UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
Departamento de Estatística e Investigación Operativa

SET ESTIMATION
UNDER CONVEXITY TYPE RESTRICTIONS

Beatriz Pateiro López

Set estimation under convexity type restrictions

Beatriz Pateiro López

Realizado el acto público de defensa y mantenimiento de esta tesis doctoral el día 1 de septiembre de 2008, en la Facultad de Matemáticas de la Universidad de Santiago de Compostela, ante el tribunal formado por:

Presidente:	Dr. D. Antonio Cuevas González
Vocales:	Dr. D. Alexandre Tsybakov
	Dr. D. Ricardo Fraiman
	Dr. D. Juan Antonio Cuesta Albertos
Secretario:	Dr. D. Manuel Febrero Bande

siendo director de la misma el Dr. D. Alberto Rodríguez Casal, obtuvo la máxima calificación de SOBRESALIENTE CUM LAUDE. Además, esta tesis ha cumplido los requisitos necesarios para la obtención del DOCTORADO EUROPEO.

Contents

Introduction

This thesis, *Set estimation under convexity type restrictions*, collects the research work done during these last years under the supervision of Prof. Alberto Rodríguez Casal. First and foremost I would like to thank him for his help and his effort. *Muchas gracias por confiar en mi, por tu dedicación y por tu ayuda.*

The title *set estimation* refers to the subject matter of the thesis, the reconstruction of an unknown set S from a random sample of points whose distribution is related to it. Apart from the set itself, we are also interested in approximating a particular characteristic of the set, the surface area. It is *under convexity type restrictions* because the problem of set estimation is so extensive that giving an efficient general solution is almost unfeasible. A traditional approach consists in assuming that the set of interest is convex. However, we restrict ourselves to a more flexible shape condition named α-convexity, which allows to handle a larger family of sets.

The essay has been organized in the following way. First sections in Chapter 1 provide an overview of set estimation results. Previous research on topics such as support estimation and surface area estimation is reviewed and the notation for some basic concepts is introduced. Last section of Chapter 1 is devoted to the statement of the main results we have obtained during the course of this research. The purpose of this chapter is to introduce the reader into the framework in which we develop our study and to present precisely our contributions. For a complete discussion of the results and their proofs, the reader is referred to Chapters 2 and 3. Chapter 2 focuses on the detailed analysis of a support estimator, the α-convex hull estimator. Chapter 3 provides an in-depth analysis of a new estimator for the surface area of a body. In Chapter 4 we present the results of a simulation study comparing some of the estimators discussed in previous chapters. Finally, we also include three useful appendices. In Appendix A we state and prove a series of geometric results that help us to relate the α-convexity with other geometric properties. In Appendix B we focus on the behaviour of a morphological operator, the closing of a random sample of points with respect to closed and open balls. Finally, we have developed a new library, named `alphahull`, for the implementation in R of the discussed estimators. Appendix C describes the functions in the library, their usage, arguments, returned values and examples.

The course of the thesis does not faithfully reproduce the time sequence in obtaining the results. In chronological order, the results in Subsection 3.3.1, providing the almost sure convergence rate of the proposed estimator for the surface area, are the starting point of our research. This work was accepted for publication in Advances in Applied Probability. The in-depth study of the α-convex hull estimator in Chapter 2 led us to complete the analysis of the statistical properties of the surface area estimator proposed in Chapter 3, see Subsection 3.3.2.

Chapter 1

On set estimation

1.1 Introduction

The problem of reconstructing a set S from a finite set of points taken into it has been addressed in different fields of research. In computational geometry, for instance, the efficient construction of convex hulls for finite sets of points has important applications in pattern recognition, cluster analysis and image processing, among others. For example, in Figure 1.1, image analysis tools could be used to recover the original set S in the left plot from the corrupt version shown in the right plot. We refer to Preparata and Shamos (1985) for an introduction to computational geometry and its applications. In a different framework, the set of points from which we try to reconstruct S is assumed to be non-deterministic. The term *set estimation* refers to the statistical problem of estimating an unknown set S from a random sample of points $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ whose distribution is closely related to S. But, what kind of sets are we talking about? We may be interested, for example, in recovering a distribution support, its boundary or a level set.

Formally, the **support estimation** problem is established as the problem of estimating the support of an absolutely continuous probability measure P_X from independent observations drawn from it.

Figure 1.1: *Original set S and corrupt version.*

Korostelëv and Tsybakov (1993) refers to Geffroy (1964), Rényi and Sulanke (1963), and

Rényi and Sulanke (1964) as the first works on support estimation. Rényi and Sulanke (1963) and Rényi and Sulanke (1964) studied the case when S is a convex support in the bidimensional euclidean space and proposed a natural estimator, the convex hull of the sample $\mathcal{X}_n$. However, if S is not convex, the convex hull of the sample is not an appropriate estimator. How can we estimate S if no assumption is made on its shape? In this setting, Chevalier (1976) and Devroye and Wise (1980) proposed to estimate the support of an unknown probability measure by means of a smoothed version of the sample $\mathcal{X}_n$. The problem of support estimation was introduced by Devroye and Wise (1980) in connection with a practical application, the detection of abnormal behaviour of a system, plant or machine. Results on the performance of the estimator were obtained, among others, by Chevalier (1976), Devroye and Wise (1980), and Korostelëv and Tsybakov (1993). Of course, there are situations in between the two described above, that is, we can assume that the set S satisfies some shape restriction, more flexible than convexity. In Rodríguez-Casal (2007), the estimation of an α-convex support is considered. The α-convexity assumption plays a mayor role in this thesis and will be studied in depth in the course of the dissertation.

