



**HAL**  
open science

## Density estimation on an unknown submanifold

Clément Berenfeld, Marc Hoffmann

► **To cite this version:**

Clément Berenfeld, Marc Hoffmann. Density estimation on an unknown submanifold. 2019. hal-02329703

**HAL Id: hal-02329703**

**<https://hal.science/hal-02329703v1>**

Preprint submitted on 23 Oct 2019

**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.

# Density estimation on an unknown submanifold

Clément Berenfeld\* and Marc Hoffmann†

October 18, 2019

## Abstract

We investigate density estimation from a  $n$ -sample in the Euclidean space  $\mathbb{R}^D$ , when the data is supported by an unknown submanifold  $M$  of possibly unknown dimension  $d < D$  under a *reach* condition. We study nonparametric kernel methods for pointwise and integrated loss, with data-driven bandwidths that incorporate some learning of the geometry via a local dimension estimator. When  $f$  has Hölder smoothness  $\beta$  and  $M$  has regularity  $\alpha$  in a sense close in spirit to (Aamari and Levrard, 2019), our estimator achieves the rate  $n^{-\alpha \wedge \beta / (2\alpha \wedge \beta + d)}$  and does not depend on the ambient dimension  $D$  and is asymptotically minimax for  $\alpha \geq \beta$ . Following Lepski’s principle, a bandwidth selection rule is shown to achieve smoothness adaptation. We also investigate the case  $\alpha \leq \beta$ : by estimating in some sense the underlying geometry of  $M$ , we establish in dimension  $d = 1$  that the minimax rate is  $n^{-\beta / (2\beta + 1)}$  proving in particular that it does not depend on the regularity of  $M$ . Finally, a numerical implementation is conducted on some case studies in order to confirm the practical feasibility of our estimators.

**Mathematics Subject Classification (2010):** 62C20, 62G05, 62G07.

**Keywords:** Point clouds, manifold reconstruction, nonparametric estimation, adaptive density estimation, kernel methods, Lepski’s method.

## 1 Introduction

### 1.1 Motivation

Suppose we observe a  $n$ -drawn  $(X_1, \dots, X_n)$  distributed on an Euclidean space  $\mathbb{R}^D$  according to some density function  $f$ . We wish to recover  $f$  at some point arbitrary point  $x \in \mathbb{R}^D$  nonparametrically. If the smoothness of  $f$  at  $x$  measured in a strong sense is of order  $\beta$  – for instance by a Hölder condition or with a prescribed number of derivatives – then the optimal (minimax) rate for recovering  $f(x)$  is of order  $n^{-\beta / (2\beta + D)}$  and is achieved by kernel or projection methods, see *e.g.* the classical textbooks Devroye and Györfi (1985); Silverman (1986) or (Tsybakov, 2008, Sec. 1.2-1.3). Extension to data-driven bandwidths (Bowman, 1984; Chiu, 1991) offers the possibility to adapt to unknown smoothness, see (Goldenshluger and Lepski, 2008, 2014; Goldenshluger et al., 2011) for a modern mathematical formulation. In many situations however, the dimension  $D$  of the ambient space is large, hitherto disqualifying such methods for practical applications. Opposite to the curse of dimensionality, a broad guiding principle in practice is that the observations  $(X_1, \dots, X_n)$  actually live on smaller dimensional structures and that the effective dimension of the problem is

---

\*Clément Berenfeld, Université Paris-Dauphine & PSL, CNRS, CEREMADE, 75016 Paris, France. email: berenfeld@ceremade.dauphine.fr

†Marc Hoffmann, Université Paris-Dauphine & PSL, CNRS, CEREMADE, 75016 Paris, France. email: hoffmann@ceremade.dauphine.fr

smaller if one can take advantage of the geometry of the data (Fefferman et al., 2016). This classical paradigm probably goes back to a conjecture of (Stone, 1982) that paved the way to the study of the celebrated single-index model in nonparametric regression, where a structural assumption is put in the form  $f(x) = g(\langle \vartheta, x \rangle)$ , where  $\langle \cdot, \cdot \rangle$  is the scalar product on  $\mathbb{R}^D$ , for some unknown univariate function  $g: \mathbb{R} \rightarrow \mathbb{R}$  and direction  $\vartheta \in \mathbb{R}^D$ . Under appropriate assumptions, the minimax rate of convergence for recovering  $f(x)$  with smoothness  $\beta$  drops to  $n^{-\beta/(2\beta+1)}$  and does not depend on the ambient dimension  $D$ , see e.g. (Gaïffas and Lecué, 2007; Lepski and Serdyukova, 2014) and the references therein. Also, in the search for significant variables, one postulates that  $f$  only depends on  $d < D$  coordinates, leading to the structural assumption  $f(x_1, \dots, x_D) = F(x_{i_1}, \dots, x_{i_d})$  for some unknown function  $F: \mathbb{R}^d \rightarrow \mathbb{R}$  and  $\{i_1, \dots, i_d\} \subset \{1, \dots, D\}$ . In an analogous setting, the minimax rate of convergence becomes  $n^{-\beta/(2\beta+d)}$  and this is also of a smaller order of magnitude than  $n^{-\beta/(2\beta+D)}$ , see (Hoffmann and Lepski, 2002) in the white noise model.

The next logical step is to assume that the data  $(X_1, \dots, X_n)$  live on a  $d$ -dimensional submanifold  $M$  of the ambient space  $\mathbb{R}^D$ . When the manifold is known prior to the experiment, nonparametric density estimation dates back to (Devroye and Györfi, 1985) when  $M$  is the circle, and on a homogeneous Riemannian manifold by (Hendriks, 1990), see also Pelletier (2005). Several results are known for specific geometric structures like the sphere or the torus involved in many applied situations: inverse problems for cosmological data (Kerkycharian et al., 2011; Kim and Koo, 2002; Kim et al., 2009), in geology (Hall et al., 1987) or flow calculation in fluid mechanics (Eugecioglu and Srinivasan, 2000). For genuine compact homogeneous Riemannian manifolds, a general setting for smoothness adaptive density estimation and inference has recently been considered by Kerkycharian et al. (2012), or even in more abstract metric spaces in Cleanthous et al. (2018). A common strategy adapts conventional nonparametric tools like projection or kernel methods to the underlying geometry, via the spectral analysis of the Beltrami-Laplace operator on  $M$ . Under appropriate assumptions, this leads to exact or approximate eigenbases (spherical harmonics for the sphere, needlets and so on) or properly modified kernel methods, according to the Riemannian metric on  $M$ .

If the submanifold  $M$  itself is unknown, getting closer in spirit to a dimension reduction approach, the situation becomes drastically different:  $M$  hence its geometry is unknown, and considered as a nuisance parameter. In order to recover the density  $f$ , one has to understand the minimal geometry of  $M$  that must be learned from the data and how this geometry affects the optimal reconstruction of  $f$ . This is the topic of the paper.

## 1.2 Main results

We construct a class of compact smooth submanifolds of dimension  $d$  of the Euclidean space  $\mathbb{R}^D$ , without boundaries, that constitute generic models for the unknown support of the target density  $f$  that we wish to reconstruct under a *reach condition*. The reach is a somehow unavoidable notion in manifold reconstruction that goes back to Federer (1959): it is a geometric invariant that quantifies both local curvature conditions and how tightly the submanifold folds on itself. It is related to the scale at which the sampling rate  $n$  can effectively recover the geometry of the submanifold, see Section 2.2 below. We consider regular manifolds  $M$  with reach bounded below that satisfy the following property:  $M$  admits a local parametrization at every point  $x \in M$  by its tangent space  $T_x M$ , and this parametrization is regular enough. A natural candidate is given by the exponential map  $\exp_x: T_x M \rightarrow M \subset \mathbb{R}^D$ . More specifically, for some regularity parameter  $\alpha \geq 1$ , we require a certain uniform bound for the  $\alpha$ -fold differential of the exponential map to hold, quantifying in some

sense the regularity of the parametrization in a minimax spirit. Our approach is close but slightly more restrictive than [Aamari and Levrard \(2019, Def. 1\)](#) that consider arbitrary parametrizations among those close to the inverse of the projection onto tangent spaces. Given a density function  $f : M \rightarrow [0, \infty)$  with respect to the volume measure on  $M$ , we have a natural extension of smoothness spaces on  $M$  by requiring that  $f \circ \exp_x : T_x M \rightarrow \mathbb{R}$  is a smooth map in any reasonable sense, see for instance [Triebel \(1987\)](#) for the characterisation of function spaces on a Riemannian manifold.

Our main result is that in order to reconstruct  $f(x)$  efficiently at a point  $x \in \mathbb{R}^D$  when  $f$  lives on an unknown smooth submanifold, it is sufficient to consider estimators of the form

$$\widehat{f}_h(x) = \frac{1}{nh^{\widehat{d}(x)}} \sum_{i=1}^n K\left(\frac{x_1}{h}, \dots, \frac{x_D}{h}\right), \quad x = (x_1, \dots, x_D) \in \mathbb{R}^D, \quad (1)$$

where  $K : \mathbb{R}^D \rightarrow \mathbb{R}$  is a kernel verifying certain assumptions and  $\widehat{d}(x) = \widehat{d}(x, X_1, \dots, X_n)$  is an estimator of the local dimension of the support of  $f$  in the vicinity of  $x$  based on a scaling estimator as introduced in [Farahmand et al. \(2007\)](#). We prove in [Theorem 1](#) that following a classical bias-variance trade-off for the bandwidth  $h$ , the rate  $n^{-\alpha \wedge \beta / (2\alpha \wedge \beta + d)}$  is achievable for pointwise and global loss functions (in  $L_p(M)$ -norm for  $p \geq 1$ ) when the dimension of  $M$  is  $d$ , irrespectively of the ambient dimension  $D$ . In particular, it is noteworthy that in terms of manifold learning, only the dimension of  $M$  needs to be estimated. When  $\alpha \geq \beta$ , we also have a lower bound ([Theorem 2](#)) showing that our result is minimax. Moreover, by implementing Lepski's principle ([Lepskii, 1992](#)), we are able to construct a data driven bandwidth  $\widehat{h} = \widehat{h}(x, X_1, \dots, X_n)$  that achieves in [Theorem 3](#) the rate  $n^{-\alpha \wedge \beta / (2\alpha \wedge \beta + d)}$  up to a logarithmic term — unavoidable in the case of pointwise loss due to the Lepski-Low phenomenon ([Lepskii, 1990](#); [Low, 1992](#)). When the dimension  $d$  is known, the estimator (1) has already been investigated in squared-error norm in [Ozakin and Gray \(2009\)](#) for a fixed manifold  $M$  and smoothness  $\beta = 2$ .

A remaining question at this stage is to understand how the regularity of  $M$  can affect the minimax rates of convergence for smooth functions, *i.e.* when  $\alpha \leq \beta$ . We restrict our attention to the one-dimensional case  $d = 1$ . When  $M$  is known, [Pelletier \(2005\)](#) studied an estimator of the form

$$\frac{1}{nh^d} \sum_{i=1}^n \frac{1}{\vartheta_x(X_i)} K\left(\frac{d_M(x, X_i)}{h}\right) \quad (2)$$

where  $K : \mathbb{R} \rightarrow \mathbb{R}$  is a radial kernel,  $d_M$  is the intrinsic Riemannian distance on  $M$  and the correction term  $\vartheta_x(X_i)$  is the volume density function on  $M$  ([Besse, 1978](#), p. 154) that accounts for the value of the density of the volume measure at  $X_i$  in normal coordinates around  $x$ , taking into account how the submanifold curves around  $X_i$ . By establishing in [Lemma 3](#) that  $\vartheta_x$  is constant (and identically equal to one) when  $d = 1$ , we have another estimator by simply learning the geometry of  $M$  via its intrinsic distance  $d_M$  in (2). This can be done by efficiently estimating  $d_M$  thanks to the Isomap method as coined by [Tenenbaum et al. \(2000\)](#). We construct an estimator that achieves in [Theorem 5](#) the rate  $n^{-\beta / (2\beta + 1)}$ , therefore establishing that in dimension  $d = 1$  at least, the regularity of the manifold  $M$  does not affect the minimax rate for estimating  $f$  even when  $M$  is unknown. However, the volume density function  $\vartheta_x$  is not constant as soon as  $d \geq 2$  and obtaining a global picture in higher dimensions remains open.

### 1.3 Organisation of the paper

In [Section 2](#), we provide with all the necessary material and notation from classical geometry for the unfamiliar reader, as well as the essential notions we need from manifold learning. We elaborate

in particular on the reach of a subset of the Euclidean space in Section 2.2 and on our statistical model for density estimation on an unknown manifold in Section 2.3. Section 3 describes in details our estimator and its minimax properties in Theorem 1 and Theorem 2 as well as its adaptation properties with respect to an unknown smoothness (Theorem 3). For clarity, we first consider the case when the dimension  $d$  and regularity  $\alpha$  of the submanifold  $M$  and the smoothness  $\beta$  of the density are known (Section 3.3). We next treat adaptation with respect to  $\alpha$  and  $\beta$  (Section 3.4). Finally, we consider adaptation with respect to  $\beta$  and  $d$  in Section 3.5 and construct an estimator in Theorem 4 that adapts to the smoothness of  $f$  and the dimension of  $M$  simultaneously. Section 4 focuses on the case on one-dimensional submanifolds, where we prove that the regularity of the underlying geometry does not affect the minimax rates of convergence for estimating  $f$  pointwise (Theorem 5). Finally, numerical examples are developed in Section 5: we elaborate on examples of non-isometric embeddings of the circle and the torus in dimension 1 and 2 and explore in particular rates of convergence on Monte-Carlo simulations, illustrating how effective Lepski's method can be in that context. The proof are delayed until Section 6.

## 2 Manifold-supported probability distributions

### 2.1 Submanifolds of $\mathbb{R}^D$

We recall some basic notions of geometry of submanifolds of the Euclidean space  $\mathbb{R}^D$  for the unfamiliar reader. We borrow material from the classical textbooks Gallot et al. (2004) and Lee (2006). We endow  $\mathbb{R}^D$  with its usual Euclidean product and norm, respectively denoted by  $\langle \cdot, \cdot \rangle$  and  $\| \cdot \|$ . We denote by  $B_D(x, r)$  the open ball of  $\mathbb{R}^D$  of center  $x$  and radius  $r$ .

#### Smooth submanifolds and tangent spaces

A *smooth submanifold*  $M$  of  $\mathbb{R}^D$  is a connected subset of  $\mathbb{R}^D$  that verifies (Gallot et al., 2004, Prp 1.3 p.3) that for any point  $x \in M$ , there exists an open neighbourhood of  $x$  in  $\mathbb{R}^D$  and a smooth map (called a *parametrization*)  $\phi : \Omega \rightarrow \mathbb{R}^D$  defined in a neighbourhood  $\Omega$  of 0 in  $\mathbb{R}^d$  for some  $d \geq 1$  and such that  $\phi(0) = x$ ;  $d\phi(0)$  is injective; and  $\phi$  is a homeomorphism from  $\Omega$  onto  $U \cap M$ . The integer  $d$ , called the dimension of  $M$ , is the same for all local parametrizations of  $M$ . A vector  $v \in \mathbb{R}^D$  is said to be *tangent* to  $M$  at point  $x \in M$  if there exists a smooth curve  $\gamma : I \rightarrow M$  defined in a neighbourhood of 0 in  $\mathbb{R}$  and such that  $\gamma(0) = x$  and  $\dot{\gamma}(0) = v$ . The set of all tangent vectors at point  $x \in M$  is called the tangent space and is denoted by  $T_x M$ . It is a subspace of  $\mathbb{R}^D$  of dimension  $d$  and it verifies that for any local parametrization  $\phi$  of  $M$  at  $x$ , we have  $T_x M = d\phi(0)[\mathbb{R}^d]$  (Gallot et al., 2004, Thm. 1.22 and 1.23 p.13). We denote by  $T_x M^\perp$  the orthonormal complement of  $T_x M$  in  $M$ , also called the *normal space* at point  $x$ .

#### Geodesics and the exponential map

A smooth curve  $\gamma : I \rightarrow M$  defined on an open set  $I$  of  $\mathbb{R}$  is called a *geodesic* if at any time  $t \in I$ , the acceleration  $\ddot{\gamma}(t)$  lies in the normal space  $T_{\gamma(t)} M^\perp$  (Gallot et al., 2004, Ex 2.80 b. p.81). In particular, the speed  $t \mapsto \|\dot{\gamma}(t)\|$  of a geodesic is constant. For any  $x \in M$  and any  $v \in T_x M$ , there exists a unique maximal geodesic  $\gamma_{x,v}$  defined in an open neighbourhood of 0 such that  $\gamma_{x,v}(0) = x$  and  $\dot{\gamma}_{x,v}(0) = v$  (Gallot et al., 2004, Cor 2.85 p.85). The set of vectors  $v \in T_x M$  such that  $\gamma_{x,v}(1)$  is well defined is a neighbourhood of 0 in  $T_x M$ . We define on this neighbourhood the map

$$\exp_x : v \mapsto \gamma_{x,v}(1),$$

called the *exponential map*. It is smooth and there always exists  $\epsilon > 0$  such that its restriction to the ball  $B_d(0, \epsilon)$  in  $T_x M$  defines a local parametrization of  $M$  at point  $x$  (Gallot et al., 2004, Cor 2.89 p.86). The supremum of all  $\epsilon$  for which the latter holds true is called the *injectivity radius* at point  $x$ .

The uniqueness of  $\gamma_{x,v}$  readily yields that  $\exp_x(tv) = \gamma_{x,v}(t)$ , for any  $t \in \mathbb{R}$  for which both expressions are well-defined, in particular in a neighbourhood of 0. Letting  $t$  go to 0 shows that  $d\exp_x(0) = \text{Id}_{T_x M}$  and that  $d^2\exp_x(0)[v \otimes v] = \ddot{\gamma}_{x,v}(0) \in T_x M^\perp$ . The map  $d^2\exp_x(0)$  is denoted by  $\mathbb{I}_x$  and is called the *second fundamental form* of  $M$  at  $x$  — see Gallot et al. (2004, Def 5.1 p. 246) for a more general definition. The quantity  $\mathbb{I}_x(v \otimes v)$  encapsulates the way the submanifold  $M$  curves around  $x$  in the direction  $v$ . For instance, if  $M$  is flat around  $x$ , then the geodesics originating from  $x$  are locally straight lines and thus  $\mathbb{I}_x = 0$ .

