Exact Rates in Density Support Estimation

Gérard BIAU [a], Benoît CADRE [b] and Bruno PELLETIER [c,*]

[a] LSTA & LPMA
Université Pierre et Marie Curie – Paris VI
Boîte 158, 175 rue du Chevaleret
75013 Paris, France
biau@ccr.jussieu.fr

[b] IRMAR, ENS Cachan Bretagne, CNRS, UEB
Campus de Ker Lann
Avenue Robert Schuman
35170 Bruz, France
Benoit.Cadre@bretagne.ens-cachan.fr

[c] Institut de Mathématiques et de Modélisation de Montpellier
UMR CNRS 5149, Equipe de Probabilités et Statistique
Université Montpellier II, CC 051
Place Eugène Bataillon, 34095 Montpellier Cedex 5, France
pelletier@math.univ-montp2.fr

Abstract

Let f be an unknown multivariate probability density with compact support S_f. Given n independent observations $X_1, \ldots, X_n$ drawn from f, this paper is devoted to the study of the estimator $\hat{S}_n$ of S_f defined as unions of balls centered at the X_i and of common radius r_n. To

[*] Corresponding author.

measure the proximity between $\hat{S}_n$ and S_f, we employ a general criterion d_g, based on some function g, which encompasses many statistical situations of interest. Under mild assumptions on the sequence (r_n) and some analytic conditions on f and g, the exact rates of convergence of $d_g(\hat{S}_n, S_f)$ are obtained using tools from Riemannian geometry. The conditions on the radius sequence are found to be sharp and consequences of the results are discussed from a statistical perspective.

Index Terms — Support estimation, Nonparametric statistics, Exact rate of convergence, Riemannian geometry, Tubular neighborhood.

AMS 2000 Classification: 62G05, 62G20.

1 Introduction

Let f be an unknown probability density function defined with respect to the Lebesgue measure on $\mathbb{R}^d$. This paper is concerned with the problem of estimating the support of f, i.e., the closed set

$$S_f = \overline{\{x \in \mathbb{R}^d : f(x) > 0\}},$$

given a random sample $X_1, \ldots, X_n$ drawn from f. Here and later, $\overline{A}$ means the closure of the set A. Since the earlier works of Rényi and Sulanke (1963, 1964) and Geffroy (1964), the problem of support estimation has been considered by several authors [see, e.g., Chevalier (1976), Devroye and Wise (1980), Grenander (1981), Cuevas (1990), Korostelev and Tsybakov (1993a, 1993b), Härdle, Park and Tsybakov (1995), Korostelev, Simar and Tsybakov (1995), Mammen and Tsybakov (1995), Cuevas and Fraiman (1997), Gayraud (1997), Baíllo, Cuevas and Justel (2000), and Klemelä (2004)]. The application scope is vast, as support estimation is routinely employed across the entire and diverse range of applied statistics, including problems in medical diagnosises, machine condition monitoring, marketing or econometrics [see the discussion in Baíllo, Cuevas and Justel (2000) and the references therein]. In closed connection with the related topic of estimating a density level set [Polonik (1995), Tsybakov (1997), Walther (1997), Cadre (2006)], the problem of support estimation has been also addressed via unsupervised learning methods, such as the one-class kernel Support Vector Machines algorithm presented in Schölkopf, Platt, Shawe-Taylor, Smola and Williamson (2001).

Among the various approaches that have been proposed to date to estimate S_f, the probably most simple and intuitive one has been considered in Devroye and Wise (1980). The estimator is defined as

$$\hat{S}_n = \bigcup_{i=1}^{n} \mathcal{B}(X_i, r_n), \tag{1.1}$$

where $\mathcal{B}(x, r)$ denotes the closed Euclidean ball centered at x and of radius r, and where (r_n) is an appropriately chosen sequence of positive smoothing parameters. Note that this approach amounts to estimate the support of the density by the support of a kernel estimate, the kernel of which has a ball-shaped support. The sequence (r_n) then plays a role analogous to that of the kernel bandwidth. The practical properties of the support estimator (1.1) are explored in Baíllo, Cuevas and Justel (2000), who argue that this estimator is a good generalist when no *a priori* information is available on S_f. Moreover, from a practical perspective, the relative simplicity of the naive strategy (1.1) arises as a major advantage in comparison with competing multidimensional set estimation techniques, that are faced with severe difficulties owing to a heavy computational burden.

To measure the performance of the support estimator, i.e., the closeness of $\hat{S}_n$ to S_f, a standard choice is to use the distance $d_1(\hat{S}_n, S_f)$ defined by

$$d_1(\hat{S}_n, S_f) = \lambda(\hat{S}_n \triangle S_f),$$

where $\triangle$ denotes the symmetric difference and λ is the Lebesgue measure on $\mathbb{R}^d$. This criterion of proximity between sets, which is *geometric* by essence, has been successfully employed for example by Korostelev and Tsybakov (1993b), Härdle, Park and Tsybakov (1995), and Mammen and Tsybakov (1995) who have considered maximum-likelihood-type estimators and have derived minimax rates of convergence under various assumptions on the boundary sharpness of f, that is, the behavior of f near the boundary of the support S_f.

The distance d_1 may be easily extended to the much more general measure-based distance d_μ defined by

$$d_\mu(\hat{S}_n, S_f) = \mu(\hat{S}_n \triangle S_f),$$

where μ is any measure on the Borel sets of $\mathbb{R}^d$. In this context, Cuevas and Fraiman (1997) discuss the d_μ-asymptotic properties of a plug-in estimator of S_f of the form $\{f_n > \alpha_n\}$, where f_n is a nonparametric density estimator of f, and where α_n is a tuning parameter converging to zero. These authors establish also asymptotic results in terms of the Hausdorff metric, which is another natural criterion of proximity between sets [Cuevas (1990), Korostelev and Tsybakov (1993b), Korostelev, Simar and Tsybakov (1995), Cuevas and Rodríguez-Casal (2004)].

Assuming for convenience that μ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^d$, with a density g, the criterion d_μ may be written as

$$d_g(\hat{S}_n, S_f) = \int_{\mathbb{R}^d} \mathbf{1}_{\hat{S}_n \triangle S_f}(x) g(x) \mathrm{d}x. \tag{1.2}$$

The proximity measure (1.2) is fairly general and encompasses several interesting cases of choices of g, depending on the problem at hand. For instance, set first $g \equiv f$, and denote by X a random variable with density f independent of the sample. This yields the criterion

$$d_f(\hat{S}_n, S_f) = \mathbb{P}(X \notin \hat{S}_n | X_1, \ldots, X_n),$$

which is a natural statistical measure of the accuracy of $\hat{S}_n$ with respect to S_f. More generally, for a random variable X with density g independent of the sample, we may write

$$d_g(\hat{S}_n, S_f) = \mathbb{P}(X \in \hat{S}_n \triangle S_f | X_1, \ldots, X_n). \tag{1.3}$$

This loss has been considered in Devroye and Wise (1980) in a concrete testing problem regarding the detection of the abnormal behavior of a system. Roughly, a machine is observed in normal operation through the sequence of independent observations $X_1, \ldots, X_n$ drawn from the density f, and the complement S_f^c of S_f is considered as a danger area. Given a new and *unique* observation X_{n+1} with density g (possibly different from f), one has to decide whether or not the system behaves abnormally, in the sense that the distribution of X_{n+1} is different from f. A natural testing strategy then consists in rejecting the null hypothesis if X_{n+1} does not belong to $\hat{S}_n$. In this context, the distances d_f and d_g have clear interpretations in terms of error of the first kind (or false alarm probability) and of the second kind, respectively. Devroye and Wise (1980) have proved consistency of the estimator

(1.1) with respect to the symmetric difference (1.3) under some conditions on the sequence (r_n) which are analogous to those imposed on the bandwidth parameter in kernel estimation. The results of Devroye and Wise (1980) have been further explored by Baíllo, Cuevas and Justel (2000), who focused more particularly on the false alarm probability and suggested data-driven strategies to select the smoothing parameter r_n.

To the best of our knowledge, no exact rates of convergence of the density support estimator (1.1) are available in the literature. In the present paper, we propose to fill this gap, using the general distance d_g defined in (1.2) as a criterion of accuracy. Our main result (Theorem 3.1) states, under some mild analytic conditions on f and g, that there exists an explicit non-negative constant c such that

$$\sqrt{nr_n^d}\,\mathbb{E}d_g(\hat{S}_n, S_f) \to c \quad \text{as } n \to \infty,$$

provided $nr_n^d \to \infty$ and $nr_n^{d+2} \to 0$. As a matter of fact, we will prove that much faster rates are achievable–including exponential ones–, depending on the relative positions and geometric characteristics of the respective supports of f and g. Moreover, we will show that the requirement $nr_n^{d+2} \to 0$ is sharp, in the sense that the condition $nr_n^{d+2} \to \infty$ implies $\sqrt{nr_n^d}\,\mathbb{E}d_1(\hat{S}_n, S_f) \to \infty$ (Theorem 3.2). We insist on the fact that, throughout the paper, the density f is supposed to be *continuous* on $\mathbb{R}^d$. Thus, we are in the case of a *non-sharp boundary*, i.e., f decreases continuously to zero at the boundary of its support.

The paper is organized as follows. Section 2 introduces notation that is used throughout. The main convergence results are exposed in Section 3 in the general setting, and next are specialized to two important sub-cases developed in Section 4. Technical lemmas necessary to the proofs of the theorems in Section 3 and Section 4 are postponed to the Appendix A. At last, the proofs of our results require integrating over a tubular neighborhood of the boundary of the support, and so a brief account to the main useful facts from differential geometry is provided by Appendix B [for further material, we refer to Gray (1990), Bredon (1993), Chavel (1993), and Kobayashi and Nomizu (1996)].

2 Notation

Let us start by introducing some general notation concerning an arbitrary smooth Riemannian submanifold (M, σ) of $\mathbb{R}^d$. The Riemannian metric σ on M is induced by the canonical embedding of M in $\mathbb{R}^d$. The Riemannian volume measure on (M, σ) will be denoted by v_σ. We shall denote by $TM^\perp$ the normal bundle of M, and the tubular neighborhood of M of radius ε will be denoted by $\mathcal{V}(M, \varepsilon)$.

Given a function h on $\mathbb{R}^d$ taking values in $\mathbb{R}_+$ and any subset A of $\mathbb{R}_+$, we use the notation

$$[h \in A] = \{x \in \mathbb{R}^d : h(x) \in A\},$$

and we let the support S_h of h be defined as

$$S_h = \overline{[h > 0]}.$$

The interior and boundary of S_h will be denoted by $\overset{\circ}{S}_h$ and $\partial S_h = S_h - \overset{\circ}{S}_h$, respectively.

