilon_n}) \subset H_r(L_{t-\varepsilon_n}).$$

When  $n$  is large enough to have all previous conditions and  $C_{t-\Delta_t} \sqrt{\ln nn^{-\frac{2}{D+4}}} \leq \Delta_t$ ,  $L_{t-\varepsilon_n}$  satisfies the  $\frac{m_t}{f_{\max}^{(2)}}$ -inside and outside rolling ball condition. Then, by Proposition 5 equation (27) (Appendix B), we obtain that (with probability one for  $n$  large enough):

$$\hat{L}_t \subset L_{t-\varepsilon_n} \oplus \frac{r^2 f_{\max}^{(2)}}{m_t} B \subset L_t \oplus \left( \frac{r^2 f_{\max}^{(2)} + 2\varepsilon_n}{m_t} \right) B. \quad (23)$$

Now, we consider reverse inclusion. Introduce  $\varepsilon'_n = C_t \sqrt{\ln nn^{-\frac{2}{D+4}}}$ . By last point of Theorem 3, with probability one for  $n$  large enough, we have  $\mathbb{X}_n \cap L_{t+\varepsilon'_n} \subset \mathbb{X}_{r,A}^+(t)$ . Introduce now

$P_t = t\omega_D \left( \frac{m_t}{f_{\max}^{(2)}} \right)^D$  and  $n' = nP_t/2$ . Recall that  $r = cn^{-\frac{1}{D+4}}$ . Thus with probability one for  $n$  large enough, we get:

$$H_{c\left(\frac{\ln n'}{n}\right)^{\frac{1}{D+1}}}(\mathbb{X}_n \cap L_{t+\varepsilon'_n}) \subset H_r(\mathbb{X}_n \cap L_{t+\varepsilon'_n}) \subset \hat{L}_t.$$

Recall that  $L_{t+\varepsilon'_n}$  has the  $\frac{m_t}{f_{\max}^{(2)}}$ -inside and outside rolling ball condition (when  $n$  is large enough). Thus we have  $|L_{t+\varepsilon'_n}| \geq \omega_D \left( \frac{m_t}{f_{\max}^{(2)}} \right)^D$  and  $\mathbb{P}(L_{t+\varepsilon'_n}) \geq P_t$ . In consequence, by Hoeffding, with probability one for  $n$  large enough, there is more than  $n'$  observations in  $\mathbb{X}_n \cap L_{t+\varepsilon'_n}$ . We can apply [2] (Theorem 2 with  $\alpha = 0$ ). More precisely, there exists a constant  $C'_t$  such that, with probability one for  $n$  large enough, we have:

$$L_{t+\varepsilon'_n} \subset H_{c\left(\frac{\ln n'}{n}\right)^{\frac{1}{D+1}}}(\mathbb{X}_n \cap L_{t+\varepsilon'_n}) \oplus C'_t \left( \frac{\ln n}{n} \right)^{\frac{2}{D+1}} B.$$

Previous chain of inclusion allows then to conclude that  $L_{t+\varepsilon'_n} \subset \hat{L}_t \oplus C'_t \left( \frac{\ln n}{n} \right)^{\frac{2}{D+1}} B$ . Thus again, by application of (22), we have  $L_t \subset \hat{L}_t \oplus \left( C'_t \left( \frac{\ln n}{n} \right)^{\frac{2}{D+1}} + \frac{2\varepsilon'_n}{m_t} \right) B$ , which, together with Equation (23) achieves the proof of the first point of Theorem 4.

**Supposenow that  $t$  is a standard level.** Choose  $\varepsilon_n = Cn^{-\frac{1}{D+4}}$ , such that for all  $X_i$  we have  $f(X_i) \geq t + \varepsilon_n \Rightarrow \hat{f}_{r,A}(X_i) \geq t \Rightarrow f(X_i) \geq t - \varepsilon_n$ . It implies that  $H_r(L_{t+\varepsilon_n} \cap \mathbb{X}_n) \subset \hat{L}_r \subset H_r(L_{t-\varepsilon_n} \cap \mathbb{X}_n)$ .

Take now  $n$  large enough to have  $\varepsilon_n \leq \Delta_t$ . We first have the trivial inclusions  $\hat{L}_t \subset H_r(L_{t-\varepsilon_n} \cap \mathbb{X}_n) \subset H_r(L_{t-\varepsilon_n}) \subset H_r(L_t \oplus \varepsilon_n \beta_t B) \subset L_t \oplus (r + \varepsilon_n \beta_t) B$  (due to  $H_r(S) \subset S \oplus rB$  for any set  $S$ ).

Focus now on the reverse inclusion. First, there exists a constant  $c_t$  such that, with probability one for  $n$  large enough,  $L_{t+\varepsilon_n} \subset (L_{t+\varepsilon_n} \cap \mathbb{X}_n) \oplus c_t (\ln n/n)^{1/D} B$ . The proof being classical, we refer the reader to Proposition 3 in Appendix B for further details. Thus, by standardness, we get  $L_t \subset (L_{t+\varepsilon_n} \cap \mathbb{X}_n) \oplus (c_t (\ln n/n)^{1/D} + \beta_t \varepsilon_n) B$ . And finally, using arguments as in Theorem 1 and Corollary 1 in [2], we have that  $d_H \left( H_{4(c_t (\ln n/n)^{1/D} + \beta_t \varepsilon_n)}(L_{t+\varepsilon_n} \cap \mathbb{X}_n), L_t \right) \leq \gamma_t (c_t (\ln n/n)^{1/D} + \beta_t \varepsilon_n)$ . Hence, we obtain

$$L_t \subset \left( H_{4(c_t (\ln n/n)^{1/D} + \beta_t \varepsilon_n)}(L_{t+\varepsilon_n} \cap \mathbb{X}_n) \right) \oplus \gamma_t (c_t (\ln n/n)^{1/D} + \beta_t \varepsilon_n) B.$$

Thus, since  $r_n = cn^{-\frac{1}{D+4}}$  with  $c > 4\beta_t C$ , we have that, with probability one for  $n$  large enough,

$$L_t \subset (H_{r_n}(L_{t+\varepsilon_n} \cap \mathbb{X}_n)) \oplus \gamma_t (c_t (\ln n/n)^{1/D} + \beta_t \varepsilon_n) B,$$

$$L_t \subset \hat{L}_t \oplus \gamma_t (c_t (\ln n/n)^{1/D} + \beta_t \varepsilon_n) B.$$