### Length of curves and the Riemannian distance

We say that a curve  $\gamma : [a, b] \rightarrow M$  with  $a < b$  is admissible if there exists a finite subdivision  $a = a_0 < \dots < a_{k-1} < a_k = b$  of  $[a, b]$  such that on any segment  $[a_i, a_{i+1}]$ ,  $\gamma$  is smooth and its derivative does not vanish. For such a curve, we define its length as

$$L(\gamma) = \int_a^b \|\dot{\gamma}(t)\| dt.$$

For any two points  $x, y \in M$ , we define the Riemannian distance  $d_M(x, y)$  between  $x$  and  $y$  as the infimum of the length of all the admissible curves that join  $x$  to  $y$ . One can show that  $d_M$  is well-defined on  $M \times M$  (recall that we only consider connected submanifolds) and that it is indeed a distance. It endows  $M$  with a topology that coincides with the one induced by the ambient space, see Lee (2006, Lem 6.2 p.94) and Arias-Castro and Le Gouic (2019, Lem 3). An admissible curve  $\gamma$  that reaches the infimum in the definition of  $d_M$  is called *minimizing*. We know that every minimizing curve is a geodesic of  $M$ , up to a reparametrization (Lee, 2006, Thm 6.6 p.100). Conversely, if  $y$  is in  $\exp_x(B_d(0, \epsilon))$  with  $\epsilon$  smaller than the injectivity radius at  $x$ , then the geodesic joining  $x$  and  $y$  is minimising (Lee, 2006, Prp 6.10). This means that for any  $v \in T_x M$  with  $\|v\| < \epsilon$ , we have  $d_M(x, \exp_x(v)) = \|v\|$ .

When  $M$  is a closed subset of  $\mathbb{R}^D$ , the maximal geodesics  $\gamma_{x,v}$  are defined on the whole line  $\mathbb{R}$ , and the exponential maps are defined on the whole tangent spaces. This is (one side of) the Hopf-Rinow theorem (Lee, 2006, Thm 6.13 p.108).

### The volume measure

We need a canonical way to define a uniform measure  $\mu_M$  on  $M$ . This can easily be done locally around  $x \in M$  by pushing forward the Lebesgue measure on  $T_x M$  through  $\exp_x : B_{T_x M}(0, \epsilon) \rightarrow M$  with  $\epsilon$  smaller than the injectivity radius at  $x$ . In order to take into account the geometry of  $M$  around  $x$ , we add a correction term, called the volume density function (Besse, 1978, p. 154). For a continuous function  $\psi : M \rightarrow \mathbb{R}$  with support in  $\exp_x(B_d(0, \epsilon))$ , we define

$$\mu_M(\psi) = \int_{B_d(0, \epsilon)} \psi \circ \exp_x(v) \sqrt{\det g_{ij}^x(v)} dv,$$

with  $g_{ij}^x(v) = \langle d\exp_x(v)[e_i], d\exp_x(v)[e_j] \rangle$  and where  $(e_1, \dots, e_d)$  is an arbitrary orthonormal basis of  $T_x M$ . The value of  $\sqrt{\det g_{ij}^x(v)}$  represents the volume of the parallelepiped in  $T_{\exp_x(v)} M$  generated by the images of  $(e_1, \dots, e_d)$  via  $d\exp_x(v)$ . We can then define  $\mu_M(\psi)$  for any continuous bounded  $\psi : M \rightarrow \mathbb{R}$  by means of partition of unity (Gallot et al., 2004, Sec 1.H), see Gallot et al.

(2004, Sec 3.H.1 and Sec 3.H.2) for further details. The volume of  $M$ , denoted by  $\text{vol} M$ , is then simply  $\mu_M(\mathbb{1})$ . It is finite if  $M$  is a compact submanifold of  $\mathbb{R}^D$ .

There is another natural candidate for a measure on  $M$ , namely the restriction of the  $d$ -dimensional Hausdorff measure  $\mathcal{H}^d$  to  $M$ , see Federer (1969, Sec 2.10.2 p.171) for a definition. Luckily, we have  $\mu_M(A) = \mathcal{H}^d(A)$  for any Borel subset  $A \subset M$  (Evans and Gariepy, 1992, Ex D p.102).

## 2.2 The reach of a subset in an Euclidean space

One of the main concern when dealing with observations sampled from a geometrically structured probability measure is to determine the suitable scale at which one should look at the data. Indeed, given finite-sized point cloud in  $\mathbb{R}^D$ , there are infinitely many submanifolds that interpolate the point cloud, see Figure 1 for an illustration. A popular notion of regularity for a subset of the Euclidean space is the *reach*, introduced by Federer (1959).

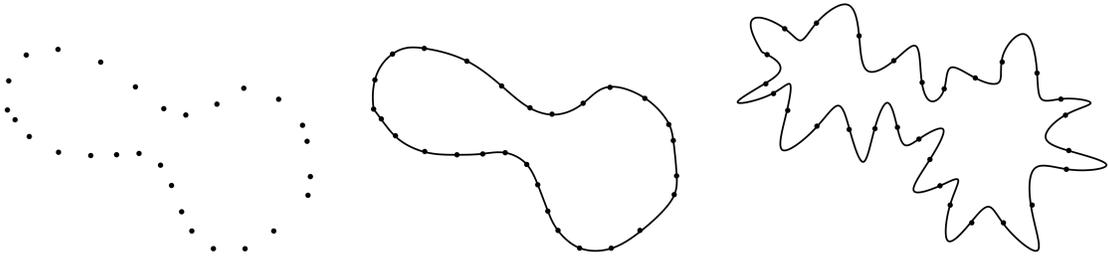


Figure 1: An arbitrary points cloud (Left) for  $D = 2$ , and two smooth one-dimensional submanifolds passing through all its points (Middle, Right). A reach condition tends to discard the Right manifold as a likely candidate among all possible submanifolds the point cloud is sampled from.

**Definition 1.** Let  $K$  be a compact subset of  $\mathbb{R}^D$ . The reach  $\tau_K$  of  $K$  is the supremum of all  $r \geq 0$  such that the orthogonal projection  $\text{pr}_K$  on  $K$  is well-defined on the  $r$ -neighbourhood  $K^r$  of  $K$ , namely

$$\tau_K = \sup\{r \geq 0 \mid \forall x \in \mathbb{R}^D, d(x, K) \leq r \Rightarrow \exists! y \in K, d(x, K) = \|x - y\|\}.$$

When  $M$  is a compact submanifold of  $\mathbb{R}^D$ , the reach  $\tau_M$  quantifies two geometric invariants: locally, it measures how curved the manifold is, and globally, it measures how close it is to intersect itself (the so-called bottleneck effect). See Figure 2 for an illustration of the phenomenon. A *reach condition*, meaning that the reach of the manifold under study is bounded below, is usually necessary in order to obtain minimax inference results in manifold learning. These include: homology inference Balakrishnan et al. (2012); Niyogi et al. (2008), curvature (Aamari and Levrard, 2019) and reach estimation itself (Aamari et al., 2019) as well as manifold estimation Aamari and Levrard (2019); Genovese et al. (2012). We follow this approach, and start with citing a few useful properties related to the reach.

**Proposition 1.** (Niyogi et al., 2008, Prp. 6.1) Let  $M$  be a compact smooth submanifold of  $\mathbb{R}^D$ . Then, for any  $x \in M$ , we have  $\|\mathbb{I}_x\|_{\text{op}} \leq 1/\tau_M$ .

Since  $\mathbb{I}_x$  is the differential of order two of the mapping  $\exp_x$ , Proposition 1 has several convenient implications. First, it gives a uniform lower bound for the injectivity radiuses of  $M$ . This is a corollary of Alexander and Bishop (2006, Thm 1.3), as explained in Aamari and Levrard (2019, Lem A.1).

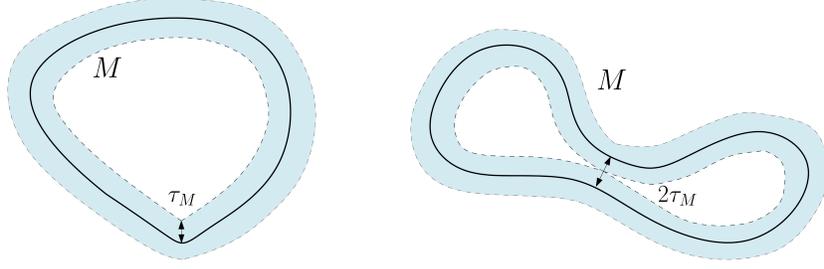


Figure 2: For the first manifold  $M$  (Left), the value of the reach  $\tau_M$  comes from its curvature. For the second one (Right), the reach is equal to  $\tau_M$  because it is close to self intersecting (a bottleneck effect). The blue area represents the tubular neighbourhood over which the orthonormal projection on each manifold is well-defined.

**Proposition 2.** *Let  $M$  be a compact smooth submanifold of  $\mathbb{R}^D$ . Then, for any  $x \in M$ , the exponential map  $\exp_x : T_x M \rightarrow M$  is a diffeomorphism from  $B_d(0, \pi\tau_M)$  to  $\exp_x(B_d(0, \pi\tau_M))$ .*

Proposition 1 also yields nice bounds on how well the Euclidean distance on  $\mathbb{R}^D$  approximates the Riemannian distance  $d_M$  on  $M \times M$ .

**Proposition 3.** (Niyogi et al., 2008, Prp. 6.3) *For any compact submanifold  $M$  of  $\mathbb{R}^D$  and any  $x, y \in M$  such that  $\|x - y\| \leq \tau_M/2$ , we have*

$$\|x - y\| \leq d_M(x, y) \leq \tau_M \left( 1 - \sqrt{1 - \frac{2\|x - y\|}{\tau_M}} \right).$$

Proposition 3 allows in turn to compare the volume measure  $\mu_M$  to the Lebesgue measure on its tangent spaces.

**Lemma 1.** *For any  $d$ -dimensional compact smooth submanifold  $M$  of  $\mathbb{R}^D$ , for any  $x \in M$  and any  $\eta \leq \tau_M/2$ , we have, for all  $h \leq \eta$*

$$(1 - \eta^2/6\tau_M^2)^d \zeta_d h^d \leq \mu_M(B_D(x, h)) \leq \left\{ (1 + (\xi(\eta/\tau_M)\eta)^2/\tau_M^2) \xi(\eta/\tau_M) \right\}^d \zeta_d h^d$$

where  $\xi(s) = (1 - \sqrt{1 - 2s})/s$  and  $\zeta_d$  is the volume of the unit Euclidean ball in  $\mathbb{R}^d$ .

*Proof.* This result already appears in (Aamari, 2017, Lem III.23) but we prove it here to make constants explicit. Let us denote by  $\text{Leb}$  the Lebesgue measure on  $T_x M$ . Using (Aamari, 2017, Prp III.22.v), we know that, as long as  $\xi(h/\tau_M)h \leq \tau_M$  (which holds if  $h \leq \tau_M/2$ ),

$$\begin{aligned} (1 - h^2/6\tau^2)^d \text{Leb}(B_d(0, h)) &\leq \mu_M(\exp_x(B_d(0, h))) \\ &\leq \mu_M(\exp_x(B_d(0, \xi(h/\tau_M)h))) \\ &\leq (1 + (\xi(h/\tau_M)h)^2/\tau_M^2)^d \text{Leb}(B_d(0, \xi(h/\tau_M)h)). \end{aligned}$$

Thanks to Proposition 3, if  $h \leq \tau_M/2$ , then  $\exp_x(B_d(0, h)) \subset M \cap B_D(x, h) \subset \exp_x(B_d(0, \xi(h/\tau_M)h))$ . These inclusions combined with the last inequalities yield the result.  $\square$

### 2.3 A statistical model for sampling on a unknown manifold

Our statistical model is characterised by two quantities: the regularity of its support and the regularity of the density defined on this support. The support belongs to a class of submanifolds

$M$ , for which we need to fix some kind of *canonical* parametrization. This is what [Aamari and Levrard \(2019\)](#) propose by asking the support  $M$  to admit a local parametrization at all point  $x \in M$  by  $T_x M$ , and that this parametrization is close to being the inverse of the projection over this tangent space. We consider here a somewhat stronger condition: among all the local parametrizations of a manifold by its tangent spaces, we impose a constraint on the exponential map up to its differentials of order  $k$ .

**Definition 2.** *Let  $1 \leq d < D$  be integers and  $\tau > 0$  be a positive number. We let  $\mathcal{C}_d(\tau)$  define the set of submanifolds  $M$  of  $\mathbb{R}^D$  satisfying the following properties:*

- (i) *(Dimension)  $M$  is a smooth submanifold of dimension  $d$  without boundaries;*
- (ii) *(Compactness)  $M$  is compact;*
- (iii) *(Reach condition) We have  $\tau_M \geq \tau$ .*

For a real number  $\alpha \geq 1$  and a positive real number  $L > 0$ , we define  $\mathcal{C}_{d,\alpha}(\tau, L)$  of as the set of  $M \in \mathcal{C}_d(\tau)$  that fulfill the additional condition:

- (iv) *For any  $x \in M$  the restriction of the exponential map  $\exp_x$  to the ball  $B_{T_x M}(0, \tau/2)$  is  $(\alpha + 1)$ -Hölder with coefficient  $L$ . Namely, setting  $k = \lceil \alpha \rceil$  and  $\delta = \alpha + 1 - k > 0$ , we have*

$$\forall v, w \in B_{T_x M}(0, \tau/2), \quad \|d^k \exp_x(v) - d^k \exp_x(w)\|_{\text{op}} \leq L \|v - w\|^\delta.$$

Condition (iv) implies a bound on derivatives of  $\exp_x$  up to order  $k$ . Indeed, we know that for any  $M \in \mathcal{C}_d(\tau)$  and  $x \in M$  the map  $v \mapsto \exp_x(v) - x$  is uniformly bounded on  $B_{T_x M}(0, \tau/2)$ . This is because for any  $v \in B_{T_x M}(0, \tau/2)$ , we have  $\|\exp_x(v) - x\| \leq d_M(\exp_x(v), x) = \|v\| \leq \tau/2$ . If now  $M$  is in  $\mathcal{C}_{d,\alpha}(\tau, L)$ , we have bounds on the derivatives  $d^j \exp_x$  up to order  $k$  of the form

$$\sup_{v \in B_{T_x M}(0, \tau/4)} \|d^j \exp_x(v)\|_{\text{op}} \leq L_j$$

with  $L_j$  depending on  $\tau$ ,  $L$  and  $\alpha$  only. This comes from Lemma 5. In that sense, the model of Definition 2 is close to the one proposed by [Aamari and Levrard \(2019\)](#).

In the following, we consider a class  $\mathcal{C}$  of submanifolds of  $\mathbb{R}^D$ , and a class  $\mathcal{F}$  of real-valued functions defined on  $\mathbb{R}^d$ .

**Definition 3.** *Given  $\mathcal{C}$  and  $\mathcal{F}$ , the statistical model  $\mathcal{P}(\mathcal{C}, \mathcal{F})$  consists of all probability measures  $P$  defined on  $\mathbb{R}^D$  (endowed with its Borel sigma-field) such that:*

- (i) *There exists  $M \in \mathcal{C}$  verifying  $P \ll \mu_M$ ;*
- (ii) *The Radon-Nikodym derivative  $f = \frac{dP}{d\mu_M}$  expressed in normal coordinates belongs to  $\mathcal{F}$ . (In other words, there exists is a version of  $f$  such that for every  $x \in M$ ,  $f \circ \exp_x$  belongs to  $\mathcal{F}$ .)*

For a model  $\Sigma$  of the form  $\mathcal{P}(\mathcal{C}, \mathcal{F})$  and any  $x \in \mathbb{R}^D$ , we define the submodel  $\Sigma(x)$  as  $\mathcal{P}(\mathcal{C}(x), \mathcal{F})$  where  $\mathcal{C}(x)$  is the subset of all the submanifolds of  $\mathcal{C}$  containing  $x$ .

Some remarks: **1)** Definition 3 models the set of all probability measures with support in  $\mathcal{C}$  and densities in  $\mathcal{F}$ . **2)** Point (ii) of Definition 3 is a natural way to define smoothness in a Hölder sense over Riemannian manifolds, see for instance [Triebel \(1987\)](#). Of course, (ii) in Definition 3 is meaningful only if the class  $\mathcal{F}$  is left invariant by the canonical action of the orthogonal group

$O(\mathbb{R}^d)$  so that the claim  $f \circ \exp_x \in \mathcal{F}$  does not depend on an arbitrary orthonormal identification we make between  $T_x M$  and  $\mathbb{R}^d$ . This will be the case for all the functional classes  $\mathcal{F}$  we will consider in the paper.

According to Definition 3, a given probability measure  $P \in \mathcal{P}(\mathcal{C}, \mathcal{F})$  defines in turn a pair of submanifold and density  $(M_P, f_P)$  abbreviated by  $(M, f)$  when no confusion may arise. If  $M_P$  is not uniquely determined by  $P$ , we simply pick an arbitrary element  $M$  among all the manifolds satisfying  $\text{supp } P \subset M$ . One classical case where  $M$  is uniquely determined is when the class  $\mathcal{F}$  consists of densities bounded away from 0, in which case  $M = \text{supp } P$  exactly.

For  $0 \leq a \leq b \leq \infty$ , we write  $\mathcal{B}(a, b)$  for the set of functions with values in  $[a, b]$ . We will further restrict our study to the following functional classes:

**Definition 4.** For a real number  $\beta \geq 1$  and constants  $r, R > 0$ , we let  $\mathcal{F}_\beta(a, b, r, R)$  denote the set of real-valued functions  $g$  on  $\mathbb{R}^d$  satisfying:

- (i)  $g$  is in  $\mathcal{B}(a, b)$ ;
- (ii) The restriction of  $g$  to the open ball  $B_d(0, r)$  is  $\beta$ -Hölder with coefficient  $R$ , namely  $g$  is  $k = \lfloor \beta - 1 \rfloor$ -times differentiable on  $B_d(0, r)$  and verifies

$$\forall v, w \in B_d(0, r), \quad \|d^k g(v) - d^k g(w)\| \leq R \|v - w\|^\delta$$

with  $\delta = \beta - k > 0$ .

## 3 Main results

### 3.1 Methodology

Let  $\Sigma = \mathcal{P}(\mathcal{C}, \mathcal{F})$  according to the construction given in Definition 3 above. Given  $x \in \mathbb{R}^D$  and an  $n$ -sample  $X_1, \dots, X_n$  arising from a distribution  $P \in \Sigma(x)$ , we measure the accuracy of an estimator  $\widehat{f}(x)$  built from a sample  $(X_1, \dots, X_n)$  with common distribution  $P = f \cdot \mu_M = f_P \cdot \mu_{M_P}$  by its pointwise risk

$$\mathcal{R}_n^{(p)}[\widehat{f}, P](x) = \left( \mathbb{E}_P[|\widehat{f}(x) - f(x)|^p] \right)^{1/p}, \quad p \geq 1, \quad (3)$$

and its associated maximal pointwise risk at  $x$ :

$$\mathcal{R}_n^{(p)}[\widehat{f}, \Sigma](x) = \sup_{P \in \Sigma(x)} \mathcal{R}_n^{(p)}[\widehat{f}, P](x),$$

where  $\mathbb{E}_P$  denotes expectation w.r.t.  $P$ . Alternatively, we may consider the global  $L_p$ -risk

$$\left( \mathbb{E}_P[\|\widehat{f} - f\|_{L_p(M)}^p] \right)^{1/p} = \|\mathcal{R}_n^{(p)}[\widehat{f}, P](\cdot)\|_{L_p(M)}. \quad (4)$$

Since now the global  $L_p$ -risk does not depend on  $x$ , it makes sense to evaluate it for any  $P \in \Sigma$  and define its maximal version

$$\sup_{f \in \Sigma} \left( \mathbb{E}_P[\|\widehat{f} - f\|_{L_p(M)}^p] \right)^{1/p} = \sup_{f \in \Sigma} \|\mathcal{R}_n^{(p)}[\widehat{f}, P](\cdot)\|_{L_p(M)}$$

provided we have nice mapping  $x \mapsto \widehat{f}(x)$ , meaning that the restriction  $\widehat{f}|_M$  of  $\widehat{f}$  can be seen as a measurable estimator with value in  $L_p(M, \mu_M)$  for all  $M \in \mathcal{C}$ .