Wherever appropriate, we shall be led to consider the unit-norm section $\{e^h_p,\, p \in \partial S_h\}$ of $T\partial S_h^\perp$ that is pointing *inwards*, i.e., for all $p \in \partial S_h$, e^h_p is the unit-norm normal vector to ∂S_h at p that is directed towards the interior of S_h. Further, whenever it exists, the k^{th} directional derivative of h at the point $p + ue_p^h$ in the direction e_p^h will be denoted by $D^k_{e_p^h} h(p + ue_p^h)$, with the conventions $D^0_{e_p^h} = \mathrm{Id}$ and $D^1_{e_p^h} = D_{e_p^h}$.

Denoting by g a real-valued function on $\mathbb{R}^d$ with support S_g (not necessarily compact), we recall that the paper is devoted to the study of the asymptotic behavior of $d_g(\hat{S}_n, S_f)$, with

$$d_g(\hat{S}_n, S_f) = \int_{\mathbb{R}^d} \mathbf{1}_{\hat{S}_n \triangle S_f}(x) g(x) \mathrm{d}x.$$

The following basic assumptions on f and g will be supposed satisfied throughout the paper:

Basic Assumptions

(a) The support S_f of f is compact, and f is of class $\mathcal{C}^2$ on $\overset{\circ}{S}_f$;

(b) g is a positive, bounded, and continuous function on $\mathbb{R}^d$;

(c) $S_f \cap S_g \neq \emptyset$.

The case where $S_f \cap S_g = \emptyset$ is excluded from the study since, for n large enough, we then have $d_g(\hat{S}_n, S_f) = 0$ with probability 1. The present study is also limited to the case of a density f of class $\mathcal{C}^2$ for the sake of simplicity. In fact, cases where f exhibits a higher regularity may also be addressed by having recourse to the same flow of arguments as those exposed in the paper, but at the expense of heavier technical developments.

Finally, we will let λ_g be the measure on $\mathbb{R}^d$ defined by

$$\lambda_g(A) = \int_A g \mathrm{d}\lambda,$$

for any Borel set $A \subset \mathbb{R}^d$. At last, the letter C will denote a positive constant, the value of which may vary from line to line.

3 The general case

3.1 Convergence

We will make the following assumption on f:

Assumption 1

(a) The boundary ∂S_f of S_f is a smooth submanifold of $\mathbb{R}^d$ of codimension 1;

(b) The set $[f > 0]$ is connected;

(c) $f > 0$ on $\overset{\circ}{S}_f$.

Note that Assumption 1 never holds when the dimension d equals 1. However, all the results stated herein are still valid in dimension one, in a sense made precise in the remark below.

Remark 3.1 *In dimension one, the set S_f is a closed interval with boundary points $a < b$. In this case, all the results of the paper, which involve integrations on ∂S_f with respect to the volume measure v_σ, still hold when v_σ is replaced by the counting measure on $\{a\} \cup \{b\}$, so that the integral may be expressed as a sum.*

Remark 3.2 *First, Assumption* 1*-(b) on the connectedness of $[f > 0]$ may be relaxed to the assumption that the boundaries of the connected components of $[f > 0]$ are submanifolds of codimension* 1 *which do not overlap; see the* *discussion in Remark 3.3 after Theorem 3.1.*

Second, in the proofs of our results, we shall be led to consider sets of the form $[f \leq t]$*, for some small positive* t*. In this respect, Assumption* 1*-(c) ensures that* f *does not vanish on the topological interior* $\overset{\circ}{S}_f$ *of* S_f*. Consequently under Assumption* 1*, the set* $[f \leq t]$ *is included in a tubular neighborhood of* ∂S_f *for* t *small enough, which allows for an identification of* $[f \leq t]$ *with a subset of the normal bundle of* ∂S_f*.*

By Assumption 1-(a), $(\partial S_f, \sigma)$ is a smooth Riemannian submanifold of $\mathbb{R}^d$. Note also that $(\partial S_f, \sigma)$ is a closed (i.e., compact and without boundary) submanifold. Consequently, there exists a tubular neighborhood of ∂S_f of radius $\rho > 0$ [see Appendix B], which implies the existence of an $\epsilon > 0$ such that, for all $p \in \partial S_f$ and all $v \in [0, \varepsilon]$, $p + v e_p^f \in S_f$.

As stated in the Basic Assumptions, the density f is of class $\mathcal{C}^2$ on $\overset{\circ}{S}_f$. Indeed, it will be demonstrated next that the convergence rate of $\hat{S}_n$ to S_f depends on the degree of smoothness of f on ∂S_f. For this reason, two cases are investigated herein:

(*i*) The case where f is of class $\mathcal{C}^0$ on $\mathbb{R}^d$ with *positive first directional derivative* $D_{e_p^f} f(p)$ for all p in ∂S_f, and

(*ii*) The case where f is of class $\mathcal{C}^1$ on $\mathbb{R}^d$ with *positive second directional derivative* $D^2_{e_p^f} f(p)$ for all p in ∂S_f.

Note that in the second case, the first directional derivative $D_{e_p^f} f(p)$ vanishes on the boundary by a continuity argument. The following assumptions, which depend on some parameter $k \in \{1, 2\}$, summarize all the smoothness constraints required on f. Despite their technical aspect, these requirements are mild.

Assumption 2

(a) f is of class $\mathcal{C}^{k-1}$ on $\mathbb{R}^d$.

(b) There exists $\varepsilon > 0$ such that, for all $p \in \partial S_f$, the map $u \mapsto f(p + ue_p^f)$ is of class $\mathcal{C}^k$ on $[0, \varepsilon]$.

(c) There exists $\varepsilon > 0$ such that $\sup_{0 \le u \le \varepsilon} \sup_{p \in \partial S_f} |D^k_{e_p^f} f(p + ue_p^f)| < \infty$.

(d) $\sup_{\varepsilon > 0} \sup_{x \in S_f : \mathrm{dist}(x, \partial S_f) \ge \varepsilon} \|\mathrm{H}f(x)\| < \infty$, where $\mathrm{H}f(x)$ denotes the hessian matrix of f at the point x.

(e) There exists $\varepsilon > 0$ such that $\inf_{0 \le u \le \varepsilon} \inf_{p \in \partial S_f} D^k_{e_p^f} f(p + ue_p^f) > 0$.

We are now in a position to state our main result.

Theorem 3.1 *Suppose that Assumption* 1 *and Assumption* 2 *hold for some* $k \in \{1, 2\}$. *Then, if* $nr_n^d \to \infty$ *and* $nr_n^{d+k} \to 0$, *we have, as* $n \to \infty$,

$$(nr_n^d)^{1/k}\, \mathbb{E}d_g(\hat{S}_n, S_f) \to \left(\frac{\pi}{2}\right)^{(k-1)/2} \omega_d^{-1/k} \int_{\partial S_f} \frac{g(p)}{\left[D^k_{e_p^f} f(p)\right]^{1/k}} \mathrm{d}v_\sigma(p),$$

where ω_d *denotes the volume of the unit ball of* $\mathbb{R}^d$.

For the related problem of estimating a density level set $[f > t]$ with $t > 0$, Cadre (2006) obtains exact rates of convergence by use of the co-area formula [Evans and Gariepy (1992)]. The limit constant turns out to be an integral over the boundary of the actual level set of the reciprocal of the norm of the gradient of f, with respect to the $(d-1)$-dimensional Hausdorff measure on $\mathbb{R}^d$. Note that the integral in Theorem 3.1 may also be expressed with respect to the $(d-1)$-dimensional Hausdorff measure. Actually, on a smooth submanifold M of $\mathbb{R}^d$ of codimension 1, the $(d-1)$-dimensional Hausdorff measure reduces to the Riemannian volume measure v_σ induced by the canonical injection $i : M \to \mathbb{R}^d$ [see e.g., Chavel (1993, p. 126)]. However, neither the proof nor the main result of Cadre (2006) apply in the present context.

Remark 3.3 *If the set $[f > 0]$ has a number $m \geq 2$ of connected components, all of whose satisfy Assumption* 1, *and if the boundaries of the connected components are mutually disjoint, then a result similar to that of Theorem 3.1 may be established. In such a case, the limit constant is expressed as the sum of m integrals on the boundaries of the connected components with respect to the induced Riemannian volume measure.*

To illustrate the result of Theorem 3.1, consider for example the Epanechnikov probability density function defined for all x in the unit closed Euclidean ball $\mathcal{B}(0,1)$ by

$$f(x) = c_0\big(1 - \|x\|^2\big),$$

and by 0 otherwise. Here, c_0 is a normalizing constant set as

$$c_0 = \left[\omega_d - d\omega_{d-1} B\Big(\frac{3}{2}, d\Big)\right]^{-1},$$

where $B(.,.)$ is the beta function, and $\omega_0 = 1$ by definition. Clearly f is of class $\mathcal{C}^2$ in the interior of S_f, and of class $\mathcal{C}^0$ on $\mathbb{R}^d$ since $D_{e_p^f} f(p) = 2c_0$ for all p in $\partial\mathcal{B}(0,1)$. For example, fix $g \equiv 1$, so that the loss reduces to the usual geometrical criterion $d_1(\hat{S}_n, S_f) = \lambda(\hat{S}_n \Delta S_f)$. In this context, Theorem 3.1 reads as

$$nr_n^d\, \mathbb{E}\lambda(\hat{S}_n \Delta S_f) \to \frac{1}{2c_0\omega_d}\, v_\sigma\big(\partial\mathcal{B}(0,1)\big) = \frac{d}{2}\left[\omega_d - d\omega_{d-1} B\Big(\frac{3}{2}, d\Big)\right],$$

since $v_\sigma(\partial\mathcal{B}(0,1)) = d\omega_d$.

