

Level set and density estimation on manifolds

Alejandro Cholaquidis *

Centro de Matemática, Facultad de Ciencias,
Universidad de la República, Uruguay.

Ricardo Fraiman

Centro de Matemática, Facultad de Ciencias,
Universidad de la República, Uruguay.

and

Leonardo Moreno

Departamento de Métodos Cuantitativos, FCEA,
Universidad de la República, Uruguay.

March 30, 2021

Abstract

We tackle the problem of the estimation of the level sets $L_f(\lambda)$ of the density f of a random vector X supported on a smooth manifold $M \subset \mathbb{R}^d$, from an iid sample of X . To do that we introduce a kernel-based estimator $\hat{f}_{n,h}$, which is a slightly modified version of the one proposed in [45], and proves its a.s. uniform convergence to f . Then, we propose two estimators of $L_f(\lambda)$, the first one is a plug-in: $L_{\hat{f}_{n,h}}(\lambda)$, which is proven to be a.s. consistent in Hausdorff distance and distance in measure, if $L_f(\lambda)$ does not meet the boundary of M . While the second one assumes that $L_f(\lambda)$ is r -convex, and is estimated by means of the r -convex hull of $L_{\hat{f}_{n,h}}(\lambda)$. The performance of our proposal is illustrated through some simulated examples. In a real data example we analyze the intensity and direction of strong and moderate winds.

1 Introduction

The statistics of functional data had an exponential growth in the last decades, being particularly important the impulse given by the early works [41] and [42]. Very different problems have been addressed in recent years, see for instance the reviews by [8], [21] [1]. Initially the results were concentrated on problems where the data took values in functional spaces, such as L^2 , L^1 and L^∞ , and more generally on Hilbert or Banach spaces. Recently a lot of interest appears for data in more general spaces (where there is no linear structure on them), like metric spaces. This is motivated by important practical applications, see

*acholaquidis@cmat.edu.uy

for instance subsection 8.1. Some relevant examples includes the case of high dimensional data (see for instance [22], [14]), directional data (see [32], [31]), cone and cylindric data (see example on subsection 8.1), and random graph data (see [19]), among others.

In what follows we will consider the case where the data take values on a Riemannian manifold.

Starting from the pioneer 1945 work of Rao (see [43]), the statistical theory for data valued on a Riemannian manifold has received a lot of interest because of its important applications. In particular, these techniques may allow to avoid the curse of dimensionality when trying to analyze data in a high dimensional ambient space. Indeed, as mentioned in [24]: “*Data belonging to some m -dimensional compact submanifold M of Euclidean space \mathbb{R}^s appear in many areas of natural science. Directional statistics, image analysis, vector cardiography in medicine, orientational statistics, plate tectonics, astronomy and shape analysis comprise a (by no means exhaustive) list of examples*”. These techniques are also applied in medical imaging applications: as it mentioned in [39] “*Examples of manifolds we routinely use in medical imaging applications are 3D rotations, 3D rigid transformations, frames (a 3D point and an orthonormal trihedron), semi- or non-oriented frames [...], positive definite symmetric matrices coming from diffusion tensor imaging*”.

The estimation of level sets $L_f(\lambda) = \{x : f(x) \geq \lambda\}$, where f is an unknown density function on \mathbb{R}^d and $\lambda > 0$ is a given constant, has been considered by many authors; see, for instance, [23], [40], [10], [33] [47], [49] for consistency results and rates of convergence, while the asymptotic distribution was derived in [6]. Some relevant applications include mode estimation [34], [40], clustering ([11], [12]) or detection of abnormal behaviour in a system ([15], [3], [2]). However, this problem is less developed when the underlying density has its support on a Riemannian manifold. In the following we address the problem of level set estimation in this setup.

More precisely, given a d' -dimensional Riemannian manifold $M \subset \mathbb{R}^d$, where M is unknown but $d' \leq d$ is assumed to be known, the aim is to estimate $L_f(\lambda)$ of the density f of a random vector X with support M from an iid sample X_1, \dots, X_n of X . In practice and for large sample sizes, if the intrinsic dimension of M is unknown it can be first estimated by means of, for instance, the classical estimator proposed in [30].

Our first proposed estimator is just the plug-in estimator $L_{\hat{f}_{n,h}}(\lambda)$, where $\hat{f}_{n,h}$ is a kernel-based estimator of f with bandwidth $h = h_n \rightarrow 0$, which is a slightly modified version of the one proposed in [45]. The almost sure (a.s.) consistency of $L_{\hat{f}_{n,h}}(\lambda)$ requires to prove the a.s. uniform convergence of the sequence of estimators $\hat{f}_{n,h}$ to f , this is done in Section 4. Regarding density estimation on manifolds, in [38] L^2 -consistency is obtained for a kernel-based density estimator (w.r.t. the L^2 norm in M , see Theorem 3.1). There are two main drawbacks to use that result to our setup, first in [38] it is assumed that the Riemannian structure in M is known, and second, L^2 consistency does not imply convergence of level sets. For the same estimator, [25] obtain the limit distribution and the a.s. consistency for the uniform metric. In [28] a different kernel-based estimator is proposed (similar to the one we will propose), and convergence in probability is obtained for the uniform metric. It does not assume that the underlying Riemannian structure is known. All the aforementioned

results are for manifolds without boundary. Other references that tackle the density estimation problem for manifolds without boundary are [35] and [29]. For manifolds with boundary point-wise L^2 consistency is obtained in [5]. Lastly, in section 6 we tackle the level set estimation problem but imposing a well known-shape restriction called r -convexity. In this case the proposed estimator is the r -convex hull of $L_{\hat{f}_{n,h}}$.

2 Roadmap

In Section 3 we introduce the basic notation and the geometric framework used throughout the manuscript. Section 4 is devoted to prove the a.s. uniform convergence of $\hat{f}_{n,h}$ to f . In Section 5 we prove that if $L_f(\lambda)$ does not meet the boundary of M , $L_{\hat{f}_{n,h}}(\lambda)$ converges a.s. in Hausdorff distance as well as in distance in measure to $L_f(\lambda)$. We also prove the convergence of its boundaries, i.e: $\partial L_{\hat{f}_{n,h}}(\lambda)$ converges in Hausdorff distance to $\partial L_f(\lambda)$. If $L_f(\lambda)$ meets the boundary of M , we prove that $L_{\hat{f}_{n,h}}(\lambda)$ converges a.s. in Hausdorff distance to $L_f(\lambda)$. Consistency in the Hausdorff metric of level sets under r -convexity is shown in Section 6. In Section 7 we provide some simulation results, while in Section 8 we consider an important application to describe the wind behavior in Uruguay. All proofs are given in the appendix.

3 Notation and geometric framework

If $B \subset \mathbb{R}^d$ is a Borel set, then we denote by $|B|$ its Lebesgue measure and by \bar{B} its closure. Given a set A on a topological space, the interior of A with respect to the underlying topology is denoted by $\overset{\circ}{A}$. The k -dimensional closed ball of radius ε centered at x will be denoted by $\mathcal{B}_k(x, \varepsilon) \subset \mathbb{R}^d$ (when $k = d$ the index will be omitted), and its Lebesgue measure is denoted by $\sigma_k = |\mathcal{B}_k(x, 1)|$. The Euclidean inner product in \mathbb{R}^d is denoted by $\langle \cdot, \cdot \rangle$, while the corresponding norm in \mathbb{R}^d is denoted by $\| \cdot \|$.

From now on, we assume that $M \subset \mathbb{R}^d$ is a compact d' -dimensional manifold of class \mathcal{C}^2 (also called a d' -regular surface of class \mathcal{C}^2). We consider the Riemannian metric on M inherited from \mathbb{R}^d . If $x \in M$, $T_x M$ denotes the tangent space at x , while $\rho(x, y)$ denotes the geodesic distance between x and y . Given a set $A \subset M$, we denote $B_\rho(A, r) = \{x \in M : \rho(x, A) < r\}$. For $f : M \rightarrow \mathbb{R}$, we denote by $\nabla f(x)$ the gradient of f at $x \in M$. When M is orientable, it has a unique associated volume form ω such that $\omega(e_1, \dots, e_{d'}) = 1$ for all oriented orthonormal bases $e_1, \dots, e_{d'}$ of $T_x M$. If $g : M \rightarrow \mathbb{R}$ is a density function, then we can define a new measure $\mu(B) = \int_B g d\omega$, where $B \subset M$ is a Borel set. In what follows we assume that M is orientable. Given a point $x \in M$, b_x is the geodesic distance from x to the boundary ∂M of M , or is ∞ if $\partial M = \emptyset$.

Recall that given two non-empty compact sets $A, C \subset \mathbb{R}^d$, the Hausdorff distance between A and C is defined as

$$d_H(A, C) = \max \left\{ \max_{a \in A} \rho(a, C), \max_{c \in C} \rho(c, A) \right\}, \text{ where } \rho(a, C) = \inf_{c \in C} \rho(a, c). \quad (3.1)$$

Given two Borel sets $A, B \subset M$, the distance in measure between them is $d_\mu(A, B) = \mu(A \setminus B) + \mu(B \setminus A)$.