*Remark 1.* Since we measure smoothness in a strong Hölder sense, it will be sufficient to restrict our study to (3). Indeed, whenever a result holds for some  $\Sigma(x)$  for (3) uniformly in  $x$ , it readily carries over to  $\Sigma$  and (4) thanks to the inequality

$$\|\mathcal{R}_n^{(p)}[\widehat{f}, P](\cdot)\|_{L_p(M)} \leq (\text{vol } M) \times \sup_{x \in M} \mathcal{R}_n^{(p)}[\widehat{f}, P](x),$$

that is always meaningful for classes  $\Sigma$  of distributions  $P$  supported on compact submanifolds, an assumption in force in the following.

In this setting, we further investigate the construction of sharp upper bounds for models of the form  $\Sigma_{\alpha, \beta}(x)$  for  $\Sigma_{\alpha, \beta} = \mathcal{P}(\mathcal{C}_\alpha, \mathcal{F}_\beta)$  where  $\mathcal{C}_\alpha = \mathcal{C}_{d, \alpha}(\tau, L)$  and  $\mathcal{F}_\beta = \mathcal{F}_\beta(a, b, r, R)$  and for fixed values of  $d, \tau, L, R, a, b$  and  $r$ , and unknown smoothness parameters  $\alpha, \beta$ .

### 3.2 Kernel estimation

Classical nonparametric density estimation methods are based on kernel smoothing (Parzen, 1962; Silverman, 1986). In this section, we combine kernel density estimation with the minimal geometric features needed in order to recover efficiently their density. In our context, it turns out that it suffices to know the dimension of the underlying manifold  $M$  where the data sit. We first work with a fixed (or known) dimension  $1 \leq d < D$  for  $M$ , and thus we further drop  $d$  in (most of) the notation. We refer to Section 3.5 for the case of an unknown  $d$ .

Let  $K : \mathbb{R}^D \rightarrow \mathbb{R}$  be a smooth function vanishing outside the unit ball  $B_D(0, 1)$ . Given an  $n$ -sample  $(X_1, \dots, X_n)$  drawn from a distribution  $P$  on  $\mathbb{R}^D$ , we are interested in the behaviour of the kernel estimator

$$\widehat{f}_h(x) = \frac{1}{nh^d} \sum_{j=1}^n K\left(\frac{X_j - x}{h}\right), \quad x \in \mathbb{R}^D, h > 0. \quad (5)$$

Notice the normalisation in  $h$  that depends on  $d$  and not on  $D$ , as one would set for a classical kernel estimator in  $\mathbb{R}^D$ . Our main result is that  $\widehat{f}_h(x)$  behaves well when  $P$  is supported on a  $d$ -dimensional submanifold of  $\mathbb{R}^D$ .

We need some notations. Let  $\Sigma_{\text{sd}} = \mathcal{P}(\mathcal{C}_d(\tau), \mathcal{B}(0, b))$  as in Section 3.1. This is quite a large model that contains in particular all the smoothness classes  $\Sigma_{\alpha, \beta}$ . For  $P \in \Sigma_{\text{sd}}(x)$  with associated pair of submanifold and density  $(M, f)$ , we let

$$\begin{aligned} f_h(P, x) &= \mathbb{E}_P[\widehat{f}_h(x)], \\ B_h(P, x) &= f_h(P, x) - f(x), \end{aligned}$$

and

$$\widehat{\xi}_h(P, x) = \widehat{f}_h(x) - f_h(P, x),$$

that correspond respectively to the mean, bias and stochastic deviation of the estimator  $\widehat{f}_h(x)$ . We also introduce the quantity

$$v(h) = \sqrt{\frac{2v}{nh^d} + \frac{\|K\|_\infty}{nh^d}} \quad \text{with } v = 4^d \zeta_d \|K\|_\infty^2 b,$$

where  $\zeta_d$  is the volume of the unit ball  $B_d(0, 1)$  in  $\mathbb{R}^d$ . The quantity  $v(h)$  will prove to be a good majorant of the stochastic deviations of  $\widehat{f}_h(x)$ . The usual bias-stochastic decomposition of  $\widehat{f}_h(x)$  leads to

$$\mathcal{R}_n^{(p)}[\widehat{f}_h, P](x) \leq |B_h(P, x)| + (\mathbb{E}_P[|\widehat{\xi}_h(P, x)|^p])^{1/p}. \quad (6)$$

for the pointwise risk (3) and

$$\|\mathcal{R}_n^{(p)}[\widehat{f}, P](\cdot)\|_{L_p(M)} \leq \|B_h(P, \cdot)\|_{L_p(M)} + \left(\mathbb{E}_P[\|\widehat{\xi}_h(P, \cdot)\|_{L_p(M)}^p]\right)^{1/p}$$

for the global risk (4). We study each term separately.

### Behaviour of the stochastic term

**Proposition 4.** *Let  $p \geq 1$ . There exists a constant  $c_p > 0$  depending on  $p$  only such that for any  $P \in \Sigma_{sd}(x)$ :*

$$\left(\mathbb{E}_P[\|\widehat{\xi}_h(P, x)\|^p]\right)^{1/p} \leq c_p v(h)$$

and for any  $P \in \Sigma_{sd}$

$$\left(\mathbb{E}_P[\|\widehat{\xi}_h(P, \cdot)\|_{L_p(M)}^p]\right)^{1/p} \leq c_p (\text{vol } M)^{1/p} v(h).$$

*Remark 2.* The last bound is uniform in  $M$  as soon as  $\text{vol } M$  is uniformly bounded. This is the case if we restrict further the class  $\Sigma_{sd}$  to  $\mathcal{P}(\mathcal{C}_d(\tau), \mathcal{B}(a, b))$  for some  $a > 0$ , so that every density  $f$  of  $P \in \Sigma$  is bounded below  $a > 0$ . Indeed, in that case  $\text{vol } M = \int_M d\mu_M \leq \frac{1}{a} \int_M f d\mu_M = \frac{1}{a}$ .

### Behaviour of the bias term

Recall that  $\Sigma_{\alpha, \beta} = \mathcal{P}(\mathcal{C}_\alpha, \mathcal{F}_\beta)$  where  $\mathcal{C}_\alpha = \mathcal{C}_{d, \alpha}(\tau, L)$  and  $\mathcal{F}_\beta = \mathcal{F}_\beta(a, b, r, R)$ . We need a certain property for the kernel  $K$ . More precisely:

**Assumption 1.** (i)  $K$  is supported on the unit ball  $B_D(0, 1)$ ;

(ii) For any  $d$ -dimensional subspace  $H$  of  $\mathbb{R}^D$ , we have  $\int_H K(v) dv = 1$ .

One way to obtain Assumption 1 is to set  $\Lambda(x) = \exp(-1/(1 - \|x\|^2))$  for  $x \in B_D(0, 1)$  and  $\Lambda(x) = 0$  otherwise. With  $\lambda_d = \int_{\mathbb{R}^d} \Lambda(v) dv$ , the function  $K(x) = \lambda_d^{-1} \Lambda(x)$  is a smooth kernel, supported on the unit ball of the ambient space  $\mathbb{R}^D$  that satisfies Assumption 1. In the following, we pick an arbitrary kernel  $K$  such that Assumption 1 is satisfied.

**Lemma 2.** *Let  $\alpha, \beta \geq 1$ . For  $P \in \Sigma_{\alpha, \beta}(x)$  and any  $h < (r \wedge \tau)/4$ , setting  $k = \lceil \alpha \wedge \beta - 1 \rceil$ , we have*

$$f_h(P, x) = f(x) + \sum_{j=1}^k h^j G_j(P, x) + R_h(P, x), \quad (7)$$

with  $|G_j(P, x)| \lesssim 1$  and  $|R_h(P, x)| \lesssim h^{\alpha \wedge \beta}$  where the symbol  $\lesssim$  means inequality up to a constant that depends on  $K, \alpha, \tau, L, \beta, r, b$  and  $R$ .

Starting from a kernel  $K$  satisfying Assumption 1, we recursively define a sequence of smooth kernels  $(K^{(d, \ell)})_{\ell \geq 1}$ , simply denoted by  $K^{(\ell)}$  in this section, with support in  $B_D(0, 1)$  as follows (see Figure 3). For  $x \in \mathbb{R}^D$ , we put

$$\begin{cases} K^{(d, 1)}(x) = K(x) \\ K^{(d, \ell+1)}(x) = 2^{1+d/\ell} K^{(d, \ell)}(2^{1/\ell} x) - K^{(d, \ell)}(x) \quad \forall \ell \geq 1. \end{cases} \quad (8)$$

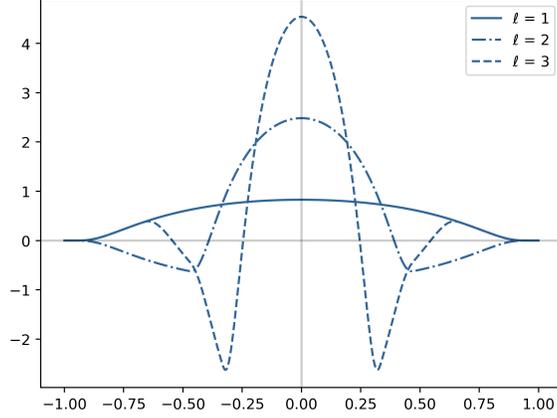


Figure 3: Plots of the kernel  $K^{(d, \ell)}$  for  $d = 1$  and  $\ell = 1, 2, 3$ .

**Proposition 5.** *Let  $\ell$  be a positive integer and let  $1 \leq \alpha, \beta \leq \ell$  be real numbers. Let  $K^{(\ell)}$  be the kernel defined in (8) starting from a kernel  $K$  satisfying Assumption 1. Then, for any  $P \in \Sigma_{\alpha, \beta}(x)$  and any bandwidth  $0 < h < (r \wedge \tau)/4$ , the estimator  $\widehat{f}_h(x)$  defined as in (5) using  $K^{(\ell)}$  is such that*

$$|B_h(P, x)| \lesssim h^{\alpha \wedge \beta} \quad (9)$$

for the pointwise bias, up to a constant depending on  $K, \ell, \alpha, \tau, L, \beta, r, b$  and  $R$  and also

$$\|B_h(P, \cdot)\|_{L_p(M)} \lesssim (\text{vol } M)^{1/p} h^{\alpha \wedge \beta}$$

up to the same constant as in (9).

The second estimate is a direct consequence of the first one: a closer inspection at the proof of Lemma 2 reveals that the estimates are indeed uniform in  $x \in M$ .

### 3.3 Minimax rates of convergence

For  $\ell \geq 1$ , we let  $K^{(\ell)}$  be defined as in (8), originating from a kernel  $K$  satisfying Assumption 1.

**Theorem 1.** *Let  $\alpha, \beta \geq 1$  and  $p \geq 1$ . Let  $\ell$  be an integer greater than  $\alpha \wedge \beta$ . For the estimator  $\widehat{f}_h(x)$  defined in (5) constructed with  $K^{(\ell)}$  and bandwidth  $h = n^{-1/(2\alpha \wedge \beta + d)}$ , we have, for  $n$  large enough (depending on  $\tau, r, K$  and  $d$ )*

$$\mathcal{R}_n^{(p)}[\widehat{f}_h, \Sigma_{\alpha, \beta}](x) \lesssim n^{-\alpha \wedge \beta / (2\alpha \wedge \beta + d)}$$

up to a constant that depends on  $p, K, \ell, \alpha, \tau, L, \beta, r, b$  and  $R$  and that does not depend on  $x$ . If moreover  $a > 0$ , the same result holds for the global risk up to a multiplication of the constant in (10) by a factor  $a^{-1/p}$ .

*Proof.* Combining Proposition 4 and Proposition 5, as soon as  $h \leq (\tau \wedge r)/4$ , we obtain, for any  $P \in \Sigma_{\alpha, \beta}(x)$ ,

$$\mathcal{R}_n^{(p)}[\widehat{f}_h, P](x) \lesssim v(h) + h^{\alpha \wedge \beta} \quad (10)$$

up to a constant that depends on  $p, K, \ell, \alpha, \tau, L, \beta, r, b$  and  $R$  and that does not depend on  $x$ . Setting  $h = n^{-1/(2\alpha \wedge \beta + d)}$  readily yields (10), while the uniformity in  $x \in M_P$  enables us to integrate (10) and obtain the result for the integrated risk.  $\square$

This rate is optimal if the regularity of  $M$  dominates the smoothness of  $f$ :

**Theorem 2.** *For any real numbers  $\alpha, \beta \geq 1$  and any  $x \in \mathbb{R}^D$ , we have, for  $L, b$  and  $1/a$  large enough (depending on  $\tau$ )*

$$\liminf_{n \rightarrow \infty} n^{\beta/(2\beta+d)} \inf_F \mathcal{R}_n^{(p)}[F, \Sigma_{\alpha,\beta}](x) \geq C_* > 0$$

where  $C_*$  only depends on  $\tau$  and  $R$ . The infimum is taken among all estimators  $F$  of  $f$  constructed from a drawn  $X_1, \dots, X_n$  with distribution  $P \in \Sigma_{\alpha,\beta}(x)$ .

Some remarks: **1)** the order  $\ell \geq 1$  of the kernel plays the role of a cancellation order as in the Euclidean case: the higher  $\alpha$  and  $\beta$ , the bigger we need to choose  $\ell$ , resulting in oscillations for  $K^\ell$ . **2)** We have a minimax rate of convergence  $n^{-\alpha \wedge \beta / (2\alpha \wedge \beta + d)}$  that is similar to the Euclidean case, provided the underlying manifold  $M$  is regular enough, namely  $\alpha \geq \beta$ , which probably covers most cases of interests. This restriction played by  $\alpha$  together with the reach condition comes as a price for recovering  $f$  when its support  $M$  is unknown. We will see in Section 4 below that for  $d = 1$ , it is possible to remove the term  $\alpha$  and achieve the rate  $n^{-\beta/(2\beta+1)}$ . This however requires extra effort on learning the geometry of  $M$ . **3)** As mentioned earlier, a reach condition  $\tau > 0$  is usually necessary in minimax geometric inference when the manifold is unknown (Aamari and Levrard, 2019; Balakrishnan et al., 2012; Genovese et al., 2012; Kim et al., 2016; Niyogi et al., 2008). Although we do not have a proof here, it is likely that having  $\tau > 0$  is necessary in order to recover  $f$  in a minimax setting.

### 3.4 Smoothness adaptation

We implement Lepski's algorithm, following closely Lepski et al. (1997) in order to automatically select the bandwidth from the data  $(X_1, \dots, X_n)$ . We know from Section 3.2 that the optimal bandwidth on  $\Sigma_{\alpha,\beta}(x)$  is of the form  $n^{-1/(2\alpha \wedge \beta + d)}$ . Hence, without prior knowledge of the value of  $\alpha$  and  $\beta$ , we can restrict our search for a bandwidth in a bounded interval of the form  $[h^-, h^+]$  for some arbitrary  $h^+ \in (0, 1]$  and consider the set

$$\mathcal{H} = \{2^{-j}h^+, \text{ for } 0 \leq j \leq \log_2(h^+/h^-)\}$$

with  $h^- = (\|K\|_\infty/2v)^{1/d}n^{-1/d}$ . This lower bound is smaller than the optimal bandwidth  $n^{-1/(2\alpha \wedge \beta + d)}$  on  $\Sigma_{\alpha,\beta}(x)$  for  $n$  large, and is such that  $v(h) \leq 2\sqrt{2v/(nh^d)}$  for all  $h \geq h^-$ . For  $h, \eta \in \mathcal{H}$ , we introduce the following quantities:

$$\begin{aligned} v(h, \eta) &= 2v(h \wedge \eta), \\ \lambda(h) &= 1 \vee \sqrt{dD_2 \log(h^+/h)}, \\ \psi(h, \eta) &= 2D_1v(h)\lambda(h) + v(h, \eta)\lambda(\eta). \end{aligned} \tag{11}$$

where  $D_1$  and  $D_2$  are positive constants (to be specified). For  $h \in \mathcal{H}$  we define the subset of bandwidths  $\mathcal{H}(h) = \{\eta \in \mathcal{H}, \eta \leq h\}$ . The selection rule for  $h$  is the following:

$$\widehat{h}(x) = \max \left\{ h \in \mathcal{H} \mid \forall \eta \in \mathcal{H}(h), |\widehat{f}_h(x) - \widehat{f}_\eta(x)| \leq \psi(h, \eta) \right\},$$

and we finally consider the estimator

$$\widehat{f}(x) = \widehat{f}_{\widehat{h}(x)}(x). \tag{12}$$

**Theorem 3.** *In the same setting as Theorem 1 for  $\widehat{f}$  defined in (12) with  $h^+ = 1$ , for any  $D_2 > p \geq 1$ , and any  $D_1 > 0$ , we have, for  $n$  large enough (depending on  $p, d, D_1, D_2, K, \ell, \alpha, \tau, L, \beta, r, b$  and  $R$ )*

$$\mathcal{R}_n^{(p)}[\widehat{f}, \Sigma_{\alpha, \beta}](x) \lesssim \left( \frac{\log n}{n} \right)^{\alpha \wedge \beta / (2\alpha \wedge \beta + d)}$$

up to a constant that depends on  $p, d, D_1, D_2, K, \ell, \alpha, \tau, L, \beta, r, b, R$  and that does not depend on  $x$ .

Some remarks: **1)** Theorem 3 provides us with a classical smoothness adaptation result in the spirit of (Lepski et al., 1997): the estimator  $\widehat{f}$  has the same performance as the estimator  $\widehat{f}_h$  selected with the optimal bandwidth  $n^{-1/(2\alpha \wedge \beta + d)}$ , up to a logarithmic factor on each model  $\Sigma_{\alpha, \beta}(x)$  without the prior knowledge of  $\alpha \wedge \beta$  over the range  $[1, \ell]$ . **2)** The extra logarithmic term is the unavoidable payment for the Lepski-Low phenomenon Lepskiĭ (1990); Low (1992) when recovering a function in pointwise or in a uniform loss. **3)** We were unable to obtain oracle inequalities in the spirit of the Goldenshluger-Lepski method, see (Goldenshluger and Lepski, 2008, 2014; Goldenshluger et al., 2011), due to the non-Euclidean character of the support of  $f$ : our route goes along the more classical approach of Lepski et al. (1997). Obtaining oracle inequalities in this framework remains an open question.