That concludes the proof of the second point of the theorem.  $\square$

## 5.5 Proofs for the manifold case

**Proof of Theorem 5 in a nutshell:** Let introduce  $g_{x,R}(z)$  the density of  $\pi_{\hat{T}_x}(X)|X \in B(x, R)$ . By applying following Lemma 6 for  $R$  and  $\Theta_n$  small enough, we are in situation similar to the one model  $\mathcal{M}_1$  so that, there exist constants  $\beta'_1, \beta'_2, C_1$  and  $C_2$  such that:

$$\mathbb{E} \left( \left| \frac{n}{k} \hat{f}_{r,R,A}(x) - g_{x,R}(0) \right| \middle| N_{x,R} = k \right) \leq \frac{C_1}{\sqrt{kr^d}} + C_2 r + \beta'_1 \exp(-\beta'_2 k r^d) + \exp(-k/8).$$

Thus, we have

$$\begin{aligned} \mathbb{E} \left( \left| \hat{f}_{r,R,A}(x) - f(x) \right| \middle| N_{x,R} = k \right) &\leq \frac{C_1 \sqrt{k}}{n \sqrt{r^d}} + \frac{k}{n} (C_2 r + \beta'_1 \exp(-\beta'_2 k r^d) + \exp(-k/8)) \\ &\quad + \mathbb{E} \left( \left| \frac{k}{n} g_{x,R}(0) - \mathbb{P}_X(B(X,R)) g_{x,R}(0) \right| \middle| N_{x,R} = k \right) \\ &\quad + \mathbb{E} ( |\mathbb{P}_X(B(X,R)) g_{x,R}(0) - f(x)| \middle| N_{x,R} = k ). \end{aligned}$$

In consequence (extensively using  $k/n \leq 1$ ), we obtain

$$\begin{aligned} \mathbb{E} \left( \left| \hat{f}_{r,R,A}(x) - f(x) \right| \right) &\leq \frac{C_1 \mathbb{E}(\sqrt{N_{x,R}})}{n \sqrt{r^d}} + C_2 r + \mathbb{E}(\beta'_1 \exp(-\beta'_2 N_{x,R} r^d) + \exp(-N_{x,R}/8)) \\ &\quad + g_{x,R}(0) \mathbb{E} \left( \left| \frac{N_{x,R}}{n} - \mathbb{P}_X(B(X,R)) \right| \right) + |\mathbb{P}_X(B(X,R)) g_{x,R}(0) - f(x)|. \end{aligned}$$

By Jensen inequality for concave functions (applied to the first term),  $\mathbb{E}(\exp(-\alpha N)) = (1 - p + p e^{-\alpha})^n \leq \exp(-np(1 - e^{-\alpha}))$  when  $N \sim \text{Binom}(n, p)$  (end of the first line), and Cauchy Schwartz and the law of  $N_{x,r}$  (the second line), it comes existence of positive  $\beta''_2$  and  $\beta_3$  such that

$$\begin{aligned} \mathbb{E} \left( \left| \hat{f}_{r,R,A}(x) - f(x) \right| \right) &\leq \frac{C_1}{\sqrt{nr^d}} + C_2 r + \beta'_1 \exp(-\beta''_2 n r^d) + \exp(-\beta_3 n) \\ &\quad + \frac{g_{x,R}(0)}{4\sqrt{n}} + |\mathbb{P}_X(B(X,R)) g_{x,R}(0) - f(x)|. \end{aligned}$$

By the following Lemma 5 second point, we have that  $|\mathbb{P}_X(B(X,R)) g_{x,R}(0) - f(x)| \leq C_3 \Theta_n$ . This achieves the proof of the existence of  $C$  such that, for  $n$  large enough,

$$\mathbb{E} \left( \left| \hat{f}_{r_n,R,A}(x) - f(x) \right| \right) \leq C \left( \frac{1}{\sqrt{nr_n^d}} + r_n + \Theta_n \right).$$

## Main Lemmas

**Lemma 5.** *Introduce  $S_{x,R} = \pi_{\hat{T}_x}(S \cap B(x,R))$ . For all  $R \leq \min\left(\frac{\tau_S}{3}, \frac{\tau_S}{2d(D-d)}, \frac{1}{4(D-d)}\right)$ , there exist  $K', r_0, C', G_{\min}$  and  $G_{\max}$  positive constants such that, for  $n$  large enough to have  $\Theta_n \leq \min\left(\frac{1}{12}, \frac{1}{2(D-d)} - 2\frac{R}{\tau_S}\right)$ , we have:*

1. for all  $z \in S_{x,R}$ ,  $G_{\min} \leq g_{x,R}(z) \leq G_{\max}$ ,
2.  $|\mathbb{P}_X(B(x,R)) g_{x,R}(0) - f(x)| \leq C' \Theta_n^2 (1 + o(1))$ ,
3. for all  $z \in B(0, r_0) \cap S_{x,r}$ ,  $|g_{x,R}(z) - g_{x,R}(0)| \leq K' (||z|| + \Theta_n^2 + \Theta_n^2 ||z||^2)$ .

*Proof.* Due to Proposition 8, we have that  $\pi_{\hat{T}_x} : S \cap B(x,R) \rightarrow S_{x,R} = \pi_{\hat{T}_x}(S \cap B(x,R))$  is one to one. Let  $(u_{x,1}, \dots, u_{x,d})$  be an orthonormalized basis of  $\hat{T}_x$ , completed by  $(u_{x,d+1}, \dots, u_{x,D})$  to have an orthonormalized basis of  $\mathbb{R}^D$ , so that

$$\begin{aligned} \pi_{\hat{T}_x}^{-1} : \quad S_{x,R} &\rightarrow S \cap B(x,R) \\ z = \sum_{i=1}^d z_i u_{x,i} &\mapsto x + z + \sum_{i=d+1}^D \varphi_{x,i-d}(z) u_{x,i} \end{aligned}$$

Define  $\phi_x(z) = (\varphi_{x,1}(z), \dots, \varphi_{x,D-d+1}(z))$  and denote by  $J_x(z)$  the Jacobian matrix of  $\phi_x(z)$  i.e.