Set estimation is also related to another interesting problem, the estimation of certain geometric characteristics of the set such as the **volume** or the **surface area**. Obviously, there are other statistical fields which also cope with problems regarding set measurements as, for example, the stereology. However, stereology focuses on the estimation of certain characteristics of S (volume, surface area, *etc.*) without needing to reconstruct the set, see, e.g., Baddeley and Jensen (2005), Cruz-Orive (2001/02), whereas the primary object of interest of set estimation is the set itself. Turning to the set estimation framework, it seems natural to think that the volume or the surface area of a good set estimator should provide good approximations of these geometrical quantities. Bräker and Hsing (1998) studied the asymptotic properties of the length and area of the convex hull of a random sample of points in $\mathbb{R}^2$. The more recent work by Cuevas et al. (2007) focuses on the surface area estimation problem from a different point of view. Assuming no shape restriction and that we have observations from both the set of interest and its complement, the surface area can be approximated, based on the notion of Minkowski content, handling two support estimators. Adding the flexible α-convexity condition, we propose in Section 1.5 a new surface area estimator, which gives a compromise between the no shape restriction considered by Cuevas et al. (2007) and the restrictive convexity assumption. The asymptotic behaviour of this new estimator was analysed by Pateiro-López and Rodríguez-Casal (2008). A complete presentation of the obtained results is provided in Chapter 3.

This chapter is organized as follows. Section 1.2 introduces some basic notions used in set estimation theory. In Section 1.3 we give a brief outline of the classical support estimators available in the literature and their properties. Subsection 1.3.1 deals with the general case, when no assumption is made on the shape of the set of interest S. Subsection 1.3.2 provides a review of the main results on support estimation under the convexity assumption, including the aforementioned works on the convex hull estimator. The notion of α-convexity is discussed in detail in Section 1.4, along with a review of the literature on set estimation under this shape restriction. Section 1.5 is devoted to the surface area estimation problem. Finally, in Section 1.6 we present the main results contained in this thesis.

1.2 Basic concepts on set estimation theory

Like in other contexts, in order to evaluate a set estimator S_n, we need certain measure of the distance between the estimator and the target S. We all are familiarized with the concept of Euclidean distance between points in the d-dimensional Euclidean space $\mathbb{R}^d$ but, what is the distance between sets? see for example Figure 1.2. How can be defined the distance between the set A and the set C? We might be persuaded to think that the distance is zero, since both sets share a common border. However, it is clear that if we want to move from point $a \in A$ to C, even to the nearest point of C, the distance we have to cover will be positive. The feeling is that, in order to give an adequate definition for the distance between A and C, we should take into account the distances from the points in A to the boundary of C and vice versa.

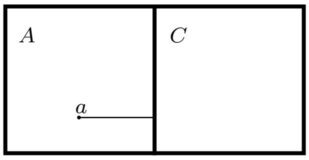


Figure 1.2: *On an adequate definition of the distance between sets.*

1.2.1 The Hausdorff distance

The Hausdorff distance can be defined over the space of the nonempty compact subsets in a given metric space. However, since it is enough for our purposes, we concentrate on the d-dimensional Euclidean space $\mathbb{R}^d$, equipped with the inner product $\langle \cdot, \cdot \rangle$ and the norm $\|\cdot\|$. See, for example, Edgar (1990) and Matheron (1975) for a more extensive discussion of the Hausdorff metric.

Definition 1.2.1. *Let A and C be nonempty compact subsets of $\mathbb{R}^d$. The Hausdorff distance between A and C is defined by*

$$d_H(A, C) = \max\left\{\sup_{a \in A} d(a, C),\ \sup_{c \in C} d(c, A)\right\},$$

where

$$d(a, C) = \inf\{\|a - c\| : c \in C\}. \tag{1.1}$$

Defining the Hausdorff distance over the collection of nonempty compact subsets of $\mathbb{R}^d$ ensures that d_H is a metric. By restricting the definition of the Hausdorff distance to nonempty and bounded subsets, d_H is well defined. On the other hand, if we do not restrict the definition of the Hausdorff distance to closed subsets, it could be the case that the distance between two sets is zero, even if the two sets are not equal.

Equation (1.1) defines the distance between a point $a \in A$ and a set C. It is worth pointing out that this distance is not necessarily equal to the Hausdorff distance between the set $\{a\}$ and the set C, see Figure 1.3. In fact,

$$d_H(\{a\}, C) = \max\left\{d(a, C),\ \sup_{c\in C} d(c, \{a\})\right\} = \sup_{c\in C} \|c - a\|.$$

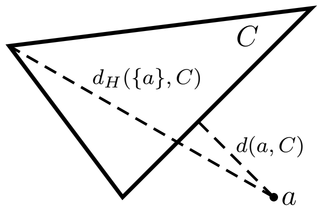


Figure 1.3: $d_H(\{a\}, C)$ *is not necessarily equal to* $d(a, C)$.

An alternative and useful way of defining the Hausdorff distance uses the concept of open ε-neighbourhood of a set.

Definition 1.2.2. *Let A be a nonempty compact subset of $\mathbb{R}^d$. The open ε-neighbourhood of A, $\mathring{B}(A, \varepsilon)$, is defined by*

$$\mathring{B}(A, \varepsilon) = \{x \in \mathbb{R}^d :\ d(x, A) < \varepsilon\}. \tag{1.2}$$

Analogously to (1.2), we can define the closed ε-neighbourhood of a set.