4 Density estimation

The aim of this section is to prove that $\hat{f}_{h,n}$, a modified version of the kernel-based density estimator, denoted by $f_{h,n}$, proposed in [5], converges uniformly to the density f when the manifold has a \mathcal{C}^2 boundary. This auxiliary result, besides the interest in itself, will be used to prove our main results regarding level set estimation in the next section. Let us recall the definition of $f_{h,n}$. We assume that K is the Gaussian kernel, (however, it can be replaced by any sub Gaussian distribution, see Remark 2 below); that is, $K(\|x\|) = \pi^{-d/2} \exp(-\|x\|^2)$. Let $h = h_n \rightarrow 0$; then,

$$f_{h,n}(x) = \frac{1}{nm_0(x)h^{d'}} \sum_{i=1}^n K\left(\frac{\|x - X_i\|}{h}\right) \text{ where } m_0(x) = \pi^{-1/2} \int_{-\infty}^{b_x/h} \exp(-z^2) dz, \quad (4.1)$$

where b_x is the distance to ∂M of the point $x \in M$. An important assumption in [5], requires that the manifold M be ‘‘uniformly tangible’’, which, roughly speaking, allows to define the projection onto the boundary of points close enough to it. More precisely:

Definition 4.1. *A d' -dimensional Riemannian manifold $M \subset \mathbb{R}^d$ is said to be uniformly tangible if*

1. *There exists $\delta > 0$ and $c > 0$ such that $R(x, y) := \|x - y\|/\rho(x, y) > c$ for all $x, y \in M$, such that $\|x - y\| < \delta$.*
2. *There exists $r > 0$ such that $N_r = \partial M \times [0, r)$ is mapped diffeomorphically onto its image via the exponential map $(x, t) \rightarrow \exp_x(tv_x)$ where $x \in \partial M$, $t \in [0, r)$, and v_x is the inward pointing unit normal vector to the boundary.*
3. *Denote by $\text{inj}(x)$ the injectivity radius of a point x (i.e the maximum radius for which $B(0, r) \subset T_x M$ is mapped diffeomorphically into M by the exponential map). Then $\inf\{\text{inj}(x) : x \notin \exp(N_r)\} > 0$.*

The following proposition states that this condition holds when M is a compact \mathcal{C}^2 manifold whose boundary (in case there exists) is also a \mathcal{C}^2 manifold. Under this condition there exists a radius $r_M > 0$ such that for any point x within a uniform geodesic distance r_M to the boundary, there exists a unique closest point on ∂M . Then it can be defined η_x the unit vector pointing in the direction of the unique closest boundary point. For points farther away than r_M from the boundary, η_x can be chosen arbitrarily.

Proposition 1. *Let $M \subset \mathbb{R}^d$ be a d' -dimensional \mathcal{C}^2 compact Riemannian manifold, whose boundary, in case there exists, is a \mathcal{C}^2 manifold. Then, M is uniformly tangible.*

Equation (5) in [5] states that, if M is uniformly tangible, and h is small enough the bias of $f_{h,n}(x)$ is

$$E(f_{h,n}(x)) - f(x) = hm_1(x)\langle \eta_x, \nabla f(x) \rangle + \mathcal{O}_x(h^2) \text{ where } m_1(x) = \frac{1}{2\sqrt{\pi}} \exp(-b_x^2/h^2). \quad (4.2)$$

Observe that if λ is such that $L_f(\lambda) \cap \partial M \neq \emptyset$, then if $x \in L_f$, $m_0(x) \rightarrow 1$. This suggest to use the following estimator:

$$\hat{f}_{h,n}(x) = \frac{1}{nh^{d'}} \sum_{i=1}^n K\left(\frac{\|x - X_i\|}{h}\right), \quad (4.3)$$

which does not depend on $m_0(x)$ or b_x .

In order to get the a.s. convergence of $L_{\hat{f}_{h,n}}(\lambda)$ to $L_f(\lambda)$ we need to prove the following auxiliary result that states the a.s. uniform convergence of $\hat{f}_{h,n}$ to f .

Theorem 4.1. *Under the hypotheses of Proposition 1. Let X be a random vector with support M whose density f is assumed to be \mathcal{C}^4 . Let $h \rightarrow 0$ such that $nh^{d'+3}/\log(n) \rightarrow \infty$ when $n \rightarrow +\infty$; then,*

$$\sup_{x \in M_n} |\hat{f}_{h,n}(x) - f(x)| = o(h/c_n) \quad a.s.$$

for any sequence of closed subsets $M_n \subset M$ such that $c_n/h \rightarrow \infty$, where $c_n = \inf_{x \in M_n} \rho(x, \partial M) > 0$.

The following result is more restrictive, but a better rate of convergence is obtained, (it holds in particular when $\partial M = \emptyset$) since the theorem does not allow the compact set M_0 to depend on n . This is proven in the same manner as Theorem 4.1,

Theorem 4.2. *Under the hypotheses of Theorem 4.1. Let $h \rightarrow 0$ and $\beta_n \rightarrow \infty$ such that $\beta_n h^2 \rightarrow 0$, $nh^{d'}/(\beta_n^2 \log(n)) \rightarrow \infty$; then,*

$$\beta_n \sup_{x \in M_0} |\hat{f}_{h,n}(x) - f(x)| \rightarrow 0 \quad a.s.$$

for any closed subset $M_0 \subset M$ such that $\inf_{x \in M_0} \rho(x, \partial M) > 0$.

Remark 1. *As a consequence, taking $\beta_n = n^\alpha$ and $h = n^{-\gamma}$ we derive that we can reach $\beta_n = n^\alpha$ for any $\alpha < 2/(d' + 4)$.*

Remark 2. *The explicit expression for m_0 and m_1 given in (4.1) and (4.2) respectively, strongly rely on the gaussianity of the kernel. If this assumption is removed, the expressions are much more involved, see [5]. Observe that the proposed estimator (4.3) does not require the computation of m_0 or m_1 , but the proofs of Theorems 4.1 and 4.2 uses them as auxiliary tools. However, the same rate of convergence is obtained in Theorems 4.1 and 4.2 if the gaussian kernel is replaced by any Lipschitz kernel fulfilling $K(\|z\|) \leq C_1 \exp(-C_2\|z\|^2)$, with a subgaussian density function, i.e. fulfilling $\mathbb{P}(\|X\| > t) \leq C_1 \exp(-C_2\|t\|^2)$, for all t and, for some positive constants C_1, C_2 , where X is a d' -dimensional vector with density $K(\|z\|)$.*

5 Level set estimation

Level set estimation is an important problem with many applications in statistics, such as in hierarchical clustering, binary classification, outliers detection, functional neuroimaging, and bioinformatics among many others. References are given in the introduction. In our setup we consider this problem when the

distribution of the data is supported on a smooth Riemannian manifold $M \subset \mathbb{R}^d$ which as mentioned in the introduction covers important applications that includes directional data, cone and cylindrical data, high dimensional data and random graph data. In what follows we state our main asymptotic results.

Once we prove the a.s. uniform convergence of the estimator $\hat{f}_{n,h}$ to f we are ready to state our main results. The first one (Theorem 5.1) tackle the case in which $L_f(\lambda)$ does not meet the boundary of the manifold. For this case we obtained not only the a.s. convergence in Hausdorff distance of the level set estimator $L_{\hat{f}_{n,h}}(\lambda)$, but also the convergence of its boundary, as well as the convergence in measure. As usual in level set estimation, we require that the boundary of $\partial L_f(\lambda)$, which is $\{f = \lambda\}$ because we assume that f is a continuous function, does not contain a plateau at level λ , see the discussion on condition f1) in Theorem 1 in [9]. The proof is based on Theorem 4.2.

Theorem 5.1. *Let M and f as in Theorem 4.1. Assume that the level $\lambda > 0$ fulfills that for all x such that $f(x) = \lambda$, there exists $a_n, b_n \rightarrow x$ such that $f(a_n) > \lambda$ and $f(b_n) < \lambda$ and $\partial L_f(\lambda) \cap \partial M = \emptyset$. Then, with probability one,*

1. $d_H(\partial L_{\hat{f}_{n,h}}(\lambda), \partial L_f(\lambda)) \rightarrow 0$;
2. $d_H(L_{\hat{f}_{n,h}}(\lambda), L_f(\lambda)) \rightarrow 0$;
3. *If, moreover, $\nabla_x f \neq 0$ for all x such that $f(x) = \lambda$, $d_\mu(L_{\hat{f}_{n,h}}(\lambda), L_f(\lambda)) \rightarrow 0$.*

If the underlying level set $L_f(\lambda)$ meets the boundary of the manifold, then we have the following result, whose proof is based on Theorem 4.1.

Theorem 5.2. *Let M and f be as in Theorem 4.1. Assume that the level $\lambda > 0$ fulfills that for all x with $f(x) = \lambda$, there exists $a_j \rightarrow x$, $a_j \in M$, such that $f(a_j) > \lambda$ for all j . Then,*

$$d_H(L_{\hat{f}_{n,h}}(\lambda), L_f(\lambda)) \rightarrow 0, \quad a.s.$$

6 Manifold level set estimation under r -convexity

In what follows we consider the level set estimation problem when we assume that the level set is r -convex set in the manifold.

In Euclidean space, a set A is said to be r -convex (for some $r > 0$) if $A = C_r(A)$, where $C_r(A)$ is the r -convex hull of A ; that is, the intersection of the complements of all open balls of radius r that does not meet A . This is a natural generalization of convexity (the half spaces are replaced by balls), and it has been widely studied in set estimation literature (see, for instance, [49, 48] [44] and [36]). Additionally, as is pointed out in [44], this concept “is closely related to the notion of alpha-shapes that arises in the literature of computational geometry”; see [18]. Departing from the idea of r -convexity, several generalizations have been given (see, for instance, [7]). If the underlying space is not Euclidean space but is rather any Riemannian manifold M endowed with the geodesic distance ρ , then the natural generalization is to replace the Euclidean balls with geodesic balls. According to this idea, given $r > 0$, we will say that a set $A \subset M$ is r -convex if it is equal to its r -convex hull in M , that

is, the intersection of the complement of all open geodesic balls of radius r that does not meet A .