$$J_x(z) = \begin{pmatrix} \frac{\partial \varphi_{x,1}}{\partial z_1}(z) & \dots & \frac{\partial \varphi_{x,1}}{\partial z_d}(z) \\ \vdots & & \vdots \\ \frac{\partial \varphi_{x,D-d}}{\partial z_1}(z) & \dots & \frac{\partial \varphi_{x,D-d}}{\partial z_d}(z) \end{pmatrix}.$$

On the one hand, when  $i \geq 1$ , then  $u_{x,d+i}$  is in  $\hat{T}_x^\perp$ . On the other hand, for all  $j \in \{1, \dots, d\}$ , we have that  $v_{x,j} := u_{x,j} + \sum_{k=1}^{D-d} \frac{\partial \varphi_{x,k}(z)}{\partial z_j} u_{x,k+d} \in T_{x+z+\phi_x(z)}S$ . Thus, we get  $|\langle u_{x,d+i}, v_{x,j} \rangle| \leq (\Theta_n + \angle T_x S, T_{x+z+\phi_x(z)}S) \|v_{x,j}\|$ . From this, since  $\|z + \phi_x(z)\| \leq R \leq \tau_S/3$ , it comes that  $|\langle u_{x,d+i}, v_{x,j} \rangle| \leq \left(\Theta_n + \frac{\pi \|z + \phi_x(z)\|}{2\tau_S}\right) \|v_{x,j}\|$ . Thus, we obtain:

$$\left(\frac{\partial \varphi_{x,i}}{\partial z_j}(z)\right)^2 \leq \left(\Theta_n + \frac{\pi \|z + \phi_x(z)\|}{2\tau_S}\right)^2 \left(1 + \sum_{k=1}^{D-d} \left(\frac{\partial \varphi_{x,k}}{\partial z_j}(z)\right)^2\right).$$

Now, summing on  $i$  gives that, if  $z \in S_{x,R}$ ,  $\sum_{k=1}^{D-d} \left(\frac{\partial \varphi_{x,k}}{\partial z_j}(z)\right)^2 \leq \frac{(\Theta_n + \frac{\pi \|z + \phi_x(z)\|}{2\tau_S})^2 (D-d)}{1 - (\Theta_n + \frac{\pi \|z + \phi_x(z)\|}{2\tau_S})^2 (D-d)}$ .

That directly implies, by Cauchy Schwartz, that  $\|{}^t J_x(z) J_x(z)\|_\infty \leq \frac{(\Theta_n + \frac{\pi \|z + \phi_x(z)\|}{2\tau_S})^2 (D-d)}{1 - (\Theta_n + \frac{\pi \|z + \phi_x(z)\|}{2\tau_S})^2 (D-d)}$ .

Then, because  ${}^t J_x J_x$  is a  $d$ -dimensional symmetric positive matrix that has a rank upper bounded by  $\min(d, D-d)$  and eigen values upper bounded by  $d \|{}^t J_x(z) J_x(z)\|_\infty$ , it comes that

$$1 \leq \det(I + {}^t J_x(z) J_x(z)) \leq \left(1 + \frac{(\Theta_n + d \frac{\pi \|z + \phi_x(z)\|}{2\tau_S})^2 (D-d)}{1 - (\Theta_n + \frac{\pi \|z + \phi_x(z)\|}{2\tau_S})^2 (D-d)}\right)^{\min(d, D-d)}, \quad (24)$$

$$g_{x,R}(z) = \sqrt{\det(I + {}^t J_x(z) J_x(z))} \frac{f(x + z + \phi_x(z))}{\mathbb{P}_X(B(x, R))} \mathbb{I}_{S_{x,R}}(z).$$

Conditions on  $\Theta_n$  and  $R$  imply that  $1 \leq \sqrt{\det(I + {}^t J_x(z) J_x(z))} \leq 2^{\frac{\min(d, D-d)}{2}}$  and thus  $G_{\min} = f_{\min}^{(0)}$  and  $G_{\max} = \frac{f_{\max}^{(0)} 2^{\frac{\min(d, D-d)}{2}}}{\mathbb{P}_X(B(x, R))}$  are suitable constants for point 1 of Lemma 5.

Also  $\frac{f(x)}{\mathbb{P}_X(B(x, R))} \leq g_{x,R}(0) \leq \frac{f(x)}{\mathbb{P}_X(B(x, R))} \left(1 + \frac{(\Theta_n)^2 (D-d)}{1 - (\Theta_n)^2 (D-d)}\right)^{\frac{\min(d, D-d)}{2}}$  allows to prove the second point with  $C' = \frac{(D-d)^2}{2}$ .

Finally, when  $z \in S_{x,r}$ , introduce  $\Delta_J = |\sqrt{\det(I + {}^t J_x(z) J_x(z))} - \sqrt{\det(I + {}^t J_x(0) J_x(0))}|$ . We have  $|g_{x,R}(z) - g_{x,R}(0)| \leq \Delta_J \frac{f_{\max}^{(0)}}{\mathbb{P}_X(B(x, R))} + \frac{\sqrt{\det(I + {}^t J_x(0) J_x(0))}}{\mathbb{P}_X(B(x, R))} K \|z + \phi_x(z)\|$ . Then, using that  $\Delta_j \lesssim \Theta_n^2 + \|z\|^2$ ,  $\sqrt{\det(I + {}^t J_x(0) J_x(0))} \lesssim (1 + \Theta_n^2)$  and Equation (37) to have  $\|z + \phi_x(z)\| \lesssim \|z\|$ , we can conclude the proof of point 3 of Lemma 5.  $\square$

**Lemma 6.** *There exist constants  $\beta'_1, \beta'_2, C_1$  and  $C_2$  such that, for  $n$  large enough,*

$$\mathbb{E} \left( \left| \frac{n}{k} \hat{f}_{r,A}(x) - g_{x,R}(0) \right| \middle| N_{x,R} = k \right) \leq \frac{C_1}{\sqrt{kr^d}} + C_2 r + \beta'_1 \exp(-\beta'_2 k r^d) + \exp(-k/8).$$

