Set estimation under convexity type assumptions

Alberto Rodríguez Casal

Departamento de Estatística e Investigación Operativa, Facultade de Matemáticas, 15706 Santiago, A Coruña, Spain

Abstract

The problem of estimating a set S from a random sample of points taken within S is considered. It is assumed that S is r-convex, which means that a ball of radius r can go around from outside the set boundary. Under this assumption, the r-convex hull of the sample is a natural estimator of S. We obtain convergence rates for this estimator under both the distance in measure and the Hausdorff metric between sets. It is also proved that the boundary of the estimator consistently estimates the boundary of S, in Hausdorff's sense.
© 2007 Elsevier Masson SAS. All rights reserved.

Résumé

Dans cet article, nous considérons le problème d'estimer un ensemble S à partir d'un échantillon aléatoire de points de S. Nous supposons que S est r-convexe, c'est à dire, une boule de rayon r peut circuler à l'extérieur de la frontière de l'ensemble. Sous cette hypothèse, l'enveloppe r-convexe de l'échantillon est un estimateur naturel pour S. Nous obtenons des vitesses de convergence pour cet estimateur avec la distance en mesure et la métrique de Hausdorff entre des ensembles. Nous montrons également que la frontière de l'estimateur estime uniformément la frontière de S, dans le sens de Hausdorff.
© 2007 Elsevier Masson SAS. All rights reserved.

1. Introduction

The problem of approximating a compact set $S \subset \mathbb{R}^d$ from a random sample of points $X_1, \ldots, X_n$ taken in S arises in connection with some applications in statistical quality control (Devroye and Wise [8], Baíllo, Cuevas and Justel [2]) clustering (Cuevas, Febrero and Fraiman [4,5]), image analysis (Marr [13], Rudemo and Stryhn [14], Bertholet, Rasson and Lissoir [3]) and statistical learning (Tsybakov [17]). This problem can be established in a more formal way as the problem of estimating the support of an absolutely continuous probability measure P_X from n independent observations drawn from P_X.

In order to assess the performance of the estimators we will consider the two usual metrics between sets $S, T \subset \mathbb{R}^d$: the *distance in measure* $d_\nu(S, T) = \nu(S \triangle T)$ (where ν is a measure in $\mathbb{R}^d$, typically the Lebesgue measure, μ, and $\triangle$ stands for the symmetric difference between sets) and the *Hausdorff metric* (see, e.g., Edgar [10]), d_H, defined by

$$d_H(S, T) = \max\left\{\sup_{x \in T} d(x, S), \sup_{y \in S} d(y, T)\right\} = \inf\{\epsilon > 0: T \subset B(S, \epsilon), S \subset B(T, \epsilon)\},$$

where $d(x, S) = \inf\{\|x - y\|: y \in S\}$, $B(S, \epsilon) = \{x \in \mathbb{R}^d: d(x, S) \leqslant \epsilon\}$ and $\|\cdot\|$ denotes the Euclidean norm.

Hausdorff's distance quantifies the *physical (or visual) proximity* between two sets whereas the distance in measure is useful to quantify their similarity *in content*. However, neither of these distances are completely useful for measuring the similarity between the *shape* of two sets. Recently (see Baíllo and Cuevas [1], Cuevas and Rodríguez-Casal [7]) the Hausdorff's distance between boundaries, $d_H(\partial S, \partial T)$, has been also used to evaluate the performance of the estimators. This idea will be also considered here. In addition, the identification of the boundary is an interesting problem on its own which appears quite often in the set estimation literature (see, e.g., Korostelev and Tsybakov [12]).

The case where S is assumed to be convex has received a special attention. In this case, there is a quite obvious estimator: the convex hull of the sample $H_n = \text{conv}(X_1, \ldots, X_n)$, that is, the smallest convex set which contains the sample; see e.g. Schneider [15] and Weil and Wieacker [20].

The asymptotic analysis of $d_H(S, H_n)$ for general dimension d is done in Dümbgen and Walther [9]. These authors show that in general, with probability one, $d_H(S, H_n) = O(\log n / n)^{1/d}$. Under some smoothness conditions on the boundary ∂S the convergence rate is faster as the exponent $1/d$ in the previous expression can be replaced by $2/(d+1)$.

In this paper we will consider a shape condition which is much less restrictive than convexity and in some sense generalizes it. We will suppose, following Walther's [19] terminology, that the set S is r-convex, that is, it exists $r > 0$, such that S can be expressed as the intersection of the complements of all open balls of radius r which are disjoint with S; see the precise definition in Section 2. This can be seen as an extension of the concept of convexity; if we let r tend to infinity, we get the convex case in the limit, since a convex set can be expressed as the intersection of the closed halfspaces which contain it. This shape condition does not only generalize the notion of convexity but it also has a clear geometric interpretation as it is equivalent to a ball of radius rolls freely in the complement of S, that is, all boundary points of S are reached by a closed ball of radius r that does not meet the interior of S. So, these sets, as opposed to what happens with convex sets, could have a boundary with concave areas as long as these are smooth.

As in literature on convexity, we will define the support estimator as the smallest r-convex set which contains the sample. Its behaviour with respect to both d_μ and d_H will be analyzed. We will show that the corresponding convergence rates are the same as those obtained for the convex case, see Dümbgen and Walther [9].

The paper is organized as follows: in Section 2, the concept of r-convex set is defined and briefly analyzed. Some notation is also established. The estimator is defined in Section 3. The main results are established in Section 4. As in Dümbgen and Walther [9], almost sure d_H-convergence rates for the proposed estimator are obtained. It is shown that, if the boundary of S is smooth enough the convergence rates are faster than those obtained in the nonsmooth case. All proofs are deferred to Section 5.

2. Notation and preliminary results

We will assume throughout that $S \subset \mathbb{R}^d$. The Euclidean norm and the corresponding inner product will be denoted by $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$, respectively, and μ will represent the d-dimensional Lebesgue measure. The closed and open ball with center x and radius r will be denoted by $B(x, r)$ and $\overset{\circ}{B}(x, r)$. If A is a subset of $\mathbb{R}^d$, then A^c, $\bar{A}$, $\text{int}(A)$ and ∂A denote the complement, closure, interior and boundary of A, respectively. Finally B and $\overset{\circ}{B}$ will stand respectively by $B(0, 1)$ and $\overset{\circ}{B}(0, 1)$.