Theorem 6.1. *Under the hypotheses of Theorem 5.2, assume also that the level sets $L_f(\lambda)$ is r -convex and $L_{\hat{f}_{n,h}}(\lambda)$ is r -convex a.s., for some $r > 0$. Then,*

$$d_H(C_r(\{X_i : \hat{f}_{h,n}(X_i) > \lambda\}), C_r(\{X_i : f(X_i) > \lambda\})) \rightarrow 0, \quad \text{a.s.}$$

and

$$d_H(C_r(\{X_i : \hat{f}_{h,n}(X_i) > \lambda\}), L_f(\lambda)) \rightarrow 0 \quad \text{a.s.}$$

7 Simulation results

To assess the performance of our proposal, we will perform a simulation example with two scenarios. In the first one, we consider a distribution on the positive cone of covariance 2×2 -matrices, which is a three dimensional manifold when endowed with the Riemannian structure given below. In the second one, we compare our density estimator with the one proposed in [26], which is specially designed for spherical data. Also, as illustrative examples, we consider the torus with the metric inherit from R^3 and the two dimensional half-sphere in \mathbb{R}^3 . In the first case we consider two distributions: the first is unimodal and the second is a mixture of distributions. In the last case we considered a Von-Mises distribution.

7.1 Positive-definite matrices

Let us denote by (\mathbb{P}_d, g) the set of positive-definite $d \times d$ -covariance matrices. Given two matrices $A, B \in \mathbb{P}_d$, the geodesic curve joining A and B is

$$\gamma(s) = A^{1/2}(A^{-1/2}BA^{-1/2})^s A^{1/2} \quad \text{for all } s \in [0, 1].$$

The geodesic distance is given by $d_g(A, B) = \|\ln(A^{-1/2}BA^{-1/2})\|$, where $\|\cdot\|$ is the Hilbert–Schmidt norm.

We consider, for $d = 2$, the Wishart distribution $\mathcal{W}_2(\Sigma, m)$ on \mathbb{P}_2 with parameters $m = 10$ and $\Sigma = (1/4)I_2$. An easy way to obtain a matrix S with this distribution is to define $S = X_1X_1' + \dots + X_mX_m'$, where X_1, \dots, X_m is an iid random sample of a multivariate Gaussian distribution with mean 0 and covariance matrix Σ .

As is well-known, (\mathbb{P}_2, g) can be represented as a cone in \mathbb{R}^3 . In Figure 1, we show the projections of a sample of size 1000, drawn from a Wishart distribution with $m = 10$ and $\Sigma = (1/4)I$, together with the convex hull of the λ level set $L_{\mathcal{W}}(\lambda)$ (in blue) and the convex hull of the level set estimator $L_{\hat{\mathcal{W}}_{n,h}}(\lambda)$ (in red) for $\lambda = 0.03$ and $h = 0.55$. The Hausdorff distance between the level sets in this case is \mathbb{R}^3 is 0.325. In Table 1, we report the mean, median, and standard deviation, over 100 replications of the Hausdorff distance (d_H) between both sets for different sample sizes $n \in \{1000, 5000, 10000, 20000\}$. The parameter h is chosen following the proposal in Appendix B of [5].

n	h	d_H	
1000	0.55	0.3388	[0.3286] (0.0616)
5000	0.4	0.2326	[0.2313] (0.0266)
10000	0.35	0.2114	[0.2110] (0.0267)
20000	0.3	0.1909	[0.1872] (0.0235)

Table 1: Mean over 100 replications for the Hausdorff distance between the true level set $L_{\mathcal{W}}(\lambda)$ and the estimator $L_{\hat{\mathcal{W}}_{n,h}}(\lambda)$ for $\lambda = 0.03$. Between [] and () the median and standard deviation respectively.

7.2 The torus

In the torus $\mathbb{T}^2 = S^1 \times S^1$, we consider the multivariate von Mises distribution, denoted by $\mathcal{MVM}(\mu, \kappa, \Delta)$. The density at $\theta \in \mathbb{T}$ is given by

$$f(\theta; \mu, \kappa, \Delta) = \frac{1}{Z(\kappa, \Delta)} \exp\{\kappa^\top c(\theta) + s(\theta)\Delta s(\theta)/2\},$$

where $\mu \in \mathbb{T}^2$ (this parameter is called mean), $\kappa \geq 0 \in \mathbb{R}^d$ (concentration parameter), $\Delta = (\lambda_{i,j})$ is a symmetric matrix on $\mathbb{R}^{d \times d}$ with null diagonal entries ($\lambda_{i,i} = 0$ for all $i \in \{1, \dots, d\}$), and $Z(\kappa, \Delta)$ is a normalization constant. The functions c_i and s_i are defined by $c_i(\theta) = \cos(\theta_i - \mu_i)$ and $s_i(\theta) = \sin(\theta_i - \mu_i)$ for all $i \in \{1, \dots, d\}$. In Figure 2 (left-hand panel), we show (in yellow) a sample of size 2000 from a $\mathcal{MVM}_1(\mu_1, \kappa_1, \Delta_1)$ distribution with

$$\mu_1 = (\pi/2, 0), \quad \kappa_1 = (20, 20), \quad \Delta_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad (7.1)$$

In the right panel of Figure 2, we show (in yellow) a sample of size 2000 from a mixture law given by

$$0.4\mathcal{MVM}_1(\mu_2, \kappa_1, \Delta_1) + 0.6\mathcal{MVM}_2(\mu_3, \kappa_1, \Delta_1), \quad (7.2)$$

where $\mu_2 = (\pi/2, 0)$ y $\mu_3 = (\pi/2, \pi/4)$. In all cases, we consider $\lambda = 0.8$ and bandwidth $h = 0.2$. The boundary of the theoretical level set is shown in red, while the boundary of the estimator is shown in magenta.

The Hausdorff distances between the theoretical curve and the estimated curve are 0.066 and 0.107.

7.3 The half-sphere

Finally, we considered the sphere $S^2 \subset \mathbb{R}^3$ endowed with the Riemannian metric inherited from \mathbb{R}^3 . The sample is drawn from a the mixture of two von Mises-Fisher distributions given by

$$f(x, \mu, \kappa) = C(\kappa)e^{\kappa^\top x} \mathbb{1}_{S^2}(x), \quad (7.3)$$

where $\kappa \geq 0$ and $\mu \in S^2$ are the concentration and directional mean parameters, respectively. $C(\kappa)$ is the normalizing constant; see [31].

The mixture is given by,

$$f_0(\cdot) = 0.5f(\cdot, v_1/\|v_1\|, 40) + 0.5f(\cdot, v_2/\|v_2\|, 40). \quad (7.4)$$

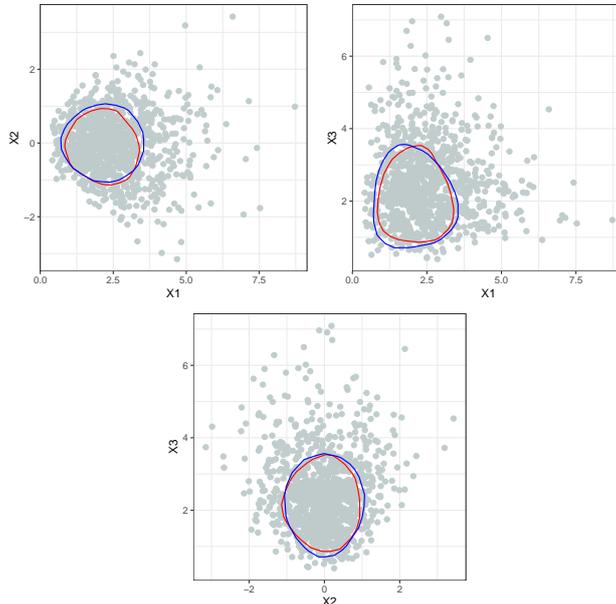


Figure 1: Projections of a sample of size 1000 drawn from a Wishart distribution with $m = 10$ and $\Sigma = (1/4)I$, together with the convex hull of the λ level set (in blue) and the convex hull of the level set estimator (in red), for $\lambda = 0.03$ and $h = 0.55$.

with $v_1 = (-1, -0.3, 0.2)$ and $v_2 = (-1, 0.3, 0.2)$. Let $X = (X_1, X_2, X_3) \in S^2$ with distribution f_0 , we consider the truncated random vector $X^{\text{tr}} = X \mathbb{I}_{S_+^2}$, where $S_+^2 = S^2 \cap \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_3 \geq 0\}$.

In Figure 3, we show (left-hand panel) a sample of size $n = 1000$ on S_+^2 of X^{tr} , together with the estimated level set (in red) and the true level set (in blue) at $\lambda = 1.1$. In the right-hand panel, we show the stereographic projections of the sample and the estimators. The Hausdorff distance between the theoretical curve and the estimated (on the stereographic projections) curve is 0.019.