### 3.5 Simultaneous smoothness and dimension adaptation

We now consider the case when the dimension  $d$  of  $M$  is unknown. We write again  $\Sigma_{\alpha, \beta}^d = \mathcal{P}(\mathcal{C}_{d, \alpha}, \mathcal{F}_\beta)$  with  $\mathcal{C}_{d, \alpha} = \mathcal{C}_{d, \alpha}(\tau, L)$  and  $\mathcal{F}_\beta = \mathcal{F}_\beta(r, a, b, R)$ , and  $\Sigma_{\text{sd}}^d = \mathcal{P}(\mathcal{C}_d, \mathcal{F})$  with  $\mathcal{C}_d = \mathcal{C}_d(\tau)$  and  $\mathcal{F} = \mathcal{B}(0, b)$ . We work with fixed values of  $\tau, L, r, a, b$  and  $R$ , and we impose this time that  $a > 0$ . For any  $1 \leq d < D$  and  $h > 0$ , we define

$$K(d; x) = \lambda_d^{-1} \Lambda(x) \quad \text{and} \quad K_h(d; x) = h^{-d} K(d; x/h)$$

where  $\Lambda$  and  $\lambda_d$  have been introduced in Section 3.2. We then define the estimator

$$\widehat{f}_h(d; x) = \frac{1}{n} \sum_{i=1}^n K_h(d; X_i - x).$$

We write  $K^{(\ell)}(d; \cdot)$  for the kernel  $K^{(d, \ell)}(d, \cdot)$  defined by recursion in (8) starting from the kernel  $K(d; \cdot)$ . We redefine all the quantities introduced before as now depending on  $d$ . Namely, for  $h, \eta > 0$ , and a given family of kernel  $K(d; \cdot)$ , we set

$$\begin{aligned} v(d; h) &= \sqrt{\frac{2v_d}{nh^d}} + \frac{\|K(d; \cdot)\|_\infty}{nh^d} \quad \text{with} \quad v_d = 4^d \zeta_d \|K(d; \cdot)\|_\infty^2 b, \\ v(d; h, \eta) &= 2v(d; h \wedge \eta), \\ \lambda(d; h) &= 1 \vee \sqrt{dD_2 \log(1/h)}, \\ \psi(d; h, \eta) &= 2D_1 v(d; h) \lambda(d; h) + v(d; h, \eta) \lambda(d; \eta), \\ h_d^- &= (\|K(d; \cdot)\|_\infty / 2v_d)^{1/d} n^{1/d}, \\ \mathcal{H}_d &= \{2^{-j}, \text{ for } 0 \leq j \leq \log_2(1/h_d^-)\}, \end{aligned}$$

where  $D_1$  and  $D_2$  are constants, unspecified at this stage. We also define

$$\widehat{h}(d; x) = \max\{h \in \mathcal{H}_d \mid \forall \eta \in \mathcal{H}_d(h), \quad |\widehat{f}_h(d; x) - \widehat{f}_\eta(d; x)| \leq \psi(d; h, \eta)\} \quad (13)$$

with  $\mathcal{H}_d(h) = \{\eta \in \mathcal{H}_d, \eta \leq h\}$ . We assume that we have an estimator  $\widehat{d}$  of the dimension  $d$  of  $M$  with values in  $\{1, \dots, D\}$ . We require the following property

**Assumption 2.** For every integer  $1 \leq d < D$ , all real numbers  $\alpha, \beta \geq 1$ , and for all  $P \in \Sigma_{\alpha, \beta}^d(x)$ , we have  $\mathbb{P}(\widehat{d} \neq d) \lesssim n^{-3p}$  up to a constant that depends on all the parameters of  $\Sigma_{\alpha, \beta}^d(x)$ .

There are various way to define such an estimator, see Farahmand et al. (2007) or even Kim et al. (2016) where an estimator with super-exponential minimax rate on a wide class of probability measures is constructed. For sake of completeness and simplicity, we will mildly adapt the work of Farahmand et al. (2007) to our setting.

**Definition 5.** For  $P \in \Sigma_{\text{sd}}^d(x)$ , we write  $P_\eta = P(B_D(x, \eta))$  for any  $\eta > 0$ , and  $\widehat{P}_\eta = \widehat{P}_n(B_D(x, \eta))$  where  $\widehat{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$  denotes the empirical measure of the sample  $(X_1, \dots, X_n)$ . Define

$$\widehat{\delta}_\eta = \frac{1}{\log 2} (\log \widehat{P}_{2\eta} - \log \widehat{P}_\eta),$$

set  $\widehat{\delta}_\eta = D$  when  $\widehat{P}_\eta = 0$  and let  $\widehat{d}_\eta$  be the closest integer of  $\{1, \dots, D\}$  to  $\widehat{\delta}_\eta$ , namely  $\widehat{d}_\eta = \lfloor \widehat{\delta}_\eta + 1/2 \rfloor$ .

**Proposition 6.** For any  $1 \leq d < D$ , any  $\alpha, \beta \geq 1$ , for any  $P \in \Sigma_{\alpha, \beta}^d(x)$  and any  $\eta$  small enough (depending on  $D, \tau, \beta, r, a, b$  and  $R$ ), we have  $|\widehat{\delta}_\eta - d| \lesssim \eta$  (up to a constant depending on  $D, \tau, \beta, r, a, b$  and  $R$ ) with  $P$ -probability at least  $1 - 4 \exp(-2n\eta^{2d+2})$ .

**Corollary 1.** In the setting of Proposition 6, for any  $P \in \Sigma_{\alpha, \beta}^d(x)$  we have

$$\mathbb{P}(\widehat{d}_\eta \neq d) \leq 4 \exp\left(-2n^{1-(d+1)/(D+1)}\right)$$

for  $\eta = n^{-1/(2D+2)}$  and  $n$  large enough (depending on  $D, \tau, \beta, r, a, b$  and  $R$ ).

*Proof.* Set  $\eta = n^{-1/(2D+2)}$ . For  $n$  large enough (depending on  $D, \tau, \beta, r, a, b$  and  $R$ ), we have  $|\widehat{\delta}_\eta - d| < 1/2$  with probability at least  $1 - 4 \exp(-2n^{1-(d+1)/(D+1)})$ , thanks to Proposition 6 and  $\widehat{d}_\eta = d$  on that event.  $\square$

The estimator  $\widehat{d}_\eta$  that we constructed satisfies thus Assumption 2. We finally obtain following adaptation result in both smoothness and dimension:

**Theorem 4.** For any integer  $1 \leq d < D$ , for any real numbers  $1 \leq \alpha, \beta \leq \ell$  and  $p \geq 1$ , for any  $P \in \Sigma_{\alpha, \beta}^d(x)$ , for  $\widehat{f}_h(d; x)$  specified with  $K^{(\ell)}(d, \cdot)$ , if  $\widehat{d}$  is an estimator satisfying Assumption 2, then, for  $n$  large enough (depending on  $p, D_1, D_2, \ell, \alpha, \tau, L, \beta, r, b$  and  $R$ ), we have

$$\mathcal{R}_n^{(p)}[\widehat{f}_h(\widehat{d}; \cdot), \Sigma_{\alpha, \beta}^d](x) \lesssim \left(\frac{\log n}{n}\right)^{\alpha \wedge \beta / (2\alpha \wedge \beta + d)}$$

up to a constant depending on  $p, D, D_1, D_2, \ell, \alpha, \tau, L, \beta, r, b, R$  and that does not depend on  $x$ . Here we set  $\widehat{h} = \widehat{h}(\widehat{d}; x)$  defined in (13) and specified with  $K^{(\ell)}(\widehat{d}, \cdot)$ ,  $D_1 > 0$  and  $D_2 > p$ .

## 4 The special case of one-dimensional submanifolds

We address the question whether the regularity  $\alpha$  of  $M$  is a necessary component in the minimax rate of convergence for estimating  $f$  at a point  $x \in \mathbb{R}^D$ . In the specific case of one-dimensional manifolds, *i.e.* closed curves in an Euclidean space, we have an answer. We restrict our attention to the model  $\Sigma_\beta^1 = \mathcal{P}(\mathcal{C}_1(\tau), \mathcal{F}_\beta)$  *i.e.* we consider the case  $d = 1$ . In this section, we do not have any restriction on the regularity of the submanifold or curve  $M$ , except for its reach that we assume bigger than  $\tau > 0$ . We have the somehow surprising result:

**Theorem 5.** For any  $\beta \geq 1$  and any  $p \geq 1$ , there exists an estimator  $\widehat{f}^*$  explicitly constructed in (15) below such that, for  $n$  large enough (depending on  $p, \tau, \beta, r$  and  $a$ ),

$$\mathcal{R}_n^{(p)}[\widehat{f}^*, \Sigma_\beta^1](x) \lesssim n^{-\beta/(2\beta+1)}$$

up to a constant that depends on  $\ell, p, \tau, a, b, R$  and that does not depend on  $x$ .

In the specific case of dimension 1, any  $M$  in  $\mathcal{C}_1(\tau)$  is a closed smooth injective curve that can be parametrised by a unit-speed path  $\gamma_M : [0, L_M] \rightarrow \mathbb{R}^D$  with  $\gamma_M(0) = \gamma_M(L_M)$  and with  $L_M = \text{vol } M$  being the length of the curve. In that case, the volume density function takes a trivial form.

**Lemma 3.** For  $M \in \mathcal{C}_1(\tau)$ , for any  $x \in M$  and any  $v \in T_x M$ , we have  $\det g^x(v) = 1$ .

*Proof.* Let  $\gamma : [0, L_M] \rightarrow M$  be a unit speed parametrization of  $M$  and extend  $\gamma$  to a smooth function on  $\mathbb{R}$  by  $L_M$ -periodicity. Suppose without loss of generality that  $\gamma(0) = x$ . For any  $t \in \mathbb{R}$ , there is a canonical identification between  $T_{\gamma(t)}M$  and  $\mathbb{R}$  through the map  $v \mapsto \langle \dot{\gamma}(t), v \rangle$ . With such an identification, we can write that for  $s \in \mathbb{R} \simeq T_x M$ ,  $\exp_x(s) = \gamma(s)$  because  $\gamma$  is unit-speed. We thus have  $d\exp_x(s)[h] = h\dot{\gamma}(s)$  for any  $h \in T_{\gamma(s)}M \simeq \mathbb{R}$ . It follows that  $\det g^x(s) = \|d\exp_x(s)[1]\|^2 = \|\dot{\gamma}(s)\|^2 = 1$  and this completes the proof.  $\square$

Thanks to Lemma 3, the estimator proposed by Pelletier (2005) takes a simpler form, which we will try to take advantage of. Indeed, in the representation (2) of Pelletier (2005), only  $d_M$  remains unknown. We now show how to efficiently estimate  $d_M$  thanks to the Isomap method as coined by Tenenbaum et al. (2000). The analysis of this algorithm essentially comes from Bernstein et al. (2000) and is pursued in Arias-Castro and Le Gouic (2019), but the bounds obtained there are manifold dependent. We thus propose a slight modification of their proofs in order to obtain uniform controls over  $\mathcal{C}_1(\tau)$ , and make use of the simplifications coming from the dimension 1. Indeed, for  $d = 1$ , we have the following simple and explicit formula for the intrinsic distance on  $M$ :

$$d_M(\gamma_M(s), \gamma_M(t)) = |t - s| \wedge (L_M - |t - s|) \quad \forall s, t \in [0, L_M].$$

The Isomap method can be described as follows: let  $\epsilon > 0$ , and let  $\widehat{G}_\epsilon(x)$  be the  $\epsilon$ -neighbourhood graph built upon the data  $(X_1, \dots, X_n)$  and  $x$ , where  $x$  is the point where we evaluate the density function. Any two vertices  $p, q$  in  $\widehat{G}_\epsilon(x)$  are adjacent if and only if  $\|p - q\| \leq \epsilon$ . For a path in  $\widehat{G}_\epsilon(x)$  (a sequence of adjacent vertices)  $s = (p_0, \dots, p_m)$ , we define its length as  $L_s = \|p_1 - p_0\| + \dots + \|p_m - p_{m-1}\|$ . The distance between two vertices  $p, q$  of the graph is then defined as

$$\widehat{d}_\epsilon(p, q) = \inf \{L_s \mid s \text{ path of } \widehat{G}_\epsilon(x) \text{ connecting } p \text{ to } q\}, \quad (14)$$

and we set this distance to  $\infty$  if  $p$  and  $q$  are not connected. We are now ready to describe our estimators  $\widehat{f}^*(x)$ . For any  $h, \epsilon > 0$ , we set

$$\widehat{f}_h^*(x) = \frac{1}{nh} \sum_{i=1}^n K(\widehat{d}_\epsilon(X_i, x)/h), \quad (15)$$

for some  $K : \mathbb{R} \rightarrow \mathbb{R}$ . Notice that the kernel  $K^{(\ell)}(1; \cdot) : \mathbb{R}^D \rightarrow \mathbb{R}$  defined in Section 3.5 only depends on the norm of its argument, so it can be put in the form  $K^{(\ell)}(1; y) = K^{(\ell)}(1; \|y\|)$  with  $K^{(\ell)}(1; \cdot)$  denoting thus (with a slight abuse of notation) both functions starting from either  $\mathbb{R}$  or  $\mathbb{R}^D$ .

**Theorem 6.** *The estimator  $\widehat{f}_h^*$ , defined above for using kernel  $K^{(\ell)}(1; \cdot)$  defined in Section 3.5, verifies that for any  $P \in \Sigma_\beta^1(x)$  with  $\beta \leq \ell$ , and any  $p \geq 1$ , for  $n$  large enough (depending on  $p, \tau$  and  $a$ ) and  $h \leq (\tau \wedge r)/4$ , we have*

$$\mathcal{R}_n^{(p)}[\widehat{f}_h^*, P](x) \lesssim \frac{\epsilon^2}{h^2} + v(h) + h^\beta + \frac{1}{nh} \quad \text{for } \epsilon = \frac{32(p+1) \log n}{an}$$

up to a constant that depends on  $\ell, p, \tau, b, R$  and that does not depend on  $x$ .

The proof of Theorem 5 readily follows from Theorem 6 using the estimator  $\widehat{f}^* = \widehat{f}_h^*$  with  $h = n^{-1/(2\beta+1)}$ .

## 5 Numerical illustration

In this section we propose a few simulations to illustrate the results presented above. For the sake of visualisation, we focus on two typical examples of submanifold of  $\mathbb{R}^D$ , namely non-isometric embeddings of the flat circle  $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z}$  and of the flat torus  $\mathbb{T}^2 = \mathbb{T}^1 \times \mathbb{T}^1$ . In particular, these embeddings will be chosen in such way that their images, as submanifolds of  $\mathbb{R}^D$ , are not homogeneous compact Riemannian manifolds, so that the work of [Kerkycharian et al. \(2012\)](#) for instance cannot be of use here.

For a given embedding  $\Phi : N \rightarrow M \subset \mathbb{R}^D$  where  $N$  is either  $\mathbb{T}^1$  or  $\mathbb{T}^2$ , we construct absolutely continuous probabilities on  $M$  by pushing forward probability densities of  $N$  w.r.t. their volume measure. Indeed, if  $P = g \cdot \mu_N$ , the push-forward measure  $Q = \Phi^* P$  has density  $f$  with respect to  $\mu_M$  given by

$$\forall x \in M, \quad f(x) = \frac{g(\Phi^{-1}(x))}{|\det d\Phi(\Phi^{-1}(x))|} \quad (16)$$

where the determinant is taken in an orthonormal basis of  $T_{\Phi^{-1}(x)}N$  and  $T_x M$ , so that, if  $\Phi$  is chosen smooth enough,  $f$  has the same regularity as  $g$ . If  $\Phi$  is an embedding of  $\mathbb{T}^1$ , we simply have  $|\det d\Phi(y)| = \|\Phi'(y)\|$  for all  $y \in \mathbb{T}^1$ . If now  $\Phi$  maps  $\mathbb{T}^2$  to  $M$ , we have

$$\begin{aligned} |\det d\Phi(y)| &= \sqrt{\det \langle d\Phi(y)[e_i], d\Phi(y)[e_j] \rangle_{1 \leq i, j \leq 2}} \\ &= \sqrt{\|d\Phi(y)[e_1]\|^2 \|d\Phi(y)[e_2]\|^2 - \langle d\Phi(y)[e_1], d\Phi(y)[e_2] \rangle^2} \end{aligned} \quad (17)$$

where  $(e_1, e_2)$  is an orthonormal basis of  $\mathbb{R}^2 \simeq T_y \mathbb{T}^2$ .

### 5.1 An example of a density supported by a one-dimensional submanifold

Let  $s \geq 1$  be a regularity parameter. We define the following function on  $[0, 1]$

$$g_s(t) = \frac{2s+4}{2s+3} \times (1 - (2t)^{s+1}) \mathbb{1}_{[0, 1/2]}(t) + \frac{2s+4}{2s+3} \mathbb{1}_{(1/2, 1]}. \quad (18)$$

The function  $g_s$  is positive and  $\int_0^1 g_s(t) ds = 1$ ; it defines a probability density over the unit interval. Also, because all derivatives of  $g$  at 0 and 1 coincide up to order  $s$ , but *not* at order  $s+1$ , the function  $g_s$  also defines a probability density on  $\mathbb{T}^1$  that is  $s$ -times differentiable at 0, but *not*  $s+1$ -times. See Figure 4 for a few plots of the functions  $g_s$ . We next consider the parametric curve

$$\Phi : \begin{cases} \mathbb{T}^1 \rightarrow \mathbb{R}^2 \\ t \mapsto (\cos(2\pi t) + a \cos(2\pi \omega t), \sin(2\pi t) + a \sin(2\pi \omega t)). \end{cases}$$

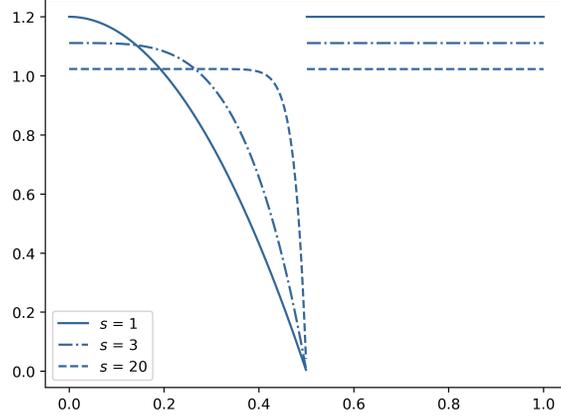


Figure 4: Plots of the densities  $g_s$  for  $s = 1, 3, 20$ .