A crucial concept to be used in this paper is the following notion of r-convexity (see Walther [19]):

Definition 1. The r-convex hull, with $r > 0$, of A is defined as

$$C_r(A) = \bigcap_{\{H_r \in \mathcal{H}_r: \ A \subset H_r\}} H_r,$$

where $\mathcal{H}_r = \{(\overset{\circ}{B}(x, r))^c: x \in \mathbb{R}^d\}$.

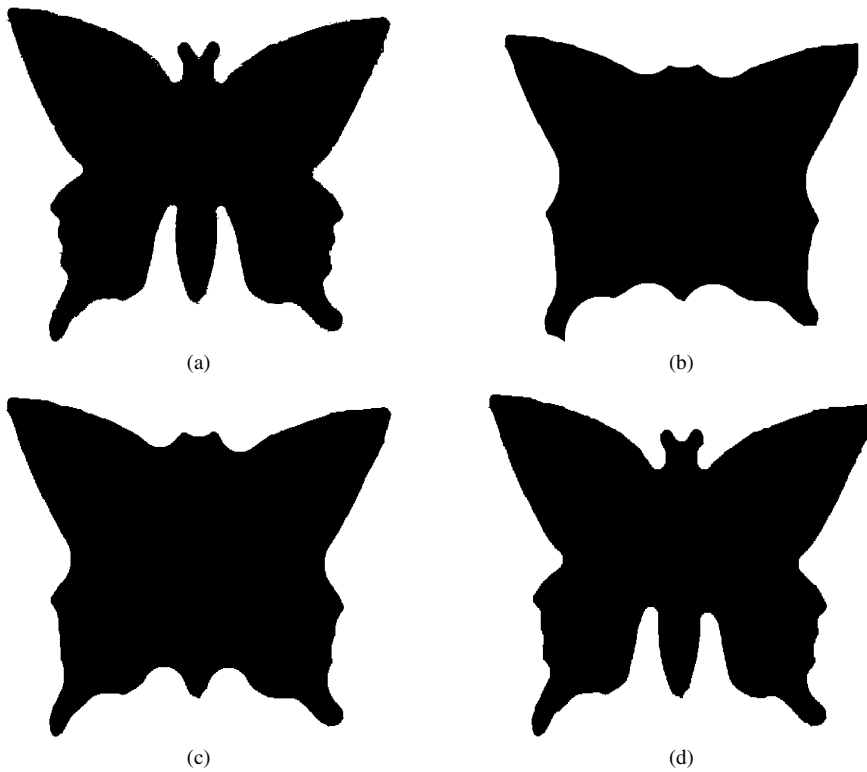The set $A \subset \mathbb{R}^d$ is called r-convex if $A = C_r(A)$.

Fig. 1. (a) The original figure; (b) 1/6-hull; (c) 1/20-hull; (d) 1/60-hull.

The r-convex hull of a set is closely related to the notion of alpha-shapes that arises in the literature of computational geometry. In Edelsbrunner and Mücke [11] alpha-shapes were introduced in order to describe the notion of shape of a point set. In fact, for large r, $C_r(A)$ becomes similar to the convex hull of A whereas it approaches to $\bar{A}$ for small values of r. This is illustrated in Fig. 1.

Apart from the existing relationships with the notion of convexity, the concept of r-convex set is closely related to the concepts of contraction and dilation from mathematical morphology (see, e.g., Serra [16]). It is easy to see that

$$C_r(A) = \left(A \oplus r\ \mathring{B}\right) \ominus r\ \mathring{B}, \tag{1}$$

where the symbols $\oplus$, $\ominus$ denote the Minkowski sum and difference, defined respectively by $A \oplus C := \{a + c \colon a \in A,\ c \in C\}$, $A \ominus C := \{x \colon \{x\} \oplus C \subset A\}$, $\lambda C := \{\lambda c \colon c \in C\}$ for any $A, C \subset \mathbb{R}^d$ and $\lambda \in \mathbb{R}$.

The family of r-convex sets is quite large. In order to obtain the convergence rates of the estimator proposed in this paper, it will be necessary to establish some additional condition concerning the shape of the set S. In particular, we will require the set to be *standard*. A Borel set $S \subset \mathbb{R}^d$ is said to be *standard* with respect to a Borel measure ν if there exists $\epsilon_0 > 0$ and $\delta > 0$ such that

$$\nu\left(B(x, \epsilon) \cap S\right) \geqslant \delta \mu\left(B(x, \epsilon)\right), \quad \forall x \in S,\ 0 < \epsilon \leqslant \epsilon_0.$$

Roughly speaking, this condition prevents the set S from being too spiky. For example, in Dümbgen and Walther [9], it is proved (but without mentioning this concept) that any convex set with nonempty interior is standard with respect to the uniform distribution in that set. Therefore, we may consider that the family of r-convex standard sets (with respect to the uniform distribution) leads to an extension of the notion of convex set without the smoothness restrictions considered in Dümbgen and Walther [9]. For more details on the use of standardness in set estimation see, for example, Cuevas and Fraiman [6].

The smoothness restriction considered in this paper will be the same as in Walther [18], that is, we will assume that S for some $r_0 > 0$

$$S = (S \oplus r_0 B) \ominus r_0 B = (S \ominus r_0 B) \oplus r_0 B. \tag{2}$$

The sets fulfilling (2) are said to belong to Serra's regular model (see Serra [16]). The condition $S = (S \oplus r_0 B) \ominus r_0 B$ is related to r-convexity whereas the second equality in (2) is an additional assumption that ensures r-convexity for the complement of S. In particular, it prevents the presence of convex corners in the boundary of S. For example, the unit square is a convex set (and, therefore, r-convex) that doesn't belong to Serra's regular model.