*Proof.* Introduce  $\hat{\Gamma}_{x,r,R} = \int_{\hat{C}_{x,r,R}} g_{x,R}(z) dz$ , and the event  $\hat{E}_{x,r,R} = \{|\hat{C}_{x,r,R}| \geq A \omega_D r^D, \tilde{\Gamma}_{x,r,R} \geq g_{x,R}(0) A \omega_D r^D \text{ and } \hat{N}_{x,r,R}^\partial \leq N_{x,R}/2\}$  and define

$$\hat{\varepsilon}_1(x) = \frac{\hat{N}_{x,r,R}^o - \hat{\Gamma}_{x,r,R}(n - \hat{N}_{x,r,R}^\partial)}{|\hat{C}_{x,r,R}|(n - \hat{N}_{x,r,R}^\partial)} \mathbb{I}_{\hat{E}_{x,r,R}} \quad \text{and} \quad \hat{\varepsilon}_2(x) = \left( \frac{\hat{\Gamma}_{x,r,R}}{|\hat{C}_{x,r,R}|} - g_{x,R}(0) \right) \mathbb{I}_{\hat{E}_{x,r,R}} \quad (25)$$

We are in the full dimensional case and we have

$$\mathbb{E} \left( \left| \frac{n}{k} \hat{f}_{r,R,A}(x) - g_{x,R}(0) \right| \middle| N_{x,R} = k \right) \leq \mathbb{E}(|\hat{\varepsilon}_1(x)| | E_{x,r}, N_{x,R} = k) + \mathbb{E}(|\hat{\varepsilon}_2(x)| | E_{x,r}, N_{x,R} = k) + G_{\max} \mathbb{P}(\hat{E}_{x,r}^c | N_{x,R} = k). \quad (26)$$

By Cauchy Schwartz and Lemma 2 Equation (13), we obtain:

$$\mathbb{E}(|\hat{\varepsilon}_1(x)| | E_{x,r}, N_{x,R} = k) \leq \sqrt{\mathbb{E}(\hat{\varepsilon}_1(x)^2 | E_{x,r}, N_{x,R} = k)} \leq \frac{2G_{\max}}{A\omega_d r^d}.$$

Exactly as in Lemma 3, we have that (deterministically)

$$|\hat{\varepsilon}_2(x)| \leq \max_{S_{x,R} \cap B(x,r)} |g_{x,R}(z) - g_{x,R}(0)| + g_{x,R}(0) \frac{|H(B(0,r) \cap S_{x,R}) \setminus S_{x,R}|}{A\omega_d r^d}.$$

Thus by application of Lemma 5, we obtain

$$|\hat{\varepsilon}_2(x)| \leq K'(r + \Theta_n^2 + r^2 \Theta_n^2) + \frac{G_{\max}}{A\omega_d r^d} |H(B(0,r) \cap S_{x,R}) \setminus S_{x,R}|.$$

When  $B(0,r) \subset S_{x,R}$ , we have  $|H(B(0,r) \cap S_{x,R}) \setminus S_{x,R}| = 0$ . To bound  $|H(B(0,r) \cap S_{x,R}) \setminus S_{x,R}|$  when  $B(0,r) \cap S_{x,R}^c \neq \emptyset$  and  $r$  is small enough, we can apply Lemma A.7 in [1] from which there exists  $R' > 0$  (depending on  $R$  and the reaches  $\tau_M$  and  $\tau_{\partial M}$ ) and  $\eta$  a unit vector such that  $B(0^* - R'\eta, R') \cap \hat{T}_x \subset S_{x,R} \subset B^c(0^* + R'\eta, R') \cap \hat{T}_x$ , where  $0^*$  is the projection of 0 on  $\pi_{\hat{T}_x}(\partial S \cap B(x,R))$ . This local version of the rolling ball condition allows, as in Proposition 4, to obtain that, for  $r < R'$ , there exists  $y \in \hat{T}_x$  such that  $B(y, r/2) \cap \hat{T}_x \subset B(0,r) \cap S_{x,R}$ . This local ball standardness and lower bound on  $g_{x,R}$  allows to bound on  $\mathbb{P}(E_{x,r}^c | N_{x,R} = k)$  as in Lemma 1.  $\square$

## Acknowledgment

Authors thank the reviewers and the editor for their constructive comments, suggestions, and questions. The authors acknowledge support of the Institut Henri Poincaré (UAR 839 CNRS-Sorbonne Université), and LabEx CARMIN (ANR-10-LABX-59-01) as well as all the colleagues met during the "Geometry and Statistics in Data Science" trimester for their insightful discussions. This work has been supported by ANR GeoDSic. We adress special thanks to Ricardo Fraiman for his nice comments and suggestions.

## References

- [1] Eddie Aamari, Catherine Aaron, and Clément Levrard. Minimax boundary estimation and estimation with boundary. *Bernoulli*, 2023.
- [2] Catherine Aaron and Olivier Bodart. Local convex hull support and boundary estimation. *J. Multivar. Anal.*, 147:82–101, 2016.
- [3] Nicolas Baldin and Markus Reiss. Unbiased estimation of the volume of a convex body.
- [4] Clément Berenfeld and Marc Hoffmann. Density estimation on an unknown submanifold. *arXiv: Statistics Theory*, 2019.

- [5] Tyrus Berry and Tim Sauer. Density estimation on manifolds with boundary. *Comput. Stat. Data Anal.*, 107:1–17, 2015.
- [6] Karine Bertin, Nicolas Klutchnikoff, José Rafael León, and Clémentine Prieur. Adaptive density estimation on bounded domains under mixing conditions. *Electronic Journal of Statistics*, 2020.
- [7] Jean-Daniel Boissonnat, André Lieutier, and Mathijs Wintraecken. The reach, metric distortion, geodesic convexity and the variation of tangent spaces. *Journal of Applied and Computational Topology*, 3:29 – 58, 2019.
- [8] Taoufik Bouezmarni and Jeroen V. K. Rombouts. Nonparametric density estimation for multivariate bounded data. 2007.
- [9] Victor-Emmanuel Brunel. Uniform deviation and moment inequalities for random polytopes with general densities in arbitrary convex bodies. *arXiv: Probability*, 2017.
- [10] José E. Chacón and Tarn Duong. Multivariate kernel smoothing and its applications. 2018.
- [11] Arthur Charpentier and Ewen Gallic. Kernel density estimation based on ripley’s correction. *GeoInformatica*, 20:95–116, 2015.
- [12] A. Cuevas. On pattern analysis in the non-convex case. *Kybernetes*, 19:26–33, 1990.
- [13] Antonio Cuevas and Alberto Rodríguez-Casal. On boundary estimation. *Advances in Applied Probability*, 36(2):340–354, 2004.
- [14] Luc Devroye. A course in density estimation. 1987.
- [15] Luc Devroye and Gábor Lugosi. Combinatorial methods in density estimation. In *Springer Series in Statistics*, 2001.
- [16] Vincent Divol. Measure estimation on manifolds: an optimal transport approach. *Probability Theory and Related Fields*, 183:581 – 647, 2021.
- [17] H. Federer. Curvature measures. *ransactions of the American Mathematical Society*, 93:418–491, 1959.
- [18] Herbert Fédérer. Geometric measure theory. 1969.
- [19] Benedikt Funke and Rafael Kawka. Nonparametric density estimation for multivariate bounded data using two non-negative multiplicative bias correction methods. *Comput. Stat. Data Anal.*, 92:148–162, 2015.
- [20] Wayne M. Getz and Christopher C. Wilmers. A local nearest-neighbor convex-hull construction of home ranges and utilization distributions. *Ecography*, 27:489–505, 2004.
- [21] M. Chris Jones, Oliver B. Linton, and Jens Perch Nielsen. A simple bias reduction method for density estimation. *Biometrika*, 82:327–338, 1995.
- [22] Yoon Tae Kim and Hyun Suk Park. Geometric structures arising from kernel density estimation on riemannian manifolds. *J. Multivar. Anal.*, 114:112–126, 2013.
- [23] Alexandre Leblanc. A bias-reduced approach to density estimation using bernstein polynomials. *Journal of Nonparametric Statistics*, 22:459 – 475, 2010.