Proof The proofs for cases $k = 1$ and $k = 2$ are similar. For the sake of simplicity, we prove the result for the case $k = 2$ only. In this context, the convergence occurs at speed $\sqrt{nr_n^d}$ under the conditions $nr_n^d \to \infty$ and $nr_n^{d+2} \to 0$. We start the proof by the equalities

$$\begin{aligned}\mathbb{E}d_g(\hat{S}_n, S_f) &= \mathbb{E}\lambda_g(\hat{S}_n \Delta S_f) \\ &= \mathbb{E}\lambda_g(\hat{S}_n \cap S_f^c) + \mathbb{E}\lambda_g(\hat{S}_n^c \cap S_f). \end{aligned} \tag{3.1}$$

Consider now the set $\tilde{S}_n$ defined as

$$\tilde{S}_n = \left\{x \in \mathbb{R}^d : \text{dist}(x, \partial S_f) \leq r_n\right\}.$$

Since $\hat{S}_n \cap S_f^c \subset \tilde{S}_n \cap S_f^c$ with probability 1, we have

$$\begin{aligned}\mathbb{E}\lambda_g(\hat{S}_n \cap S_f^c) &= \int_{S_f^c} \mathbb{P}(x \in \hat{S}_n) g(x) \mathrm{d}x \\ &= \int_{S_f^c \cap \tilde{S}_n} \mathbb{P}(x \in \hat{S}_n) g(x) \mathrm{d}x \\ &\leq \lambda(S_f^c \cap \tilde{S}_n) \sup_{\mathbb{R}^d} g.\end{aligned}$$

By the Tubular Neighborhood Theorem [cf. Appendix B], there exists a tubular neighborhood $\mathcal{V}(\partial S_f, \rho)$ of ∂S_f of radius $\rho > 0$. Consequently, as long as $r_n < \rho$, which occurs for n larger than some integer n_0, we have $S_f^c \cap \tilde{S}_n \subset \mathcal{V}(\partial S_f, \rho)$. In this case, using (B.1), the volume of $S_f^c \cap \tilde{S}_n$ is bounded above by $\sup_{\mathcal{V}(\partial S_f, \rho)} \Theta(p, u) v_\sigma(\partial S_f) r_n$. Thus, we have just proved that there exists a constant $C > 0$ such that

$$\mathbb{E}\lambda_g(\hat{S}_n \cap S_f^c) \leq C r_n. \tag{3.2}$$

Since $nr_n^{d+2} \to 0$, we conclude that

$$\sqrt{nr_n^d}\, \mathbb{E}\lambda_g(\hat{S}_n \cap S_f^c) \to 0 \quad \text{as } n \to \infty. \tag{3.3}$$

Let us now examine the second term in equality (3.1), namely $\mathbb{E}\lambda_g(\hat{S}_n^c \cap S_f)$. To this aim, we introduce the function ψ_n, defined for all $x \in S_f$ by

$$\psi_n(x) = r_n^d \omega_d f(x) + r_n^{d+2} K_n(x),$$

where K_n is a function defined in Lemma A.1 satisfying

$$\sup_n \sup_{x \in S_f} |K_n(x)| < \infty. \tag{3.4}$$

We have

$$\begin{aligned}
\mathbb{E}\lambda_g(\hat{S}_n^c \cap S_f) &= \int_{S_f} \mathbb{P}(x \notin \hat{S}_n) g(x) \mathrm{d}x \\
&= \int_{S_f} \Big[1 - \mathbb{P}\big(X \in \mathcal{B}(x, r_n)\big)\Big]^n g(x) \mathrm{d}x \\
&= \int_{S_f} \big[1 - \psi_n(x)\big]^n g(x) \mathrm{d}x,
\end{aligned}$$

where the last equality follows from Lemma A.1. Denote by (ε_n) a sequence of positive real numbers satisfying $\varepsilon_n \to 0$ and $\sqrt{n r_n^d} \varepsilon_n \to \infty$. Using the notation

$$I = \int_{S_f \cap [f \leq \varepsilon_n]} \big[1 - \psi_n(x)\big]^n g(x) \mathrm{d}x, \tag{3.5}$$

we obtain

$$\begin{aligned}
\big|\mathbb{E}\lambda_g\big(\hat{S}_n^c \cap S_f\big) - I\big| &= \int_{[f > \varepsilon_n]} \big[1 - \psi_n(x)\big]^n g(x) \mathrm{d}x \\
&\leq \int_{[f > \varepsilon_n]} \exp\big[-n\psi_n(x)\big] g(x) \mathrm{d}x,
\end{aligned}$$

where, in the last inequality, we have used the fact that $1 - t \leq \exp(-t)$ for $t \in \mathbb{R}$. This leads, using the definition of $\psi_n(x)$ and (3.4), to

$$\begin{aligned}
\big|\mathbb{E}\lambda_g(\hat{S}_n^c \cap S_f) - I\big| &\leq \exp(-n r_n^d \varepsilon_n \omega_d) \int_{[f > \varepsilon_n]} \exp\big(n r_n^{d+2} |K_n(x)|\big) g(x) \mathrm{d}x \\
&\leq C \exp(-n r_n^d \varepsilon_n \omega_d), \qquad (3.6)
\end{aligned}$$

since $nr_n^{d+2} \to 0$ as $n \to \infty$. Consequently, for n large enough,

$$\begin{aligned}
\sqrt{nr_n^d}\left|\mathbb{E}\lambda_g(\hat{S}_n^c \cap S_f) - I\right| &\leq C\sqrt{nr_n^d}\exp(-nr_n^d\varepsilon_n\omega_d) \\
&\leq C\sqrt{nr_n^d}\exp(-\sqrt{nr_n^d}), \qquad (3.7)
\end{aligned}$$

and this latter term tends to 0 since $nr_n^d \to \infty$. Therefore, we only need to deal with the asymptotic behavior of the term I.

Let $\mathcal{V}(\partial S_f, \rho)$ be a tubular neighborhood of ∂S_f of radius $\rho > 0$, the existence of which follows from the Tubular Neighborhood Theorem under Assumption 1.a. From Assumption 1.b, it follows that the set $[f \leq \varepsilon_n]$ is included in $\mathcal{V}(\partial S_f, \rho)$ for all n large enough. From now on, it is assumed in the remainder of the proof that n is large enough for this inclusion to hold. Next, since n is large enough, for all $p \in \partial S_f$, we define, as in (A.4), $\kappa_p^f(\varepsilon_n)$ as the distance between p and the points x of $[f = \varepsilon_n]$ such that the vector $x - p$ is orthogonal to ∂S_f. To simplify the notation, we write $\kappa_p(\varepsilon)$ instead of $\kappa_p^f(\varepsilon)$, and e_p for the normal vector field instead of e_p^f. From the identity (B.1), and since n is larger enough, it follows that the integral I may be expressed as

$$I = \int_{\partial S_f} I(p)\mathrm{d}v_\sigma(p), \qquad (3.8)$$

where, for all $p \in \partial S_f$, the term $I(p)$ is defined as

$$\begin{aligned}
I(p) &= \int_0^{\kappa_p(\varepsilon_n)} \left[1 - \psi_n(p + ve_p)\right]^n g(p + ve_p)\Theta(p, v)\mathrm{d}v \\
&= \frac{1}{\sqrt{nr_n^d}} \int_0^{\sqrt{nr_n^d}\kappa_p(\varepsilon_n)} \exp\left[n\log\Big(1 - \psi_n\big(p + \frac{u}{\sqrt{nr_n^d}}e_p\big)\Big)\right] \\
&\qquad \times g\big(p + \frac{u}{\sqrt{nr_n^d}}e_p\big)\Theta\big(p, \frac{u}{\sqrt{nr_n^d}}\big)\mathrm{d}u. \qquad (3.9)
\end{aligned}$$

According to Lemma A.2, for n large enough, $\sup_{p\in\partial S_f}\kappa_p(\varepsilon_n)\le\rho$. Applying Lemma A.3, we obtain

$$\begin{aligned} I(p) & \\ &= \frac{1}{\sqrt{nr_n^d}}\int_0^{\sqrt{nr_n^d}\kappa_p(\varepsilon_n)} \exp\Big[-\frac{u^2\omega_d}{2}D^2_{e_p}f(p+\xi e_p)-\frac{u^4\omega_d^2}{8n}\big(D^2_{e_p}f(p+\xi e_p)\big)^2 \\ &\qquad + nr_n^{d+2}R_n(p,u)\Big]g\big(p+\frac{u}{\sqrt{nr_n^d}}e_p\big)\Theta\big(p,\frac{u}{\sqrt{nr_n^d}}\big)\mathrm{d}u, \qquad (3.10)\end{aligned}$$

where

$$\xi=\xi(n,p,u)\in\big(0,\kappa_p(\varepsilon_n)\big),$$

and $R_n(p,u)$ satisfies

$$\sup_n\sup_{p\in\partial S_f}\sup_{0\le u\le\sqrt{nr_n^d}\kappa_p(\varepsilon_n)}|R_n(p,u)|<\infty.$$

Using the fact that, for each $p\in\partial S_f$, $0\le\xi\le\kappa_p(\varepsilon_n)$ and $\sup_{p\in\partial S_f}\kappa_p(\varepsilon_n)\to 0$ as $n\to\infty$ [by Lemma A.2], we are sure that, for n large enough, all points $p+\xi e_p$ fall in $\mathcal{V}(\partial S_f,\rho)$. Consequently, by Assumption 2.e, there exists some $\alpha>0$ independent of n such that, for n large enough,

$$\inf_{p\in\partial S_f} D^2_{e_p}f(p+\xi e_p)\ge\alpha. \qquad (3.11)$$

Thus, the Lebesgue dominated convergence Theorem may be applied to the integral in (3.10). Since, by Lemma A.2, $\sqrt{nr_n^d}\kappa_p(\varepsilon_n)\to\infty$, since g is continuous, and since Θ is $\mathcal{C}^\infty$ with $\Theta(p,0)=1\,\forall p\in\partial S_f$, we obtain, for each $p\in\partial S_f$,

$$\sqrt{nr_n^d}\,I(p)\to\int_0^{+\infty}\exp\Big[-\frac{u^2\omega_d}{2}D^2_{e_p}f(p)\Big]g(p)\mathrm{d}u \quad \text{as } n\to\infty.$$

The limit above is equal to

$$\sqrt{\frac{\pi}{2\omega_d}}\frac{g(p)}{\sqrt{D^2_{e_p}f(p)}}.$$

Using once again inequality (3.11) yields to

$$\sup_n \sup_{p \in \partial S_f} \sqrt{nr_n^d}\, I(p) < \infty.$$

As ∂S_f is compact, it has finite volume, i.e., $v_\sigma(\partial S_f) < \infty$, and we conclude by the Lebesgue Theorem that

$$\sqrt{nr_n^d}\, I = \int_{\partial S_f} \sqrt{nr_n^d}\, I(p) \mathrm{d}v_\sigma(p) \to \int_{\partial S_f} \sqrt{\frac{\pi}{2\omega_d}} \frac{g(p)}{\sqrt{D^2_{e_p} f(p)}} \mathrm{d}v_\sigma(p) \quad \text{as } n \to \infty.$$

Putting together (3.3), (3.7) and the limit above leads to the desired result. $\square$

3.2 Necessary condition on the radius

Theorem 3.2 *Suppose that S_f is the closed unit Euclidean ball of $\mathbb{R}^d$, and that Assumption 2 holds for some $k \in \{1, 2\}$. Then, if $nr_n^{d+k} \to \infty$, as $n \to \infty$,*

$$(nr_n^d)^{1/k}\, \mathbb{E} d_1(\hat{S}_n, S_f) \to \infty.$$

Theorem 3.2 shows that the assumption $nr_n^{d+k} \to 0$ of Theorem 3.1 is sharp. If we restrict ourselves to choices of radius $r_n = \mathrm{O}(n^{-s})$ for some $s > 0$, then the condition of Theorem 3.1 becomes $1/(d+k) < s < 1/d$. In this context, the best convergence rate, which corresponds to values of s close to $1/(d+k)$, must be slower than $\mathrm{O}(n^{1/(d+k)})$.