7.4 The two dimensional sphere

To assess the performance of (4.3) and compare it with the density estimator proposed in [26] (which will be denoted by \hat{f}_n), specially designed for spherical data, we considered a Von-Mises distribution in S^2 , whose density is given by (7.3), with parameters $\kappa = 40$ and $\mu = (0, 0, 1)$. Four sample sizes were considered: $n \in \{500, 1000, 5000, 10000\}$ and the whole procedure was repeated 100 times. For each replication the bandwidth h for our estimator was selected as the one minimizing $\|\hat{f}_n - f\|_\infty := \sup_{x \in S^2} |\hat{f}_n(x) - f(x)|$ on a grid of values, whereas the parameter κ in [26], was selected as the one minimizing $\|\tilde{f}_n - f\|_\infty$ on a grid of values. The mean over 100 replications is shown in Table 2, while between \square and $()$ we report the median and standard deviation respectively. The results are very encouraging. The performances of both estimators are very similar, with a slightly better behaviour of \tilde{f}_n for $n = 500, 1000$ while for

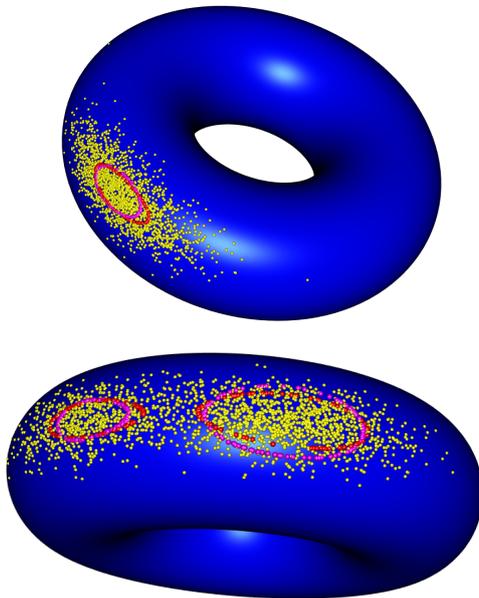


Figure 2: Left-hand panel: a sample of size 2000 from a $\mathcal{MVM}_1(\mu_1, \kappa_1, \Delta_1)$ distribution with μ_1 , κ_1 and Δ_1 given in (7.1). Right-hand panel: a sample of size 2000 from the mixture law given in (7.2). In both cases, the data are shown in yellow, whereas the boundary of the true level sets is shown (in red) together with the estimated boundary (in magenta).

$n = 5000, 10000$ \hat{f}_n outperforms \tilde{f}_n slightly.

n	$\ \hat{f}_n - f\ _\infty$	$\ \tilde{f}_n - f\ _\infty$
500	1.1067 [1.0640] (0.2983)	1.1062 [1.0664] (0.2992)
1000	0.9651 [0.9596] (0.2079)	0.9631 [0.9601] (0.2092)
5000	0.6282 [0.6155] (0.1329)	0.6285 [0.6118] (0.1327)
10000	0.5262 [0.5018] (0.0910)	0.5358 [0.5164] (0.0915)

Table 2: Mean over 100 replications for $\|\hat{f}_n - f\|_\infty$ and $\|\tilde{f}_n - f\|_\infty$. Between [] and () the median and standard deviation respectively.

8 A real data example: extreme and non-extreme winds in Uruguay

As an example of real-data on a manifold, we will study the behavior of extreme and non-extreme winds, measured at the meteorological station located in the international airport of Carrasco, Montevideo Uruguay. The aim is to characterize the direction of these winds, with respect to the time period at

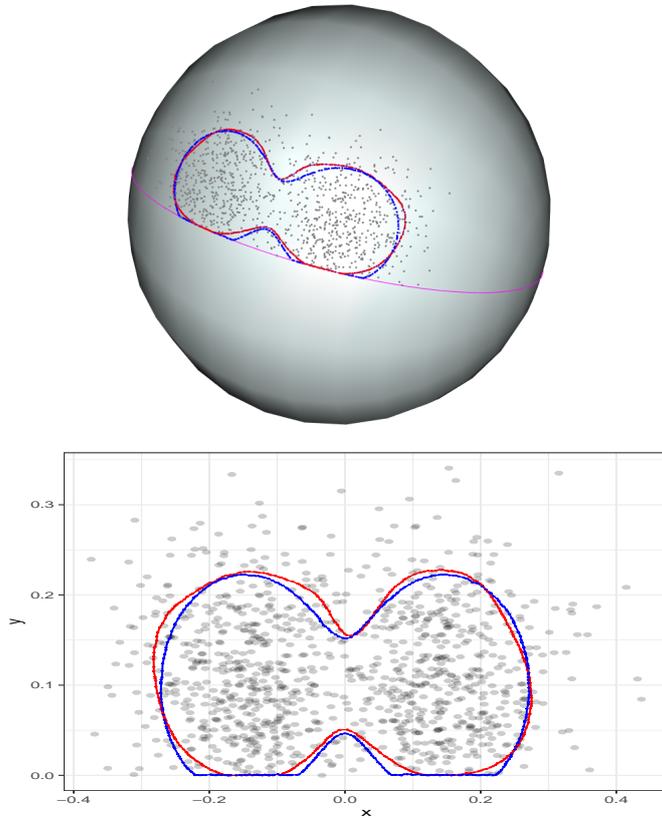


Figure 3: Left-hand panel: A sample of size 1000 from the mixture of two von Mises–Fisher distributions given in equation (7.4). Right-hand panel: the stereographic projections of the sample and the level sets. In both cases, the estimator is shown in red, while the true underlying level set is shown in blue.

which the data were obtained (divided in two groups, corresponding to cold or hot seasons respectively). The data are stored as points in a cylinder \mathcal{C} (i.e., $\mathcal{C} = S^1 \times \mathbb{R}^+$), with two parameters, the wind intensity $a \geq 0$ (in m/s), and the wind direction ($\theta \in [0, 2\pi)$), see Figure 6. The angles are measured clockwise, with zero located at North. The original database consisted of 149040 wind measures obtained between the dates 01/01/2000 and 31/12/2016. However, we ignored some missing data (0.86% in total) and we will study separately extreme and moderate (non-extreme winds).

The intensities recorded are the average of the horizontal component of the wind in the last ten minutes of each hour, while the directions recorded are the average of the directions, also in the last ten minutes of each hour. Gust are not recorded (sudden increases in the instantaneous wind, that exceed the average wind in more than 5 m / s.)

8.1 Extreme winds

At large scale there exists two kinds of wind: the synoptic extreme wind, which are produced by the passage of cold or warm fronts, with intensive convective activity, and the extreme non-synoptic winds, produced by extratropical cyclones associated with low and high pressure systems (see [16]). Extreme winds are not classified according to the casuistry of the phenomenon. According to [17], extreme maximum speeds are considered to be those that exceed 22 m/s, because this is the threshold at which different kind of risks for buildings and the population can exist. Measures below this threshold were removed. This is a typical procedure in the theory of extreme winds, and is called peak over threshold, (POT). We obtained 256 measure above this value, 39.8% correspond to the warm period and 60.2% to the cold period. The mean intensities are similar in both periods (29.40 and 29.38 for the warm and cold period respectively). In Figure 4, we show the estimated marginal densities of the extreme wind speed in every period and their boxplots. As it can be observed, there is no evident significant difference between them.

However, if we take into account the direction of the wind, the situation is different. Figure 5 shows the circular histograms of the marginal wind directions for the two periods of time considered. As it is seen, extreme winds in the warm period come more frequently from direction SW and SSW, with a strong trend to the south. At the cold period they come from the SW also, but with a trend to the West. This is also supported with the results obtained with the estimation of the level sets at each period.

Figure 6 shows at left the extreme winds on the cylinder and at right the estimation of four level sets for each period considered plot together, as it was proposed in Section 4. The chosen levels were those containing 90%, 75%, 50% and 25% of the data. As it can be seen most of the extreme winds in the cold period comes from SW and W, and most of the extreme winds in the warm period comes from S and SW.

As can be seen in Figure 6, there is a shift in the direction of the extreme winds in the cold period with respect to the warm period in all level sets. In addition, the shapes of the level sets are quite different. For the 90% level set we found in both cases two connected components, which are more separated for the warm period than for the cold period. The small component in the warm period is centered around the North-East, while the small component in the cold period is centered around East. In addition, the 10% of the extreme winds with intensity above 40 m/s are located at very different directions in the cold period and in the warm period. In Table ?? we provide the mean values for intensity and direction for each of the chosen level sets and for each period.

8.2 Moderate winds

Uruguay is currently one of the countries with the largest presence of wind energy in its electricity supply. Today almost 700 wind turbines are distributed in about thirty public and private parks. Uruguay occupies the second place in the world with 40.1% of wind generation. Analyzing the behavior of moderate winds is important in many aspects in particular for wind energy generation. In this subsection we replicate the previous analysis for the case of non-extreme winds, which corresponds to intensities below 22 m/s.

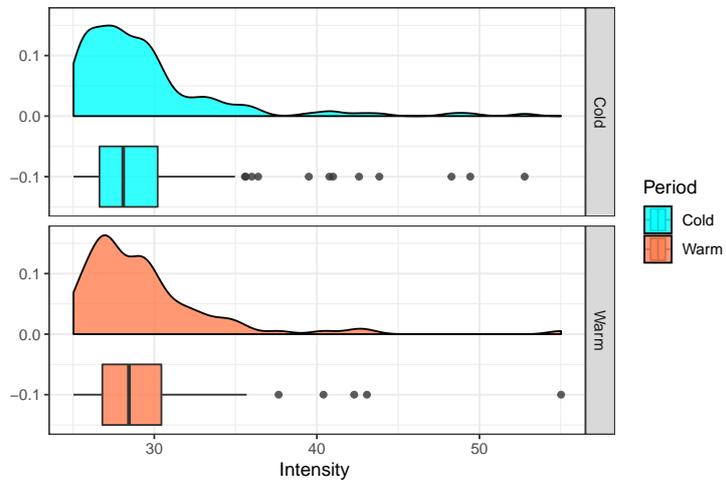


Figure 4: Estimated marginal densities of extreme wind speed. Top panel: warm period. Bottom panel: cold period

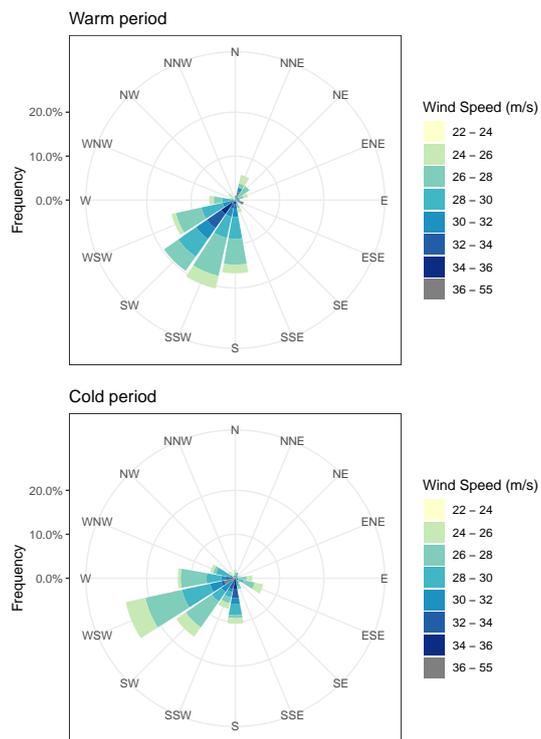


Figure 5: Circular histograms. Left panel: warm period. Right panel: cold period.