Short computations show that  $\Phi$  is indeed an embedding as soon as  $a\omega < 1$ , in which case  $M = \Phi(\mathbb{T}^1)$  is indeed a smooth compact submanifold of  $\mathbb{R}^2$ . For the rest of this section, we set  $a = 1/8$  and  $\omega = 6$ . See Figure 5 for a plot of  $M$  with these parameters. We are interested in estimating the density

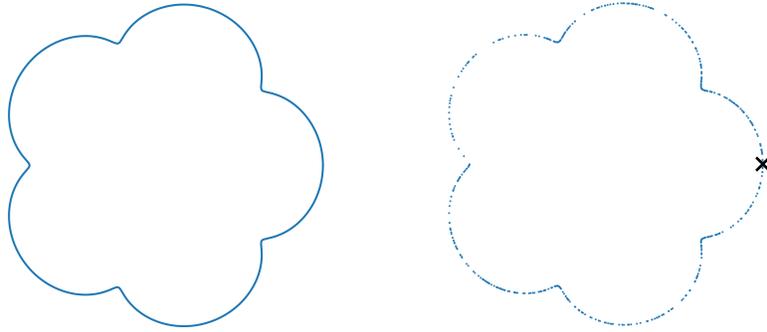


Figure 5: Plot of the submanifold  $M$  (Left) for parameters  $a = 1/8$  and  $\omega = 6$ , and 500 points sampled independently from  $\Phi^* g_s \cdot \mu_{\mathbb{T}^1}$  for  $s = 3$  (Right). The black cross denotes the point  $x$ .

$f_s$  with respect to  $d\mu_M$  of the push-forward measure  $\Phi^* g_s \cdot \mu_{\mathbb{T}^1}$ , at point  $x = (1 + a, 0) \in M$ . We set  $P_s = \Phi^* g_s \cdot \mu_{\mathbb{T}^1}$ . We use formula (16) to compute  $f_s(x)$ : We have  $\Phi^{-1}(x) = 0$  and  $\|\Phi'(0)\| = 2\pi(1 + a\omega)$  hence

$$f_s(x) = \frac{1}{2\pi(1 + a\omega)} \times \frac{2s + 4}{2s + 3} \text{ at } x = (1 + a, 0).$$

Our aim here is to provide an empirical measure for the convergence of the risk  $n \mapsto \mathcal{R}_n^{(p)}[\widehat{f}_h, P](x)$  when  $h$  is tuned optimally (in an oracle way). We pick  $p = 2$ . Our numerical procedure is detailed in Algorithm 1 below, and the numerical results are presented in Figure 6.

## 5.2 An example of a density supported by a two-dimensional submanifold

We consider a non-isometric embedding of the flat torus  $\mathbb{T}^2$ . We first construct a density function. For an integer  $s \geq 1$ , define

$$G_s : (t, u) \in [0, 1]^2 \mapsto g_s(t)g_s(u) \quad (19)$$

where  $g_s$  is defined as in (18). Obviously,  $G_s$  defines a density function on  $\mathbb{T}^2$  that is  $s$  times differentiable at  $(0, 0) \in \mathbb{T}^2$ , but *not*  $s + 1$  times. See Figure 7 for a plot of  $G_s$ .

---

**Algorithm 1** MSE rate of convergence estimation
 

---

- 1: Fix arbitrary integers  $s \geq 1$  and  $\ell \geq s$ .
  - 2: Set a grid of increasing number of points  $\mathbf{n} = (n_1, \dots, n_k) \in \mathbb{N}^k$ .
  - 3: **for**  $n_i \in \mathbf{n}$  **do**
  - 4:   Sample  $n_i$  points independently of  $P_s$  (using an acceptance-rejection method),
  - 5:   Compute  $\widehat{f}_h(x)$  with kernel  $K_\ell$  and bandwidth  $h = n_i^{-1/(2s+1)}$ ,
  - 6:   Compute the square error  $(f_s(x) - \widehat{f}_h(x))^2$ ,
  - 7:   Repeat the three previous steps  $N$  times,
  - 8:   Average the errors to get a Monte-Carlo approximation  $\widehat{\mathcal{R}}_{n_i}^{(2)}[\widehat{f}_h, P_s](x)^2$  of  $\mathcal{R}_{n_i}^{(2)}[\widehat{f}_h, P_s](x)^2$ .
  - 9: **end for**
  - 10: Perform a linear regression on the curve  $\log n_i \mapsto \log \widehat{\mathcal{R}}_{n_i}^{(2)}[\widehat{f}_h, P_s](x)^2$ .
  - 11: **return** The coefficient of the linear regression.
- 

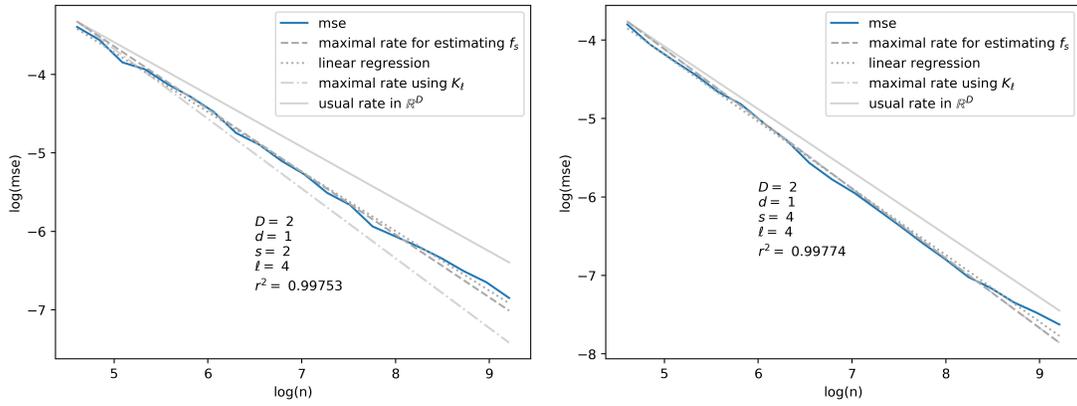


Figure 6: Plot of the empirical mean square error (blue) for a density supported by a one-dimensional submanifold case with parameters  $s = 2$ ,  $\ell = 4$  (Left) and  $s = 10$ ,  $\ell = 1$  (Right). We use a log-regular grid  $\mathbf{n}$  of 20 points ranging from 100 to  $10^4$ . Each experiment is repeated  $N = 500$  times.

We next consider the parametric surface

$$\Phi : \begin{cases} \mathbb{T}^2 \rightarrow \mathbb{R}^3 \\ (t, u) \mapsto \begin{pmatrix} (b + \cos(2\pi t)) \cos(2\pi u) + a \sin(2\pi \omega t) \\ (b + \cos(2\pi t)) \sin(2\pi u) + a \cos(2\pi \omega t) \\ \sin(2\pi t) + a \sin(2\pi \omega u) \end{pmatrix} \end{cases}$$

for some  $a, b, \omega \in \mathbb{R}$ . In the remaining of the section, we set  $a = 1/8$ ,  $b = 3$  and  $\omega = 5$ . We show that  $\Phi$  indeed defines an embedding. See Figure 8 for a plot of the submanifold  $M = \Phi(\mathbb{T}^2)$ . For an integer  $s \geq 1$ , we denote by  $F_s$  the density of the push forward measure  $Q_s = \Phi^* G_s \cdot \mu_{\mathbb{T}^2}$  with respect to the volume measure  $\mu_M$ . Let  $x = (b+1, a, 0)$  be the image of 0 by  $\Phi$ . Simple calculations show that the differential of  $\Phi$  at 0 evaluated at  $e_1 = (1, 0) \in T_0 \mathbb{T}^2$  and  $e_2 = (0, 1) \in T_0 \mathbb{T}^2$  is equal to respectively  $d\Phi(0)[e_1] = 2\pi(a\omega, 0, 1)$  and  $d\Phi(0)[e_2] = 2\pi(0, b+1, a\omega)$ . Hence formula (17) yields

$$\det d\Phi(0) = (2\pi)^2 \left( (1 + a^2\omega^2) \left( (b+1)^2 + a^2\omega^2 \right) - a^2\omega^2 \right)^{1/2} \quad (20)$$

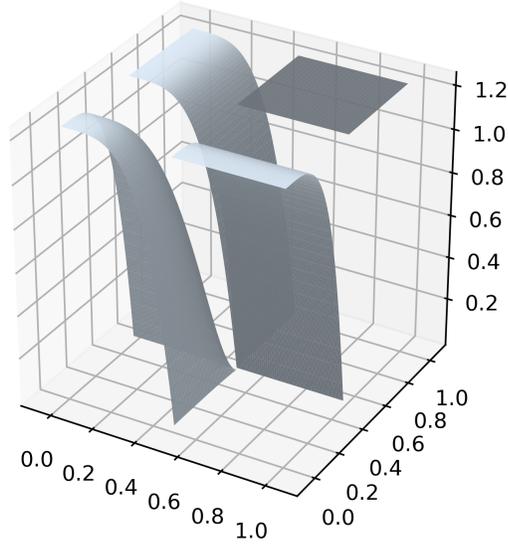


Figure 7: Plot of the probability density function  $G_s$  for  $s = 3$ .

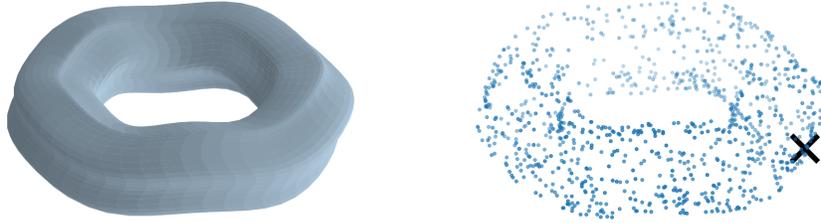


Figure 8: Plot of the submanifold  $M$  (Left) for parameters  $a = 1/8$ ,  $b = 3$  and  $\omega = 6$ , and 500 points sampled independently from  $\Phi^*G_s \cdot \mu_{\mathbb{T}^2}$  with  $s = 3$  (Right). The black cross marks the point  $x$ .

and we obtain

$$F_s(x) = \left( \frac{2s+4}{2s+3} \right)^2 \frac{1}{4\pi^2} \left( (1+a^2\omega^2) \left( (b+1)^2 + a^2\omega^2 \right) - a^2\omega^2 \right)^{-\frac{1}{2}}.$$

In the same way as in the previous section, we aim at providing an empirical measure for the rate of convergence of the risk  $\mathcal{R}_n^{(p)}[\widehat{f}_h, P_s]$  when  $h$  is suitably tuned with respect to  $n$  and  $s$ . This is done using again Algorithm 1. The results are presented in Figure 9.

### 5.3 Adaptation

In this section we estimate a density when its regularity is unknown, contrary to the previous simulation where the regularity parameter  $s$  is pugged in the bandwidth choice  $n^{-1/(2s+d)}$ . This is performed using Lepski's method presented in Section 3.4. The MSE rate is computed using Algorithm 1, for both the one-dimensional and the two-dimensional synthetic datasets.

For the adaptive estimation on the two-dimensionnal manifold, we observed that the corrective term  $\det d\Phi(0)$  computed in (20) made the resulting density  $F_s(x)$  very small, while the function  $\psi$

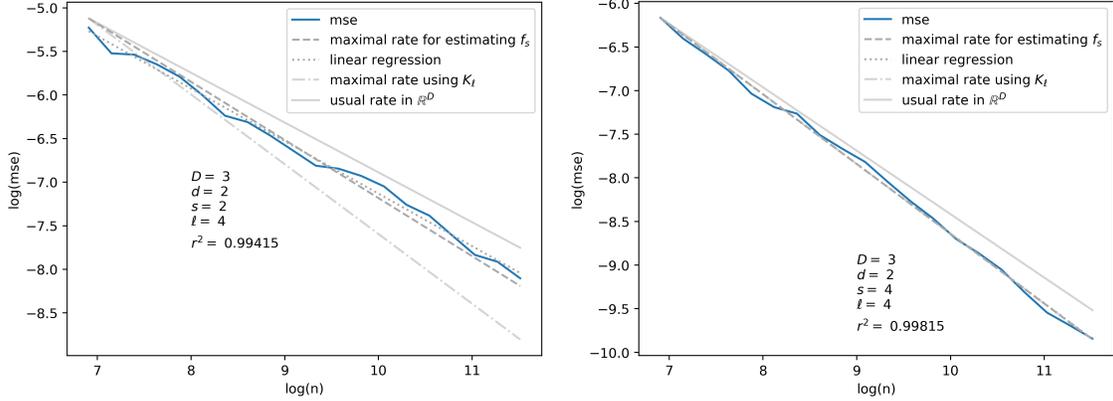


Figure 9: Plot of the empirical mean square error (blue) for a 2-dimensional submanifold with parameters  $s = 2$ ,  $\ell = 4$  (Left) and  $s = 4$ ,  $\ell = 4$  (Right). We use a log-regular grid  $\mathbf{n}$  of 20 points ranging from  $10^3$  to  $10^5$ . Each experiment is repeated  $N = 500$  times.

defined at (11) and used to tune the bandwidth soared dramatically because of the retained value of  $v_d = 4^d \zeta_d \|K^{(d,\ell)}\|_\infty^2 b$ , so that the values of  $\hat{f}_h$  and  $\psi(h, \cdot)$  are not of the same order anymore at this scale (using maximum  $10^6$  observations). To circumvent this effect, we use a different kind of density  $g_{s,\lambda}$  defined as

$$g_{s,\lambda}(t) = C_s \lambda (1 - (2\lambda t)^{s+1}) \mathbb{1}_{[0,1/2\lambda]}(t) + C_s \lambda (1 - (2\lambda(1-t))^{s+2}) \mathbb{1}_{[1-1/2\lambda,1]}(t)$$

for some  $\lambda \geq 1$  and with  $C_s = 1 - 1/(2s+2) - 1/(2s+6)$ . We plot the function  $g_{s,\lambda}$  below in Figure 10 for some value of  $s$  and  $\lambda = 2$ .

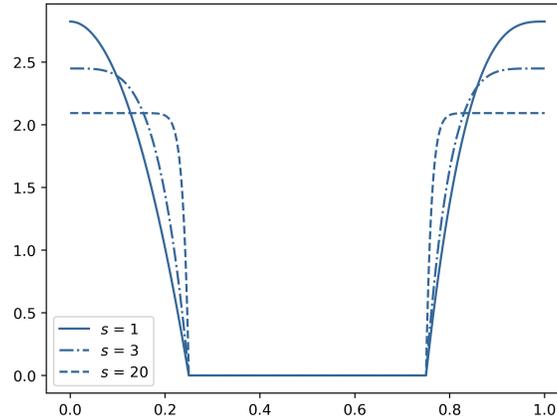


Figure 10: Plot of the probability density function  $g_{s,\lambda}$  for  $\lambda = 2$  and  $s \in \{1, 3, 20\}$ .

Again,  $g_{s,\lambda}$  defines a probability density on  $\mathbb{T}^1$  that is  $s$ -times differentiable at 0 but not  $s+1$ -times. Like before, we consider the density  $G_{s,\lambda} : (t, u) \mapsto g_{s,\lambda}(t)g_{s,\lambda}(u)$  on the torus  $\mathbb{T}^2$  and the push-forwarded probability measure  $\Phi^* G_{s,\lambda} \cdot \mu_{\mathbb{T}^2}$  which has density  $F_{s,\lambda}$  with respect to  $\mu_M$ . For  $\lambda = 10$ , we find that  $F_{s,\lambda}(x) \simeq 1$  for most value of  $s$ , so that we set  $v = 1$  in the definition of  $\psi$  for the sake of the estimation procedure. We have no theoretical guarantee that such a method work but we recover nonetheless the right rate in the estimation of the value of the density, see Figure 11. The numerical results are presented in Table 1.

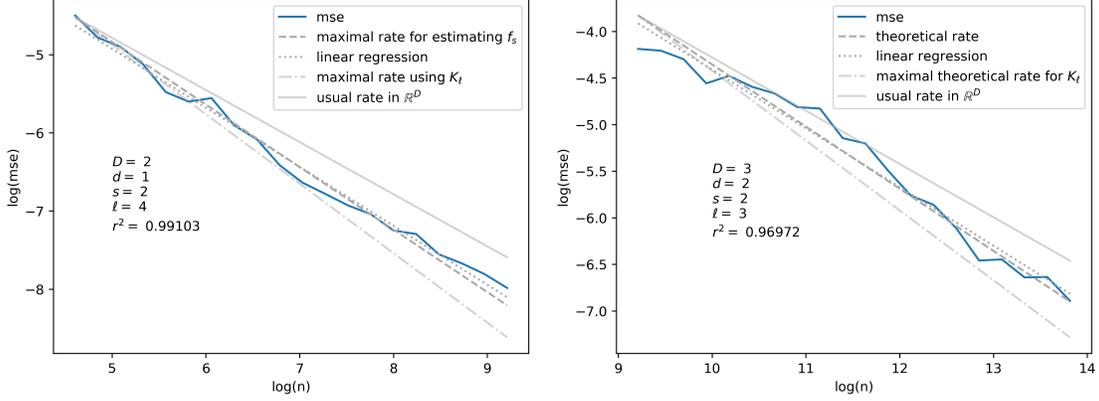


Figure 11: Plot of the empirical mean square error (blue) for  $s = 2$ ,  $\ell = 4$  for a one-dimensional submanifold (Left) and a two-dimensional submanifold for  $s = 2$  and  $\ell = 3$  (Right). The bandwidth  $h$  is chosen adaptively using Lepski's method as in Section 3.4. Each experience is repeated  $N = 100$  times.

$D$	$d$	$\ell$	$s$	Choice of $h$	$\frac{2s}{2s+D}$	$\frac{2s}{2s+d}$	$\frac{2\ell}{2\ell+d}$	Linear regression	Error (%)
2	1	4	2	Deterministic	0.667	0.8	0.889	0.759	5.08
2	1	4	4	Deterministic	0.8	0.889	0.889	0.851	4.25
3	2	4	2	Deterministic	0.571	0.667	0.8	0.602	9.65
3	2	4	4	Deterministic	0.727	0.8	0.8	0.798	2.48
2	1	4	2	Adaptive	0.667	0.8	0.889	0.756	5.65
3	2	3	2	Adaptive	0.571	0.667	0.75	0.630	5.55

Table 1: This table sums up the results of the simulation carried on in this section. In light blue is the expected rate for the mean square error. The column *Linear regression* contains the outcome of Algorithm 1 for the given set of parameters. We then compute the relative error between this coefficient and the expected rate in the last column.

## 6 Proofs

### 6.1 Proofs of Section 3.2

We set  $K_h(x) = h^{-d}K(x/h)$  and start with bounding the variance of  $K_h(X-x)$  when  $X$  is distributed according to  $P \in \Sigma_{sd}$ . Let first observe that

$$|K_h(X-x)| \leq \frac{\|K\|_\infty}{h^d} \mathbb{1}_{B_D(x,h)}(X) \leq \frac{\|K\|_\infty}{h^d}$$

**Lemma 4.** *If  $P \in \Sigma_{sd}$  then, for all  $x \in M_P$  and all  $h \leq \tau/2$ ,*

$$\text{Var}_P(K_h(X-x)) \leq \frac{v}{h^d} \quad \text{where } v = 4^d \zeta_d \|K\|_\infty^2 b.$$

with  $\zeta_d$  being the volume of the unit ball in  $\mathbb{R}^d$ .