Definition 1.2.3. *Let A be a nonempty compact subset of $\mathbb{R}^d$. The closed ε-neighbourhood of A, $B(A, \varepsilon)$, is defined by*

$$B(A, \varepsilon) = \{x \in \mathbb{R}^d :\ d(x, A) \leq \varepsilon\}.$$

Definition 1.2.4. *Let A and C be nonempty compact subsets of $\mathbb{R}^d$. The Hausdorff distance between A and C is defined by*

$$d_H(A, C) = \inf\left\{\varepsilon > 0 :\ \ A \subset \mathring{B}(C, \varepsilon) \text{ and } C \subset \mathring{B}(A, \varepsilon)\right\}.$$

Figure 1.4 illustrates how to compute the Hausdorff distance between two sets. It can be easily proved that Definitions 1.2.1 and 1.2.4 are equivalent. There is a third definition for the Hausdorff distance in $\mathbb{R}^d$. Its formulation is based on mathematical morphology theory. Mathematical morphology can be defined as the theory for the analysis of the shape of spatial structures, based on set theory, integral geometry, and lattice algebra. It is an extremely powerful image analysis methodology that has been applied to numerous scientific fields such as biology, quality control, and medical imaging. For a more comprehensive presentation on this topic, we refer to Serra (1984). Morphological operators aim to extract relevant structures of the set under study from its interaction with another set of known shape called structuring element. The dilation and the erosion of a set by a structuring element are the two fundamental morphological operators. They are closely related to the Minkowski addition and subtraction defined below.

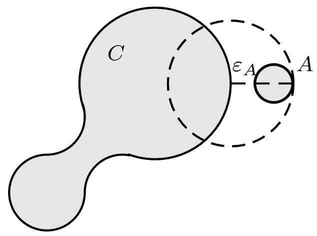


Figure 1.4: $\varepsilon_A = \inf\left\{\varepsilon > 0 : \ \ A \subset \mathring{B}(C,\varepsilon)\right\}$.

Definition 1.2.5. *Let A, C be subsets of $\mathbb{R}^d$. The Minkowski addition, $\oplus$, is defined by*

$$A \oplus C = \{a + c : a \in A, c \in C\}.$$

The Minkowski subtraction, $\ominus$, is defined by

$$A \ominus C = \{x : \{x\} \oplus C \subset A\}.$$

For $\lambda \in \mathbb{R}$,

$$\lambda C = \{\lambda c : c \in C\}.$$

Note that, according to this notation, $A \oplus \{x\}$ is the translation of A by the vector x. The dilation and erosion operators are formally defined from the Minkowski addition and Minkowski substraction, respectively. They are always performed by applying a structuring element to the set of interest. Thus, the result of the dilation or erosion of a set A is the result of the interaction between the set and the structuring element. Dilation allows the set to expand while erosion shrinks the set by eroding its boundary. The way in which the set is dilated or eroded depends on the structuring element. Although, a priori, the structuring element could be any set, in our context it is usual to consider d-dimensional balls. We denote by $B(x, r)$ and $\mathring{B}(x, r)$ the closed and open ball with centre x and radius r, respectively. In order to simplify the notation B and $\mathring{B}$ will stand for $B(0, 1)$ and $\mathring{B}(0, 1)$. Moreover, from now on, A^c, int(A), $\overline{A}$ and ∂A will denote the complement, interior, closure and boundary of A, respectively.

Definition 1.2.6. *The dilation of a set $A \subset \mathbb{R}^d$ by the structuring element $\mathring{B}(0, r)$ is defined as the union of open balls of radius r with centres in A, that is,*

$$\bigcup_{x \in A} \mathring{B}(x, r).$$

Definition 1.2.7. *The erosion of a set $A \subset \mathbb{R}^d$ by the structuring element $\mathring{B}(0, r)$ is defined as the locus of points x such that $\mathring{B}(x, r)$ is included in A, that is,*

$$\{x : \ \ \mathring{B}(x, r) \subset A\}.$$

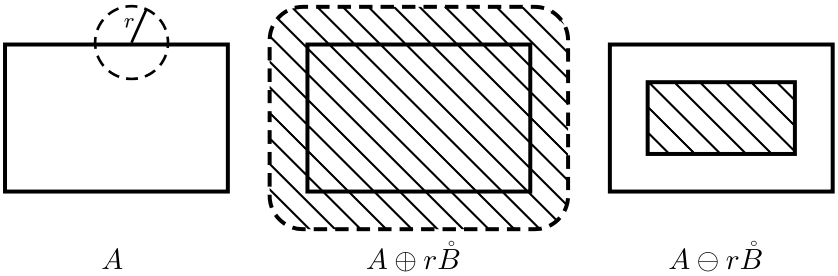


Figure 1.5: *Dilation and erosion of the set A by the structuring element $\mathring{B}(0, r)$.*

The dilation and erosion of a set A by the structuring element $\mathring{B}(0, r)$, see Figure 1.5, fit in with the above defined Minkowski addition $A \oplus \mathring{B}(0, r)$ and Minkowski subtraction $A \ominus \mathring{B}(0, r)$, respectively. It is worth mentioning that this relation cannot be generalized to all structuring elements. For the definition of the dilation and erosion of a set A by a general structuring element C, see Serra (1984). As an example, Figure 1.6 shows the dilation of the set A by the triangle C, and the set $A \oplus C$. Since the triangle is not symmetric with respect to the origin, both operations do not lead to the same result.

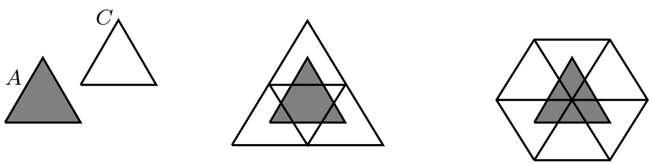


Figure 1.6: *(a) Set A and structuring element C. (b) $A \oplus C$. (c) Dilation of A by C.*

As previously mentioned, there is an alternative definition of the Hausdorff distance, formulated in terms of the dilation.