Proof According to decomposition (3.1), it suffices to prove that

$$(nr_n^d)^{1/k}\, \mathbb{E}\lambda(\hat{S}_n \cap S_f^c) \to \infty.$$

For simplicity, for all $p \in \partial S_f$, we write e_p instead of e_p^f. Denote by S^{d-1} the unit sphere in $\mathbb{R}^d$. Clearly, by (B.1),

$$\begin{aligned}
\mathbb{E}\lambda(\hat{S}_n \cap S_f^c) &= \int_{S_f^c} \mathbb{P}(x \in \hat{S}_n)\mathrm{d}x \\
&= \int_{S^{d-1}} \int_{-r_n}^{0} \Big[1 - \big(1 - p_n(p + ue_p)\big)^n\Big] \Theta(p, u)\mathrm{d}u\mathrm{d}v_\sigma(p),
\end{aligned}$$

where

$$p_n(p + ue_p) = \mathbb{P}\big(\mathrm{dist}(p + ue_p, X) \leq r_n\big),$$

and where X is a random variable with density f. Taking the inner integral from $-r_n/2$ yields the lower bound

$$\mathbb{E}\lambda(\hat{S}_n \cap S_f^c) \geq \int_{S^{d-1}} \int_{-r_n/2}^{0} \Big[1 - \big(1 - p_n(p + ue_p)\big)^n\Big] \Theta(p, u)\mathrm{d}u\mathrm{d}v_\sigma(p).$$

Clearly, for each fixed $p \in S^{d-1}$, the map $[-r_n/2, 0] \ni u \mapsto p_n(p + ue_p)$ is increasing. Thus, for each $u \in [-r_n/2, 0]$ and each $p \in S^{d-1}$, the quantity $p_n(p + ue_p)$ is bounded from below by $p_n(p - (r_n/2)e_p)$, which in turn is bounded from below, and uniformly in p, by a sequence p_n such that $p_n \geq Cr_n^{d+k}$ for some constant $C > 0$ by Lemma A.5. Consequently,

$$\begin{aligned}
\mathbb{E}\lambda(\hat{S}_n \cap S_f^c) &\geq Cv_\sigma\big(S^{d-1}\big)\frac{r_n}{2}\Big[1 - \exp\big(n\log(1 - p_n)\big)\Big] \\
&= \frac{C}{2}v_\sigma\big(S^{d-1}\big)r_n\Big[1 - \exp\Big(-np_n\frac{\log(1 - p_n)}{-p_n}\Big)\Big],
\end{aligned}$$

and so, for n large enough,

$$\Big[1 - \exp\Big(-np_n\frac{\log(1 - p_n)}{-p_n}\Big)\Big] \geq \frac{1}{2},$$

since $nr_n^{d+k} \to \infty$ by assumption. Hence, for large n,

$$\mathbb{E}\lambda(\hat{S}_n \cap S_f^c) \geq \frac{C}{4}v_\sigma\big(S^{d-1}\big)r_n,$$

and thus

$$(nr_n^d)^{1/k}\,\mathbb{E}\lambda(\hat{S}_n \cap S_f^c) \geq \frac{C}{4}v_\sigma\big(S^{d-1}\big)(nr_n^{d+k})^{1/k},$$

from which the result follows. $\square$

4 The case $S_g \subset S_f$

An inspection of the limit term in Theorem 3.1 reveals that

$$(nr_n^d)^{1/k}\, \mathbb{E} d_g(\hat{S}_n, S_f) \to 0 \quad \text{as } n \to \infty,$$

as soon as $\partial S_f \subset (\overset{\circ}{S}_g)^c$. In this case, the rate $(nr_n^d)^{1/k}$ is therefore sub-optimal, and this section aims at investigating the true convergence rate. For the same reason that the case $S_f \cap S_g = \emptyset$ was excluded, the requirement $\partial S_f \subset (\overset{\circ}{S}_g)^c$ means that we can assume that $S_g \subset S_f$. Thus, from now on, this latter condition will be supposed fulfilled. At this stage, two sub-cases, leading to different limit theorems, have to be considered:

(*i*) The case $\partial S_f \cap \partial S_g \neq \emptyset$, and

(*ii*) The case $\partial S_f \cap \partial S_g = \emptyset$.

From a statistical perspective, the sub-case (*i*), which allows for $g \equiv f$, is the most important. Indeed, recall that if X denotes a random variable with density f independent of the sample, the choice $g \equiv f$ yields the criterion

$$\mathbb{E} d_f(\hat{S}_n, S_f) = \mathbb{P}(X \notin \hat{S}_n).$$

However, for the sake of completeness, we will also discuss in detail the sub-case (*ii*).

4.1 The sub-case $\partial S_f \cap \partial S_g \neq \emptyset$

We first introduce some smoothness assumptions on g, depending on a parameter $k \in \{1, 2\}$.

Assumption 3

(a) There exists $\varepsilon > 0$ such that, for all $p \in \partial S_f$, the map $u \mapsto g(p + ue_p^f)$ is of class $\mathcal{C}^k$ on $[0, \varepsilon]$.

(b) There exists $\varepsilon > 0$ such that $\sup_{0 \le u \le \varepsilon} \sup_{p \in \partial S_f} |D^k_{e_p^f} g(p + ue_p^f)| < \infty$.

As explained in Remark 3.2 and Remark 3.3, Assumption 1 may be relaxed.

Theorem 4.1 *Suppose that $\partial S_f \cap \partial S_g \neq \emptyset$, and that Assumption 1, Assumption 2, and Assumption 3 hold for some $k \in \{1, 2\}$. Moreover, suppose that, for all $p \in \partial S_f$, $D^{k-1}_{e_p^f} g(p) = 0$. Then, if $nr_n^d \to \infty$ and $nr_n^{d+k/(k+1)} \to 0$, we have, as $n \to \infty$,*

$$(nr_n^d)^{(k+1)/k}\, \mathbb{E} d_g(\hat{S}_n, S_f) \to \left(\sqrt{\frac{\pi}{8}}\right)^{k-1} \omega_d^{-(k+1)/k} \int_{\partial S_f} \frac{D^k_{e_p^f} g(p)}{\left[D^k_{e_p^f} f(p)\right]^{(k+1)/k}} \mathrm{d}v_\sigma(p).$$

Set $g \equiv f$ and denote by X a random variable with density f independent of the sample. In this case, Theorem 4.1, applied for example with $k = 1$, yields the following simple result

$$(nr_n^d)^2\, \mathbb{P}(X \notin \hat{S}_n) \to \omega_d^{-2} \int_{\partial S_f} \left[D_{e_p^f} f(p)\right]^{-1} \mathrm{d}v_\sigma(p). \tag{4.1}$$

We emphasize that the consistency result (4.1) has interesting statistical consequences regarding the detection problem stated in the Introduction. Indeed, it allows for a control of the asymptotic behavior of the false alarm probability. For example, to guarantee a false alarm level $\alpha \in (0, 1)$ given beforehand, with a radius $r_n \approx 1/n^{1/(d+1/2)}$ (up to a logarithmic factor), the number of observations should approximately satisfy

$$n \approx \left(\frac{1}{\alpha\, \omega_d^2} \int_{\partial S_f} \left[D_{e_p^f} f(p)\right]^{-1} \mathrm{d}v_\sigma(p)\right)^{d+1/2}.$$

Theorem 4.1 may be obtained by recursing to arguments similar to the ones advanced in the proof of Theorem 3.1. For this reason, we only sketch the proof.

Sketch of proof According to (3.1) and (3.2), one only needs to prove that

$$(nr_n^d)^{(k+1)/k}\,\mathbb{E}\lambda_g(\hat{S}_n^c \cap S_f) \to \left(\sqrt{\frac{\pi}{8}}\right)^{k-1} \omega_d^{-(k+1)/k} \int_{\partial S_f} \frac{D^k_{e_p^f} g(p)}{\left[D^k_{e_p^f} f(p)\right]^{(k+1)/k}} \mathrm{d}v_\sigma(p).$$

Denote by (ε_n) a sequence of positive real numbers satisfying $\varepsilon_n \to 0$ and $(nr_n^d)^{1/k}\varepsilon_n \to \infty$. For such an ε_n, let I be defined as in (3.5) for n large enough. Since $nr_n^{d+k/(k+1)} \to 0$, inequality (3.6) remains true. Therefore, we only need to deal with the asymptotic behavior of the term I. Following (3.8), I may be written as

$$I = \int_{\partial S_f} I(p)\mathrm{d}v_\sigma(p),$$

where $I(p)$ is defined by (3.9). For the sake of simplicity, we now consider the case $k = 2$ as in the proof of Theorem 3.1. Then, representation (3.10) of $I(p)$ also holds in this context for n large enough. Since $g(p) = D_{e_p} g(p) = 0$ for all $p \in \partial S_f$, we deduce from Assumption 3 and an expansion of g that for all $p \in \partial S_f$:

$$\begin{aligned} I(p) & \\ &= \frac{1}{\sqrt{nr_n^d}} \int_0^{\sqrt{nr_n^d}\kappa_p(\varepsilon_n)} \exp\Big[-\frac{u^2\omega_d}{2} D^2_{e_p} f(p+\xi e_p) - \frac{u^4\omega_d^2}{8n}\big(D^2_{e_p} f(p+\xi e_p)\big)^2 \\ &\qquad + nr_n^{d+2} R_n(p,u)\Big] \frac{u^2}{2nr_n^d} D^2_{e_p} g(p+\chi e_p)\Theta\big(p, \frac{u}{\sqrt{nr_n^d}}\big)\mathrm{d}u, \end{aligned}$$

where

$$\xi = \xi(n,p,u) \in \big(0, \kappa_p(\varepsilon_n)\big), \quad \chi = \chi(n,p,u) \in \big(0, \kappa_p(\varepsilon_n)\big),$$

and $R_n(p,u)$ satisfies

$$\sup_n \sup_{p \in \partial S_f} \sup_{0 \le u \le \sqrt{nr_n^d}\kappa_p(\varepsilon_n)} |R_n(p,u)| < \infty.$$

Using similar arguments as in the end of proof of Theorem 3.1, we obtain for all $p \in \partial S_f$:

$$(nr_n^d)^{3/2} I(p) \to \frac{1}{2} D_{e_p}^2 g(p) \int_0^\infty u^2 \exp\Big[-\frac{u^2 \omega_d}{2} D_{e_p}^2 f(p) \Big] \mathrm{d}u.$$

The limit above is equal to

$$\sqrt{\frac{\pi}{8}}\, \omega_d^{-3/2} \frac{D_{e_p}^2 g(p)}{\big[D_{e_p}^2 f(p)\big]^{3/2}}.$$

We then conclude as in the proof of Theorem 3.1. □

4.2 The sub-case $\partial S_f \cap \partial S_g = \emptyset$

We introduce the function $\overline{f}$ defined on S_g by

$$\overline{f}(x) = f(x) - \inf_{S_g} f, \quad x \in S_g.$$

The support $S_{\overline{f}}$ of $\overline{f}$ is itself compact. Moreover,

$$\partial S_{\overline{f}} = \{x \in S_g \,:\, f(x) = \inf_{S_g} f\} = \arg\min_{S_g} f_{|S_g}.$$

We will need the following assumptions on f.