*Proof.* We have

$$\text{Var}_P(K_h(X-x)) \leq \mathbb{E}_P[K_h(X-x)^2] \leq \frac{\|K\|_\infty^2}{h^{2d}} P(B(x,h)) \leq \frac{4^d \zeta_d b \|K\|_\infty^2}{h^d}$$

where we used (21) and Lemma 1 with  $\eta = \tau/2$ .  $\square$

Using Bernstein inequality (Boucheron et al., 2013, Thm. 2.10 p.37), for any  $P \in \Sigma_{\text{sd}}$ , any  $x \in M_P$ , and any  $t > 0$ , we infer

$$\mathbb{P}\left(|\widehat{\xi}_h(P,x)| \geq \sqrt{\frac{2vt}{nh^d}} + \frac{\|K\|_\infty t}{nh^d}\right) \leq 2e^{-t}, \quad (21)$$

where  $\mathbb{P}$  is a short-hand notation for the distribution  $P^{\otimes n}$  of the  $n$ -sample  $X_1, \dots, X_n$  taken under  $P$ . The bound (21) is the main ingredient needed to bound the  $L_p$ -norm of the stochastic deviation of  $\widehat{f}_h$ .

*Proof of Proposition 4.* We denote by  $y_+ = \max\{y, 0\}$  the positive part of a real number  $y$ . We start with

$$\mathbb{E}_P[|\widehat{\xi}_h(P,x)|^p] \leq 2^{p-1} (v(h)^p + \mathbb{E}_P[(|\widehat{\xi}_h(P,x)| - v(h))_+^p]).$$

The first term has the right order. For the second one, we make use of (21) to infer

$$\begin{aligned} \mathbb{E}_P[(|\widehat{\xi}_h(P,x)| - v(h))_+^p] &= \int_0^\infty \mathbb{P}(|\widehat{\xi}_h(P,x)| > v(h) + u) p u^{p-1} du \\ &= p v(h)^p \int_0^\infty \mathbb{P}(|\widehat{\xi}_h(P,x)| > v(h)(1+u)) u^{p-1} du \\ &\leq p v(h)^p \left(1 + \int_1^\infty \mathbb{P}\left(|\widehat{\xi}_h(P,x)| > \sqrt{\frac{2v(1+u)}{nh^d}} + \frac{\|K\|_\infty(1+u)}{nh^d}\right) u^{p-1} du\right) \\ &\leq p v(h)^p \left(1 + \int_1^\infty 2e^{-1-u} u^{p-1} du\right) \\ &\leq p v(h)^p (1 + \Gamma(p)) \end{aligned}$$

which ends the proof.  $\square$

The proof of Lemma 2 partly relies on the following technical lemma.

**Lemma 5.** *Let  $\beta \geq 1$  be a real number and let  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  verifying that  $\|g\|_\infty \leq b$  and that the restriction of  $g$  to  $B_d(0,r)$  is  $\beta$ -Hölder, meaning that*

$$\forall v, w \in B_d(0,r), \quad \|d^k g(v) - d^k g(w)\| \leq R \|v - w\|^\delta$$

for some  $R \geq 0$  with  $k = \lceil \beta - 1 \rceil$  and  $\delta = \beta - k$ . Then there exists a constant  $C$  (depending on  $\beta, r, b$  and  $R$ , and depending on  $\beta$  when  $r = \infty$ ) such that, for all  $1 \leq j \leq k$ ,

$$\sup_{v \in B(0,r/2)} \|d^j g(v)\|_{\text{op}} \leq C b^{1-j/\beta} R^{j/\beta}.$$

*Proof.* Let  $v \in B_d(0,r/2)$ . Since  $g$  is  $\beta$ -Hölder on  $B_d(0,r)$ , we know that there exists a function  $R_v$  such that, for any  $z$  such that  $v+z \in B(0,r)$ , we have

$$g(v+z) - \sum_{j=0}^k \frac{1}{j!} d^j g(v)[z^{\otimes j}] = R_v(z)$$

with  $|R_v(z)| \leq R\|z\|^\beta/k!$ . Let  $h = (2bk!/R)^{1/\beta}$ , and  $z_0 \in \mathbb{R}^d$  be unit-norm. Pick  $a_1, \dots, a_k \in (0, 1)$  all distincts and small enough such that  $ha_k z_0 \in B(0, r/2)$  for all  $k$  (if  $r = \infty$ , then we can pick the  $a_i$  independently from  $R, b$  and  $\beta$ ). Introducing the vectors of  $\mathbb{R}^k$

$$X = \left( hdg(v)[z_0], \dots, \frac{h^k}{k!} d^k g(v)[z_0^{\otimes k}] \right) \quad \text{and}$$

$$Y = (g(v + ha_1 z_0) - g(v) - R_v(ha_1 z_0), \dots, g(v + ha_k z_0) - g(v) - R_v(ha_k z_0))$$

we have  $Y = VX$  with  $V$  being the Vandermonde matrix associated with the real numbers  $(a_1, \dots, a_k)$ . The former being invertible, we have  $\|X\| \leq \|V^{-1}\|_{\text{op}} \|Y\|$  and thus, for any  $1 \leq j \leq k$

$$\left| \frac{h^j}{j!} d^j g(v)[z_0^{\otimes j}] \right| \leq \|V^{-1}\|_{\text{op}} \left( 2b + \frac{R}{k!} h^\beta \right).$$

Substituting the value of  $h$  and noticing that the former inequality holds for every unit-norm vector  $z_0$ , we can conclude.  $\square$

*Proof of Lemma 2.* We set  $B_h = B_D(x, h)$ . Since  $\tau/4$  is smaller than the injectivity radius of  $\exp_x$  (see Proposition 2) we can write

$$f_h(P, x) = \int_{B_h} K_h(p-x) f(p) d\mu_M(p) = \int_{\exp_x^{-1} B_h} K_h(\exp_x v - x) f(\exp_x v) \zeta(v) dv \quad (22)$$

with  $\zeta(v) = \sqrt{\det g^x(v)}$ . We set  $\gamma = \alpha \wedge \beta$  and  $k = \lceil \gamma - 1 \rceil$ . Let  $F$  denote the map  $f \circ \exp_x$ . For  $h$  smaller than  $(\tau \wedge r)/4$ , we have  $\exp_x^{-1} B_h \subset B_{T_x M}(0, 2h) \subset B_{T_x M}(0, (r \wedge \tau)/2)$  (see Proposition 3). We can thus write the following expansion, valid for all  $v \in \exp_x^{-1} B_h$  and all  $w \in T_x M$ ,

$$\exp_x(v) = x + v + \sum_{j=2}^{k+1} \frac{h^j}{j!} d^j \exp_x(0)[v^{\otimes j}] + R_1(v) \quad \text{with} \quad \|R_1(v)\| \leq C_1 \|v\|^{\gamma+1}, \quad (23)$$

$$F(v) = f(x) + \sum_{j=1}^k \frac{h^j}{j!} d^j F(0)[v^{\otimes j}] + R_2(v) \quad \text{with} \quad |R_2(v)| \leq C_2 \|v\|^\gamma, \quad (24)$$

$$K(v+w) = \mathbb{1}_{\{\|v+w\| \leq 1\}} \left( K(v) + \sum_{j=1}^k \frac{h^j}{j!} d^j K(v)[w^{\otimes j}] + R_3(v, w) \right) \quad \text{with} \quad |R_3(v, w)| \leq C_3 \|w\|^\gamma, \quad (25)$$

with  $C_1$  depending on  $\alpha, \tau$  and  $L$ ,  $C_2$  depending on  $\beta, r, b$  and  $R$  (see Lemma 5), and  $C_3$  depending on  $K$ . Since now we know that  $g_{ij}^x(v) = \langle d\exp_x(v)[e_i], d\exp_x(v)[e_j] \rangle$ , we have a similar expansion for the mapping  $\zeta(v) = \sqrt{\det g^x(v)}$

$$\zeta(v) = 1 + \sum_{j=1}^k \frac{h^j}{j!} d^j \zeta(0)[v^{\otimes j}] + R_4(v) \quad \text{with} \quad |R_4(v)| \leq C_4 \|v\|^\gamma \quad (26)$$

with  $C_4$  depending on  $\alpha, \tau$  and  $L$ . Making the change of variable  $v = hw$  in (22), we get

$$f_h(P, x) = \sum_{k=0}^k G_j(h, P, x) + R_h(P, x)$$

with  $G_j$  corresponding to the integration of the  $j$ -th order terms in the expansion around 0 of the function  $v \mapsto K\left(\frac{p-\exp_p(hw)}{h}\right)F(hw)\zeta(hw)$ . In particular  $G_j$  can be written as a sum of terms of the type

$$I = h^j \int_{\frac{1}{h} \exp_x^{-1} B_h} d^m K(w) [\phi(w)^{\otimes m}] \psi(w) dw$$

where  $\psi$  and  $\phi$  are monomials in  $w$  satisfying  $m \deg \phi + \deg \phi = j$ , with coefficients bounded by constants depending on  $\alpha, \tau, L, \beta, r, b$  and  $R$  (again, use Lemma 5 to bound the derivatives). Since now  $B_{T_x M}(0, 1) \subset \frac{1}{h} \exp_x^{-1} B_h$ , and since  $d^j K$  is zero outside of  $B(0, 1)$ , we have that  $G_j$  can actually be written  $G_j(h, P, x) = h^j G(P, x)$  with  $|G(P, x)| \leq C$  for some  $C$  depending on  $K, \alpha, \tau, L, \beta, r, b$  and  $R$ . Similar reasoning leads to  $R_h(P, x) \leq Ch^\gamma$  with  $C$  depending again on  $K, \alpha, \tau, L, \beta, r, b$  and  $R$ . To conclude, it remains to compute  $G_0(P, x)$ . Looking at the zero-th order terms in the expansions (23) to (26), we find that

$$G_0(P, x) = \int_{B_{T_x M}(0, 1)} K(w) f(x) dw = f(x)$$

where we used Assumption 1. The proof of Lemma 2 is complete.  $\square$

*Proof of Proposition 5.* For a positive integer  $\ell \geq 1$ , let  $f_h^{(\ell)}(P, x)$  be the mean of the estimator  $\widehat{f}_h(x)$  computed using  $K^{(\ell)}$ . Let  $\gamma = \alpha \wedge \beta$  and  $k = \lceil \gamma - 1 \rceil$ . We recursively prove on  $1 \leq \ell < \infty$  the following identity

$$\forall h \leq (\tau \wedge r)/4, \quad f_h^{(\ell)}(P, x) = f(x) + \sum_{j=\ell}^k h^j G_j^{(\ell)}(P, x) + R_h^{(\ell)}(P, x) \quad (27)$$

where  $|R_h^{(\ell)}(P, x)| \leq C^{(\ell)} h^\gamma$  for some constant  $C^{(\ell)}$  depending on  $\tau, \ell, L, R, b, \beta$  and  $r$ . The initialisation step  $\ell = 1$  has been proven in Lemma 2. Let now  $1 \leq \ell \leq k$ . By linearity of  $f_h(P, x)$  with respect to  $K$ , we have

$$f_h^{(\ell+1)}(P, x) = 2f_{2^{-1/\ell}h}^{(\ell)}(P, x) - f_h^{(\ell)}(P, x).$$

Since  $2^{-1/\ell}h \leq h$ , we can use our induction hypothesis (27) and find

$$f_h^{(\ell+1)}(P, x) = f(x) + \sum_{j=\ell}^k (2^{1-j/\ell} - 1) h^j G_j^{(\ell)}(P, x) + 2R_{2^{-1/\ell}h}^{(\ell)}(P, x) - R_h^{(\ell)}(P, x).$$

We conclude noticing that  $2^{1-j/\ell} - 1 = 0$  for  $j = \ell$ , and setting  $G_j^{(\ell+1)}(P, x) = (2^{1-j/\ell} - 1) G_j^{(\ell)}(P, x)$  and  $R_h^{(\ell+1)}(P, x) = 2R_{2^{-1/\ell}h}^{(\ell)}(P, x) - R_h^{(\ell)}(P, x)$ . The new remainder term verifies

$$|R_h^{(\ell+1)}(P, x)| \leq (2^{1-\gamma/\ell} + 1) C^{(\ell)} h^\gamma \leq 3C^{(\ell)} h^\gamma \quad (28)$$

ending the induction by setting  $C^{(\ell+1)} = 3C^{(\ell)}$ . When  $\ell \geq k + 1$ , the induction step is trivial.  $\square$

## 6.2 Proofs of Section 3.3

We go along a classical line of arguments, thanks to a Bayesian two-point inequality by means of Le Cam's lemma (Yu, 1997, Lem. 1), restated here in our context. For two probability measures  $P_1, P_2$ , we write  $\text{TV}(P_1, P_2) = \sup_A |P_1(A) - P_2(A)|$  for their variational distance and  $H^2(P_1, P_2) = \int (\sqrt{dP_1} - \sqrt{dP_2})^2$  for their (squared) Hellinger distance.

**Lemma 6.** (Le Cam) For any  $P_1, P_2 \in \Sigma_{\alpha, \beta}(x)$ , we have,

$$\begin{aligned} \inf_{\widehat{f}} \mathcal{R}_n^{(p)}[\widehat{f}, \Sigma_{\alpha, \beta}](x) &\geq \frac{1}{2} |f_1(x) - f_2(x)| (1 - \text{TV}(P_1^{\otimes n}, P_2^{\otimes n})) \\ &\geq \frac{1}{2} |f_1(x) - f_2(x)| \left(1 - \sqrt{2 - 2(1 - H^2(P_1, P_2)/2)^n}\right). \end{aligned} \quad (29)$$

*Proof.* The proof of (29) can be found in Yu (1997, Lem. 1). It only remains to see that  $\text{TV}(P_1^{\otimes n}, P_2^{\otimes n}) \leq \sqrt{2 - 2(1 - H^2(P_1, P_2)/2)^n}$ . This comes from classical inequalities on the Hellinger distance, see Tsybakov (2008, Lem. 2.3 p.86) and Tsybakov (2008, Prp.(i)-(iv) p.83).  $\square$

*Proof of Theorem 2.* Suppose without loss of generality that  $x = 0$  and consider a smooth submanifold  $M$  of  $\mathbb{R}^{d+1} \subset \mathbb{R}^D$  that contains the disk  $B_d(0, 1) \subset \mathbb{R}^d$  with reach is greater than  $\tau$ , see Figure 12 for a diagram of such an  $M$ . By smoothness and compactness of  $M$ , there exists  $L_*$  (depending on  $\tau$ ) such that  $M \in \mathcal{C}_{d, \alpha}(\tau, L_*)$ . Let  $P$  be the uniform probability measure over  $M$ , with density  $f : x \mapsto 1/\text{vol } M$ . We have  $P \in \Sigma_{\alpha, \beta}(0)$  as long as  $L_* \leq L$  and  $a \leq 1/\text{vol } M \leq b$  an assumption we make from now on. For  $0 < \delta \leq 1$ , let  $P_\delta = f_\delta \cdot \mu_M$  with

$$f_\delta(y) = \begin{cases} f(y) + \delta^\beta G(y/\delta) & \text{if } y \in B(0, \delta) \\ f(y) & \text{otherwise} \end{cases}$$

with  $G : \mathbb{R}^d \rightarrow \mathbb{R}$  a smooth function with support in  $B_d(0, 1)$  and such that  $\int_{\mathbb{R}^d} G(y) dy = 0$ . We pick  $G$  such that  $f_\delta \in \mathcal{F}_\beta$  for small enough  $\delta$ , depending on  $r$  and  $\tau$ . Such a  $G$  can be chosen to depend on  $R$  only.

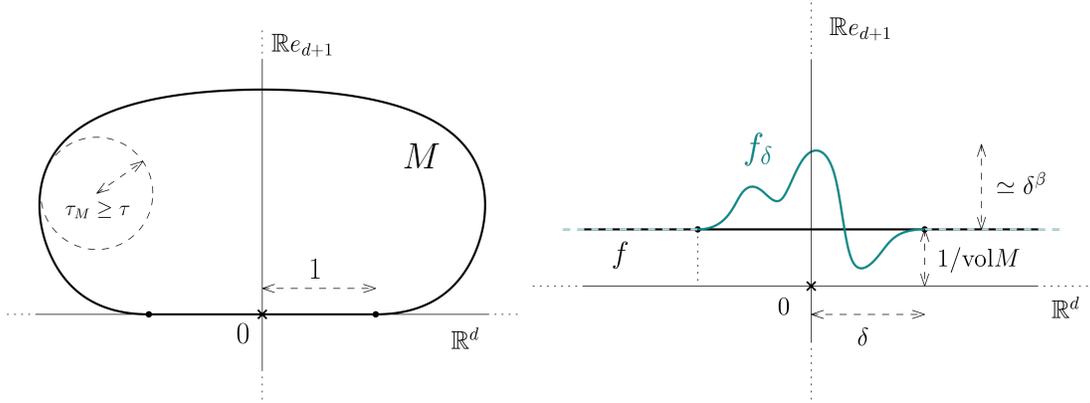


Figure 12: Diagram of a candidate for  $M$  (Left) and of the densities  $f$  and  $f_\delta$  around 0 (Right).