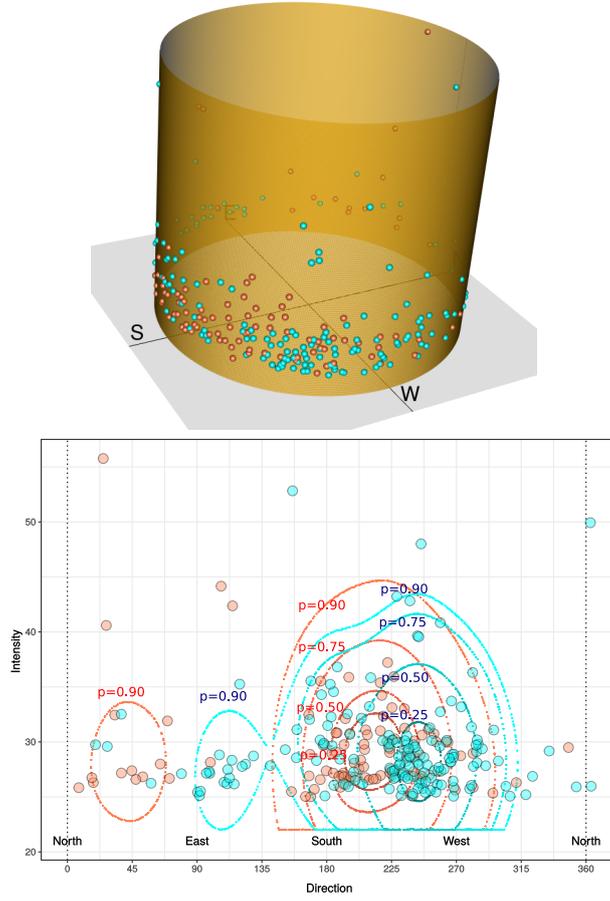


Figure 6: Left-hand panel: the data on the cylinder \mathcal{C} , in sky-blue the data points corresponding to the cold period. Right-hand panel: the cylindrical coordinate map of data and the level sets for each period.

	Direction	Intensity		Direction	Intensity
90%	200	28.92	90%	220	28.88
	(57.54)	(2.76)		(53.14)	(3.28)
75%	214	29.29	75%	236	28.93
	(24.99)	(2.78)		(32.92)	(3.06)
50%	214	29.00	50%	247	27.99
	(17.22)	(2.38)		(17.30)	(1.94)
25%	212	29.04	25%	241	27.65
	(10.96)	(1.84)		(8.23)	(1.66)

Table 3: Mean value and standard deviation for the direction and the intensity, for each level set. Left: warm period. Right: Cold period.

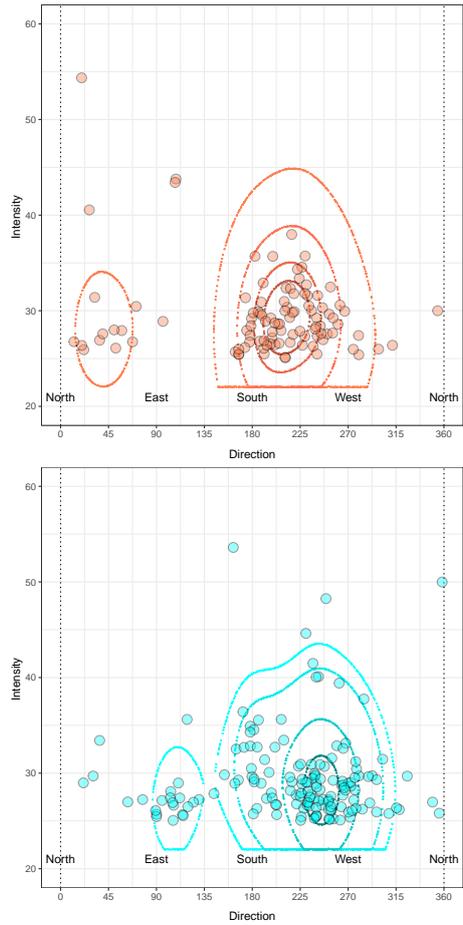


Figure 7: Left-hand panel: the cylindrical coordinate map of data and the level sets for the warm period. Right-hand panel: the cylindrical coordinate map of data and the level sets for the cold period.

	Direction	Intensity		Direction	Intensity
90%	126 (87.92)	9.04 (3.72)	90%	163 (109.01)	8.08 (3.27)
75%	100 (63.85)	8.89 (3.50)	75%	151 (111.44)	(7.74) (2.95)
50%	99 (40.83)	8.98 (2.96)	50%	131 (113.19)	7.53 (2.59)
25%	119 (13.94)	9.49 (2.64)	25%	92 (102.48)	7.41 (2.12)

Table 4: Non-extreme winds. Mean value and standard deviation for the direction and the intensity, for each level set. Left: warm period. Right: Cold period.

In Figure 9 we show the estimated marginal densities of the non-extreme wind speed at every period and their boxplots, as can be seen they are quite similar. Figure 10 shows the circular histograms of the marginal wind directions for the two periods of time considered. As can be seen, the directions of non-extreme winds in the warm period corresponds more frequently with ESE direction, with a strong trend to the south. During cold period the main directions corresponds to NNE. This is also supported by the results obtained with the estimation of the level sets at each period.

Figure 11 shows the four level sets for each period considered, as proposed in Section 4. The chosen levels were those containing 90%, 75%, 50% and 25% of the data. As can be seen the directions of most of the non-extreme winds in the cold period corresponds to SW and W, while those of the extreme winds in the warm period corresponds to directions S and SW (see Figure ?? and Table ??).

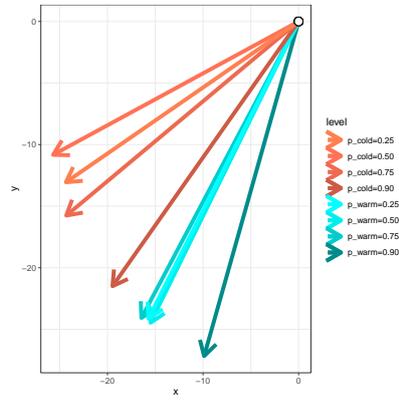


Figure 8: For each level we draw an arrow whose norm is equal to mean intensity in that level, pointing toward the mean of the direction in that level.

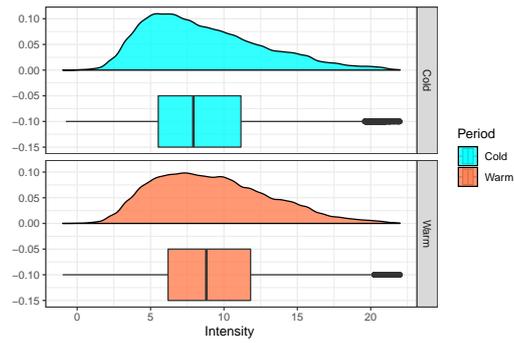


Figure 9: Non-extreme winds. Estimated marginal densities of extreme wind speed. Top panel: warm period. Bottom panel: cold period.

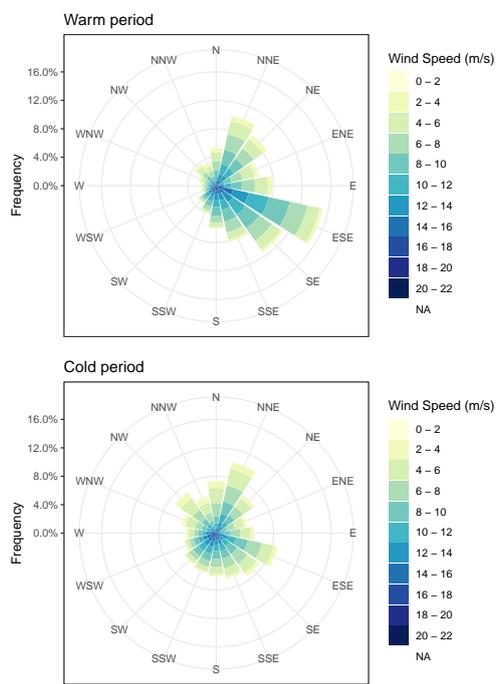


Figure 10: Circular histograms for non-extreme winds. Left-hand panel: warm period. Right-hand panel: cold period.