Assumption 4

(a) The boundary $\partial S_{\overline{f}}$ of $S_{\overline{f}}$ is a smooth submanifold of $\mathbb{R}^d$ of codimension 1;

(b) The set $[\overline{f} > 0]$ is connected;

(c) $\overline{f} > 0$ on $\mathring{S}_g$.

Assumption 5 There exists $\varepsilon > 0$ such that $\inf_{0 \le u \le \varepsilon} \inf_{p \in \partial S_{\overline{f}}} D_{e_p^{\overline{f}}} f(p + u e_p^{\overline{f}}) > 0$.

By analogy with Assumption 1, Assumption 4 does not hold when the dimension d equals one, but the result remains valid in the sense exposed in Remark 3.1 with $\overline{f}$ in place of f. Furthermore, the simple connectedness assumption of $[\bar{f} > 0]$ may be relaxed as explained in Remark 3.2 and Remark 3.3. Basically, Assumption 5 means that f has no flat part on the boundary of $S_{\overline{f}}$.

Theorem 4.2 *Suppose that $\partial S_f \cap \partial S_g = \emptyset$, and that Assumption 4 and Assumption 5 hold. Then, if $nr_n^d \to \infty$, $nr_n^{d+2} \to 0$, and $nr_n^{2d} \to 0$, we have, as $n \to \infty$,*

$$nr_n^d \exp(nr_n^d \omega_d \inf_{S_g} f)\, \mathbb{E} d_g(\hat{S}_n, S_f) \to \omega_d^{-1} \int_{\partial S_{\overline{f}}} \frac{g(p)}{D_{e_p^{\overline{f}}} f(p)} \mathrm{d}v_\sigma(p).$$

Observe that the limit vanishes when $S_{\overline{f}} = S_g$ since, in such a case, we have $g(p) = 0$ for all $p \in \partial S_{\overline{f}}$. In this context, for a sufficiently smooth g, it is straightforward to improve the result and to obtain the exact rate of convergence, which just differs from above by a power of nr_n^d. We leave the details to the reader.

Note that if Assumption 4 proves useful for establishing Theorem 4.2, it does not allow for those situations where, in dimension 2, the set $\partial S_{\overline{f}} = \arg\min_{S_g} f_{|S_g}$ is finite. However, when this occurs, it is still possible to derive an exponential rate of convergence. For example, suppose that f has no flat part on $\overline{f}^{-1}([0, \varepsilon])$, for some $\varepsilon > 0$. Then one can deduce from the coarea Formula [see Evans and Gariepy (1992)], from some of the arguments used in the proof of Theorem 4.2, and under the same conditions on the radius, that

$$\limsup_n nr_n^d \exp(nr_n^d \omega_d \inf_{S_g} f)\, \mathbb{E} d_g(\hat{S}_n, S_f) \leq C\mathcal{H}\big(\partial S_{\overline{f}}\big),$$

for some $C > 0$. Here, $\mathcal{H}$ stands for the $(d-1)$-dimensional Hausdorff measure on $\mathbb{R}^d$. When $\mathcal{H}(\partial S_{\overline{f}}) = 0$, a situation which occurs for instance when $d \geq 2$ and $\partial S_{\overline{f}}$ is finite, the result reads

$$nr_n^d \exp(nr_n^d \omega_d \inf_{S_g} f)\, \mathbb{E} d_g(\hat{S}_n, S_f) \to 0.$$

Proof We have

$$\begin{aligned}
\mathbb{E} d_g(\hat{S}_n, S_f) &= \mathbb{E}\lambda_g(\hat{S}_n \Delta S_f)\\
&= \mathbb{E}\lambda_g\big(\hat{S}_n^c \cap S_f\big)\\
&\quad \text{(since } S_g \subset S_f\text{)}\\
&= \int_{S_g} \mathbb{P}(x \notin \hat{S}_n) g(x) \mathrm{d}x\\
&= \int_{S_g} \Big[1 - \mathbb{P}\big(X \in \mathcal{B}(x, r_n)\big)\Big]^n g(x) \mathrm{d}x.
\end{aligned}$$

According to Lemma A.6, for all $x \in S_g$,

$$\varphi_n(x) = \mathbb{P}\big(X \in \mathcal{B}(x, r_n)\big) = r_n^d \omega_d f(x) + r_n^{d+2} J_n(x),$$

where the quantity $J_n(x)$ satisfies $\sup_n \sup_{x \in S_g} |J_n(x)| < \infty$. Now, let (ε_n) be a sequence of positive real numbers satisfying $\varepsilon_n \to 0$, $\sqrt{nr_n^d}\varepsilon_n \to \infty$, and denote by I the integral

$$I = \int_{[\overline{f} \leq \varepsilon_n]} \big[1 - \varphi_n(x)\big]^n g(x) \mathrm{d}x.$$

Recalling that $\overline{f}$ is only defined on S_g, we obtain

$$\begin{aligned}
|\mathbb{E} d_g\big(\hat{S}_n, S_f\big) - I| &= \int_{[\overline{f} > \varepsilon_n]} \big[1 - \varphi_n(x)\big]^n g(x) \mathrm{d}x\\
&\leq \int_{[\overline{f} > \varepsilon_n]} \exp\big[-n\varphi_n(x)\big] g(x) \mathrm{d}x\\
&\quad \text{(since } 1 - t \leq \exp(-t) \text{ for } t \in \mathbb{R}\text{)},\\
&\leq v_n^{-1} \exp(-nr_n^d \omega_d \varepsilon_n) \int_{[\overline{f} > \varepsilon_n]} \exp\big(nr_n^{d+2} |J_n(x)|\big) g(x) \mathrm{d}x,
\end{aligned}$$

where

$$v_n = \exp(nr_n^d \omega_d \inf_{S_g} f).$$

Since $nr_n^{d+2} \to 0$, since $\sup_n \sup_{x \in S_g} |J_n(x)| < \infty$, and since g is bounded, we obtain

$$\begin{aligned} v_n |\mathbb{E} d_g(\hat{S}_n, S_f) - I| &\leq C \exp(-nr_n^d \omega_d \varepsilon_n) \\ &\leq C \exp(-\sqrt{nr_n^d}), \end{aligned} \tag{4.2}$$

because $\sqrt{nr_n^d}\varepsilon_n \to \infty$ as $n \to \infty$. Therefore, as $nr_n^d \exp(-\sqrt{nr_n^d}) \to 0$, we only need to focus on the term I.

As in the proof of Theorem 3.1, for n large enough, the set $[\overline{f} \leq \varepsilon_n]$ is contained in a tubular neighborhood of $\partial S_{\overline{f}}$. In this case, any $x \in [\overline{f} \leq \varepsilon_n]$ may be expressed in the form $p + ue_p^{\overline{f}}$, where $p \in \partial S_{\overline{f}}$. For ease of notation, we will write, for $p \in \partial S_{\overline{f}}$, e_p instead of $e_p^{\overline{f}}$ and, for $\varepsilon > 0$, $\kappa_p(\varepsilon)$ instead of $\kappa_p^{\overline{f}}(\varepsilon)$ [recall the definition of $\kappa_p^{\overline{f}}(\varepsilon)$ in (A.4)].

According to Assumption 4 and identity (B.1), we have

$$I = \int_{\partial S_{\overline{f}}} I(p) \mathrm{d}v_\sigma(p), \tag{4.3}$$

where, for all $p \in \partial S_{\overline{f}}$,

$$I(p) = \int_0^{\kappa_p(\varepsilon_n)} \big[1 - \varphi_n(p + ve_p)\big]^n g(p + ve_p) \Theta(p, v) \mathrm{d}v.$$

Using a change of variable leads to the equality

$$I(p) = \frac{1}{nr_n^d} \int_0^{nr_n^d \kappa_p(\varepsilon_n)} \exp\left[n \log\Big(1 - \varphi_n\big(p + \frac{u}{nr_n^d} e_p\big)\Big)\right] g\big(p + \frac{u}{nr_n^d} e_p\big) \Theta\big(p, \frac{u}{nr_n^d}\big) \mathrm{d}u.$$

By Lemma A.7, $\sup_{p\in\partial S_{\overline{f}}}\kappa_p(\varepsilon_n)\to 0$ as $n\to\infty$. Thus, the equality above together with Lemma A.8 show that

$$nr_n^d I(p) = v_n^{-1}\int_0^{nr_n^d\kappa_p(\varepsilon_n)} \exp\big[-u\omega_d D_{e_p}f(p+\xi^{'}e_p) + n(r_n^{2d}+r_n^{d+2})R_n^{'}(p,u)\big] \times g\big(p+\frac{u}{nr_n^d}e_p\big)\Theta\big(p,\frac{u}{nr_n^d}\big)\mathrm{d}u,$$

where $\xi^{'}=\xi^{'}(n,p,u)\in\big(0,\kappa_p(\varepsilon_n)\big)$ and where

$$\sup_n \sup_{p\in\partial S_{\overline{f}}} \sup_{0\le u\le nr_n^d\kappa_p(\varepsilon_n)} |R_n^{'}(p,u)|<\infty.$$

Consequently,

$$nr_n^d v_n I(p) = \int_0^{nr_n^d\kappa_p(\varepsilon_n)} \exp\big[-u\omega_d D_{e_p}f(p+\xi^{'}e_p) + n(r_n^{2d}+r_n^{d+2})R_n^{'}(p,u)\big] \times g\big(p+\frac{u}{nr_n^d}e_p\big)\Theta\big(p,\frac{u}{nr_n^d}\big)\mathrm{d}u. \tag{4.4}$$

We deduce from Lemma A.7 and Assumption 5 that there exists an $\alpha>0$ such that, for n large enough,

$$\inf_{p\in\partial S_{\overline{f}}} D_{e_p}f(p+\xi^{'}e_p)\ge\alpha. \tag{4.5}$$

Recall that g is bounded, that $n(r_n^{2d}+r_n^{d+2})\to 0$, and that Θ is $\mathcal{C}^\infty$ with $\Theta(p,0)=1$. In particular, this implies that the domination condition of Lebesgue Theorem is satisfied by the function under the integral in (4.4). Moreover, g is continuous, and $nr_n^d\kappa_p(\varepsilon_n)\to\infty$ [Lemma A.7]. These facts, together with Lebesgue Theorem show that, for all $p\in\partial S_{\overline{f}}$,

$$nr_n^d v_n I(p)\to\int_0^\infty \exp\big[-u\omega_d D_{e_p}f(p)\big]g(p)\mathrm{d}u = \omega_d^{-1}\frac{g(p)}{D_{e_p}f(p)}.$$