The Serra's regular model has been extensively analyzed in Walther [19]. This author has established some deep and valuable characterizations that, given its importance in the rest of the paper, we are going to state here in precise terms. First, we will need to establish a geometrical concept with an obvious intuitive interpretation. A ball of radius r is said to roll freely in a closed set A if for each point $a \in \partial A$ there exists $x \in \mathbb{R}^d$ such that $a \in B(x, r) \subset A$. The precise statement of the theorem is as follows.

Theorem 1. (*Walther* [19]) *Let $S \subset \mathbb{R}^d$ be a nonempty compact set and $r > 0$. Then, the following conditions are equivalent*

(1) *The condition*

$$S = (S \ominus \lambda B) \oplus \lambda B, \quad 0 \leqslant \lambda \leqslant r, \qquad S = (S \oplus \lambda B) \ominus \lambda B, \quad 0 \leqslant \lambda < r,$$

holds.
(2) *S and $\overline{S^c}$ are r-convex and $\mathrm{int}(S_i) \neq \emptyset$ for each path-connected component $S_i \subset S$.*
(3) *A ball of radius λ rolls freely inside S and $\overline{S^c}$ for all $0 \leqslant \lambda \leqslant r$.*
(4) *∂S is a $(d-1)$-dimensional C^1 submanifold with the outward pointing unit normal vector $\eta(s)$ at $s \in \partial S$ satisfying the condition*

$$\big\| \eta(s) - \eta(t) \big\| \leqslant \frac{1}{r} \| s - t \|, \quad s, t \in \partial S.$$

(5) *S belongs to Serra's regular model.*

The proof of this theorem can be found in Walther [19]. It should be noted that some of the conditions in this theorem could also be partially used to characterize the r-convex sets in general. For instance, it is not hard to prove that S is r-convex if and only if a ball of radius r rolls freely in $\overline{S^c}$.

3. The estimator

Eq. (1) suggests a natural way to propose a estimator for the support S of a distribution P_X when S is r-convex. Using the same ideas as those in the convex case a reasonable estimator of S is the r-convex hull of the sample $\mathcal{X}_n := \{X_1, \ldots, X_n\}$ which can be expressed as

$$C_r(\mathcal{X}_n) := \left(\mathcal{X}_n \oplus r \, \mathring{B} \right) \ominus r \, \mathring{B} . \tag{3}$$

Notice that if S is r-convex then $C_r(\mathcal{X}_n) \subset S$. We are implicitly supposing that the value of r is known and, in practice, this could not be true. However, this is not a very important problem since it is easy to prove that if S is r-convex then S is also λ-convex for all $\lambda \leqslant r$. Therefore, the only situation that we should worry about is when a value r_0 larger that the true parameter r is chosen. Nevertheless, even in this case, our estimator makes sense because $C_{r_0}(\mathcal{X}_n)$ should approximate $C_{r_0}(S)$. Intuitively r_0 would represent the chosen level of resolution to observe S as $C_{r_0}(S)$ could be considered as a smoothed version of S. In any case, when the value of r is not known, $r \equiv r_n$ with $r_n \to 0$ could be selected to guarantee that the set is r_n-convex for n large enough. The case of unknown r will be considered in Section 4.

In Fig. 2 it is shown the estimator for four different values of r from a uniform sample of size 1000 in the butterfly (see Fig. 1(a)). Note that the estimator is more irregular when r becomes smaller. We think that the role of this *smoothing factor* should be analyzed in greater detail but the problem of optimal selection of r is beyond the scope of the present paper. In the example for $r = 1/60$ the estimator with $n = 10\,000$ seems quite reasonable (see Fig. 3).
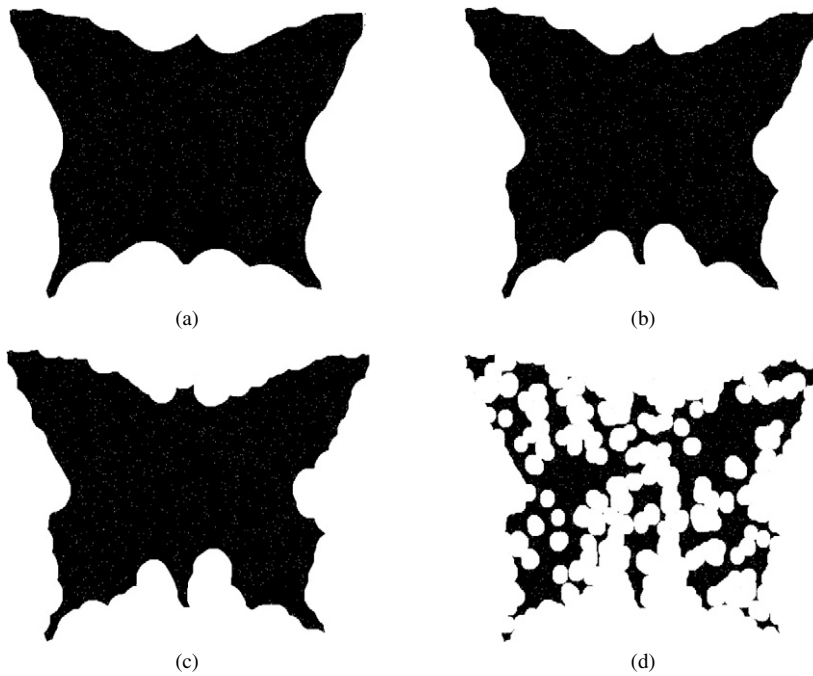
Fig. 2. The proposed estimator for $n = 1000$ and (a) $r = 1/12$; (b) $r = 1/20$; (c) $r = 1/30$; (d) $r = 1/60$.
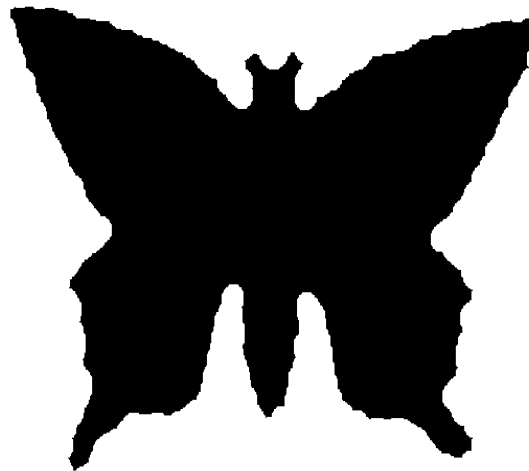

Fig. 3. The proposed estimator for $r = 1/60$ and $n = 10\,000$.