- [24] J. S. Marron and David Ruppert. Transformations to reduce boundary bias in kernel density estimation. *Journal of the royal statistical society series b-methodological*, 56:653–671, 1994.
- [25] Emanuel Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.
- [26] Bruno Pelletier. Kernel density estimation on riemannian manifolds. *Statistics & Probability Letters*, 73:297–304, 2005.
- [27] Alberto Rodríguez-Casal and Paula Saavedra-Nieves. A data-adaptive method for estimating density level sets under shape conditions. *The Annals of Statistics*, 2022.
- [28] Murray Rosenblatt. Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27:832–837, 1956.
- [29] David Ruppert and Daren B. H. Cline. Bias reduction in kernel density estimation by smoothed empirical transformations. *Annals of Statistics*, 22:185–210, 1994.
- [30] David W. Scott. Multivariate density estimation, theory, practice and visualization. *The Statistician*, 43:218–219, 1992.
- [31] Bernard Walter Silverman. Density estimation for statistics and data analysis. 1987.
- [32] G. Walther. Granulometric smoothing. *The Annals of Statistics*, 25:2273–2299, 1997.
- [33] G. Walther. On a generalization of blaschke’s rolling theorem and the smoothing of surfaces. *Mathematical Methods in The Applied Sciences*, 22:301–316, 1999.

## A Basic Tools

**Proposition 1.** *Let  $p \in ]0, 1[$  and  $(n, q, D) \in (\mathbb{N}^*)^3$ . We have*

$$\sum_{k=1}^n \binom{n}{k} k^{-\frac{2q}{D+1}} p^k (1-p)^{n-k} \leq \exp\left(-\frac{np}{10}\right) + \left(\frac{np}{2}\right)^{-\frac{2q}{D+1}}.$$

*Proof.*

$$\sum_{k=1}^n \binom{n}{k} k^{-\frac{2q}{D+1}} p^k (1-p)^{n-k} \leq \sum_{k=1}^{\lfloor np/2 \rfloor} \binom{n}{k} k^{-\frac{2q}{D+1}} p^k (1-p)^{n-k} + \sum_{k=\lceil np/2 \rceil}^n \binom{n}{k} k^{-\frac{2q}{D+1}} p^k (1-p)^{n-k}.$$

Due to Bennet’s Inequality that implies that  $\mathbb{P}(K \leq np/2) \leq \exp(-np/10)$ , we have that

$$\sum_{k=1}^n \binom{n}{k} k^{-\frac{2q}{D+1}} p^k (1-p)^{n-k} \leq \exp\left(-\frac{np}{10}\right) + \left(\frac{np}{2}\right)^{-\frac{2q}{D+1}}.$$

That concludes the proof of the proposition. □

**Proposition 2.** *Let  $g : \mathring{S} \subset \mathbb{R}^D \mapsto \mathbb{R}$  be a  $\mathcal{C}^2$  function such that  $\sup_{\mathring{S}} \|\mathcal{H}_g(z)\|_{op} < +\infty$ . For all  $x \in \mathring{S}$  and all  $y$  such that  $(x, y) \subset S$ , we have:  $|g(y) - g(x) - (y-x) \uparrow \nabla_g(x)| \leq \sup_{\mathring{S}} \|\mathcal{H}_g(z)\|_{op} \|y-x\|^2$ .*

*Proof.* Let  $x \in \mathring{S}$ , let  $u$  be an unit vector and  $\lambda \in (0, \lambda_0)$  with  $\lambda_0 = \sup\{z, (x, z) \in \mathring{S}\}$ . Taylor Lagrange applied to  $G_{x,u} : \mathbb{R} \mapsto \mathbb{R}$   $G_{x,u}(\lambda) = g(x + \lambda u)$  provides the existence of  $\mu \in (0, \lambda)$  such that  $G_{x,u}(\lambda) - G_{x,u}(0) = \lambda G'_{x,u}(0) + \frac{\lambda^2}{2} G''_{x,u}(\mu)$ . Thus, we have  $|G_{x,u}(\lambda) - G_{x,u}(0) - \lambda u \cdot \nabla g(x)| \leq \frac{\lambda^2}{2} |{}^t u \mathcal{H}_f(x + \mu u) u| \leq \frac{\lambda^2}{2} \sup_{\mathring{S}} \|\mathcal{H}_f(z)\|_{op}$ .  $\square$

**Proposition 3.** *If  $t$  is a standard level, there exists  $c_t$  such that, with probability one for  $n$  large enough, for all  $\varepsilon > 0$ ,  $L_{t+\varepsilon} \subset (L_{t+\varepsilon} \cap \mathbb{X}_n) \oplus (\frac{4}{t\delta_t\omega_D})^{1/D} (\ln n/n)^{1/D} B$ .*