Moreover, using (4.4) and (4.5), we have

$$\sup_n \sup_{p\in\partial S_{\overline{f}}} nr_n^d v_n I(p)<\infty.$$

Since $v_\sigma(\partial S_{\overline{f}}) < \infty$ by compacity of $\partial S_{\overline{f}}$, it follows from Lebesgue Theorem and identity (4.3) that

$$nr_n^d v_n I = \int_{\partial S_{\overline{f}}} nr_n^d v_n I(p) \mathrm{d}v_\sigma(p) \to {\omega_d}^{-1} \int_{\partial S_{\overline{f}}} \frac{g(p)}{D_{e_p} f(p)} \mathrm{d}v_\sigma(p) \quad \text{as } n \to \infty.$$

Finally, using (4.2), we conclude that

$$nr_n^d v_n \, \mathbb{E} d_g\big(\hat{S}_n, S_f\big) \to {\omega_d}^{-1} \int_{\partial S_{\overline{f}}} \frac{g(p)}{D_{e_p} f(p)} \mathrm{d}v_\sigma(p),$$

as desired. □

A Some auxiliary results

A.1 Auxiliary results for the proof of Theorem 3.1

Lemma A.1 *Suppose that Assumption* 1 *and Assumption* $2.a - 2.d$ *hold for some* $k \in \{1, 2\}$. *Then, for all* $x \in S_f$, *there exists a quantity* $K_n(x)$ *such that* $\sup_n \sup_{x \in S_f} |K_n(x)| < \infty$ *and*

$$\mathbb{P}\big(X \in \mathcal{B}(x, r_n)\big) = r_n^d \omega_d f(x) + r_n^{d+k} K_n(x),$$

where X *is a random variable with density* f.

Proof Let us define the set $\mathcal{I}_n$ as

$$\mathcal{I}_n = \{x \in S_f : \mathrm{dist}(x, \partial S_f) > r_n\}.$$

Suppose first that $x \in \mathcal{I}_n$. Since f is twice continuously differentiable on $\overset{\circ}{S}_f$ and $\mathcal{B}(x, r_n) \subset \overset{\circ}{S}_f$, one has, for all $u \in \mathcal{B}(x, r_n)$, by Taylor Formula,

$$f(u) = f(x) + (u - x)^t \nabla f(x) + \frac{1}{2}(u - x)^t \mathrm{H} f(\xi)(u - x),$$

for some $\xi = \xi(x,u)$ in the interior of $\mathcal{B}(x, r_n)$, where $\nabla f(x)$ stands for the gradient of f at the point x. Observe that, by symmetry,

$$\int_{\mathcal{B}(x,r_n)} (u-x)^t \nabla f(x) \mathrm{d}u = 0,$$

so that

$$\mathbb{P}\big(X \in \mathcal{B}(x, r_n)\big) = \int_{\mathcal{B}(x,r_n)} f(u)\mathrm{d}u = r_n^d \omega_d f(x) + r_n^{d+2} J_n(x), \tag{A.1}$$

where

$$J_n(x) = \frac{1}{r_n^{d+2}} \int_{\mathcal{B}(x,r_n)} (u-x)^t \mathrm{H} f(\xi)(u-x)\mathrm{d}u$$

satisfies $\sup_n \sup_{x \in \mathcal{I}_n} |J_n(x)| < \infty$ according to Assumption 2.d.

On the other hand, suppose now that $x \in S_f - \mathcal{I}_n$. By Assumption 1, each $u \in \mathcal{B}(x, r_n) \cap S_f$ may be expressed as $u = p + \alpha e_p^f$, where $p \in \partial S_f$ and $0 \leq \alpha \leq C r_n$. Using Assumption 2.a and Assumption 2.b, we deduce that

$$f(u) = \frac{\alpha^k}{k} D^k_{e_p^f} f(p + \xi e_p^f),$$

for some $\xi \in (0, \alpha)$. But, by Assumption 2.c,

$$\sup_n \sup_{p \in \partial S_f} |D^k_{e_p^f} f(p + \xi e_p^f)| < \infty,$$

so that, consequently,

$$f(u) \leq C r_n^k \tag{A.2}$$

for some constant $C > 0$. Therefore, in this case,

$$\mathbb{P}\big(X \in \mathcal{B}(x, r_n)\big) = \int_{\mathcal{B}(x,r_n) \cap S_f} f(u)\mathrm{d}u \leq C r_n^{d+k}. \tag{A.3}$$

Using (A.1), we can now write, for all $x \in S_f$,

$$\begin{aligned}
\mathbb{P}\big(X \in \mathcal{B}(x, r_n)\big) &= \big[r_n^d \omega_d f(x) + r_n^{d+2} J_n(x)\big] \mathbf{1}_{\mathcal{I}_n}(x) \\
&\quad + \mathbb{P}\big(X \in \mathcal{B}(x, r_n)\big) \mathbf{1}_{\mathcal{I}_n^c}(x) \\
&= r_n^d \omega_d f(x) + r_n^{d+k} K_n(x),
\end{aligned}$$

where K_n is defined, for all $x \in S_f$, by

$$K_n(x) = r_n^{2-k} J_n(x) \mathbf{1}_{\mathcal{I}_n}(x) - r_n^{-k} \omega_d f(x) \mathbf{1}_{\mathcal{I}_n^c}(x) + r_n^{-d-k} \mathbb{P}\big(X \in \mathcal{B}(x, r_n)\big) \mathbf{1}_{\mathcal{I}_n^c}(x).$$

Clearly, K_n satisfies the required condition $\sup_n \sup_{x \in S_f} |K_n(x)| < \infty$ according to (A.2) and (A.3). $\square$

Definition of $\kappa_p^h(\varepsilon)$. Let $\mathcal{D} \subset \mathbb{R}^d$ and $h : \mathcal{D} \to \mathbb{R}_+$ be a function with compact support S_h and smooth boundary ∂S_h. Fix $\varepsilon_0 > 0$, small enough such that there exists a tubular neighborhood of ∂S_h of radius ρ containing the set $[h \leq \varepsilon_0]$. For all $p \in \partial S_h$ and $0 < \varepsilon < \varepsilon_0$, we define $\kappa_p^h(\varepsilon)$ by

$$\kappa_p^h(\varepsilon) = \mathrm{dist}\Big(p, [h = \varepsilon] \cap \big\{x \in \mathbb{R}^d \,:\, x = p + v e_p^h,\, v \in [0, \rho]\big\}\Big). \tag{A.4}$$

In other words, $\kappa_p^h(\varepsilon)$ represents the minimum distance between p and the points x of $[h = \varepsilon]$ such that the vector $x - p$ is orthogonal to ∂S_h. Note that when $h > 0$ on $\overset{\circ}{S}_h$ and ε_0 is small enough, such a point x is unique.

Whenever $h \equiv f$, the behavior of $\kappa_p^f(\varepsilon)$ with respect to p and ε is controlled by the following lemma.

Lemma A.2 *Suppose that Assumption* 1*, Assumption* 2.*a* $-$ 2.*c, and Assumption* 2.*e hold for some* $k \in \{1, 2\}$*. Then,*

$$\sup_{p \in \partial S_f} \kappa_p^f(\varepsilon) \to 0 \quad \textit{as } \varepsilon \to 0.$$

Moreover, if $\varepsilon_0 > 0$ *is small enough, there exists* $C > 0$ *such that, for all* $p \in \partial S_f$*,*

$$\kappa_p^f(\varepsilon) \geq C\varepsilon, \ \forall \varepsilon \in (0, \varepsilon_0).$$

Proof By Assumption 1, there exists a tubular neighborhood of ∂S_f of radius $\rho > 0$. Without loss of generality, ε_0 may be chosen in such a way that the set $[f \le \varepsilon_0] \subset \mathcal{V}(\partial S_f, \rho)$ [This comes from the simple-connectedness of $[f > 0]$ by Assumption 1]. Furthermore, by Assumption 2.c and Assumption 2.e, one can assume that $\sup_{0 \le u \le \rho} \sup_{p \in \partial S_f} |D^k_{e^f_p} f(p + u e^f_p)| < \infty$ and $\inf_{0 \le u \le \rho} \inf_{p \in \partial S_f} D^k_{e^f_p} f(p + u e^f_p) > 0$. Then, for all $\varepsilon \le \varepsilon_0$ and all $p \in \partial S_f$, we have $\kappa^f_p(\varepsilon) \le \rho$.

Observe now that, for all $p \in \partial S_f$ and $\varepsilon \le \varepsilon_0$, $f(p + \kappa^f_p(\varepsilon) e^f_p) = \varepsilon$. Consequently, according to Assumption 2.a and Assumption 2.b, we deduce from Taylor Formula that

$$\varepsilon = f(p + \kappa^f_p(\varepsilon) e^f_p) = \frac{\kappa^f_p(\varepsilon)^k}{k} D^k_{e^f_p} f(p + \xi e^f_p),$$

for some $\xi \in (0, \kappa^f_p(\varepsilon))$. Taking the infimum over ξ, and next, the supremum over p in the above equation, yields the existence of a constant $C > 0$ such that, for all $\varepsilon \le \varepsilon_0$,

$$\varepsilon \ge C \sup_{p \in \partial S_f} \kappa^f_p(\varepsilon)^k.$$

This proves the first statement of the lemma. Finally, using the fact that ρ can be chosen smaller than 1, a similar argument shows that there exists a constant $C > 0$ such that

$$C\varepsilon \le \sup_{p \in \partial S_f} \kappa^f_p(\varepsilon),$$

for all $\varepsilon \le \varepsilon_0$. □

Lemma A.3 *Suppose that Assumption* 1 *and Assumption* 2.$a - 2.d$ *hold for some* $k \in \{1, 2\}$. *Let* $\gamma_0 > 0$ *be small enough. Then, for all* $p \in \partial S_f$ *and*

$0 \le u \le (nr_n^d)^{1/k}\gamma_0$, *one has, with the notation of Lemma A.1,*

$$\log\Big[1 - r_n^d \omega_d f\big(p + \frac{u}{(nr_n^d)^{1/k}} e_p^f\big) - r_n^{d+k} K_n\big(p + \frac{u}{(nr_n^d)^{1/k}} e_p^f\big)\Big]$$
$$= -\frac{u^k \omega_d}{kn} D_{e_p^f}^k f\big(p + \xi e_p^f\big) - \frac{1}{2}\frac{u^{2k}\omega_d^2}{k^2 n^2}\Big[D_{e_p^f}^k f\big(p + \xi e_p^f\big)\Big]^2 + r_n^{d+k} R_n(p,u),$$

for some $\xi = \xi(n,p,u) \in (0,\gamma_0)$ *and some* $R_n(p,u)$ *satisfying*

$$\sup_n \sup_{0 \le u \le (nr_n^d)^{1/k}\gamma_0} \sup_{p \in \partial S_f} \big|R_n(p,u)\big| < \infty.$$