4. Main results

In this section we will establish the main results. Firstly, we will present a general result that will allow us to obtain as corollaries the almost sure convergence rates in Hausdorff distance for the r-convex hull of the sample. We will see that our results extend those obtained by Dümbgen and Walther [9] in the convex case.

Theorem 2. *Let* $X_1, \ldots, X_n, \ldots$ *be independent identically distributed* (*iid*) *random variables whose common distribution,* P_X, *has support* S. *Assume that* S *is compact and standard with respect to* P_X. *Let* $\widehat{S}_n$ *a estimator of* S *satisfying, with probability one, for* n *large enough,*

$$\mathcal{X}_n \subset \widehat{S}_n \subset S.$$

Then, with probability one

$$d_H\big(S, \widehat{S}_n\big) = O\bigg(\bigg(\frac{\log n}{n}\bigg)^{1/d}\bigg).$$

Remark 1. The theorem above allows us to conclude that, if S is r-convex and standard with respect to P_X, then, with probability one, $d_H(S, C_r(\mathcal{X}_n)) = O((\log n/n)^{1/d})$. Although, the family of r-convex set is much wider than the convex family, $C_r(\mathcal{X}_n)$ achieves the *same* convergence rates as $\text{conv}(X_1, \ldots, X_n)$ in the general convex case.

Remark 2. Theorem 2 allows us to show that if S is a convex set with nonempty interior and the sample distribution is uniform on S then both $d_H(S, C_r(\mathcal{X}_n))$ and $d_H(S, \text{conv}(X_1, \ldots, X_n))$ are of order $(\log n/n)^{1/d}$. Therefore, Theorem 2 not only shows that the r-convex hull of the sample reaches the same converge rates as the convex hull but it also provides us with an alternative proof of the result obtained for the convex hull in Dümbgen and Walther [9]. Note that this alternative proof does not use the properties of the convex sets.

The rates obtained in the above mentioned theorem can be improved if we assume that the support belongs to Serra's regular model. This is established in the following theorem which constitutes the main result in this paper.

Theorem 3. *Let* $X_1, X_2, \ldots$ *be a sequence of iid* $\mathbb{R}^d$-*valued random vectors whose common distribution,* P_X, *is absolutely continuous with respect to the Lebesgue measure,* μ. *Let* f *be the density of* P_X *and* S *the support of* f. *Assume that* S *is under the conditions of Theorem* 1. *Assume, also, that* f *is essentially bounded (in S) from below by a positive constant. Then with probability one*

$$d_H\big(S, C_r(\mathcal{X}_n)\big) = O\bigg(\frac{\log n}{n}\bigg)^{2/(d+1)}.$$

The same convergence order holds for $d_H(\partial S, \partial C_r(\mathcal{X}_n))$ *and* $d_\mu(S, C_r(\mathcal{X}_n))$.

Remark 3. Notice that the smoothness condition of S that appears in Dümbgen and Walther [9] is, by Theorem 1, exactly the same that we have imposed here. In particular the estimator proposed here would reach the same rates as those achieved by the convex hull under this smoothness condition when the support is convex (see Corollary 1 in Dümbgen and Walther [9]).

Remark 4. Different methods of estimation of an r-convex set that belongs to Serra's regular model are also proposed in Walther [18], obtaining the same convergence rates both for Hausdorff distance and distance in measure. However, note that Walther [18] deals with level set estimation while we are concerned with support estimation. The motivation of both problems is considerably different. The interest in level sets arises almost exclusively from statistical motivations while support estimation can be seen as a problem in stochastic geometry: we want to recover a set from random points *inside* it. This distinctive feature enhances important technical differences (for example in the construction of estimators) that prevent us from considering the support estimation as a "particular" case of level sets estimation. In fact the support estimation problem is not included in Theorem 3 in Walther [18]. However, Theorem 4 in that paper includes the support as a limit case which leads to suboptimal results. For example, the convergence rates (for distance in measure) given here for the important case $d = 2$ are better than those obtained with Walther's estimator and convergence rates in Hausdorff distance for support estimation are not obtained in Walther [18]. Let us also note that Walther [18] does not consider the boundary estimation problem.

To conclude this section we will analyze the case when S belongs to Serra's regular model, the value of r is unknown and $C_{r_n}(\mathcal{X}_n)$, with $r_n \to 0$, is used to estimate S. Rates will be almost the same in this case except for a penalty factor of order r_n^{-1}, provided that r_n does not converge too fast to zero. We will use the notation $a_n \gg b_n$ to indicate that the sequence $\{a_n\}$ is asymptotically larger than the sequence $\{b_n\}$, that is, $\lim a_n/b_n = \infty$.

Theorem 4. *Under the same conditions of Theorem* 3 *let* $\{r_n\}$ *be a sequence of positive numbers that converge to zero such that* $r_n \gg (\log n/n)^{1/d}$. *Then with probability one*

$$d_H\big(S, C_{r_n}(\mathcal{X}_n)\big) = \mathrm{O}\left(r_n^{-1}\left(\frac{\log n}{n}\right)^{2/(d+1)}\right).$$

The same convergence order holds for $d_H(\partial S, \partial C_r(\mathcal{X}_n))$ *and* $d_\mu(S, C_r(\mathcal{X}_n))$.