Definition 1.2.8. *Let A and C be nonempty compact subsets of $\mathbb{R}^d$. The Hausdorff distance between A and C is defined by*

$$d_H(A, C) = \inf\{\varepsilon > 0 : \ A \subset C \oplus \varepsilon\mathring{B} \text{ and } C \subset A \oplus \varepsilon\mathring{B}\}.$$

It is straightforward to prove that Definitions 1.2.1, 1.2.4 and 1.2.8 are equivalent. Note that the open ε-neighbourhood in (1.2) satisfies

$$\mathring{B}(A, \varepsilon) = \{x \in \mathbb{R}^d : \ d(x, A) < \varepsilon\} = \bigcup_{x \in A} \mathring{B}(x, \varepsilon) = A \oplus \varepsilon\mathring{B}.$$

The Hausdorff distance gives us an idea of the proximity of two sets and, in this sense, it is an appropriate tool to evaluate the performance of a set estimator S_n. It is desirable that

$$d_H(S, S_n) \to 0. \tag{1.3}$$

However, the convergence in (1.3) is not sufficient in general to ensure that the estimator S_n performs well. For example, from the sets in Figure 1.7, it is apparent that the Hausdorff distance $d_H(A, C)$ is close to zero, even though we do not have the feeling that both sets are similar. If we are concern, not only about the proximity of the sets, but also about the shape similarity, we should ask the estimator to satisfy both (1.3) and

$$d_H(\partial S, \partial S_n) \to 0. \tag{1.4}$$

Conditions that guarantee the convergence (1.3) and (1.4) (in probability, almost surely, ...) of some existing estimators can be found, for example, in Cuevas and Rodríguez-Casal (2004).

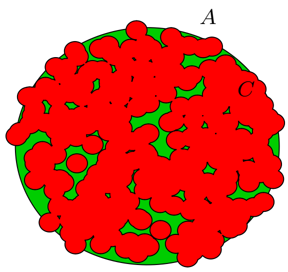


Figure 1.7: *In green $A = B(0,1)$. In red a set C, visually close to A. Although both sets are close in terms of the Hausdorff distance, their shapes are quite different.*

1.2.2 The distance in measure

The distance in measure is useful to quantify the similarity in content of two sets. Again, it can be defined in any measure space but it is enough for our purposes to consider the measure space $(\mathbb{R}^d, \mathcal{B}, \mu)$, where $\mathcal{B}$ denotes the Borel σ-algebra of $\mathbb{R}^d$ and μ denotes the Lebesgue measure.

Definition 1.2.9. *Let A and $C \subset \mathcal{B}$. The distance in measure between A and C is defined by*

$$d_\mu(A, C) = \mu(A \Delta C),$$

where $A \Delta C$ denotes the symmetric difference between A and C, that is,

$$A \Delta C = (A \setminus C) \cup (C \setminus A).$$

The distance in measure is specially useful when we are interested in the content of the sets rather than in their proximity. Furthermore, the distance in measure is closely related to the L_1 functional distance since

$$d_\mu(A, C) = \int |\mathbb{I}_A - \mathbb{I}_C| \, d\mu,$$

where $\mathbb{I}_A$ and $\mathbb{I}_C$ denote the indicator functions of A and C, respectively.

1.3 Support estimation

Returning to the subject matter of this chapter, let us assume that we are given a random sample $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ of i.i.d. observations from a random variable X with absolutely continuous probability distribution P_X and nonempty compact support $S \subset \mathbb{R}^d$. The goal is to reconstruct the set by using the available information. As usual in estimation, the problem changes substantially depending on the model assumptions. In Subsection 1.3.1 we tackle the most general framework, when no assumptions are made on the shape of S. In this situation we need to define a flexible estimator in order to effectively estimate S whatever its shape. More sophisticated estimators can be considered if we are given some additional information on the set. For instance if we know that S in convex. In that case we can ensure that, at least, the convex hull of the sample is contained in the set. The convexity of S is one of the classical assumptions in the literature on set estimation. Because of its importance, we have considered appropriate to include an independent subsection devoted to this subject. Thus, in Subsection 1.3.2 the estimation of a convex set S is discussed. The main drawback of the convexity assumption is that it rules a large number of sets out.

1.3.1 The general case

If no assumption is made on the shape of S, then the only information we have comes from the sample. Actually, this is the first estimator we shall consider. The estimator $\mathcal{X}_n$ is d_H-consistent, that is, with probability one, $d_H(S, \mathcal{X}_n) \to 0$ (it is understood that the limit as $n \to \infty$ is taken). However, it can be easily seen that $d_\mu(S, \mathcal{X}_n) = \mu(S) > 0$. As mentioned in the introduction of this chapter, Chevalier (1976) and Devroye and Wise (1980) proposed a very intuitive estimator based on an smoothed version of the sample that achieves better results for d_μ. More precisely, let

$$S_n = B(\mathcal{X}_n, \varepsilon_n) = \bigcup_{i=1}^{n} B(X_i, \varepsilon_n), \tag{1.5}$$