Proof In the sequel, e_p stands for e_p^f and $\gamma_0 > 0$ is chosen such that $p + ue_p \in S_f$ for all $p \in \partial S_f$ and $0 \le u \le \gamma_0$. Let ψ_n be the function defined for all $x \in S_f$ by

$$\psi_n(x) = r_n^d \omega_d f(x) + r_n^{d+k} K_n(x),$$

where $K_n(x)$ is as in Lemma A.1. Using the expansion

$$\log(1-\varepsilon) = -\varepsilon - \frac{\varepsilon^2}{2} + \mathrm{O}(\varepsilon^3),$$

we obtain, for all $p \in \partial S_f$ and $0 \le u \le (nr_n^d)^{1/k}\gamma_0$,

$$\log\Big[1 - \psi_n\big(p + \frac{u}{(nr_n^d)^{1/k}} e_p\big)\Big]$$
$$= -\psi_n\big(p + \frac{u}{(nr_n^d)^{1/k}} e_p\big) - \frac{1}{2}\psi_n^2\big(p + \frac{u}{(nr_n^d)^{1/k}} e_p\big) + r_n^{d+k} R_{n,1}(p,u), \quad \text{(A.5)}$$

where $R_{n,1}(p,u)$ satisfies

$$\sup_n \sup_{0 \le u \le (nr_n^d)^{1/k}\gamma_0} \sup_{p \in \partial S_f} \big|R_{n,1}(p,u)\big| < \infty$$

according to Lemma A.1.

On the one hand, by Assumption 2.a, by Assumption 2.b, and by Taylor Formula, we obtain, for all $p \in \partial S_f$,

$$\begin{aligned}
&\psi_n\big(p + \frac{u}{(nr_n^d)^{1/k}} e_p\big) \\
&= r_n^d \omega_d f\big(p + \frac{u}{(nr_n^d)^{1/k}} e_p\big) + r_n^{d+k} K_n\big(p + \frac{u}{(nr_n^d)^{1/k}} e_p\big) \\
&= \frac{u^k \omega_d}{kn} D_{e_p}^k f(p + \xi e_p) + r_n^{d+k} K_n\big(p + \frac{u}{(nr_n^d)^{1/k}} e_p\big),
\end{aligned} \tag{A.6}$$

for some $\xi = \xi(n, p, u) \in (0, \gamma_0)$.

On the other hand, employing Lemma A.1,

$$\psi_n^2\big(p + \frac{u}{(nr_n^d)^{1/k}} e_p\big) = r_n^{2d} \omega_d^2 f^2\big(p + \frac{u}{(nr_n^d)^{1/k}} e_p\big) + r_n^{d+k} R_{n,2}(p, u),$$

where the quantity $R_{n,2}(p, u)$ satisfies

$$\sup_n \sup_{0 \le u \le (nr_n^d)^{1/k}\gamma_0} \sup_{p \in \partial S_f} \big|R_{n,2}(p, u)\big| < \infty.$$

An application of Taylor Formula leads to

$$\begin{aligned}
\psi_n^2\big(p + \frac{u}{(nr_n^d)^{1/k}} e_p\big) &= r_n^{2d} \omega_d^2 \Big[\frac{u^k}{knr_n^d} D_{e_p}^k f(p + \xi e_p)\Big]^2 + r_n^{d+k} R_{n,2}(p, u) \\
&= \frac{u^{2k} \omega_d^2}{k^2 n^2} \Big[D_{e_p}^k f(p + \xi e_p)\Big]^2 + r_n^{d+k} R_{n,2}(p, u).
\end{aligned} \tag{A.7}$$

Finally, setting

$$R_n(p, u) = R_{n,1}(p, u) - K_n\big(p + \frac{u}{(nr_n^d)^{1/k}} e_p\big) - \frac{1}{2} R_{n,2}(p, u),$$

we deduce from (A.5), (A.6), and (A.7) that

$$\begin{aligned}
&\log\Big[1 - \psi_n(p + \frac{u}{(nr_n^d)^{1/k}} e_p)\Big] \\
&= -\frac{u^k \omega_d}{kn} D_{e_p}^k f\big(p + \xi e_p\big) - \frac{u^{2k} \omega_d^2}{2k^2 n^2} \Big[D_{e_p}^k f\big(p + \xi e_p\big)\Big]^2 + r_n^{d+k} R_n(p, u),
\end{aligned}$$

where $R_n(p, u)$ satisfies

$$\sup_n \sup_{0 \le u \le (nr_n^d)^{1/k}\gamma_0} \sup_{p \in \partial S_f} \big|R_n(p, u)\big| < \infty.$$

□

A.2 Auxiliary results for the proof of Theorem 3.2

We consider in this section a probability density f with support the closed unit Euclidean ball of $\mathbb{R}^d$. Let S^{d-1} be the unit sphere of $\mathbb{R}^d$. The following technical lemma, which is stated without proof, is elementary and may be obtained by tedious but easy calculus.

Lemma A.4 *Let $a \in (0, r_n)$, where $r_n < 1$. Then,*

(i) The trace of $\mathcal{B}(p - ae_p^f, r_n)$ in S^{d-1} is the geodesic ball, further denoted by $\mathcal{B}_\sigma(p, \rho(a, r_n))$, in S^{d-1} of center p and radius $\rho(a, r_n)$ given by:

$$\rho(a, r_n) = \arccos\left(\frac{(1+a)^2 + 1 - r_n^2}{2(1+a)}\right).$$

In particular, $\rho\left(\frac{r_n}{2}, r_n\right) = \arccos\left(1 - \frac{3}{4}\frac{r_n^2}{2+r_n}\right)$ and there exists a constant $C > 0$ such that

$$\rho\Big(\frac{r_n}{2}, r_n\Big) > Cr_n,$$

for all $0 < r_n < 1$.

(ii) Let $\rho_n = \rho\left(\frac{r_n}{2}, r_n\right)$. For all $p \in S^{d-1}$, and for all $q \in \mathcal{B}_\sigma(p, \rho_n)$, the trace of the half-line $\{q + ve_q^f : v \geq 0\}$ in $\partial\mathcal{B}(p - \frac{r_n}{2}e_p^f, r_n)$ is the point $q + \omega(q, r_n)e_q^f$ of $\mathbb{R}^d$, where $\omega(q, r_n) \geq 0$ does not depend on p. For a fixed value of r_n, the map $q \mapsto \omega(q, r_n)$ is a decreasing function of the geodesic distance $d_\sigma(q, p)$ on S^{d-1}. Moreover, there exists a constant $C > 0$ such that

$$\omega(q, r_n) > Cr_n,$$

for all q with $d_\sigma(q, p) \leq \rho_n/2$ and for all $0 < r_n < 1$.

Lemma A.5 *Suppose that Assumption 2.a − 2.e hold for some $k \in \{1, 2\}$. Let $0 < r_n < 1$. For all $p \in S^{d-1}$, let $p_n(p - \frac{r_n}{2}e_p^f) = \mathbb{P}(\mathrm{dist}(p - \frac{r_n}{2}e_p^f, X) \leq r_n)$,* *where X is a random variable with density f. Then there exists a constant $C > 0$ such that*

$$p_n\left(p - \frac{r_n}{2}e_p^f\right) \geq C r_n^{d+k},$$

for all $p \in S^{d-1}$ and for all $0 < r_n < 1$.

Proof In the sequel, for all $q \in S^{d-1}$, e_q stands for e_q^f. Using the notation of Lemma A.4, we have by (B.1),

$$\begin{aligned} p_n\left(p - \frac{r_n}{2}e_p^f\right) &= \int_{\mathcal{B}_\sigma(p,\rho_n)} \int_0^{\omega(q,r_n)} f(q + ue_q)\Theta(q,u)\mathrm{d}u\mathrm{d}v_\sigma(q) \\ &\geq \int_{\mathcal{B}_\sigma(p,\rho_n/2)} \int_0^{\omega(q,r_n)} f(q + ue_q)\Theta(q,u)\mathrm{d}u\mathrm{d}v_\sigma(q). \end{aligned}$$

By Lemma A.4, there exists a constant $C > 0$ such that $\omega(q, r_n) \geq Cr_n$ for all $q \in \mathcal{B}_\sigma(p, \rho_n/2)$, and for all r_n. Consequently,

$$p_n\left(p - \frac{r_n}{2}e_p^f\right) \geq \int_{\mathcal{B}_\sigma(p,\rho_n/2)} \int_0^{Cr_n} f(q + ue_q)\Theta(q,u)\mathrm{d}u\mathrm{d}v_\sigma(q).$$

Now differentiating k-times yields the inequalities

$$\begin{aligned} p_n\left(p - \frac{r_n}{2}e_p^f\right) &\geq \int_{\mathcal{B}_\sigma(p,\rho_n/2)} \int_0^{Cr_n} D_{e_q}^k f(q + \xi e_q)u^k\Theta(q,u)\mathrm{d}u\mathrm{d}v_\sigma(q) \\ &\geq \int_{\mathcal{B}_\sigma(p,\rho_n/2)} D_{e_q}^k f(q + \xi e_q)\frac{C^{k+1}}{k+1}r_n^{k+1}\mathrm{d}v_\sigma(q) \\ &\geq Cv_\sigma\big(\mathcal{B}_\sigma(p,\rho_n/2)\big)r_n^{k+1} \\ &\geq C\omega_\sigma^{d-1}(\rho_n/2)^{d-1}r_n^{k+1}, \end{aligned}$$

where ω_σ^{d-1} is the volume of the geodesic ball in S^{d-1} with radius 1. By Lemma A.4, there exists another constant $C > 0$ such that $\rho > Cr_n$ and $\omega(p_\rho) > Cr_n$. This leads to the desired result. $\square$

A.3 Auxiliary results for the proof of Theorem 4.2

Lemma A.6 *Suppose that $S_g \subset S_f$ and $\partial S_f \cap \partial S_g = \emptyset$. Then, for all $x \in S_g$, there exists a quantity $J_n(x)$ such that $\sup_n \sup_{x \in S_g} |J_n(x)| < \infty$ and*

$$\mathbb{P}\big(X \in \mathcal{B}(x, r_n)\big) = r_n^d \omega_d f(x) + r_n^{d+2} J_n(x),$$

where X is a random variable with density f.