5. Proofs

Proof. (Theorem 2) By assumption, with probability one, for n large enough, $\mathcal{X}_n \subset \widehat{S}_n \subset S$ and therefore $d_H(S, \widehat{S}_n) \leqslant d_H(\mathcal{X}_n, S)$. Now the result is an immediate consequence of Theorem 3 in Cuevas and Rodríguez-Casal [7] where it is proved that if S is compact and standard with respect to P_X then $d_H(\mathcal{X}_n, S) = \mathrm{O}((\log n/n)^{1/d})$. $\quad\square$

Proof. (Theorem 3) We will essentially follow the outline of the proof of Corollary 1 in Dümbgen and Walther [9]. However notice that due to r-convexity assumption some intermediate results will require more elaborated proofs. These intermediate results will be established as auxiliary lemmas and propositions. The fundamental results in this proof are Propositions 1 and 2. The theorem is an immediate consequence of both. Before proving these results, we are going to present (at a heuristic level) the ideas of their proofs.

As $C_r(\mathcal{X}_n) \subset S$ the Hausdorff distance (or the distance in measure) between the estimator and the set is essentially determined when the $S \backslash C_r(\mathcal{X}_n)$ points are identified. These are the points of S that are contained in a open ball of radius r which does not meet the sample. It is obvious that, if n is large enough, these balls must have their center outside S (notice that S is the support of the observed distribution).

Let us assume for a while an ideal situation (which in fact is outside the theorem hypothesis) where the boundary of S is a hyperplane and there is a subset, $\mathcal{Z}_n^*$, of sample points which satisfies $\mathcal{Z}_n^* \subset \partial S \subset B(\mathcal{Z}_n^*, \ell)$, for some $\ell > 0$.

An elementary geometric reasoning shows that any ball of radius r which meets S (from outside) without touching the sample points gets into S a distance of order ℓ^2, see Fig. 4. In particular $d_H(S, C_r(\mathcal{X}_n)) = \mathrm{O}(\ell^2)$. The same hold for the other distances considered in the statement. In Proposition 1 we show that (for smooth r-convex sets) the convergence order does not change if we replace $\mathcal{Z}_n^*$ by $\mathcal{Z}_n = \{X_i \in \mathcal{X}_n : d(X_i, S) \leqslant \ell^2\}$.

The reasoning above is based on the condition $\partial S \subset B(\mathcal{Z}_n, \ell)$. The aim of Proposition 2 is to ensure that this condition is met for a suitable sequence of ℓ's. More precisely it is proved that if we take ϵ_n of the same order as the one selected in Dümbgen and Walther [9] for convex sets we can also guarantee that, with probability one, for n large enough, $\partial S \subset B(\mathcal{Z}_n, 2\epsilon_n)$, which is what we need, by virtue of Proposition 1, to conclude that the convergence order of all the distances considered in the statement of the theorem is $\mathrm{O}(\epsilon_n^2)$.
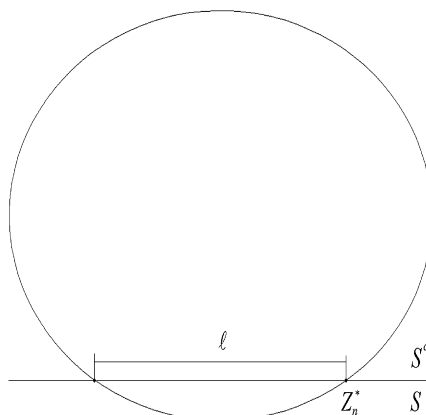

Fig. 4. Heuristic interpretation.

Proposition 1. *Let S be a set under the conditions of Theorem* 1. *Let $\epsilon_n \to 0$ satisfying*

$$P\big(\partial S \subset B(\mathcal{Z}_n, 2\epsilon_n), \ eventually\big) = 1,$$

where $\mathcal{Z}_n = \{X_i \in \mathcal{X}_n: \ d(X_i, \partial S) \leqslant \epsilon_n^2\}$. Then, with probability one, $d_H(S, C_r(\mathcal{X}_n))$, $d_H(\partial S, \partial C_r(\mathcal{X}_n))$ and $d_\mu(S, C_r(\mathcal{X}_n))$ are of order ϵ_n^2.

Proof. (Proposition 1) The proof of this proposition is based in three lemmas. The first is a very useful geometric lemma.

Lemma 1. *Let S be a set under the conditions of Theorem* 1. *Let $y \in S^c$ such that $d(y, S) = r - \delta$ for some $0 \leqslant \delta < r$. If $x \in S$ satisfies $d(x, \partial S) \leqslant \delta/2$ and $\|x - y\| \geqslant r$, then the metric projection s of y onto S satisfies*

$$\|x - s\| \geqslant \sqrt{\frac{r\delta}{2}}.$$

Proof. Let η be the outer normal vector of S in the point $s \in \partial S$ (see statement (4) in Theorem 1). It is not hard to show that

$$y = s + (r - \delta)\eta,$$

and

$$B(t, r) \subset S \quad \text{with } t = s - r\eta.$$

Note that by the assumptions on x

$$r^2 \leqslant \|x - y\|^2 = \big\|x - s - (r - \delta)\eta\big\|^2 = \|x - s\|^2 - 2(r - \delta)\langle x - s, \eta\rangle + (r - \delta)^2,$$

$$\left(r - \frac{\delta}{2}\right)^2 \leqslant \|x - t\|^2 = \|x - s + r\eta\|^2 = \|x - s\|^2 + 2r\langle x - s, \eta\rangle + r^2.$$

The second inequality follows from $d(x, \partial S) \leqslant \delta/2$ and $B(t, r) \subset S$. These two inequalities may be rewritten as

$$\|x - s\|^2 - 2(r - \delta)\langle x - s, \eta\rangle \geqslant 2r\delta - \delta^2,$$

$$\|x - s\|^2 + 2r\langle x - s, \eta\rangle \geqslant -r\delta + \frac{\delta^2}{4}.$$

Adding the first inequality times r to the second inequality times $r - \delta$ and dividing the resulting inequality by $2r - \delta$ yields

$$\|x - s\|^2 \geqslant \frac{2r^2\delta - r\delta^2 - r(r - \delta)\delta + (r - \delta)\delta^2/4}{2r - \delta} = \frac{r^2\delta + (r - \delta)\delta^2/4}{2r - \delta} \geqslant \frac{r\delta}{2},$$

which concludes the proof of the lemma. $\square$

The next lemma is a straightforward consequence of the preceding one. It states in precise terms the heuristic fact mentioned at the beginning of the proof of this theorem: a ball that does not meet any sample point cannot go inside the set too much.