where ε_n is a number depending only upon n. We shall refer to this estimator as Devroye-Wise estimator, see Figure 1.8. Devroye and Wise (1980) establishes the d_μ-consistency in probability and almost surely of (1.5). If $\varepsilon_n \to 0$ and $n\varepsilon_n^d \to \infty$, then $d_\mu(S, S_n) \to 0$ in probability. In fact, the result is proved not only for the Lebesgue measure μ and compact sets, but for any measure whose restriction to a general set S is absolutely continuous with respect to the distribution P_X. It is worth noting that the assumptions on ε_n are identical to those imposed on the bandwidth parameter in nonparametric density estimation, to ensure the consistency. Moreover, we release that similar conditions also imply the d_μ- consistency in mean of a more sophisticated support estimator, the r_n-convex hull of $\mathcal{X}_n$, see Chapter 2. Regarding other works on the Devroye-Wise estimator, Korostelëv and Tsybakov (1993) obtained the minimax convergence rates of S_n by assuming that the boundary of S satisfies some piecewise Lipschitz conditions. Cuevas and Rodríguez-Casal (2004) are concerned with the estimation of ∂S with respect to the Hausdorff metric. Although the almost sure d_H-consistency of S_n can be straightforwardly obtained under the assumption that $\varepsilon_n \to 0$, consistency results of the form $d_H(\partial S, \partial S_n) \to 0$ are not so immediate. Note in Figure 1.8 that, as ε_n is smaller, the estimator becomes more and more fragmented with holes in the midst of the sample points. In order to consistently estimate ∂S, it is useful to take larger values of ε_n to guarantee that $S \subset S_n$. The precise result, established by Cuevas and Rodríguez-Casal (2004), states that $\varepsilon_n \to 0$ almost surely together with $S \subset S_n$ imply the almost sure d_H- consistency of ∂S_n. Another way to ensure the almost sure d_H-consistency is by assuming certain shape restriction on S. In this situation it seems natural to select ε_n such that S_n fulfills the same shape restriction as S. For example, if we assume that S is star-shaped, we can incorporate this additional information to the Devroye-Wise estimator in such a way that S_n is also star-shaped. This provides a method of choosing ε_n from the sample that ensures $d_H(\partial S, \partial S_n) \to 0$ almost surely, see Baíllo and Cuevas (2001).

(a) (b) (c)

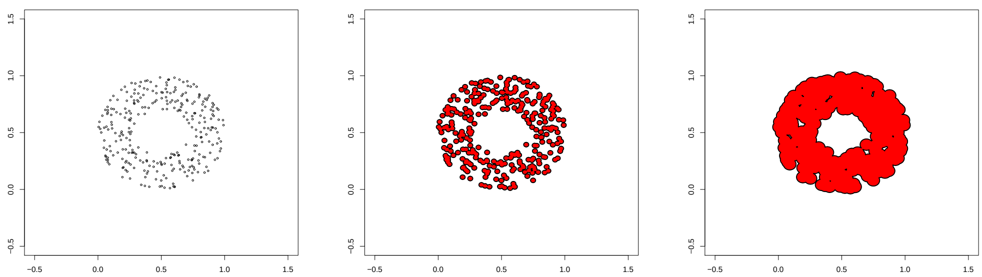


Figure 1.8: *(a) Sample in the disc* $B(0, 0.5) \setminus \mathring{B}(0, 0.2)$ *of size* $n = 300$. *(b) Devroye-Wise estimator for* $\varepsilon_n = 0.02$. *(c) Devroye-Wise estimator for* $\varepsilon_n = 0.05$.

1.3.2 On the estimation of a convex set

When the set S is assumed to be convex there is a natural estimator, the convex hull of the sample

$$H_n = \text{conv}(X_1, \ldots, X_n). \tag{1.6}$$

Note that, a priori, H_n is a reasonable choice since the convex hull fulfills the convexity shape restriction assumed on S. Moreover, H_n is the maximum likelihood estimator in the family of all closed convex sets, see Korostelëv and Tsybakov (1993). Now, how closely is S approximated by the convex hull H_n of the sample $\mathcal{X}_n$? This problem is posed in Dümbgen and Walther

(1996). The proximity between the set and the convex hull is studied in terms of the Hausdorff distance in an arbitrary dimension d. More precisely, it is proved that $d_H(S, H_n) = O((\log n/n)^{1/d})$ almost surely. Furthermore, if ∂S satisfies an additional smoothness condition, it is proved that $d_H(S, H_n)$ is of order $(\log n/n)^{2/(d+1)}$.

Also in connection with the convex hull estimator, there are a series of papers concerned with certain statistics of H_n such as the number of vertices, the number of facets, the volume, and the surface area. For example, Bräker and Hsing (1998) studied the asymptotic behaviour of the expected area and perimeter of H_n in the bidimensional case under more general conditions than those considered by Rényi and Sulanke (1963) and Rényi and Sulanke (1964). See Schneider (1988) for a extensive review of classical references in this line.

1.4 Relaxing the convexity assumption

The convexity assumption may be too restrictive in practice and the estimator H_n, given in (1.6), is not the best possible choice when S is not convex. Notice that H_n tends to fill in the space in the midst of the observations. The result is a convex set when the original one had not even to be connected. This section focuses on the introduction of a more flexible assumption than convexity, named α-convexity.

Definition 1.4.1. *A set $A \subset \mathbb{R}^d$ is said to be α-convex, for $\alpha > 0$, if*

$$A = C_\alpha(A),$$

where

$$C_\alpha(A) = \bigcap_{\{\mathring{B}(x,\alpha):\ \mathring{B}(x,\alpha)\cap A=\emptyset\}} \left(\mathring{B}(x,\alpha)\right)^c \tag{1.7}$$

is called the α-convex hull of A.

The α-convex hull of a set A satisfies that $C_\alpha(A) \subset C_{\alpha'}(A)$ for $\alpha \leq \alpha'$. Furthermore, it can be proved that, under certain conditions of A, see Walther (1999), $C_\alpha(A)$ tends to the closure of A as α tends to zero and it tends to the convex hull of A as α tends to infinity, see Figure 1.9.