*Proof.* Deterministically cover  $L_{t+\varepsilon}$  with  $\nu \leq C_t \frac{n}{a^D \ln n}$  balls of radius  $a(\ln n/n)^{1/D}$  centered at some  $x_1, \dots, x_\nu$  in  $L_t$ . If there exists  $x \in L_{t+\varepsilon}$  such that  $d(x, \mathbb{X}_n) \geq b(\ln n/n)^{1/D}$ , then there exists a  $x_i$  with  $B(x_i, (b-a)(\ln n/n)^{1/D}) \cap \mathbb{X}_n = \emptyset$ . By standardness of the level and classical calculation, we obtain  $\mathbb{P}(d(x, \mathbb{X}_n) \geq b(\ln n/n)^{1/D}) \leq \nu(1 - t\delta_t\omega_D(b-a)^D \ln n/n)^n = O(n^{1-t\delta_t\omega_D(b-a)^D})$ . The choice of  $b = (\frac{4}{t\delta_t\omega_D})^{1/D}$  and  $a = \frac{4^{1/D} - 3^{1/D}}{(t\delta_t\omega_D)^{1/D}}$  allow to have  $\sum \mathbb{P}(d(x, \mathbb{X}_n) \geq b(\ln n/n)^{1/D}) < +\infty$ . That concludes the proof by application of Borrel-Cantelli Lemma.  $\square$

## B Technical results on sets with rolling ball property

**Proposition 4.** *If  $E \subset \mathbb{R}^D$  satisfies the  $r_0$ -inside and outside rolling ball condition, then  $E$  and  $\bar{E}^c$  are  $(2r_0, 1/2)$ -ball standard.*

*Proof.* First let prove that, for all  $r \leq 2r_0$ , there exists  $y$  such  $B(y, r/2) \subset E \cap B(x, r)$ . If  $d(x, \partial S) \geq r/2$ , one can trivially choose  $y = x$ . If now  $d(x, \partial S) < r/2 \leq r_0$ , introduce  $x^* = \pi_{\partial S}(x)$ . We have that  $x = x^* - d(x, \partial S)\eta_{x^*}$  and  $O_x = x^* - r_0\eta_{x^*}$ . By rolling ball condition, we have that  $B(O_x, r_0) \cap B(x, r) \subset E \cap B(x, r)$ . The choice of  $y = x^* - (r/2)\eta_{x^*}$  is convenient since, if  $z \in B(y, r/2)$ , we have

1.  $\|z - x\| \leq \|z - y\| + \|y - x\| = \|z - y\| + (r/2 - d(x, \partial S)) \leq r$ , thus  $z \in B(x, r)$ ;
2.  $z = y + \alpha\eta_{x^*} + \beta v$  with  $v$  a unit vector of  $\eta_{x^*}^\perp$  and  $\alpha^2 + \beta^2 \leq (r/2)^2$ . Thus we have  $\|O_y - z\|^2 = (r_0 - (r/2) + \alpha)^2 + \beta^2 \leq r_0^2 - (2r_0 - r)(\alpha - r/2) \leq r_0^2$ . And so we obtain  $z \in B(O_x, r_0)$ .

Suppose that  $E$  has more than one connected component and consider two connected components of  $E$ :  $E_1$  and  $E_2$ . Suppose that  $\inf_{x \in E_1, y \in E_2} \|x - y\| = \ell < 2r_0$ . Compacity of  $E_1$  and  $E_2$  allows to turn the inf into a min which is realized for some  $x^*$  and  $y^*$  respectively located at the boundary of  $E_1$  and  $E_2$ , and consider now  $O = \frac{x^* + y^*}{2}$ . By inside rolling ball condition we must have  $O - x^* = (\ell/2)\eta_{x^*}$  and  $O - y^* = (\ell/2)\eta_{y^*}$  (otherwise we can contradict the minimum distance) which, in turn contradicts the outside rolling ball condition because  $y^* \in B(x^* + r_0\eta_{x^*}, r_0)$ . Thus, if  $E$  has many path connected component  $E_1, \dots, E_k$ ,  $x \in E_i$  and  $r \leq 2r_0$ , then  $B(x, r) \cap E = B(x, r) \cap E_i$ .  $E_i$  has the  $r_0$  rolling ball property thus positive reach  $r_0$ . So, by application of Corollary 1 in [7],  $B(x, r)$  is geodesically convex and then path connected since  $r < r_0$ .  $\square$

**Proposition 5.** *Let  $S \subset \mathbb{D}^D$  be a set that has the  $R_0$  inside and outside ball property. For all  $r \leq R_0/2$  and all  $x \in S$ , we have:*

$$d(H(B(x, r) \cap S), S) \leq r^2/R_0, \quad (27)$$

$$|H(B(x, r) \cap S) \setminus S| \leq \frac{4\omega_{D-1}}{3R_0} r^{D+1}. \quad (28)$$

*Proof.* If  $d(x, \partial S) \geq r$ , we have  $H(B(x, r) \cap S) \setminus S = \emptyset$ , thus it has a null volume. Consider now  $x$  such that  $d(x, \partial S) = h < r$ . As possible due to rolling ball condition, for  $r \leq R_0$ , introduce  $x^* = \pi_{\partial S}(x)$ ,  $O_{\text{in}} = x^* - R_0 \eta_{x^*}$  and  $O_{\text{out}} = x^* + R_0 \eta_{x^*}$ . Let  $v$  be a unit vector of  $\eta_{x^*}^\perp$  and  $z = x + \alpha \eta_{x^*} + \beta v \in B(x, r) \cap S$ . By rolling ball property,  $\|z - O_{\text{out}}\|^2 > R_0^2$ , and we have

$$\begin{cases} \|z - x\|^2 = \alpha^2 + \beta^2 \leq r^2 \\ \|z - O_{\text{out}}\|^2 = (\alpha - (h + R_0))^2 + \beta^2 > R_0^2 \end{cases} \Rightarrow \alpha \leq \frac{(h + R_0)^2 - R_0^2 + r^2}{2(h + R_0)} = h + \frac{r^2 - h^2}{2(R_0 + h)}.$$