Lemma 2. *Let us assume that $\partial S \subset B(\mathcal{Z}_n, 2\epsilon_n)$. If $y \in S^c$ satisfies that $d(y, S) = r - \delta$ with $\max(2, 8/r)\epsilon_n^2 < \delta < r$ then*

$$\mathring{B}(y, r) \cap \mathcal{X}_n \neq \emptyset.$$

Proof. Let s be the metric projection of y onto S. The assumption $\partial S \subset B(\mathcal{Z}_n, 2\epsilon_n)$ implies that there is $X_i \in \mathcal{Z}_n$ such that $\|X_i - s\| \leqslant 2\epsilon_n$. Remember that $d(X_i, \partial S) \leqslant \epsilon_n^2 < \delta/2$. Therefore if $\|X_i - y\| \geqslant r$, by Lemma 1, we would have

$$\|X_i - s\| \geqslant \sqrt{\frac{r\delta}{2}} > 2\epsilon_n,$$

which is a contradiction. So we have proved that $\|X_i - y\| < r$ which concludes the proof of the lemma. $\square$

The following lemma gives a "*lower bound*" for $C_r(\mathcal{X}_n)$.

Lemma 3. *Let us assume that $\partial S \subset B(\mathcal{Z}_n, 2\epsilon_n)$ and $d_H(\mathcal{X}_n, S) < r$. Then*

$$S \ominus (L\epsilon_n^2 B) \subset C_r(\mathcal{X}_n), \quad \text{for } L \geqslant \max(2, 8/r).$$

Proof. Let us suppose that there is $x \in S \ominus (L\epsilon_n^2 B)$ which does not belong to $C_r(\mathcal{X}_n)$. In this case, there is $y \in S^c$ such that $\overset{\circ}{B}(y, r) \cap \mathcal{X}_n = \emptyset$ and $x \in \overset{\circ}{B}(y, r)$. By Lemma 2, we know that $d(y, S) \geqslant r - \max(2, 8/r)\epsilon_n^2$. Let $z \in [x, y] \cap \partial S$ where $[x, y]$ denotes the segment between the points x and y. We have

$$\|x - z\| \geqslant L\epsilon_n^2,$$
$$\|z - y\| \geqslant r - \max(2, 8/r)\epsilon_n^2,$$

which leads to a contradiction since

$$r > \|x - y\| = \|x - z\| + \|z - y\| \geqslant r - \max(2, 8/r)\epsilon^2 + L\epsilon^2 \geqslant r.$$

Therefore, every $x \in S \ominus (L\epsilon_n^2 B)$ is in $C_r(\mathcal{X}_n)$. $\quad \square$

Now we are in position to complete the proof of Proposition 1. Lemma 3 ensures that with probability one (for n large enough)

$$S \ominus L\epsilon_n^2 B \subset C_r(\mathcal{X}_n) \subset S, \quad L \geqslant \max(2, 8/r). \tag{4}$$

By Theorem 1 the set S satisfies (for $\epsilon \leqslant r$) the condition

$$S = (S \ominus \epsilon B) \oplus \epsilon B. \tag{5}$$

If $L\epsilon_n^2 \leqslant r$ we have

$$S = (S \ominus L\epsilon_n^2 B) \oplus L\epsilon_n^2 B \subset C_r(\mathcal{X}_n) \oplus L\epsilon_n^2 B.$$

Hence (as $C_r(\mathcal{X}_n) \subset S$, by r-convexity of S)

$$d_H\big(S, C_r(\mathcal{X}_n)\big) \leqslant L\epsilon_n^2.$$

The distance in measure can be bounded in a similar way. From (4) we have

$$\mu\big(C_r(\mathcal{X}_n) \triangle S\big) \leqslant \mu\big(S \backslash (S \ominus L\epsilon_n^2 B)\big).$$

As $\mu(S) - \mu(S \ominus \epsilon B) = \mathrm{O}(\epsilon)$ (see Eq. (6) in Walther [18]) we immediately get that $\mu(C_r(\mathcal{X}_n) \triangle S) = \mathrm{O}(\epsilon_n^2)$.

The bound for the distance between the boundaries can be also easily found from (4) and (5). $\quad \square$

Proposition 2. *For some $c > 0$ large enough*

$$P\big(\partial S \subset B(\mathcal{Z}_n, 2\epsilon_n), \text{ eventually}\big) = 1, \tag{6}$$

where $\epsilon_n = (c \log n / n)^{1/(d+1)}$ and $\mathcal{Z}_n = \{X_i \in \mathcal{X}_n : d(X_i, \partial S) \leqslant \epsilon_n^2\}$.

Proof. It can be easily proved that (see Theorem 1 in Dümbgen and Walther [9])

$$P\big(\partial S \nsubseteq B(\mathcal{Z}_n, 2\epsilon_n)\big) \leqslant D(S, \epsilon_n) \Pi(S, \mathcal{Z}_n, \epsilon_n), \quad \epsilon > 0,$$

where $D(S, \epsilon) = \max\{\sharp F : F \subset \partial S, \|x - y\| > \epsilon, \text{ for different } x, y \in F\}$ and $\Pi(S, \mathcal{Z}_n, \epsilon) = \sup_{s \in \partial S} P(B(s, \epsilon) \cap \mathcal{Z}_n = \emptyset)$ where $\sharp F$ denotes the cardinal of set F. It is easy to prove that $D(S, \epsilon)$ is of order ϵ^{-d}. Hence to show (6) it is enough to prove that

$$\sum_{n=1}^{\infty} \epsilon_n^{-d} \Pi(S, \mathcal{Z}_n, \epsilon_n) < \infty. \tag{7}$$

For each $s \in \partial S$ we have

$$
\begin{aligned}
P\big(B(s, \epsilon_n) \cap \mathcal{Z}_n = \emptyset\big) &= P\left(\bigcap_{i=1}^{n} \{X_i \notin B(\partial S, \epsilon_n^2) \cap B(s, \epsilon_n)\}\right) \\
&= \big(1 - P_X\big(B(\partial S, \epsilon_n^2) \cap B(s, \epsilon_n)\big)\big)^n \leqslant \exp\big(-n P_X\big(B(\partial S, \epsilon_n^2) \cap B(s, \epsilon_n)\big)\big),
\end{aligned}
$$

the last equality following from $(1 - x)^n \leqslant \exp(-nx)$ for $0 \leqslant x \leqslant 1$. Therefore, as the density f is essentially bounded from below in S, we have that

$$
\mu\big(S \cap B(\partial S, \epsilon^2) \cap B(s, \epsilon)\big) \geqslant \alpha \epsilon^{d+1}, \quad \forall \epsilon \in [0, \beta], \ \forall s \in \partial S, \tag{8}
$$

for some constants $\alpha, \beta > 0$ implies (7) and in consequence (6) holds. This fact is proved in the following lemma.