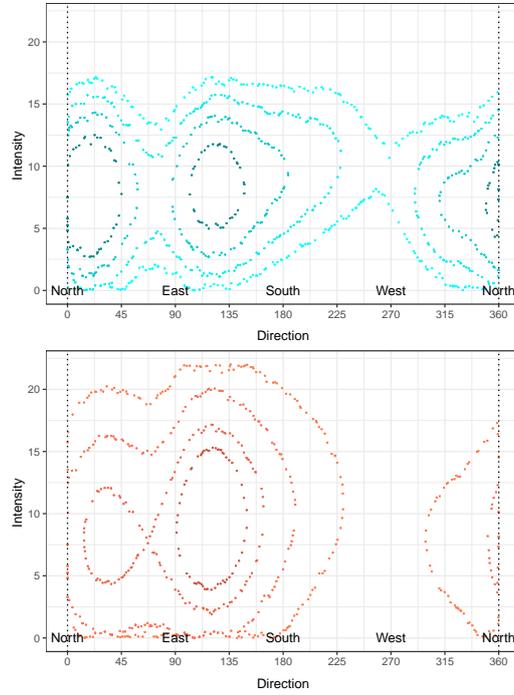


Figure 11: Non-extreme winds. Left-hand panel: the level four sets for cold period. Right-hand panel: the four level sets for the warm period.

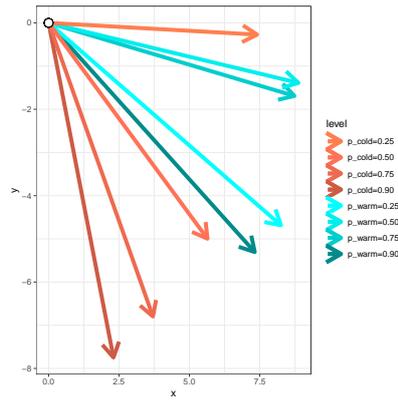


Figure 12: Non-extreme winds. For each level we draw an arrow whose norm is equal to mean intensity in that level, pointing toward the mean of the direction in that level.

Appendix

Proof of Proposition 1

From Proposition 14 in [46] ∂M and M have positive reach. Denote by $\text{reach}(M)$ the reach of M , (it is also known as the condition number, see [20]). By point 4 in Lemma 3, in [20] $R(x, y) = \|x - y\|/\rho(x, y)$ is bounded away from zero if $\|x - y\| < \text{reach}(M)/2$. Since $\text{reach}(\partial M) > 0$ w.r.t. to the intrinsic metric in M , (see remark 12 in [46]), M admits a normal collar N_r . Since $M \setminus \exp(N_r)$ is compact and of class \mathcal{C}^2 the injectivity radius on $M \setminus \exp(N_r)$ is bounded away from zero, from where it follows that M is uniformly tangible.

Proof of Theorem 4.1

Let us bound

$$\begin{aligned} \sup_{x \in M_n} |\hat{f}_{n,h}(x) - f(x)| &\leq \sup_{x \in M_n} |\hat{f}_{n,h}(x) - f_{n,h}(x)| + \\ &\sup_{x \in M_n} |f_{n,h}(x) - E(f_{n,h}(x))| + \sup_{x \in M_n} |E(f_{n,h}(x)) - f(x)| = A_1 + A_2 + A_3 \end{aligned}$$

First we prove that $h^{-(1+\delta)}A_2 \rightarrow 0$ a.s., for all $\delta < 1/2$. We will follow the same ideas used in [25]. First, we define the random variables

$$V_j(x) = \frac{1}{m_0(x)} K\left(\frac{\|x - X_j\|}{h}\right) - \frac{1}{m_0(x)} E\left[K\left(\frac{\|x - X_j\|}{h}\right)\right]$$

and let $S_n(x) = \sum_{j=1}^n V_j(x)$. Observe that from (8.2), $m_0(x) > 1/2$ for all x and n large enough (independent of x). Because K is bounded, it follows that $|V_j(x)| \leq C_2$ for all x . Let $\beta_n = h^{-(1+\delta)} \rightarrow \infty$, where $\delta < 1/2$, then from Bernstein's inequality,

$$\sup_{x \in M_n} P\left(\beta_n \frac{1}{nh^{d'}} |S_n(x)| > \epsilon\right) \leq 2 \exp\left(-\frac{\epsilon^2}{4C_2^2} \frac{nh^{d'}}{\beta_n^2}\right) \quad (8.1)$$

For n large enough fixed, we consider a finite collection of balls $B_i = B(p_i, h^\gamma)$ centered at $p_i \in M_n$, with $\gamma > d' + 1$ such that $\beta_n h^{\gamma-d'-1} \rightarrow 0$, and $M_n \subset \cup_{i=1}^l B_i$. Because M_n is compact and $\gamma > d' + 1$, $l \leq C_3 h^{-\gamma}$.

$$\sup_{x \in M_n} \frac{1}{nh^{d'}} |S_n(x)| \leq \max_{1 \leq j \leq l} \sup_{p \in B_j} \frac{1}{nh^{d'}} |S_n(x) - S_n(p_j)| + \max_{1 \leq j \leq l} \frac{1}{nh^{d'}} |S_n(p_j)| = I_1 + I_2$$

Because K is Lipschitz, $\beta_n I_1 \leq C_4 \beta_n h^{\gamma-(d'+1)} \rightarrow 0$ for some positive constant C_4 . Then $I_1 < \epsilon/2$ for n large enough. From (8.1) we get that for n large enough

$$P\left(\beta_n \sup_{x \in M_n} \frac{1}{nh^{d'}} |S_n(x)| > \epsilon\right) \leq P\left(\beta_n I_2 > \frac{\epsilon}{2}\right) \leq 2C_3 h^{-\gamma} \exp\left(-\frac{C_5 nh^{d'}}{\beta_n^2}\right),$$

C_5 being a positive constant. Now from Borel-Cantelli's lemma, together with condition $nh^{d'}/(\beta_n^2 \log(n)) \rightarrow \infty$, it follows that $\beta_n A_2 \rightarrow 0$ a.s.

To bound A_3 , first we use that the term $\mathcal{O}_x(h^2)$ can be bounded independently of x , from above by $C_1 h^2$ for some constant $C_1 > 0$ (see the proof of Theorem 3.1 in [5]). Let $c_n = \inf_{x \in M_n} \rho(x, \partial M)$. Now if we bound $m_1(x) \leq (1/(2\sqrt{\pi})) \exp(-c_n^2/h^2)$ and using that $\|\nabla f(x)\| \leq C$, then it follows from (4.2) that A_3 is of the order $o(h)$.

To bound A_1 observe that $\hat{f}_{h,n}(x) = m_0(x)f_{n,h}(x)$,

$$\sup_{x \in M_n} |\hat{f}_{n,h}(x) - f_{n,h}(x)| \leq \sup_{x \in M_n} |m_0(x) - 1| \sup_{x \in M_n} f_{n,h}(x)$$

From $b_x \geq c_n > 0$ for all $x \in M_n$ it follows that

$$|m_0(x) - 1| = \frac{1}{\sqrt{\pi}} \int_{\frac{b_x}{h}}^{+\infty} \exp(-z^2) dz \leq \frac{1}{\sqrt{\pi}} \int_{\frac{c_n}{h}}^{+\infty} \exp(-z^2) dz \leq \frac{h}{2c_n \sqrt{\pi}} \exp\left(-\frac{c_n^2}{h^2}\right) \quad (8.2)$$

To bound $\sup_{x \in M_n} f_{n,h}(x)$, we proceed as we did with A_3 , and it follows that

$$\beta_n \sup_{x \in M_n} |f_{n,h}(x) - E(f_{n,h}(x))| \rightarrow 0 \quad a.s.$$

so it is enough to bound $|E(f_{n,h}(x))|$, but we have proven that $\sup_{x \in M_n} |E(f_{n,h}(x)) - f(x)| \rightarrow 0$. Then, because f is continuous and M is compact, it is bounded. So, for n to be large enough, $\sup_{x \in M_n} |E(f_{n,h}(x))| < 2 \sup_{x \in M} f(x) < \infty$. Finally, we have proven that

$$A_1 = \sup_{x \in M_n} |\hat{f}_{n,h}(x) - f_{n,h}(x)| = o(h/c_n).$$

Proof of Theorem 5.1

Let us prove point 1, because $L_f(\lambda) \cap \partial M = \emptyset$, we can take $\delta > 0$ small enough such that $L_f(\lambda) \subset M_\delta$, then condition M1 in [9] is fulfilled in M_δ ; that is, $B_\rho(x, r)$ is connected for all $x \in M_\delta$ and for all $0 < r < \delta$. Because M is compact, condition f2 in [9] is fulfilled. Then, Theorem 4.2, and Theorem 2 in [9] entails that $d_H(\partial L_{\hat{f}_{n,h}}(\lambda), \partial L_f(\lambda)) \rightarrow 0$. To prove 2, observe that Theorem 2.1 in [33] implies that $d_H(L_{\hat{f}_{n,h}}(\lambda), L_f(\lambda)) \rightarrow 0$ (observe that $\partial L_f(\lambda) = \{x : f(x) = \lambda\}$). Finally, to prove point 3, observe that if $\nabla_x f \neq 0$ for all $x : f(x) = \lambda$, then $\partial L_f(\lambda) = \{x : f(x) = \lambda\}$ is a $d'-1$,dimensional submanifold of M , and then $\mu(\partial L_f(\lambda)) = 0$, then point 3 is a consequence of Theorem 2 in [13], which still holds for any metric space.