For  $\delta$  small enough (depending on  $r$  and  $\tau$ ), we thus have  $P_\delta \in \Sigma_{\alpha, \beta}(0)$  as well. By Lemma 6, we infer

$$\inf_{\widehat{f}} \mathcal{R}_n^{(p)}[\widehat{f}, \Sigma_{\alpha, \beta}](0) \geq \frac{1}{2} \delta^\beta |G(0)| \left(1 - \sqrt{2 - 2(1 - H^2(P, P_\delta))^n}\right)$$

so that it remains to compute  $H^2(P, P_\delta)$ . We have the following bound

$$\begin{aligned} H^2(P, P_\delta) &= \int_{B_d(0, \delta)} (1 - \sqrt{1 + \text{vol } M \delta^\beta G(x/\delta)})^2 dx \\ &\leq \int_{B_d(0, \delta)} (\text{vol } M)^2 \delta^{2\beta} G^2(x/\delta) dx \\ &\leq (C \vee 1) \delta^{2\beta + d} \end{aligned}$$

with  $C = \text{vol } M \int_{B(0,1)} G(x)^2 dx$  depending on  $\tau$  and  $R$  only. Taking  $\gamma = (1/(C \vee 1)n)^{1/(2\beta+d)}$  we obtain, for large enough  $n$  (depending on  $\tau$ )

$$\inf_{\widehat{f}} \mathcal{R}_n^{(p)}[\widehat{f}, \Sigma_{\alpha,\beta}](0) \geq \frac{1}{2}((C \vee 1)n)^{-\beta/(2\beta+d)} \sqrt{2 - 2(1 - 1/n)^n} \geq C_* n^{-\beta/(2\beta+d)},$$

with  $C_* = (C \vee 1)^{-1/2}$  depending on  $\tau$  and  $R$ . □

### 6.3 Proofs of Section 3.4

**Lemma 7.** *For any  $p \geq 1$ ,  $P \in \Sigma_{\text{sd}}(x)$ , and  $D_2 > p$ , we have*

$$\mathcal{R}_n^{(p)}[\widehat{f}, P](x) \lesssim v(h^*(P, x))\lambda(h^*(P, x))$$

up to a constant depending on  $p, D_1$  and  $D_2$ , with

$$h^*(P, x) = \max \{h \in \mathcal{H} \mid \forall \eta \in \mathcal{H}(h), |f_\eta(P, x) - f(x)| \leq D_1 v(h)\lambda(h)\}.$$

*Proof.* We fix  $P \in \Sigma_{\alpha,\beta}(x)$  and write  $\widehat{h}$  and  $h^*$  for  $\widehat{h}(x)$  and  $h^*(P, x)$  respectively. Let  $\mathcal{A} = \{\widehat{h} \geq h^*\}$ . We can write  $(\mathcal{R}_n^{(p)}[\widehat{f}, P](x))^p = \mathcal{R}_{\mathcal{A}} + \mathcal{R}_{\mathcal{A}^c}$ , where

$$\mathcal{R}_{\mathcal{A}} = \mathbb{E}_P[|\widehat{f}(x) - f(x)|^p \mathbb{1}_{\mathcal{A}}] \quad \text{and} \quad \mathcal{R}_{\mathcal{A}^c} = \mathbb{E}_P[|\widehat{f}(x) - f(x)|^p \mathbb{1}_{\mathcal{A}^c}].$$

We start with bounding  $\mathcal{R}_{\mathcal{A}}$ . Firstly,

$$\begin{aligned} \mathcal{R}_{\mathcal{A}} &\leq 3^{p-1} (\mathbb{E}_P[|\widehat{f}_{\widehat{h}}(x) - \widehat{f}_{h^*}(x)|^p \mathbb{1}_{\mathcal{A}}] + \mathbb{E}_P[|\widehat{f}_{h^*}(x) - f_{h^*}(P, x)|^p \mathbb{1}_{\mathcal{A}}] \\ &\quad + \mathbb{E}_P[|f_{h^*}(P, x) - f(x)|^p \mathbb{1}_{\mathcal{A}}]). \end{aligned}$$

Next, by definition of  $\widehat{h}$  and  $\mathcal{A}$ , we have

$$|\widehat{f}_{\widehat{h}}(x) - \widehat{f}_{h^*}(x)| \mathbb{1}_{\mathcal{A}} \leq \psi(\widehat{h}, h^*) \mathbb{1}_{\mathcal{A}} \leq (2D_1 + 2)v(h^*)\lambda(h^*).$$

By definition of  $h^*$ , we also have  $|f_{h^*}(P, x) - f(x)| \leq D_1 v(h^*)\lambda(h^*)$ . Finally, using Proposition 4

$$\mathbb{E}_P[|\widehat{f}_{h^*}(x) - f_{h^*}(P, x)|^p \mathbb{1}_{\mathcal{A}}] \leq c_p v(h^*)^p \leq c_p (v(h^*)\lambda(h^*))^p$$

holds as well. Putting all three inequalities together yields

$$\mathcal{R}_{\mathcal{A}} \leq C_{\mathcal{A}} (v(h^*)\lambda(h^*))^p \quad \text{with} \quad C_{\mathcal{A}} = 3^{p-1} ((2D_1 + 2)^p + c_p + D_1^p).$$

We now turn to  $\mathcal{R}_{\mathcal{A}^c}$ . Notice that for any  $h \in \mathcal{H}(h^*)$ , we have

$$|f_h(P, x) - f(x)| \leq D_1 v(h^*)\lambda(h^*) \leq D_1 v(h)\lambda(h),$$

hence

$$|\widehat{f}_h(x) - f(x)| \leq D_1 v(h^*)\lambda(h^*) + |\widehat{\xi}_h(P, x)|.$$

We can thus write

$$\mathcal{R}_{\mathcal{A}^c} = \sum_{h \in \mathcal{H}(h^*)} \mathbb{E}_P[|\widehat{f}_h(x) - f(x)|^p \mathbb{1}_{\{\widehat{h} < h\}}] \leq \sum_{h \in \mathcal{H}(h^*)} \mathbb{E}_P[(D_1 v(h^*)\lambda(h^*) + |\widehat{\xi}_h(P, x)|)^p \mathbb{1}_{\{\widehat{h} < h\}}].$$

Now, for any  $h \in \mathcal{H}$ ,  $h < h^*$ , we have

$$\begin{aligned} \{\widehat{h} = h\} &\subset \{\exists \eta \in \mathcal{H}(h), |\widehat{f}_{2h}(x) - \widehat{f}_\eta(x)| > \psi(2h, \eta)\} \\ &\subset \bigcup_{\eta \in \mathcal{H}(h)} \{2D_1v(h^*)\lambda(h^*) + |\widehat{\xi}_{2h, \eta}(P, x)| > \psi(2h, \eta)\}, \end{aligned}$$

where  $\widehat{\xi}_{2h, \eta}(P, x) = \widehat{\xi}_{2h}(P, x) - \widehat{\xi}_\eta(P, x)$ , and where we used the triangle inequality and the definition of  $h^*$ . Now, we have  $2D_1v(h^*)\lambda(h^*) \leq 2D_1v(2h)\lambda(2h)$  since  $2h \leq h^*$  and by definition of  $\psi(2h, \eta)$ , we infer

$$\{\widehat{h} = h\} \subset \bigcup_{\eta \in \mathcal{H}(h)} \{|\widehat{\xi}_{2h, \eta}(P, x)| > v(2h, \eta)\lambda(\eta)\}$$

so that

$$\mathbb{P}(\widehat{h} = h) \leq \sum_{\eta \in \mathcal{H}(h)} \mathbb{P}(|\widehat{\xi}_{2h, \eta}(P, x)| > v(2h, \eta)\lambda(\eta)) \quad (30)$$

$$\begin{aligned} &\leq \sum_{\eta \in \mathcal{H}(h)} \mathbb{P}\left(|\widehat{\xi}_{2h, \eta}(P, x)| > \sqrt{\frac{8v\lambda(\eta)}{n\eta^d}} + \frac{2\|K\|_\infty\lambda(\eta)}{n\eta^d}\right) \\ &\leq \sum_{\eta \in \mathcal{H}(h)} 2\exp(-\lambda(\eta)^2). \end{aligned} \quad (31)$$

For (30) we use the fact that  $\lambda(\eta) \geq 1$  and Bernstein's inequality on the random variable  $\widehat{\xi}_{2h, \eta}(P, x)$  for (31). Noticing now that  $\lambda(\eta)^2 \geq dD_2 \log(h^*/\eta)$ , we further obtain

$$\mathbb{P}(\widehat{h} = h) \leq 2 \left(\frac{h}{h^+}\right)^{D_2d} \times \sum_{j=0}^{\lfloor \log_2(h^*/h^-) \rfloor} 2^{-jD_2d} \leq \frac{2}{1-2^{-D_2d}} \left(\frac{h}{h^+}\right)^{D_2d}.$$

For any  $h \in \mathcal{H}$ ,  $h < h^*$ , we thus get the following bound, using Cauchy-Schwarz inequality

$$\begin{aligned} &\mathbb{E}_P\left[\left(D_1v(h^*)\lambda(h^*) + |\widehat{\xi}_h(P, x)|\right)^p \mathbb{1}_{\{\widehat{h}=h\}}\right] \\ &\leq \mathbb{P}(\widehat{h} = h)^{1/2} E_P\left[\left(D_1v(h^*)\lambda(h^*) + |\widehat{\xi}_h(P, x)|\right)^{2p}\right]^{1/2} \\ &\leq 2^{(2p-1)/2} \sqrt{\frac{2}{1-2^{-D_2d}} \left(\frac{h}{h^+}\right)^{D_2d/2}} \left(D_1^p v(h^*)^p \lambda(h^*)^p + c_{2p}^{1/2} v(h)^p\right). \end{aligned} \quad (32)$$

We plan to sum over  $h \in \mathring{\mathcal{H}}(h^*)$  the RHS of (32). Notice first that

$$\sum_{h < h^*} h^{D_2d/2} \leq (h^*)^{D_2d/2} (1 - 2^{-D_2d/2})^{-1}.$$

Moreover, for any  $h \geq h^-$ , we have  $v(h) \leq 2\sqrt{2v/(nh^d)}$  by definition of  $h^-$ . It follows that

$$v(h^*) \leq v(h) \leq 2v(h^*) \left(\frac{h^*}{h}\right)^{d/2}.$$

for any  $h \leq h^*$ . This enables us to bound the following sum

$$\begin{aligned} \sum_{\mathring{\mathcal{H}}(h^*)} h^{D_2d/2} v(h)^p &\leq 2^p v(h^*) (h^*)^{pd/2} \sum_{\mathring{\mathcal{H}}(h^*)} h^{D_2d/2 - pd/2} \\ &\leq \frac{2^p}{1 - 2^{-(p-D_2)d/2}} v(h^*) (h^*)^{D_2d/2} \end{aligned}$$

where we used that  $D_2 > p$ . Putting all these estimates together, using that  $h^*/h^+ \leq 1$  and  $\lambda(h^*) \geq 1$ , we eventually obtain

$$\mathcal{R}_{\mathcal{A}^c} \leq C_{\mathcal{A}^c} v(h^*)^p \lambda(h^*)^p \quad \text{with} \quad C_{\mathcal{A}^c} = \frac{2^p}{\sqrt{1-2^{-D_2 d}}} \left( \frac{D_1^p}{1-2^{-D_2 d/2}} + \frac{\sqrt{c_{2p}} 2^p}{1-2^{-(p-D_2)d/2}} \right).$$

In conclusion  $(\mathcal{R}_n^{(p)}[\widehat{f}(x), P])^p \leq (C_{\mathcal{A}} + C_{\mathcal{A}^c}) v(h^*)^p \lambda(h^*)^p$  which completes the proof.  $\square$

*Proof of Theorem 3.* Let  $P \in \Sigma_{\alpha, \beta}(x)$  and let  $\bar{h} = (\omega \log n / n)^{1/(2\gamma+d)}$  with  $\gamma = \alpha \wedge \beta$  and for some constant  $\omega$  to be specified later. By Proposition 5 we know that for  $n$  large enough (depending on  $\omega, \alpha, \beta, d$ ) such that  $\bar{h} \leq (r \wedge \tau)/4$ , we have  $|f_\eta(P, x) - f(x)| \leq C_1 \eta^\gamma$  for all  $\eta \leq \bar{h}$  with  $C_1$  depending on  $K, \ell, \alpha, \tau, L, \beta, r, b$  and  $R$ . Moreover, we also have

$$\frac{D_1^2 v(\bar{h})^2 \lambda(\bar{h})^2}{C_1^2 \bar{h}^{2\gamma}} \geq \frac{D_1^2 d D_2 2v \log(1/\bar{h})}{C_1^2 n \bar{h}^{2\gamma+d}} = \frac{D_1^2 d D_2 2v (2\gamma+d)^{-1} \log n - \log \log n - \log \omega}{C_1^2 \omega \log n}.$$

Thus, picking  $\omega = D_1^2 d D_2 v (2\gamma+d)^{-1} / C_1^2$  yields  $C_1 \bar{h}^\delta \leq D_1 v(\bar{h}) \lambda(\bar{h})$  for  $n$  large enough (depending on  $\omega$ ), and therefore  $\bar{h} \leq h^*(P, x)$ . By Lemma 7 this implies

$$\mathcal{R}_n^{(p)}[\widehat{f}, P](x) \leq C_2 v(\bar{h}) \lambda(\bar{h})$$

where  $C_2$  depends on  $p, D_1$  and  $D_2$ . But using that both  $\bar{h} \geq h^-$  and  $\lambda(\bar{h})^2 = d D_2 \log(1/\bar{h})$  for  $n$  large enough (depending on  $\omega, d, K$  and  $D_2$ ), we also obtain

$$v(\bar{h})^2 \lambda(\bar{h})^2 \leq \frac{8vdD_2 \log(1/\bar{h})}{n \bar{h}^d} = \frac{8vdD_2 (2\gamma+d)^{-1} \log n - \log \log n - \log \omega}{\omega \log n} \bar{h}^{2\gamma} \leq \frac{8C_1^2}{D_1^2} \bar{h}^{2\gamma}.$$

This last estimate yields

$$\mathcal{R}_n^{(p)}[\widehat{f}(x), P] \leq \left( \frac{4C_1 C_2}{D_1} \omega^{\gamma/(2\gamma+d)} \right) \left( \frac{\log n}{n} \right)^{\gamma/(2\gamma+d)}$$

for  $n$  large enough depending on  $\omega, \alpha, \beta, d, K$  and  $D_2$ , which completes the proof.  $\square$

## 6.4 Proofs of Section 3.5

*Proof of Proposition 6.* Let  $P \in \Sigma_{\alpha, \beta}^d(x)$  and  $\eta > 0$ . Assume that  $\widehat{P}_\eta > 0$ . We have

$$\begin{aligned} |\widehat{\delta}_\eta - d| &\leq \frac{1}{\log 2} |\log \widehat{P}_{2\eta} - \log \widehat{P}_{2\eta}| + \frac{1}{\log 2} |\log \widehat{P}_\eta - \log \widehat{P}_\eta| + \frac{1}{\log 2} |\log \widehat{P}_{2\eta} - \log P_\eta - d \log 2| \\ &\leq \frac{1}{\log 2} \left( \frac{|\widehat{P}_{2\eta} - P_{2\eta}|}{\widehat{P}_{2\eta} \wedge P_{2\eta}} + \frac{|\widehat{P}_\eta - P_\eta|}{\widehat{P}_\eta \wedge P_\eta} + |\log(P_{2\eta}/(2^d P_\eta))| \right) \end{aligned}$$

We first consider the determinist term. For  $\eta \leq \tau/2$ , we have, writing  $r_\eta = \xi(\eta/\tau)\eta$  and using Lemma 1,

$$L_{2\eta}(1 - \eta^2/6\tau^2)(2\eta)^2 \zeta_d \leq P_{2\eta} \leq U_{2\eta}(1 + r_{2\eta}^2/\tau^2) r_{2\eta}^d \zeta_d$$

and

$$L_\eta(1 - \eta^2/6\tau^2)\eta^d \zeta_d \leq P_\eta \leq U_\eta(1 + r_\eta^2/\tau^2) r_\eta^d \zeta_d,$$

where  $L_\eta = \inf_{B_D(x,\eta)} f$  and  $U_\eta = \sup_{B_D(x,\eta)} f$ . Using the fact that  $f \in \mathcal{F}_\beta$ , and by Lemma 5, we know that for  $\eta \leq r/4$ , there exists  $R_1 > 0$  (depending on  $\beta, r, b$  and  $R$  such that  $f(x) - R_1\eta \leq L_\eta \leq U_\eta \leq f(x) + R_1\eta$ ). We thus obtain

$$(f(x) - R_1 2\eta)(1 - \eta^2/6\tau^2)(2\eta)^d \zeta_d \leq P_{2\eta} \leq (f(x) + R_1 2\eta)(1 + r_{2\eta}^2/\tau^2)r_{2\eta}^d \zeta_d$$

and

$$(f(x) - R_1\eta)(1 - \eta^2/6\tau^2)\eta^d \zeta_d \leq P_\eta \leq (f(x) + R_1\eta)(1 + r_\eta^2/\tau^2)r_\eta^d \zeta_d. \quad (33)$$

Using these two inequalities, and the fact that  $r_\eta/\eta \rightarrow 1$  as  $\eta \rightarrow 0$ , we find that  $|P_{2\eta}/(2^d P_\eta) - 1| \lesssim \eta$  up to a constant that depends on  $R_1, \tau$  and  $a$ , for  $\eta$  small enough (depending on  $R_1, \tau$  and  $a$  as well). For the other terms, a simple use of Hoeffding's inequality yields for any  $\eta, \epsilon > 0$ ,

$$\mathbb{P}(|\widehat{P}_\eta - P_\eta| > \epsilon) \leq 2 \exp(-2n\epsilon^2).$$

On the event  $\mathcal{A}_\eta = \{|\widehat{P}_\eta - P_\eta| \leq \epsilon\}$ , we have moreover  $\widehat{P}_\eta \wedge P_\eta \geq P_\eta - \epsilon$ . Setting  $\epsilon = \eta^{d+1}$ , and using (33), we see that  $P_\eta - \epsilon \gtrsim \eta^d$  for  $\eta$  small enough (depending on  $R_1, \tau$  and  $a$ ). Thus, on the event  $\mathcal{A}_\eta \cap \mathcal{A}_{2\eta}$ , with probability at least  $1 - 4 \exp(-2n\eta^{2d+2})$ , we derive

$$|\widehat{\delta}_\eta - d| \lesssim \eta + \frac{\epsilon}{P_\eta - \epsilon} + \frac{\epsilon}{P_{2\eta} - \epsilon} \lesssim \eta,$$

for  $\eta$  small enough (depending on  $R_1, \tau$  and  $a$ ), up to a constant that depends on  $R_1, \tau$  and  $a$ , ending the proof.  $\square$

*Proof of Theorem 4.* By triangle inequality, for any  $P \in \Sigma_{\alpha,\beta}^d(x)$ , we write

$$\mathcal{R}_n^{(p)}[\widehat{f}_h(\widehat{d}; \cdot), P](x) \leq \left( \mathbb{E}_P \left[ |\widehat{f}_h(\widehat{d}; x) - f(x)|^p \mathbb{1}_{\{\widehat{d}=d\}} \right] \right)^{1/p} + \left( \mathbb{E}_P \left[ |\widehat{f}_h(\widehat{x}; \cdot) - f(x)|^p \mathbb{1}_{\{\widehat{d} \neq d\}} \right] \right)^{1/p}.$$

The first term in the right-hand side has the right order thanks to Theorem 3. For the second one, Cauchy-Schwarz inequality yields

$$\left( \mathbb{E}_P \left[ |\widehat{f}_h(\widehat{d}; x) - f(x)|^p \mathbb{1}_{\{\widehat{d} \neq d\}} \right] \right)^{1/p} \leq \mathbb{P}(\widehat{d} \neq d)^{1/2p} \left( \mathbb{E}_P \left[ |\widehat{f}_h(\widehat{d}; x) - f(x)|^{2p} \right] \right)^{1/2p}.$$

Using that  $|f(x)| \leq b$  and

$$|\widehat{f}_h(\widehat{d}, x)| \leq \sup_{1 \leq d < D} \frac{\|K^{(\ell)}(d, \cdot)\|_\infty}{(h_d^-)^d} \lesssim n$$

up to a constant that depend on  $D, K$  and  $\ell$ , we infer

$$\mathcal{R}_n^{(p)}[\widehat{f}_h(\widehat{d}, \cdot), P](x) \lesssim \left( \frac{\log n}{n} \right)^{\alpha \wedge \beta / (2\alpha \wedge \beta + d)} + n^{-3/2}(n + b)$$

for  $n$  large enough depending on  $p, D_1, D_2, K, \ell, \alpha, \tau, L, \beta, r, b$  and  $R$ , so that the result indeed holds up to a constant depending on  $p, D, D_1, D_2, K, \ell, \alpha, \tau, L, \beta, r, b$  and  $R$ .  $\square$

## 6.5 Proofs of Section 4

We write  $\widehat{V} = \{X_1, \dots, X_n\} \cup \{x\}$  for the vertices of  $\widehat{G}_\epsilon(x)$  and  $\widehat{\eta} = \sup_{p \in M} d(p, \widehat{V})$ . For small enough  $\widehat{\eta}$  we have that  $\widehat{G}_\epsilon(x)$  is connected, therefore the distance  $\widehat{d}_\epsilon$  is well-defined on  $\widehat{V}$ . We have in that case a good reverse control of  $d_M$  by  $\widehat{d}_\epsilon$ , as shown in the next two lemmata.