Introduce  $H_+ = \left\{ z, \langle z - x, \eta_* \rangle \leq h + \frac{r^2 - h^2}{2(R_0 + h)} \right\}$ . We so have  $B(x, r) \cap S \leq B(x, r) \cap H_+$  and

$$H(B(x, r) \cap S) \subset B(x, r) \cap H_+.$$

Similarly  $\{z, \|z - x\|^2 \leq r^2, \|z - O_{\text{in}}\|^2 \leq R_0^2\} \subset B(x, r) \cap S$  and  $B(x, r) \cap H_- \subset B(x, r) \cap S$  with  $H_- = \left\{ z, \langle z - x, \eta_* \rangle \leq h - \frac{r^2 - h^2}{2(R_0 - h)} \right\}$ . Thus we have

$$H(B(x, r) \cap S) \setminus S \subset (B(x, r) \cap H_+) \setminus S = (B(x, r) \cap H_+) \setminus (B(x, r) \cap S) \subset (B(x, r) \cap H_+) \setminus (B(x, r) \cap H_-^c).$$

This finally implies that  $H(B(x, r) \cap S) \setminus S \subset B(x, r) \cap (H_+ \setminus H_-)$ . Thus  $H(B(x, r) \cap S) \setminus S$  is included in the cylinder  $\mathcal{C}$  defined as follows:

$$\mathcal{C} = \left\{ z = x + \alpha \eta_{x^*} + \beta v, \|v\| = 1, v \in \eta_{x^*}^\perp, |\beta| r, h - \frac{r^2 - h^2}{2(R_0 - h)} \leq \alpha \leq h + \frac{r^2 - h^2}{2(R_0 + h)} \right\}.$$

We then can conclude the proof arguing that  $|\mathcal{C}| = \omega_{D-1} r^{D-1} \frac{(r^2 - h^2) R_0}{R_0^2 - h^2} \leq \frac{4\omega_{D-1}}{3R_0} r^{D+1}$ .  $\square$

## C Technical results to bound $\|\nabla f(x)\| \frac{|B(x, r) \setminus C_{x, r}|}{|B(x, r)|}$

Introduce, for  $x$  a point where  $f(x) > 0$ ,  $g_{x, r}(y) = r^D \frac{f(x + ry)}{\int_{B(x, r)} f(z) dz} \mathbb{I}_{B(0, 1)}(y)$ , which is the density of a local re-scaling of the sample. When exists, also introduce  $M_{x, r} = \max_{y \in B(0, 1)} g_{x, r}(y)$ . From the results in Theorem 1, Corollary 1 and Theorem 3 in [9], we have:

**Proposition 6.**

$$\mathbb{P} \left[ k \left( \frac{\int_{B(x, r) \setminus C_{x, r}} f(z) dz}{\int_{B(x, r)} f(z) dz} - C_1 (M_{x, r} + 1) k^{-2/(D+1)} \right) > t | N_{x, r} = k \right] \leq C_2 e^{-t}, \quad (29)$$

$$\mathbb{E} \left[ \left( \frac{\int_{B(x, r) \setminus C_{x, r}} f(z) dz}{\int_{B(x, r)} f(z) dz} \right)^q | N_{x, r} = k \right] \leq A_q (M_{x, r} + 1)^q k^{-2q/(D+1)}, \quad (30)$$

$$\mathbb{E} \left[ (N_{x, r}^\partial)^q | N_{x, r} = k \right] \leq B_q (M_{x, r} + 1)^q k^{q(D-1)/(D+1)}, \quad (31)$$

where  $C_1$  and  $C_2$  only depend on the dimension  $D$ ,  $A_q$  and  $B_q$  depend on the dimension  $D$  and  $q \in \mathbb{N}^*$ .

**Proposition 7.** Let  $f$  be a  $\mathcal{C}^2$  density supported in  $S$  such that, for all  $x \in S$ ,

$$g(x) = \|\nabla f(x)\| \exp(3\|\nabla \ln(f)(x)\|) \leq L_{\max}.$$

Additionally, assume that, for all  $x$ , we have that  $B(x, r) \cap S$  is path connected and that, for all  $(x, y) \in S^2$  which are path connected,  $d_S(x, y) \leq K_{\max} \|x - y\|$ .

First, we have that  $f_{\max}^{(1)} := \sup_S \|\nabla_f(x)\|$  is well defined.

Second, for all  $r \leq \frac{1}{2K_{\max}}$ , for all  $x \in S$  such that  $B(x, r) \cap S \subset \{z, f(z) > 0\}$ , we have:

$$G_r(x) := \|\nabla_f(x)\| \frac{\max_{z \in B(x, r) \cap S} f(z)}{\min_{z \in B(x, r) \cap S} f(z)} \left( \frac{1}{\omega_d} \frac{\max_{z \in B(x, r) \cap S} f(z)}{\min_{z \in B(x, r) \cap S} f(z)} + 1 \right) \leq \max \left( \frac{2L_{\max}}{\omega_d}, f_{\max}^{(1)} \right).$$

*Proof.* Preliminary introduce  $S_- = \{x \in S, f(x) < 1\}$  and  $S_+ = \{x \in S, f(x) \geq 1\}$ . Because  $S_+$  is a compact, regularity condition on  $f$  implies that  $\max_{S_+} \|\nabla_f(x)\|$  exists. Suppose now that  $x \in S_-$ . We have that:  $\|\nabla_f(x)\| \exp(3\|\nabla_f(x)\|) \leq \|\nabla_f(x)\| \exp(3\|\nabla_f(x)/f(x)\|) \leq L_{\max}$ . Because  $\varphi : \mathbb{R}^+ \mapsto \mathbb{R}^+$  defined by  $\varphi(x) = xe^{3x}$  is an increasing one to one function, we have  $\|\nabla_f(x)\| \leq \varphi^{-1}(L_{\max})$  and thus  $\sup_{x \in S_-} \|\nabla_f(x)\|$  is well defined. Therefore,  $f_{\max}^{(1)} = \sup_{x \in S} \|\nabla_f(x)\|$  is well defined.