Lemma 4. *Let us suppose that S satisfies the conditions of Theorem* 1. *Then there are constants $\alpha, \beta > 0$ for which* (8) *holds.*

Proof. Let $s \in \partial S$. For $0 < h < r$ we define the spherical cap of height h as the set $\mathcal{C}(h) = \mathcal{R}(h) \cap B(s - r\eta(s), r)$ where $\mathcal{R}(h)$ is the (infinite) rectangle $\mathcal{R}(h) = \{x \in \mathbb{R}^d : -h \leqslant \langle x - s, \eta(s) \rangle \leqslant 0\}$ and $\eta(s)$ is the outer normal vector of S in s. We will prove that there is a constant $A > 0$, depending only on d and r, such that, for h small enough

$$
\mu\big(\mathcal{C}(h)\big) \geqslant A h^{(d+1)/2}. \tag{9}
$$

Also, we will prove that if $h_\epsilon = \min\{\epsilon^2/2, \epsilon^2/2r\} = A'\epsilon^2$ is smaller than r (that is, $\epsilon < \sqrt{r/A'}$) then

$$
\mathcal{C}(h_\epsilon) \subset B(s, \epsilon) \cap S \cap B(\partial S, \epsilon^2). \tag{10}
$$

The proof of this lemma is an immediate consequence of (9) and (10) since, in this case, for ϵ small enough, we have that

$$
\mu\big(B(s, \epsilon) \cap S \cap B(\partial S, \epsilon^2)\big) \geqslant \mu\big(\mathcal{C}(h_\epsilon)\big) \geqslant A h_\epsilon^{(d+1)/2} = A(A')^{(d+1)/2} \epsilon^{d+1} = \alpha \epsilon^{d+1},
$$

where, in the first inequality we have used (9) whereas in the second inequality we have used (10). Therefore the proof of the lemma will be completed if (9) and (10) are proved.

In order to prove (9) we must note that there is at least one affine transformation, $\mathcal{T}$, which transforms set $\mathcal{C}(h)$ in the set

$$
\mathcal{C}_0(h) = \mathcal{T}\big(\mathcal{C}(h)\big) = \big\{x = (x_1, \ldots, x_d) \in \mathbb{R}^d : 0 \leqslant x_d \leqslant h\big\} \cap B\big(-(r - h)e_d, r\big),
$$

where $e_d = (0, 0, \ldots, 1)^{\mathrm{T}}$. By the translation and rotation invariance of Lebesgue measure, we have that $\mu(\mathcal{C}(h)) = \mu(\mathcal{C}_0(h))$. From this it is clear that (9) is trivial for $d = 1$. In order to prove (9) for $d \geqslant 2$ we define (for $0 \leqslant l \leqslant h$) the section of $\mathcal{C}_0(h)$ to the height l as $\mathcal{C}_0(h, l) = \{x = (x_1, \ldots, x_{d-1}, l) \in \mathbb{R}^d : x \in \mathcal{C}_0(h)\}$. As the Lebesgue measure is a product measure, we have that

$$
\mu\big(\mathcal{C}_0(h)\big) = \int_0^h \mu_{d-1}\big(\mathcal{C}_0(h, l)\big) \, \mathrm{d}l,
$$

where μ_{d-1} denotes the $(d-1)$-dimensional Lebesgue measure. Note that $\mathcal{C}_0(h, l)$ is a $(d-1)$-dimensional sphere where the radius $r(l)$ satisfies the equation

$$
r(l) = \sqrt{r^2 - (r - h + l)^2} = \sqrt{2r(h - l) - (h - l)^2}.
$$

Hence

$$
\mu\big(\mathcal{C}_0(h)\big) = \omega_{d-1} \int_0^h r(l)^{d-1} \, \mathrm{d}l \geqslant \omega_{d-1} r^{d-1} \int_0^h l^{(d-1)/2} = \frac{2\omega_{d-1} r^{d-1}}{d+1} h^{(d+1)/2},
$$

where ω_{d-1} is the volume of the $(d-1)$-dimensional unit ball. This concludes the proof of (9).

It remains to prove (10). Since $B(s - r\eta(s), r) \subset S$ it is clear that for all h the spherical cap of height h is contained in S. On the other hand, for $x \in \mathcal{C}(\epsilon^2/2)$ we have

$$d(x, \partial S) \leqslant d\big(x, B\big(s + r\eta(s), r\big)\big) \leqslant 2\big\|x - P_{H_s}(x)\big\| = 2\big|\langle x - s, \eta(s)\rangle\big| \leqslant \frac{2\epsilon^2}{2} = \epsilon^2,$$

where $P_{H_s}(x)$ is the projection of x onto $H_s = \{x\colon \langle x - s, \eta(s)\rangle = 0\}$. Notice that H_s is the hyperplane tangent to ∂S passing by s. The first inequality follows from $\overset{\circ}{B}(s + r\eta(s), r) \subset S^c$ (see Lemma 1). Hence we have

$$\mathcal{C}\left(\frac{\epsilon^2}{2}\right) \subset S \cap B(\partial S, \epsilon^2). \tag{11}$$

Therefore to conclude the proof of (10) we only need to prove

$$\mathcal{C}\left(\frac{\epsilon^2}{2r}\right) \subset B(s, \epsilon). \tag{12}$$