**Lemma 8.** *If  $\epsilon \leq 8\tau$  and  $16\widehat{\eta} \leq \epsilon$ , then  $\widehat{d}_\epsilon(p, q) \leq d_M(p, q)$  for any  $p, q \in \widehat{V}$ .*

**Lemma 9.** *If  $\epsilon \leq \tau/2$ , then  $d_M(p, q) \leq \widehat{d}_\epsilon(p, q) \left(1 + \frac{\pi^2}{48\tau^2} \epsilon^2\right)$  for any  $p, q \in \widehat{V}$ .*

*Proof of Lemma 8.* We can suppose without loss of generality that the shortest path between  $p$  and  $q$  is given by  $\gamma : [0, \ell] \rightarrow \mathbb{R}^D$  with  $\ell = d_M(p, q) \leq L_M/2$ . We let  $\delta = \ell/(4\lfloor \ell/\epsilon \rfloor)$  and  $N = 4\lfloor \ell/\epsilon \rfloor$ . Notice that  $\epsilon/4 \leq \delta \leq \epsilon/2$ . Let us define  $p_j = \gamma(j\delta)$ , so that  $p_0 = p$  and  $p_N = q$ . Since  $\widehat{\eta} \leq \epsilon/16$ , for every  $1 \leq j \leq N-1$ , there exists among our vertices  $\widehat{V}$  a point denoted by  $\widehat{p}_j$  such that  $\|p_j - \widehat{p}_j\| \leq \epsilon/16$ . We set  $\widehat{t}_j \in [0, L_M]$  for its coordinate, namely  $\widehat{p}_j = \gamma(\widehat{t}_j)$ .

Let us show first that for  $1 \leq j < N$ , we have  $\widehat{t}_j \in [0, \ell]$ . Indeed, thanks to Proposition 3, since  $\epsilon/16 \leq \tau/2$ , we have  $|t_j - \widehat{t}_j| \leq 2\|p_j - \widehat{p}_j\| \leq \epsilon/8$ . Since  $\delta \geq \epsilon/4$ , we thus have  $0 \leq \widehat{t}_1 \leq \dots \leq \widehat{t}_{N-1} \leq \ell$ . Furthermore, writing  $\widehat{p}_0 = p$  and  $\widehat{p}_N = q$ , we have

$$\|\widehat{p}_j - \widehat{p}_{j+1}\| \leq \|\widehat{p}_j - p_j\| + \|p_j - p_{j+1}\| + \|p_{j+1} - \widehat{p}_{j+1}\| \leq \epsilon$$

for any  $0 \leq j < \ell$ . The sequence  $\widehat{s} = (\widehat{p}_0, \dots, \widehat{p}_N)$  is thus a path in  $\widehat{G}_\epsilon(x)$  and so

$$\widehat{d}_\epsilon(p, q) \leq L_{\widehat{s}} = \|\widehat{p}_1 - \widehat{p}_0\| + \dots + \|\widehat{p}_N - \widehat{p}_{N-1}\| \leq \|\widehat{t}_1 - \widehat{t}_0\| + \dots + \|\widehat{t}_N - \widehat{t}_{N-1}\| = \widehat{t}_N - \widehat{t}_0$$

where we set  $\widehat{t}_0 = 0$  and  $\widehat{t}_N = \ell = d_M(p, q)$ , ending the proof.  $\square$

*Proof of Lemma 9.* Following the proof of (Arias-Castro and Le Gouic, 2019, Lem. 5) if there exists  $\delta > 0$  such that  $\|x - y\| \leq \delta$  implies  $d_M(x, y) \leq \pi\tau$  for all  $x, y \in M$ , then we must have that for any  $x, y \in M$  verifying  $\|x - y\| \leq \delta$ ,

$$d_M(x, y) \leq \|x - y\| \left(1 + \frac{\pi^2}{48\tau^2} \|x - y\|^2\right).$$

Thanks to Proposition 3, this must hold for  $\delta = \tau/2$ . Now let  $p_0, \dots, p_m$  be the shortest path in  $\widehat{G}_\epsilon(x)$  between  $p$  and  $q$ . Since  $\epsilon \leq \tau/2$ , we have

$$\begin{aligned} d_M(p, q) &\leq \sum_{j=1}^m d_M(p_j, p_{j-1}) \leq \sum_{j=1}^m \|p_j - p_{j-1}\| \left(1 + \frac{\pi^2}{48\tau^2} \|p_j - p_{j-1}\|^2\right) \\ &\leq \widehat{d}_\epsilon(p, q) \left(1 + \frac{\pi^2}{48\tau^2} \epsilon^2\right) \end{aligned}$$

which ends the proof.  $\square$

In view of Lemma 8 and Lemma 9, we want to tune  $\epsilon$  so that it is the smallest possible and so that  $16\widehat{\eta} \leq \epsilon$  holds with high probability. This is achieved for  $\epsilon$  of order  $\log n/n$ .

**Lemma 10.** *Setting  $\epsilon = \frac{32(p+1)\log n}{an}$ , for every  $n \geq 3$ , we have  $\mathbb{P}(16\widehat{\eta} \leq \epsilon) \geq 1 - 1/n^p$ .*

*Proof.* Let  $\delta > 0$ , and let  $N = \lfloor L_M/\delta \rfloor$ . We split  $[0, L_M]$  into  $N$  intervals  $I_1, \dots, I_N$  of length  $L_M/N$ . We denote  $\mathcal{A}$  the event for which each  $I_j$  contains at least one coordinate among those of the sample of observations  $(X_1, \dots, X_n)$ . On  $\mathcal{A}$ , we have  $\widehat{\eta} \leq L_M/N \leq 2\delta$ . Moreover,

$$\begin{aligned} \mathbb{P}(\mathcal{A}) &= 1 - \mathbb{P}(\exists j, \gamma(I_j) \text{ contains no observation}) \\ &\geq 1 - N \left( 1 - \min_{1 \leq j \leq N} P(\gamma(I_j)) \right)^n \geq 1 - N \left( 1 - \frac{aL_M}{N} \right)^n. \end{aligned}$$

Using that  $N \leq L_M/\delta$  and that  $L_M \leq 1/a$  we infer

$$\mathbb{P}(\widehat{\eta} \leq 2\delta) \geq 1 - \frac{1}{a\delta} (1 - a\delta)^n \geq 1 - \frac{e^{-a\delta n}}{a\delta}.$$

Setting  $\delta = \frac{(p+1)\log n}{an}$  and  $\epsilon = 32\delta$  yields

$$\mathbb{P}(16\widehat{\eta} \leq \epsilon) \geq 1 - \frac{n}{(p+1)n^{p+1}\log n} \geq 1 - \frac{1}{n^p}$$

as soon as  $\log n \geq 1$ , *i.e.* for  $n \geq 3$ . □

*Proof of Theorem 6.* We write  $K^{(\ell)}$  in short for  $K^{(\ell)}(1; \cdot)$ . Let  $\mathcal{A} = \{16\widehat{\eta} \leq \epsilon\}$ . By triangle inequality,  $\mathcal{R}_n^{(p)}[\widehat{f}_h^*, P] \leq \mathcal{R}_{\mathcal{A}} + \mathcal{R}_{\mathcal{A}^c}$ , with

$$\mathcal{R}_{\mathcal{A}} = \left( \mathbb{E}_P[|\widehat{f}_h^*(x) - f(x)|^p \mathbb{1}_{\mathcal{A}}] \right)^{1/p} \quad \text{and} \quad \mathcal{R}_{\mathcal{A}^c} = \left( \mathbb{E}_P[|\widehat{f}_h^*(x) - f(x)|^p \mathbb{1}_{\mathcal{A}^c}] \right)^{1/p}.$$

On  $\mathcal{A}$ , we have, for  $n$  large enough (depending on  $p, a$  and  $\tau$ ) such that  $\epsilon \leq \tau/2$  holds,  $|\widehat{d}_\epsilon(X_i, x) - d_M(X_i, x)| \leq C_1\epsilon^2$  with  $C_1$  depending on  $\tau$  only. This is inferred by Lemmas 8 and 9. We deduce that, on this event,

$$|\widehat{f}_h^*(x) - \widehat{g}_h^*(x)| \leq \frac{C_1 \|K^{(\ell)'}\|_\infty \epsilon^2}{h^2} \quad \text{with} \quad \widehat{g}_h^*(x) = \frac{1}{n} \sum_{i=1}^n K_h^{(\ell)}(d_M(X_i, x)).$$

It follows that

$$\mathcal{R}_{\mathcal{A}} \leq \frac{C_1 \|K^{(\ell)'}\|_\infty \epsilon^2}{h^2} + \left( \mathbb{E}_P[|\widehat{g}_h^*(x) - f(x)|^p] \right)^{1/p} \leq \frac{C_1 \|K^{(\ell)'}\|_\infty \epsilon^2}{h^2} + \left( \mathbb{E}_P[|\widehat{\xi}_h^*(P, x)|^p] \right)^{1/p} + |B_h^*(P, x)|$$

with  $B_h^*$  and  $\widehat{\xi}_h^*$  denoting the bias and stochastic deviation of estimator  $\widehat{g}_h^*(x)$ . Following the same arguments as in proof of Proposition 4, we have  $\mathbb{E}[|\widehat{\xi}_h^*(x)|^p] \leq c_p v(h)^p$  with  $c_p$  depending only on  $p$ . For the bias term, as soon as  $h \leq (\tau \wedge r)/4$ , we have

$$\begin{aligned} B_h^*(P, x) &= \mathbb{E}_P[\widehat{g}_h^*(x)] - f(x) = \int_M K_h^{(\ell)}(d(p, x)) f(p) d\mu_M(p) - f(x) \\ &= \int_{B_1(0,1)} K^{(\ell)}(\|v\|) (f \circ \exp_x(hv) - f(x)) dv. \end{aligned}$$

Since now  $f \circ \exp_x \in \mathcal{F}_\beta$  with  $\ell \geq \beta$ , we know that all the term in the development of  $B_h^*(P, x)$  up to order  $\lceil \beta - 1 \rceil$  cancels. We deduce  $|B_h^*(x)| \leq C_2 h^\beta$  with  $C_2$  depending on  $\ell$  and  $R$  only. For the other term  $\mathcal{R}_{\mathcal{A}^c}$ , we write  $|\widehat{g}_h^*(x) - f(x)| \leq \frac{\|K^{(\ell)}\|_\infty}{h} + b$ , so that, according to Lemma 10,

$$\mathcal{R}_{\mathcal{A}^c} \leq \left( \frac{\|K^{(\ell)}\|_\infty}{h} + b \right) \mathbb{P}(\mathcal{A}^c)^{1/p} \leq C_3 \frac{1}{nh}$$

with  $C_3$  depending on  $\ell$  and  $b$ . Putting all these estimates together yields the result. □

**Acknowledgements** *We are grateful to Krishnan (Ravi) Shankar for insightful discussions and comments.*

## References

- Aamari, E. (2017). *Vitesses de convergence en inférence géométrique*. Ph. D. thesis, Paris Saclay.
- Aamari, E., J. Kim, F. Chazal, B. Michel, A. Rinaldo, and L. Wasserman (2019). Estimating the reach of a manifold. *Electron. J. Stat.* *13*(1), 1359–1399.
- Aamari, E. and C. Levrard (2019). Nonasymptotic rates for manifold, tangent space and curvature estimation. *Ann. Statist.* *47*(1), 177–204.
- Alexander, S. B. and R. L. Bishop (2006). Gauss equation and injectivity radii for subspaces in spaces of curvature bounded above. *Geometriae Dedicata* *117*(1), 65–84.
- Arias-Castro, E. and T. Le Gouic (2019). Unconstrained and curvature-constrained shortest-path distances and their approximation. *Discrete & Computational Geometry* *62*(1), 1–28.
- Balakrishnan, S., A. Rinaldo, D. Sheehy, A. Singh, and L. Wasserman (2012). Minimax rates for homology inference. In *Artificial Intelligence and Statistics*, pp. 64–72.
- Bernstein, M., V. De Silva, J. C. Langford, and J. B. Tenenbaum (2000). Graph approximations to geodesics on embedded manifolds. Technical report, Technical report, Department of Psychology, Stanford University.
- Berry, T. and T. Sauer (2017). Density estimation on manifolds with boundary. *Computational Statistics & Data Analysis* *107*, 1–17.
- Besse, A. (1978). *Manifolds all of whose Geodesics are Closed*, Volume 93. Springer Science & Business Media.
- Boucheron, S., G. Lugosi, and P. Massart (2013). *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press.
- Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* *71*(2), 353–360.
- Chiu, S.-T. (1991). Bandwidth selection for kernel density estimation. *Ann. Statist.* *19*(4), 1883–1905.
- Cleanthous, G., A. Georgiadis, G. Kerkycharian, P. Petrushev, and D. Picard (2018). Kernel and wavelet density estimators on manifolds and more general metric spaces. *arXiv preprint arXiv:1805.04682*.
- Devroye, L. and L. Györfi (1985). *Nonparametric density estimation*. Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics. John Wiley & Sons, Inc., New York. The  $L_1$  view.
- Dubins, L. E. (1957). On curves of minimal length with a constraint on average curvature, and with prescribed initial and terminal positions and tangents. *American Journal of mathematics* *79*(3), 497–516.
- Eugeciouglu, Ö. and A. Srinivasan (2000). Efficient nonparametric density estimation on the sphere with applications in fluid mechanics. *SIAM Journal on Scientific Computing* *22*(1), 152–176.
- Evans, L. C. and R. F. Gariepy (1992). *Measure theory and fine properties of functions*. Chapman and Hall/CRC.
- Farahmand, A. M., C. Szepesvári, and J.-Y. Audibert (2007). Manifold-adaptive dimension estimation. In *Proceedings of the 24th international conference on Machine learning*, pp. 265–272. ACM.
- Federer, H. (1959). Curvature measures. *Transactions of the American Mathematical Society* *93*(3), 418–491.
- Federer, H. (1969). *Geometric measure theory*. Springer.
- Fefferman, C., S. Mitter, and H. Narayanan (2016). Testing the manifold hypothesis. *Journal of the American Mathematical Society* *29*(4), 983–1049.

- Gaïffas, S. and G. Lecué (2007). Optimal rates and adaptation in the single-index model using aggregation. *Electron. J. Stat.* 1, 538–573.
- Gallot, S., D. Hulin, and J. Lafontaine (2004). *Riemannian geometry*, Volume 3. Springer.
- Genovese, C., M. Perone-Pacifco, I. Verdinelli, and L. Wasserman (2012). Minimax manifold estimation. *Journal of machine learning research* 13(May), 1263–1291.
- Goldenshluger, A. and O. Lepski (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli* 14(4), 1150–1190.
- Goldenshluger, A. and O. Lepski (2014). On adaptive minimax density estimation on  $R^d$ . *Probab. Theory Related Fields* 159(3-4), 479–543.
- Goldenshluger, A., O. Lepski, et al. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics* 39(3), 1608–1632.
- Hall, P., G. Watson, and J. Cabrera (1987). Kernel density estimation with spherical data. *Biometrika* 74(4), 751–762.
- Hartman, P. (1951). On geodesic coordinates. *American Journal of Mathematics* 73(4), 949–954.
- Hendriks, H. (1990). Nonparametric estimation of a probability density on a Riemannian manifold using Fourier expansions. *The Annals of Statistics*, 832–849.
- Henry, G. S., A. L. Muñoz, and D. A. Rodriguez (2013). Locally adaptative density estimation on Riemannian manifolds.
- Hoffmann, M. and O. Lepski (2002). Random rates in anisotropic regression. *Ann. Statist.* 30(2), 325–396. With discussions and a rejoinder by the authors.
- Kerkyacharian, G., R. Nickl, and D. Picard (2012). Concentration inequalities and confidence bands for needlet density estimators on compact homogeneous manifolds. *Probability Theory and Related Fields* 153(1-2), 363–404.
- Kerkyacharian, G., T. M. Pham Ngoc, and D. Picard (2011). Localized spherical deconvolution. *Ann. Statist.* 39(2), 1042–1068.
- Kim, J., A. Rinaldo, and L. Wasserman (2016). Minimax rates for estimating the dimension of a manifold. *arXiv preprint arXiv:1605.01011*.
- Kim, P. T. and J.-Y. Koo (2002). Optimal spherical deconvolution. *J. Multivariate Anal.* 80(1), 21–42.
- Kim, P. T., J.-Y. Koo, and Z.-M. Luo (2009). Weyl eigenvalue asymptotics and sharp adaptation on vector bundles. *J. Multivariate Anal.* 100(9), 1962–1978.
- Lee, J. M. (2006). *Riemannian manifolds: an introduction to curvature*, Volume 176. Springer Science & Business Media.
- Lepski, O. and N. Serdyukova (2014). Adaptive estimation under single-index constraint in a regression model. *Ann. Statist.* 42(1), 1–28.
- Lepski, O. V., E. Mammen, V. G. Spokoiny, et al. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *The Annals of Statistics* 25(3), 929–947.
- Lepskii, O. (1992). Asymptotically minimax adaptive estimation. i: Upper bounds. optimally adaptive estimates. *Theory of Probability & Its Applications* 36(4), 682–697.
- Lepskii, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen.* 35(3), 459–470.
- Low, M. G. (1992). Nonexistence of an adaptive estimator for the value of an unknown probability density. *Ann. Statist.* 20(1), 598–602.
- Nikol’skii, S. M. (2012). *Approximation of functions of several variables and imbedding theorems*, Volume 205. Springer Science & Business Media.
- Niyogi, P., S. Smale, and S. Weinberger (2008). Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry* 39(1-3), 419–441.

- Ozakin, A. and A. G. Gray (2009). Submanifold density estimation. In *Advances in Neural Information Processing Systems*, pp. 1375–1382.
- Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* 33, 1065–1076.
- Pelletier, B. (2005). Kernel density estimation on Riemannian manifolds. *Statistics & probability letters* 73(3), 297–304.
- Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.* 9(2), 65–78.
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.
- Silverman, B. W. (1998). *Density estimation for statistics and data analysis*. Routledge.
- Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* 10(4), 1040–1053.
- Tenenbaum, J. B., V. De Silva, and J. C. Langford (2000). A global geometric framework for nonlinear dimensionality reduction. *science* 290(5500), 2319–2323.
- Triebel, H. (1987). Characterizations of function spaces on a complete Riemannian manifold with bounded geometry. *Mathematische Nachrichten* 130(1), 321–346.
- Tsybakov, A. (2008). *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer New York.
- Van der Vaart, A. W. (2000). *Asymptotic statistics*, Volume 3. Cambridge university press.
- Watson, G. S. and E. J. Williams (1956). On the construction of significance tests on the circle and the sphere. *Biometrika* 43, 344–352.
- Yu, B. (1997). Assouad, Fano, and Le Cam. In *Festschrift for Lucien Le Cam*, pp. 423–435. Springer.