Proof of Theorem 5.2

Let $\epsilon_n \rightarrow 0$ such that $\epsilon_n/h \rightarrow \infty$ and define the sequence of sets $M_{\epsilon_n} = \{x \in M : \rho(x, \partial M) \geq \epsilon_n\}$, observe that M_{ϵ_n} is compact for all n , then

$$\begin{aligned} d_H(L_{\hat{f}_{n,h}}(\lambda), L_f(\lambda)) &\leq d_H(L_{\hat{f}_{n,h}}(\lambda), L_{\hat{f}_{n,h}}(\lambda) \cap M_{\epsilon_n}) + \\ d_H(L_{\hat{f}_{n,h}}(\lambda) \cap M_{\epsilon_n}, L_f(\lambda) \cap M_{\epsilon_n}) &+ d_H(L_f(\lambda) \cap M_{\epsilon_n}, L_f(\lambda)) = I_1 + I_2 + I_3. \end{aligned}$$

To prove that $I_2 \rightarrow 0$ a.s. let us denote $\gamma_n = \sup_{x \in M_{\epsilon_n}} |\hat{f}_{h,n}(x) - f(x)|$, then

$$L_f(\lambda + \gamma_n) \cap M_{\epsilon_n} \subset L_{\hat{f}_{h,n}}(\lambda) \cap M_{\epsilon_n} \subset L_f(\lambda - \gamma_n) \cap M_{\epsilon_n}.$$

Then $I_2 \leq \sup_{x \in L_f(\lambda - \gamma_n) \cap M_{\epsilon_n}} \rho(x, L_f(\lambda + \gamma_n) \cap M_{\epsilon_n}) =: R_n$. To prove that $R_n \rightarrow 0$ assume by contradiction that there exists $\delta > 0$ and $x_n \in L_f(\lambda - \gamma_n) \cap M_{\epsilon_n}$ such that $\delta < \rho(x_n, L_f(\lambda + \gamma_n) \cap M_{\epsilon_n})$. We can assume that $x_n \rightarrow x_0$ and for all n , $\delta/2 \leq \rho(x_0, L_f(\lambda + \gamma_n) \cap M_{\epsilon_n})$. If $f(x_0) > \lambda$ there exists N_{x_0} such

that $f(z) > \lambda$ for all $z \in N_{x_0}$, fix $\epsilon < \delta/2$ and $z \in N_{x_0}$ with $\rho(x_0, z) < \epsilon$ and n large enough such that $z \in M_{\epsilon_n}$, then for n large enough $z \in L_f(\lambda + \gamma_n)$, which is a contradiction. Then, $f(x_0) = \lambda$. Fix a_j with $\rho(x_0, a_j) < \delta/2$ and $f(a_j) > \lambda$, then for n large enough $f(a_j) > \lambda + \gamma_n$ and $a_j \in M_{\epsilon_n}$ which is again a contradiction. This proves that $R_n \rightarrow 0$ and then $I_2 \rightarrow 0$.

Let us prove that $I_1 \rightarrow 0$ a.s., as $n \rightarrow \infty$. Because $L_{\hat{f}_{n,h}}(\lambda) \cap M_{\epsilon_n} \subset L_{\hat{f}_{n,h}}(\lambda)$, it follows that,

$$d_H(L_{\hat{f}_{n,h}}(\lambda), L_{\hat{f}_{n,h}}(\lambda) \cap M_{\epsilon_n}) = \sup_{x \in L_{\hat{f}_{n,h}}(\lambda)} \rho(x, L_{\hat{f}_{n,h}}(\lambda) \cap M_{\epsilon_n}).$$

Suppose by contradiction that I_1 does not converge to 0 a.s., then there exists $\delta > 0$ such that with positive probability there exists $x_n \in L_{\hat{f}_{n,h}}(\lambda)$ such that $\rho(x_n, L_{\hat{f}_{n,h}}(\lambda) \cap M_{\epsilon_n}) > \delta$. Because M is compact, we can assume that there exists $x \in M$ such that $x_n \rightarrow x$ (by taking a subsequence if it is necessary), with positive probability. Observe that $x \in \partial M$ because $x_n \in M \setminus M_{\epsilon_n}$ and $\epsilon_n \rightarrow 0$. Now let us prove that $f(x) \leq \lambda$ (recall that $\hat{f}_{h,n}(x_n) \geq \lambda$). Suppose that $f(x) > \lambda$, then there exists a neighbourhood B_x on M , of x , such that $f(z) > \lambda$ for all $z \in B_x$. Observe that for all n large enough $B_x \cap M_{\epsilon_n} \neq \emptyset$, then for all $y \in B_x \cap M_{\epsilon_n}$, $\hat{f}_{h,n}(y) \geq \lambda$ because $\sup_{x \in M_{\epsilon_n}} \|\hat{f}_{h,n}(x) - f(x)\| \rightarrow 0$ a.s. But then we can choose a sequence $y_n \in B_x \cap M_{\epsilon_n}$ with $y_n \rightarrow x$ such that $\hat{f}_{h,n}(y_n) \geq \lambda$. So $\rho(x_n, L_{\hat{f}_{n,h}}(\lambda) \cap M_{\epsilon_n}) \leq \rho(x_n, y_n) \leq \rho(x_n, x) + \rho(x, y_n) \rightarrow 0$, which contradict that $\delta < \rho(x_n, L_{\hat{f}_{n,h}}(\lambda) \cap M_{\epsilon_n})$. This proves that $f(x) \leq \lambda$.

Observe that $\hat{f}_{h,n}(x_n) = m_0(x_n)f_{h,n}(x_n) \geq \lambda$. We will prove that $f_{h,n}(x_n) \rightarrow f(x) \leq \lambda$ a.s., which is a contradiction because $m_0(x_n) \rightarrow 1/2$. Let us bound,

$$|f_{h,n}(x_n) - f(x)| \leq |f_{h,n}(x_n) - E(f_{h,n}(x_n))| + |E(f_{h,n}(x_n)) - f(x_n)| + |f(x_n) - f(x)|$$

and $|f_{h,n}(x_n) - E(f_{h,n}(x_n))| \leq \sup_{x \in M} |f_{h,n}(x) - E(f_{h,n}(x))|$. Now the convergence $\sup_{x \in M} |f_{h,n}(x) - E(f_{h,n}(x))| \rightarrow 0$ a.s., is proved following the same ideas used to prove that $A_2 \rightarrow 0$ as in the proof of Theorem 4.2). Regarding the second term, we also bound $|E(f_{h,n}(x_n)) - f(x_n)| \leq \sup_{x \in M} |E(f_{h,n}(x)) - f(x)|$, which converges to 0 following the same ideas used to prove $A_3 \rightarrow 0$ in Theorem 4.2. Finally, $f(x_n) \rightarrow f(x)$ because f is a continuous function.

To prove that $I_3 \rightarrow 0$ as $n \rightarrow \infty$ assume by contradiction that this is not true, then there exists $\delta > 0$ and a sequence x_n such that $f(x_n) \geq \lambda$, $x_n \in M \setminus M_{\epsilon_n}$ and $\rho(x_n, M_{\epsilon_n} \cap L_f(\lambda)) > \delta$. Because $\epsilon_n \rightarrow 0$, there exists a subsequence of x_n (which will be denoted x_n for ease of writing), such that $x_n \rightarrow x \in \partial M$. Because f is continuous $f(x) \geq \lambda$. If $f(x) > \lambda$, then there exists N_x a neighborhood of x such that for all $y \in N_x \cap M_{\epsilon_n}$ $f(y) \geq \lambda$. Let us choose $y_n \rightarrow x$ and $y_n \in N_x \cap M_{\epsilon_n}$, then $\delta < \rho(x_n, M_{\epsilon_n} \cap L_f(\lambda)) \leq \rho(x_n, y_n) \rightarrow 0$, which is a contradiction. The other case is $f(x) = \lambda$, let $a_j \rightarrow x$ such that $f(a_j) > \lambda$ for all j . For all j , we can choose $n(j) \rightarrow \infty$ as $j \rightarrow \infty$, such that $a_j \in M_{\epsilon_{n(j)}}$. Then

$$\delta < \rho(x_{n(j)}, M_{\epsilon_{n(j)}} \cap L_f(\lambda)) \leq \rho(x_{n(j)}, a_{n(j)}) \leq \rho(x_{n(j)}, x) + \rho(x, a_{n(j)}) \rightarrow 0 \text{ as } j \rightarrow \infty.$$

Proof of Theorem 6.1

Let us denote $\mathcal{X}_n = \{X_1, \dots, X_n\}$, then

$$\begin{aligned} d_H(C_r(\{X_i : \hat{f}_{h,n}(X_i) > \lambda\}), C_r(\{X_i : f(X_i) > \lambda\})) \leq \\ d_H(C_r(\mathcal{X}_n \cap L_{\hat{f}_{h,n}}(\lambda)), L_{\hat{f}_{h,n}}(\lambda)) + \\ d_H(L_{\hat{f}_{h,n}}(\lambda), L_f(\lambda)) + d_H(L_f(\lambda), C_r(\mathcal{X}_n \cap L_f(\lambda))) = A + B + C. \end{aligned}$$

From Theorem 5.2, $B \rightarrow 0$ a.s. Because $L_f(\lambda)$ is r -convex, $\mathcal{X}_n \cap L_f(\lambda) \subset C_r(\mathcal{X}_n \cap L_f(\lambda)) \subset L_f(\lambda)$ and then,

$$C \leq d_H(\mathcal{X}_n \cap L_f(\lambda), L_f(\lambda)) \rightarrow 0 \quad \text{a.s., as } n \rightarrow \infty.$$