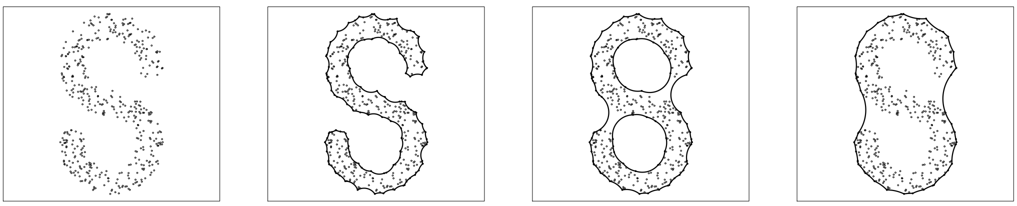


Figure 1.9: *Finite set of points and α-convex hull for increasing values of α.*

Regarding the relation between convexity and α-convexity, if A is convex and closed then it is also α-convex for all $\alpha > 0$, see Figure 1.10. On the other hand, Walther (1999) proved that if the interior of the convex hull is not empty, then the reciprocal is also true.

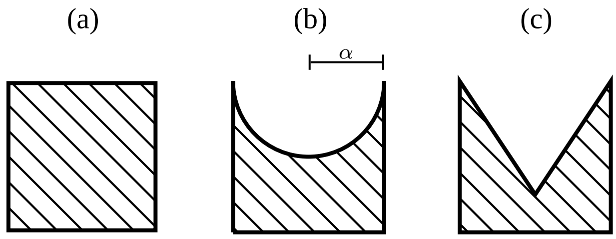


Figure 1.10: *(a) Set convex and α-convex for all $\alpha > 0$. (b) Set non convex but α-convex. (c) Set neither convex nor α-convex for any $\alpha > 0$.*

The α-convex hull of a set is intimately related to the dilation and erosion operators through the closing of the set, whose precise definition is given below. The idea behind the morphological closing is to define an operator that tends to recover the original shape of a set that has been previously dilated. This is achieved by eroding the dilated set. Note that the closing may not coincide with the original set since dilation and erosion are not inverse operators. In the same manner, once a set has been eroded, there exists in general no inverse transformation to recover the initial set. The morphological opening tries to recover as much as possible the original shape of an eroded set by dilating it.

Definition 1.4.2. *The closing of a set A with respect to $\mathring{B}(0,r)$ is defined as*

$$(A \oplus r\mathring{B}) \ominus r\mathring{B}.$$

Definition 1.4.3. *The opening of a set A with respect to $\mathring{B}(0,r)$ is defined as*

$$(A \ominus r\mathring{B}) \oplus r\mathring{B}.$$

As it occurred with the dilation and erosion, the closing and opening are the result of the interaction between the set of interest and a structuring element. Definitions 1.4.2 and 1.4.3 correspond with the particular case in which the structuring element is the open ball $\mathring{B}(0,r)$. See Matheron (1975) for the definition of opening and closing with respect to a general structuring element. Figures 1.11 and 1.12 show the closing and opening of a given set A, respectively. Closing and opening operations are increasing, idempotent, and dual to each other with respect to taking complements, that is, $(A^c \ominus r\mathring{B}) \oplus r\mathring{B} = ((A \oplus r\mathring{B}) \ominus r\mathring{B})^c$ and $(A^c \oplus r\mathring{B}) \ominus r\mathring{B} = ((A \ominus r\mathring{B}) \oplus r\mathring{B})^c$. We say that a set A is morphologically close with respect to $\mathring{B}(0,r)$ if $A = (A \oplus r\mathring{B}) \ominus r\mathring{B}$, and morphologically open with respect to $\mathring{B}(0,r)$ if $A = (A \ominus r\mathring{B}) \oplus r\mathring{B}$.

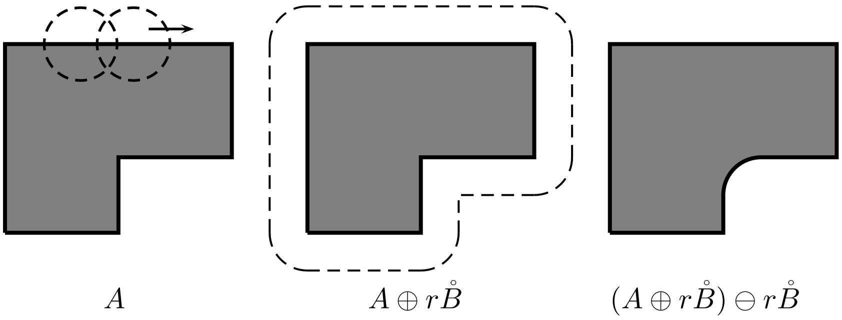


Figure 1.11: *Dilation an erosion leading to the closing of A.*

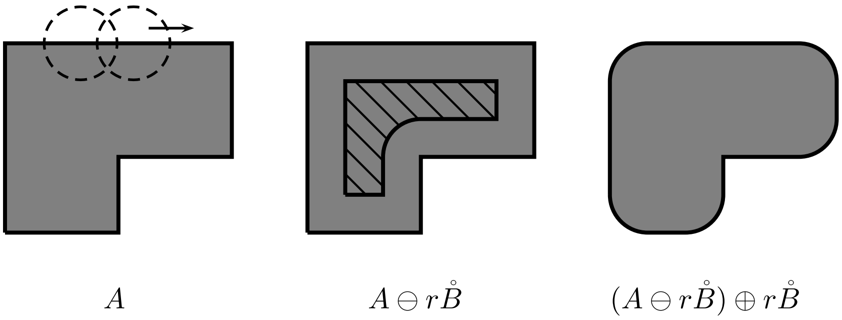


Figure 1.12: *Erosion and dilation leading to the opening of A.*

It is easy to prove that the opening of A coincides with the points of all balls $\mathring{B}(x, r)$ which are completely contained in A, that is,

$$(A \ominus r\mathring{B}) \oplus r\mathring{B} = \bigcup_{\mathring{B}(y,r) \subset A} \mathring{B}(y, r). \tag{1.8}$$