Indeed, if $x \in \mathcal{C}(\epsilon^2/(2r))$ then $x \in B(s - r\eta(s), r)$ and

$$r^2 \geqslant \big\|x - \big(s - r\eta(s)\big)\big\|^2 = \big\|x - s + P_{H_s}(x) - P_{H_s}(x) + r\eta(s)\big\|^2$$
$$= \big\|P_{H_s}(x) - s\big\|^2 + \big\|x - P_{H_s}(x) + r\eta(s)\big\|^2 = \big\|P_{H_s}(x) - s\big\|^2 + \big(r + \langle x - s, \eta(s)\rangle\big)^2.$$

Also, as $\langle x - s, \eta(s)\rangle \geqslant -\epsilon^2/(2r) > -r$, we have $(r + \langle x - s, \eta(s)\rangle)^2 \geqslant (r - \epsilon^2/(2r))^2$ and, therefore,

$$\big\|P_{H_s}(x) - s\big\|^2 \leqslant r^2 - \left(r - \frac{\epsilon^2}{2r}\right)^2.$$

Hence

$$\|x - s\|^2 = \big\|x - P_{H_s}(x) + P_{H_s}(x) - s\big\|^2 = \big\|P_{H_s}(x) - s\big\|^2 + \langle x - s, \eta(s)\rangle^2$$
$$\leqslant r^2 - \left(r - \frac{\epsilon^2}{2r}\right)^2 + \left(\frac{\epsilon^2}{2r}\right)^2 = \epsilon^2,$$

that is, $x \in B(s, \epsilon)$, which concludes the proof of (12) and the proof of the lemma. $\square$

Proposition 2 is also proved. $\square$

It is clear that Theorem 3 is an immediate consequence of Propositions 1 and 2. $\square$

Proof. (Theorem 4) The proof is very similar to the r constant case above. In fact, Proposition 2 only depends on the sample and support properties and therefore is independent on the estimator. Hence we can use it here. Finally it is possible to prove an extension of Proposition 1 by changing the auxiliary lemmas in obvious details. In fact the Theorem is a straightforward consequence of the following lemma which extends Lemma 3 for $C_{r_n}(\mathcal{X}_n)$. It can be proved replacing r by r_n elsewhere in Lemmas 1, 2. We will omit the details.

Lemma 5. *Let us assume that* $\partial S \subset B(\mathcal{Z}_n, 2\epsilon_n)$ *and* $d_H(\mathcal{X}_n, S) < r_n \leqslant r$. *Then*

$$S \ominus (L_n \epsilon_n^2 B) \subset C_{r_n}(\mathcal{X}_n), \quad L_n \geqslant \max(2, 8/r_n).$$

Now the theorem can be proved using the Theorem 3 arguments. We should mention than the condition $r_n \gg (\log n/n)^{1/d}$ ensures

$$P\big(d_H(\mathcal{X}_n, S) < r_n, \text{ eventually }\big) = 1,$$

which is needed to apply Lemma 5. In the above result we are using again Theorem 3 in Cuevas and Rodríguez-Casal [7]. $\square$

Acknowledgements

I thank Prof. Antonio Cuevas for his valuable help. I am also very grateful to a referee whose deep and constructive remarks led to an improved version of the paper. This work was partially carried out when I was at University of Vigo. This research was supported by grants MTM2005-00820 and PGIDIT06PXIB207009PR.

References

[1] A. Baíllo, A. Cuevas, On the estimation of a star-shaped set, Adv. Appl. Probab. 33 (2001) 1–10.

[2] A. Baíllo, A. Cuevas, A. Justel, Set estimation and nonparametric detection, Canad. J. Statist. 28 (2000) 765–782.

[3] V. Bertholet, J.P. Rasson, S. Lissoir, About the automatic detection of training sets for multispectral images classification, in: A. Rizzi, M. Vichi, H.H. Bock (Eds.), Advances in Data Science and Classification, Springer, Berlin, 1988, pp. 221–226.

[4] A. Cuevas, M. Febrero, R. Fraiman, Cluster analysis: a further approach based on density estimation, Computational Statistics and Data Analysis 36 (2001) 441–459.

[5] A. Cuevas, M. Febrero, R. Fraiman, Estimating the number of clusters, Canad. J. Statist. 28 (2000) 367–382.

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

[7] A. Cuevas, A. Rodríguez-Casal, On boundary estimation, Adv. Appl. Probab. 36 (2004) 340–354.

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

[9] L. Dümbgen, G. Walther, Rates of convergence for random approximations of convex sets, Adv. Appl. Probab. 28 (1996) 384–393.

[10] G.A. Edgar, Measure, Topology and Fractal Geometry, Springer-Verlag, 1990.

[11] H. Edelsbrunner, E.P. Mücke, Three dimensional alpha shapes, ACM Trans. Graph. 13 (1994) 43–72.

[12] A.P. Korostelev, A.B. Tsybakov, Minimax Theory of Image Reconstruction, Springer-Verlag, 1993.

[13] D. Marr, Vision, Freeman and Co, 1982.

[14] M. Rudemo, H. Stryhn, Approximating the distribution of maximum likelihood contour estimators in two-region images, Scand. J. Statist. 21 (1994) 41–55.

[15] R. Schneider, Convex Bodies: The Brunn–Minkowski Theory, Cambridge University Press, 1993.

[16] J. Serra, Image Analysis and Mathematical Morphology, Academic Press, 1982.

[17] A.B. Tsybakov, Optimal aggregation of classifiers in statistical learning, Ann. Statist. 1 (2004) 135–166.

[18] G. Walther, Granulometric smoothing, Ann. Statist. 25 (1997) 2273–2299.

[19] G. Walther, On a generalization of Blaschke's rolling theorem and the smoothing of surfaces, Math. Methods Appl. Sci. 22 (1999) 301–316.

[20] W. Weil, J.A. Wieacker, Stochastic geometry, in: P.M. Wills (Ed.), Handbook of Convex Geometry, vol. B, Elsevier, Amsterdam, 1993, pp. 1391–1438.