Regarding A , observe that $A = \sup_{x \in L_{\hat{f}_{h,n}}(\lambda)} \rho(x, C_r(\mathcal{X}_n \cap L_{\hat{f}_{h,n}}(\lambda)))$. Let us proceed by contradiction, assume that with positive probability A does not converge to 0, then there exists a sequence $x_n \in L_{\hat{f}_{h,n}}(\lambda)$ and $\delta > 0$ such that $\delta < \rho(x_n, C_r(\mathcal{X}_n \cap L_f(\lambda)))$ for all $n > n_0$. Because M is compact, there exists a convergent subsequence of x_n (which we will denote x_n) such that $x_n \rightarrow x$. Because $B \rightarrow 0$, it follows that $f(x) \geq \lambda$ but with positive probability $\delta/2 < \rho(x, C_r(\mathcal{X}_n \cap L_{\hat{f}_{h,n}}(\lambda)))$ for all n large enough. If $f(x) > \lambda$, then there exists $\eta > 0$ such that for all $z \in B_\rho(x, \eta)$ $f(z) > \lambda$. Let us take $0 < \eta < \delta/2$, then with probability one, for n large enough $\hat{f}_{h,n}(z) > \lambda$ for all $z \in B(x, \eta)$. Let us take n large enough such that $d_H(\mathcal{X}_n \cap L_f(\lambda), L_f(\lambda)) < \eta$, then $\mathcal{X}_n \cap B(x, \eta) \neq \emptyset$ but then $\rho(x, C_r(\mathcal{X}_n \cap L_{\hat{f}_{h,n}}(\lambda))) < \eta$ which is a contradiction. The case $f(x) = \lambda$ is proved in the same way, let a_j such that $f(a_j) > \lambda$ and $0 < \eta < \delta$ such that for all $z \in B_\rho(a_j, \eta)$ $f(z) > \lambda$. Let n be large enough such that $\hat{f}_{h,n}(z) > \eta$ for all $z \in B(a, \eta)$ and $d_H(\mathcal{X}_n \cap L_f(\lambda), L_f(\lambda)) < \eta$. Again $\rho(x, C_r(\mathcal{X}_n \cap L_{\hat{f}_{h,n}}(\lambda))) < \eta$, which is a contradiction.

References

- [1] Aneiros, G., Cao, R., Fraiman, R., Genest, C., and Vieu, P. (2019). Recent advances in functional data analysis and high-dimensional statistics. *Journal of Multivariate Analysis*. 170, 3–9.
- [2] Baillo, A.(2003). Total error in a plug-in estimator of level sets. *Statist. Probab. Lett.* **65**, 441–417.
- [3] Baillo, A., Cuesta-Albertos, J.A., and Cuevas, A.(2001). Convergence rates in nonparametric estimation of level sets. *Statist. Probab. Lett.* **53**, 27–35.
- [4] Bhattacharya A, and Bhattacharya R. (2012) Nonparametric inference on manifolds: with applications to shape spaces. *Vol. 2. Cambridge University Press*
- [5] Berry, T. and Sauer, T. (2017) Density estimation on manifolds with boundary, *Computational Statistics & Data Analysis*, 107, pp. 1–17.
- [6] Chen, Y-C., Genovese, C.R., and Wasserman, Larry (2017) Density level sets: asymptotics, inference, and visualization *JASA*, **112**, 1684-1696.
- [7] Cholaquidis, A., Cuevas, A. and Fraiman, R. (2014) On Poincaré cone property. *Ann. Statist.*, **42**, 255–284.

- [8] Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference* 147, 1–23.
- [9] Cuevas, A., González-Manteiga, W., and Rodríguez-Casal, A. (2006) Plug-in estimation of general level sets. *Aust. N. Z. J. Stat.* **48**(1), 7–19.
- [10] Cuevas, A. and Fraiman, R. (1997) A plug-in approach to support estimation. *Ann. Statist.* **25**, 2300–2312.
- [11] Cuevas, A., Febrero, M. and Fraiman, R. (2000) Estimating the number of clusters. *Canad. J. Statist.* **28**, 367–382.
- [12] Cuevas, A., Febrero, M. and Fraiman, R. (2000) Cluster analysis: a further approach based on density estimation. *Comput. Statist. Data Anal.* **36**, 441–459.
- [13] Cuevas, A., Fraiman, R. and Pateiro-López, B. (2012). On statistical properties of sets fulfilling rolling-type conditions. *Adv. in Appl. Probab.* **44** 311–329.
- [14] Delicado, P. (2001) Another look at principal curves and surfaces. *J. Multivariate Anal.* **77**, 84–116.
- [15] Devroye, L. and Wise, G.L. (1980). Detection of abnormal behaviour via nonparametric estimation of the support. *SIAM J. Appl. Math.* **38** 480–488.
- [16] Durañona, V., Sterling, M. and Baker, C.J. (2007) An analysis of extreme non-synoptic winds. *Journal of wind engineering and industrial aerodynamics*, **95**(9-11), 1007-1027.
- [17] Durañona, V. (2012) Actualización de la estadística de vientos extremos para Uruguay. *Memorias del II Congreso Latinoamericano de Ingeniería de Viento, La Plata, Argentina*, (CD-ROM),.
- [18] Edelsbrunner, H. and Mücke, E. P. (1994). Three-dimensional alpha-shapes. *ACM Transactions on Graphics* **13**, 43–72.
- [19] Fraiman, D. Fraiman, N., and Fraiman, R. (2014). Non parametric statistics of dynamic random networks with distinguishable nodes. *Test* 26 (3), 546–573.
- [20] Genovese, C.R., Perone-Pacifico, M., Verdinelli, I. and Wasserman, L. (2012b). Minimax Manifold Estimation. *J. Mach. Learn. Res.* **13**, 1263–1291.
- [21] Goia, A., and Vieu, P. (2016) An introduction to recent advances in high/infinite dimensional Statistics *Journal of Multivariate Analysis* 146, 1–6.
- [22] Hastie, T. and Stuetzle, W. (1989). Principal curves. *J. Amer. Statist. Assoc.* **84**, 502-516.
- [23] Hartigan, J.A. (1987) Estimation of convex density contour in two dimensions. *J. Amer. Statist. Assoc.* **82**, 267–270

- [24] Hendriks, H. and Landsman, Z. (2007) Asymptotic data analysis on manifolds. *The Annals of Statistics* 35(1), 109-131.
- [25] Henry, G. and Rodriguez, D. (2009) Kernel density estimation on Riemannian manifolds: Asymptotic results.(2009) *Journal of Mathematical Imaging and Vision*, 34(3), 235-239.
- [26] Hall, P., Watson, G. S., and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika*, 74(4), 751-762.
- [27] Heinrich, H.(2017) Uniform Density level set estimation on manifolds with dbscan *Proceedings of the 34th International Conference on Machine Learning*, 1684–1693.
- [28] Heinrich, H.(2017) Uniform convergence rates for kernel density estimation. *International Conference on Machine Learning*, 1694–1703.
- [29] Kim, J., Shin, J., Rinaldo, A. and Wasserman, L.(2019) Uniform Convergence of the Kernel Density Estimator Adaptive to Intrinsic Volume Dimension In:*Proceedings of the 36th International Conference on Machine Learning*, 3398–3407
- [30] Levina, E., and Bickel, P. J. (2005). Maximum likelihood estimation of intrinsic dimension. In *Advances in neural information processing systems*, pp. 777–784.
- [31] Mardia K.V. (1972) Statistics of directional data. *Academic Press*
- [32] Mardia, K.V. and Jupp, P.E. (2000). *Directional Statistics*. Wiley, Chichester.
- [33] Molchanov, I. (1998) A limit theorem for solutions of inequalities. *Scand. J Stat.* **25**, 235–242.
- [34] Müller, D.W., and Sawitzki, G. (1991). Excess mass estimates and test of multimodality *J. Amer. Statist. Assoc.* **86**, 738–746.
- [35] Ozakin, A., and Gray, A. G. (2009) Submanifold density estimation. In *Advances in Neural Information Processing Systems*, 1375–1382.
- [36] Pateiro-López, B. and Rodríguez-Casal, A. (2008) Length and surface area estimation under smoothness restrictions. *Adv. in Appl. Probab.* **40** 348–358.
- [37] Patrangenaru, V., and Ellingson, L.(2015) Nonparametric statistics on manifolds and their applications to object data analysis. *CRC Press*.
- [38] Pelletier, B. (2006) Non-parametric regression estimation on closed riemannian manifolds. *Journal of Nonparametric Statistics*, **18**(1):57–67, 2006.
- [39] Pennec, X. (2006) Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, **25**, no 1.

- [40] Polonik, W. (1995) Measuring mass concentration and estimating density contour clusters -an excess mass approach. *Ann. Statist.* **23**, 855–881.
- [41] Ramsay, J. O. and Silverman, B. W. (2002). Applied functional data analysis. Methods and case studies. Springer, New York.
- [42] Ramsay, J. O. and Silverman, B. W. (2005). Functional Data Analysis. Second edition. Springer, New York.
- [43] Rao, C. R. (1945) Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **37**, 81–91.
- [44] Rodríguez-Casal, A. (2007). Set estimation under convexity-type assumptions. *Ann. Inst. H. Poincaré Probab. Statist.* **43** 763–774.
- [45] Rodríguez-Casal, A. and Saavedra-Nieves, P. (2014). A fully data-driven method for estimating density level sets. <https://arxiv.org/pdf/1411.7687v1.pdf>.
- [46] Thäle, C. (2008). 50 years sets with positive reach. A survey. *Surv. Math. Appl.* **3**, 123–165.
- [47] Tsybakov, A.B. (1997). On nonparametric estimation of density level sets. *Ann. Statist.* **25**, 948–969.
- [48] Walther, G. (1999). On a generalization of Blaschke’s rolling theorem and the smoothing of surfaces, *Math. Meth. Appl. Sci.* **22**, 301–316.
- [49] Walther, G. (1997). Granulometric smoothing. *Ann. Statist.* **25** 2273–2299.