Equation (1.8), together with the duality with respect to the complement of opening and closing, leads to

$$(A \oplus r\mathring{B}) \ominus r\mathring{B} = \bigcap_{\{\mathring{B}(x,r):\ \mathring{B}(x,r) \cap A = \emptyset\}} \left(\mathring{B}(x, r)\right)^c, \tag{1.9}$$

which coincides with the definition of the α-convex hull in (1.7), for $\alpha = r$. Therefore, the α-convexity of a set A can be defined in terms of the closing with respect to $\mathring{B}(0, \alpha)$. Thus, the set A is said to be α-convex if

$$A = (A \oplus \alpha\mathring{B}) \ominus \alpha\mathring{B}.$$

Once we have introduced all these concepts we are in a position to return to the subject matter, the estimation of a set S from a sample $\mathcal{X}_n = \{X_1, \ldots, X_n\}$. If the set S is α-convex, then the α-convex hull of the sample

$$C_\alpha(\mathcal{X}_n) = (\mathcal{X}_n \oplus \alpha\mathring{B}) \ominus \alpha\mathring{B} \tag{1.10}$$

seems to be the natural estimator. The estimator in (1.10) was first studied by Rodríguez-Casal (2007) under the assumption that the set S belongs to Serra's regular model. We refer to Serra (1984) for a complete description of this class of sets.

Definition 1.4.4. *Serra's regular model is the class of compact sets A that are morphologically open and closed with respect to the compact ball αB of radius α for some $\alpha > 0$, that is,*

$$A = (A \ominus \alpha B) \oplus \alpha B = (A \oplus \alpha B) \ominus \alpha B.$$

The Serra's regular model was studied in depth by Walther (1999). Indeed, Walther (1999) provided a generalization of the Blaschke's Rolling Theorem that gives an exact geometric characterization of Serra's regular model in terms of α-convexity and free rolling conditions. Before stating the theorem in question, we introduce the free rolling condition and some comments concerning its definition.

Definition 1.4.5. *Let $A \subset \mathbb{R}^d$ be a closed set. The ball αB is said to roll freely in A if for each boundary point $a \in \partial A$ there exists some $x \in A$ such that $a \in B(x, \alpha) \subset A$.*

It should be mentioned that the free rolling condition in Theorem 1.4.1 is not exactly the same as the one given in Definition 1.4.5. In Walther (1999) it is also required that $A \ominus \alpha B$ is path-connected in order to preserve the physical meaning of rolling freely. This additional requirement will not be necessary for our purposes and that's the reason why it is not included in Definition 1.4.5. Next, we present Theorem 1.4.1 as it is stated in Walther (1999).

Theorem 1.4.1 (Walther (1999))**.** *Let $S \neq \emptyset$ be a compact and path-connected subset of $\mathbb{R}^d$ and $\alpha > 0$. Then, the following conditions are equivalent:*

i) The conditions

$$\begin{aligned} S &= (S \ominus \lambda B) \oplus \lambda B, \quad 0 \leq \lambda \leq \alpha, \\ S &= (S \oplus \lambda B) \ominus \lambda B, \quad 0 \leq \lambda < \alpha, \end{aligned}$$

hold.

ii) S and $\overline{S^c}$ are α-convex and $\mathrm{int}(S_i) \neq \emptyset$.

iii) A ball of radius λ rolls freely inside S and $\overline{S^c}$ for all $0 \leq \lambda \leq \alpha$.

iv) For every $r_1 \in [0, \alpha]$, $r_2 \in [0, \alpha)$ there exist A, $D \subset \mathbb{R}^d$ with $S = A \oplus r_1 B = D \ominus r_2 B$.

v) ∂S is a $(d-1)$-dimensional C^1 submanifold in $\mathbb{R}^d$ with the outward pointing unit normal vector $\eta(x)$ at $x \in \partial S$, satisfying the Lipschitz condition

$$\|\eta(x) - \eta(t)\| \leq \frac{1}{\alpha} \|x - t\|, \quad \textit{for all } x, t \in \partial S.$$

Moreover, for some $\alpha > 0$ above is equivalent to

vi) S belongs to Serra's regular model.

The importance of Theorem 1.4.1 lies in the fact that it links the notions of α-convexity, free rolling condition and Serra's regular model and it relates geometric properties to analytic concepts whose mathematical treatment is, in principle, quite different. Turning to the α-convex hull estimator, it is common in the literature to assume that S satisfies the conditions of Theorem 1.4.1. However, for our purposes it suffices to assume that