Suppose that  $r \leq (2K_{\max})^{-1}$  is fixed and consider  $x$  such that  $B(x, r) \subset S$ . First, notice that:

$$\text{If } \frac{\max_{z \in B(x, r)} f(z)}{\min_{z \in B(x, r)} f(z)} \leq \omega_d, \text{ then } G_r(x) \leq 2\|\nabla_f(x)\|. \quad (32)$$

Second, suppose that  $x$  is such that  $\frac{\max_{z \in B(x, r)} f(z)}{\min_{z \in B(x, r)} f(z)} > \omega_d$ , then  $G_r(x) \leq \frac{2}{\omega_d} \|\nabla_f(x)\| \left( \frac{\max_{z \in B(x, r)} f(z)}{\min_{z \in B(x, r)} f(z)} \right)^2$ . Denote by  $z_+$  (resp  $z_-$ ) the point in  $B(x, r) \cap S$  where  $\max_{B(x, r) \cap S} f(z)$  (resp.  $\min_{B(x, r) \cap S} f(z)$ ) is realized. Because  $z_+$  and  $z_-$  are path connected, by application of the mean value theorem, we get

$$\frac{\ln(f(z_+)) - \ln(f(z_-))}{d_S(z_+, z_-)} \leq \max_{B(x, r)} \|\nabla_{\ln(f)}(y)\|.$$

Because  $r \leq \frac{1}{2K_{\max}}$  and the geodesic condition, we have

$$\frac{\max_{z \in B(x, r)} f(z)}{\min_{z \in B(x, r)} f(z)} \leq \exp \left( \max_{B(x, r)} \|\nabla_{\ln(f)}(y)\| \right). \quad (33)$$

Thus, we obtain  $G_r(x) \leq \frac{2}{\omega_d} \|\nabla_f(x)\| \exp(2 \max_{B(x, r)} \|\nabla_{\ln(f)}(y)\|)$ , from which we have

$$G_r(x) \leq \frac{2}{\omega_d} f(x) \frac{\|\nabla_f(x)\|}{f(x)} \exp \left( 2 \max_{B(x, r)} \|\nabla_{\ln(f)}(y)\| \right),$$

and finally, we get

$$G_r(x) \leq \frac{2}{\omega_d} f(x) \|\nabla_{\ln(f)}(x)\| \exp \left( 2 \max_{B(x, r)} \|\nabla_{\ln(f)}(y)\| \right).$$

In the sequel  $z$ , is a point of  $B(x, r)$  where  $\|\nabla_{\ln(f)}(z)\| = \max_{y \in B(x, r)} \|\nabla_{\ln(f)}(y)\|$ . We clearly have  $G_r(x) \leq \frac{2}{\omega_d} f(x) \|\nabla_{\ln(f)}(z)\| \exp(2\|\nabla_{\ln(f)}(z)\|)$ . Thus, we get:

$$G_r(x) \leq \frac{2}{\omega_d} \frac{f(x)}{f(z)} \|\nabla_f(z)\| \exp(2\|\nabla_{\ln(f)}(z)\|),$$

and, by new application of (33), we have  $G_r(x) \leq \frac{2}{\omega_d} \|\nabla_f(z)\| \exp(3\|\nabla_{\ln(f)}(z)\|)$ . Thus, we obtain

$$\text{if } \frac{\max_{z \in B(x, r)} f(z)}{\min_{z \in B(x, r)} f(z)} \geq \omega_d, \text{ then } G_r(x) \leq 2\|\nabla_f(z)\| \exp(3\|\nabla_{\ln(f)}(z)\|) \leq \frac{2L_{\max}}{\omega_d}. \quad (34)$$

This and Equation (32) conclude the proof of the second part of the proposition.  $\square$

## D Manifolds with positive reach

We recall here some properties of manifolds with positive reach that we will use in the following. First, by Lemma 3 in [7], by prop 5.2 and prop 5.4 in [1], we have that, when  $\|x - y\| \leq 2\tau_S$ ,

$$\|x - y\| \leq d_S(x, y) \leq \frac{\pi}{2}\|x - y\|; \quad \|\pi_{\hat{T}_x}(y - x)\| \geq \|y - x\| \left(1 - \angle \hat{T}_x, T_x S - \frac{\|x - y\|}{2\tau_S}\right); \quad (35)$$

$$\angle T_x S, T_y S \leq \frac{d_S(x, y)}{\tau_S}; \quad \|\pi_{\hat{T}_x^\perp}(y - x)\| \leq \|y - x\| \left(\angle \hat{T}_x, T_x S + \frac{\|x - y\|}{2\tau_S}\right). \quad (36)$$

In particular, when  $\|\pi_{\hat{T}_x}(y - x)\| < \frac{\tau_S(1 - \angle \hat{T}_x, T_x S)^2}{2}$ , by (35) it comes that either

$$\|y - x\| \leq \tau_S(1 - \angle \hat{T}_x, T_x S) \left(1 - \sqrt{1 - \frac{2\|\pi_{\hat{T}_x}(y - x)\|}{\tau_S(1 - \angle \hat{T}_x, T_x S)^2}}\right) \leq \frac{2\|\pi_{\hat{T}_x}(y - x)\|}{(1 - \angle \hat{T}_x, T_x S)}, \quad (37)$$

or

$$\|y - x\| \geq \tau_S(1 - \angle \hat{T}_x, T_x S) \left(1 + \sqrt{1 - \frac{2\|\pi_{\hat{T}_x}(y - x)\|}{\tau_S(1 - \angle \hat{T}_x, T_x S)^2}}\right) \geq \tau_S(1 - \angle \hat{T}_x, T_x S). \quad (38)$$

This last equation (38) implies, again by (35), that we also have  $\|\pi_{\hat{T}_x}(y - x)\| > \frac{\tau_S(1 - \angle \hat{T}_x, T_x S)^2}{2}$  and thus, by contradiction, only (37) is true.

**Proposition 8.** *If for all  $x \in S$ ,  $\angle T_x S, \hat{T}_x \leq \theta < 1/12$ , then for all  $x \in S$ ,  $\pi_{\hat{T}_x}$  is injective on  $B(x, \tau_S/3)$ .*

*Proof.* For  $y$  and  $y'$  in  $B(x, \tau_S/3)$ , first  $\angle T_y S, \hat{T}_x \leq \angle T_y S, T_x S + \theta$ . Thus by (36) and (35), we have  $\angle T_y S, \hat{T}_x \leq \frac{d_S(x, y)}{\tau_S} + \theta \leq \frac{\pi\|x - y\|}{2\tau_S} + \theta \leq \frac{\pi}{6} + \theta$ . From that and Equation (35), we obtain

$$\|\pi_{\hat{T}_x}(y' - y)\| \geq \|y' - y\| \left(1 - \angle \hat{T}_x, T_y S - \frac{\|y' - y\|}{2\tau_S}\right).$$

With that, equation (35) gives :

$$\|\pi_{\hat{T}_x}(y' - y)\| \geq \|y' - y\| \left(\frac{4 - \pi}{6} - \theta\right) \geq \|y' - y\| \frac{7 - 2\pi}{12},$$

from which we obtain the injectivity.  $\square$