Proof Since $S_g \subset S_f$ and $\partial S_f \cap \partial S_g = \emptyset$, for all $x \in S_g$ and all n large enough, the balls $\mathcal{B}(x, r_n)$ are contained in $\mathring{S}_f$. Recalling equality (A.1), the result is a straightforward consequence from the fact that f is twice continuously differentiable on $\mathring{S}_f$. □

The proofs of Lemma A.7 and Lemma A.8 below are similar to the proofs of Lemma A.2 and Lemma A.3, respectively. Recall that $\kappa_p^{\overline{f}}(\varepsilon)$ is defined in (A.4).

Lemma A.7 *Suppose that $S_g \subset S_f$ and $\partial S_f \cap \partial S_g = \emptyset$, and that Assumption 4 and Assumption 5 hold. Then,*

$$\sup_{p \in \partial S_{\overline{f}}} \kappa_p^{\overline{f}}(\varepsilon) \to 0 \quad \text{as } \varepsilon \to 0.$$

Moreover, if $\varepsilon_0 > 0$ is small enough, there exists $C > 0$ such that, for all $p \in \partial S_{\overline{f}}$,

$$\kappa_p^{\overline{f}}(\varepsilon) \geq C\varepsilon, \ \forall \varepsilon \in (0, \varepsilon_0).$$

Lemma A.8 *Suppose that $S_g \subset S_f$ and $\partial S_f \cap \partial S_g = \emptyset$, and that Assumption 4 holds. Let $\gamma_0 > 0$ be small enough. Then, for all $p \in \partial S_{\overline{f}}$ and $0 \leq u \leq n r_n^d \gamma_0$, one has, with the notation of Lemma A.6,*

$$\begin{aligned}
&\log\Big[1 - r_n^d \omega_d f\big(p + \frac{u}{n r_n^d} e_p^{\overline{f}}\big) - r_n^{d+2} J_n\big(p + \frac{u}{n r_n^d} e_p^{\overline{f}}\big)\Big] \\
&\quad = -r_n^d \omega_d \inf_{S_g} f - \frac{u \omega_d}{n} D_{e_p^{\overline{f}}} f\big(p + \xi' e_p^{\overline{f}}\big) + (r_n^{2d} + r_n^{d+2}) R_n'(p, u),
\end{aligned}$$

for some $\xi' = \xi'(n,p,u) \in (0,\gamma_0)$ *and some* $R_n'(p,u)$ *satisfying*

$$\sup_n \sup_{0\le u\le nr_n^d\gamma_0} \sup_{p\in\partial S_{\overline{f}}} \left|R_n'(p,u)\right| < \infty.$$

B Geometry

Let (M,σ) be a smooth, closed (i.e., compact and without boundary), Riemannian submanifold of $\mathbb{R}^d$, with Riemannian metric σ taken as $\sigma = i^*\delta$, where $i : M \to \mathbb{R}^d$ is the canonical injection, and where δ is the Euclidean metric on $\mathbb{R}^d$, i.e., σ is the pullback of δ by i. The Riemannian volume measure on (M,σ) will be denoted by v_σ.

Let T_pM be the tangent space to M at p, and let TM be the tangent bundle of M. For all $p \in M$, T_pM may be considered as a subspace of $\mathbb{R}^d$ via the canonical identification of $T_p\mathbb{R}^d$ with $\mathbb{R}^d$ itself. Via this identification, the normal space $T_pM^\perp$ to M at p is the orthogonal complement of T_pM in $\mathbb{R}^d$. The normal bundle of M in $\mathbb{R}^d$ is defined by $TM^\perp = \cup_{p\in M} T_pM^\perp$, with bundle projection map $\pi : TM^\perp \to M$ defined by $\pi\langle p,v\rangle = p$, i.e., each element $< p,v >$ of $TM^\perp$ is mapped on p by π.

Now let $\theta : TM^\perp \to \mathbb{R}^d$ be given by $\theta\langle p,v\rangle = p+v$. Also let $TM_\varepsilon^\perp = \{\langle p,v\rangle \in TM^\perp : \|v\| < \varepsilon\}$. Then the *Tubular Neighborhood Theorem* [see e.g., Bredon (1993, p. 93)] states that there exists an $\varepsilon > 0$ such that $\theta : TM_\varepsilon^\perp \to \mathbb{R}^d$ is a diffeomorphism onto the neighborhood $\mathcal{V}(M,\varepsilon) = \{x \in \mathbb{R}^d : \mathrm{dist}(x,M) < \varepsilon\}$ of M in $\mathbb{R}^d$, which is called a *tubular neighborhood* of radius ε of M in $\mathbb{R}^d$.

Denote by λ^d the Lebesgue measure on $\mathbb{R}^d$. On $TM_\varepsilon^\perp$, there is the canonical measure $v_g \otimes \lambda^1$ defined by

$$(v_g \otimes \lambda^1)(B) = \int_{\pi(B)} \lambda^1\big(\pi^{-1}(p)\big)\mathrm{d}v_\sigma(p),$$

for all Borel set $B \subset TM_\varepsilon^\perp$. There is also on $TM_\varepsilon^\perp$ the measure $\theta^*\lambda^d$, i.e., the pullback of λ^d on $\mathbb{R}^d$ by θ. Now let $\Theta \in \mathcal{C}^\infty(TM_\varepsilon^\perp)$ be the function such that $\mathrm{d}(\theta^*\lambda^d) = \Theta\mathrm{d}(v_\sigma \otimes \lambda^1)$. This function satisfies $\Theta(< p, 0 >) = 1$. Then, given an integrable function φ on $\mathbb{R}^d$, its integral on a tubular neighborhood of M with respect to λ^d may be expressed as

$$\begin{aligned}\int_{\mathcal{V}(M,\varepsilon)} \varphi(x)\mathrm{d}\lambda^d(x) &= \int_{TM_\varepsilon^\perp} (\varphi \circ \theta)(< p, v >)\mathrm{d}(v_\sigma \otimes \lambda^1)(< p, v >) \\ &= \int_M \int_{u \in T_pM^\perp : \|u\| < \varepsilon} \varphi(p + u)\Theta(p, u)\mathrm{d}\lambda^1(u)\mathrm{d}v_\sigma(p).\end{aligned}$$

Introducing a unit-norm section $\{e_p : p \in M\}$ of $TM^\perp$, i.e., a continuous unit-norm normal vector field on M, yields the more convenient expression:

$$\int_{\mathcal{V}(M,\varepsilon)} \varphi(x)\mathrm{d}\lambda^d(x) = \int_M \int_{-\varepsilon}^{\varepsilon} \varphi(p + ue_p)\Theta(p, u)\mathrm{d}u\mathrm{d}v_\sigma(p). \tag{B.1}$$

Acknowledgements The authors are indebted to two referees for a very careful reading of the paper and stimulating questions and remarks.

References

[1] Baíllo, A., Cuevas, A. and Justel, A. (2000). Set estimation and non-parametric detection, *Canad. J. Statist.*, Vol. 28, pp. 765–782.

[2] Bredon, G.E. (1993). *Topology and Geometry*, Volume 139 of *Graduate Texts in Mathematics*, Springer-Verlag, New York.

[3] Cadre, B. (2006). Kernel estimation of density level sets, *J. Multivariate Anal.*, Vol. 97, pp. 999–1023.

[4] Chavel, I. (1993). *Riemannian Geometry: A Modern Introduction*, Cambridge University Press, Cambridge.

[5] Chevalier, J. (1976). Estimation du support et du contenu du support d'une loi de probabilité, *Ann. Inst. H. Poincaré Sec. B*, Vol. 12, pp. 339–364.

[6] Cuevas, A. (1990). On pattern analysis in the non-convex case, *Kybernetes*, Vol. 19, pp. 26–33.

[7] Cuevas, A. and Fraiman, R. (1997). A plug-in approach to support estimation, *Ann. Statist.*, Vol. 25, pp. 2300–2312.

[8] Cuevas, A. and Rodríguez-Casal, A. (2004). On boundary estimation, *Adv. Appl. Prob.*, Vol. 36, pp. 340–354.

[9] Devroye, L. and Wise, G. (1980). Detection of abnormal behavior via nonparametric estimation of the support, *SIAM J. Appl. Math.*, Vol. 38, pp. 480–488.

[10] Evans, L.C. and Gariepy, R.F. (1992). *Measure Theory and Fine Properties of Functions*, CRC Press, Boca Raton, Florida.

[11] Gayraud, G. (1997). Estimation of functional of density support, *Math. Methods Statist.*, Vol. 6, pp. 26–46.

[12] Geffroy, J. (1964). Sur un problème d'estimation géométrique, *Publ. Inst. Statist. Univ. Paris*, Vol. 13, pp. 191–210.

[13] Gray, A. (1990). *Tubes*, Addison-Wesley, Redwood City, California.

[14] Grenander, U. (1981). *Abstract Inference*, Wiley, New York.

[15] Härdle, W., Park, B.U. and Tsybakov, A.B. (1995). Estimation of non-sharp support boundaries, *J. Multivariate Anal.*, Vol. 55, pp. 205–218.

[16] Klemelä, J. (2004). Complexity penalized support estimation, *J. Multivariate Anal.*, Vol. 88, pp. 274–297.

[17] Kobayashi, S. and Nomizu, K. (1996). *Foundations of Differential Geometry*, Vol I & II, Second Edition, Wiley, New York.

[18] Korostelev, A.P., Simar, L. and Tsybakov, A.B. (1995). Efficient estimation of monotone boundaries, *Ann. Statist.*, Vol. 23, pp. 476–489.

[19] Korostelev, A.P. and Tsybakov, A.B. (1993a). Estimation of the density support and its functionals, *Problems Inform. Transmission*, Vol. 29, pp. 1–15.

[20] Korostelev, A.P. and Tsybakov, A.B. (1993b). *Minimax Theory of Image Reconstruction*, Lecture Notes in Statistics, Vol. 82, Springer-Verlag, New York.

[21] Mammen, E. and Tsybakov, A.B. (1995). Asymptotical minimax recovery of sets with smooth boundaries, *Ann. Statist.*, Vol. 23, pp. 502–524.

[22] Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters–an excess mass approach, *Ann. Statist.*, Vol. 23, pp. 855–881.

[23] Rényi, A. and Sulanke, R. (1963). Über die konvexe Hülle von n zufällig gewählten Punkten, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, Vol. 2, pp. 75–84.

[24] Rényi, A. and Sulanke, R (1964). Über die konvexe Hülle von n zufällig gewählten Punkten, II, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, Vol. 3, pp. 138–147.

[25] Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J. and Williamson, R.C. (2001). Estimating the support of a high-dimensional distribution, *Neural Comput.*, Vol. 13, pp. 1443–1471.

[26] Tsybakov, A.B. (1997). On nonparametric estimation of density level sets, *Ann. Statist.*, Vol. 25, pp. 948–969.

[27] Walther, G. (1997). Granulometric smoothing, *Ann. Statist.*, Vol. 25, pp. 2273–2299.