(A1) S is a nonempty compact subset of $\mathbb{R}^d$ such that a ball of radius $\alpha > 0$ rolls freely in S and in $\overline{S^c}$,

where the free rolling condition must be understood in the sense of Definition 1.4.5. We do not need the set S to be path-connected nor the more restrictive free rolling condition assumed in Theorem 1.4.1. Assumption (A1) is enough to guarantee that S is α-convex, see Lemma A.0.8 in Appendix A. It also guarantees the existence at each point $s \in \partial S$ of a unique outward pointing unit normal vector $\eta(s)$ such that $B(s-\alpha\eta(s),\alpha) \subset S$ and $B(s+\alpha\eta(s),\alpha) \subset \overline{S^c}$, see Lemma A.0.5 for the precise statement and proof. In some sense these results can be thought as an alternative proof for Remark 3 in Walther (1999) referring to the validity of Theorem 1.4.1 when the set S is not assumed to be path-connected. Lemma A.0.5 will be very useful when studying the convergence rate of the α-convex hull estimator. Another implication of Assumption (A1) has to do with the concept of positive reach of a set, not mentioned so far. Federer (1959) defines the reach of a set S, $\text{reach}(S)$, as the largest α, possibly infinity, such that if $x \in \mathbb{R}^d$ and $d(x,S) < \alpha$, then the metric projection of x onto S is unique. Federer (1959) provides a generalization of the Steiner's formula for sets with positive reach. Recall that, roughly speaking, the Steiner's formula establishes that the d-dimensional measure of the closed r-neighbourhood of a convex set in $\mathbb{R}^d$ can be expressed as a polynomial of degree at most d in r. Although the characterization of the sets of positive reach is beyond the scope of this work, Lemma A.0.7 relates the free rolling condition and the reach. More precisely that result states that under Assumption (A1), the reach of S is greater or equal to α. Lemma A.0.7 will be useful in Chapters 2 and 3.

We end this section with a review of the main existing results on the behaviour of the α-convex hull estimator. The proximity between a set S and the α-convex hull of a sample of points taken into it is studied in Rodríguez-Casal (2007). If no assumption is made on S, apart for the α-convexity, it can be proved that $d_H(S, C_\alpha(\mathcal{X}_n)) = O((\log n/n)^{1/d})$ almost surely. Note that, although the family of α-convex sets is much wider than the family of convex sets, the convergence rates of $d_H(C_\alpha(\mathcal{X}_n), S)$ and $d_H(H_n, S)$ are of the same order, see Dümbgen and Walther (1996). If S is under the conditions of Theorem 1.4.1, it is proved that $d_H(S, C_\alpha(\mathcal{X}_n)) = O((\log n/n)^{2/(d+1)})$ almost surely. Again, the order of convergence of $d_H(S, C_\alpha(\mathcal{X}_n))$ is equal to that obtained for $d_H(S, H_n)$ when S is convex and satisfies the smoothness conditions of Theorem 1.4.1. The same order of convergence $(\log n/n)^{2/(d+1)}$ is obtained for $d_H(\partial S, \partial C_\alpha(\mathcal{X}_n))$ and $d_\mu(S, C_\alpha(\mathcal{X}_n))$. We must be aware, however, that the estimator (1.10) suffers from an inherent limitation since, in practice, the parameter α is typically unknown. When this is the case, $C_{r_n}(\mathcal{X}_n)$ is proposed to estimate S, with $r_n > 0$. Note that if S is α-convex, then it is also r_n-convex for $r_n \leq \alpha$ and, therefore, the estimator $C_{r_n}(\mathcal{X}_n)$ seems to be a sensible choice whenever r_n is small enough. This can be guaranteed if we choose $r_n \to 0$. Rodríguez-Casal (2007) provides the convergence rates for $d_H(S, C_{r_n}(\mathcal{X}_n))$, $d_\mu(S, C_{r_n}(\mathcal{X}_n))$

and $d_H(\partial S, \partial C_{r_n}(\mathcal{X}_n))$. Further details on the r_n-convex hull estimator are given in Chapter 2, where we analyse the asymptotic behaviour of $\mathbb{E}(d_\mu(S, C_{r_n}(\mathcal{X}_n)))$.

1.5 When the target is the surface area

Until now we have discussed the problem of estimating a set from a random sample of points taken into it. When studying this problem, one immediately realizes that, apart from the set itself, there are geometrical characteristics that may be of interest. Everybody remembers the perimeter and area of the square, the triangle or the circle. And almost everybody remembers the formulas for the volume or surface area of the sphere, for example. These are some of the geometric characteristics we referred to. They provide us with important additional information about the shape of the set and, therefore, it is useful to know them. Consider for example, in the bidimensional case, the ratio between the perimeter and the squared root of the area of a set. This measure, known in the literature as contour index, provides a scale invariant measurement of boundary roughness. Its minimal value, $2\sqrt{\pi}$, is attained by the circle and it increases as the set becomes more fragmented. The contour index has been used as an auxiliary diagnosis criterion in medical imaging. For example, in oncology the irregularity in the border of a tumor may suggest a bad prognosis since the damage is highly disseminated, see Cuevas et al. (2007) for more details. In this section we focus our attention on the estimation of the surface area of a set. The estimation of the surface area of a set S in the Euclidean space $\mathbb{R}^d$ has been extensively considered in the literature. Some of the most relevant results in this field, obtained by using tools of nonparametric statistics, have been published in recent years. When one has to confront the problem of estimating the surface area of a body from which we only have a sample of points $\mathcal{X}_n = \{X_1, \ldots, X_n\}$, many questions arises in a moment. The first and at the same time most naive one is how to do it. One immediately visualizes the problem in $\mathbb{R}^2$, selects those points of the sample which are closer to the boundary of the set and adds up the length of the segments that join the selected points. The first dilemma we face with is how to determine which points of the sample are closer to the boundary of the set. This has to do with the intuitive idea of what an extreme point is. Thus, we could consider the convex hull of the sample H_n as starting point, recall (1.6). However, the convex hull does not always work well and it is not difficult to picture situations where the perimeter of H_n systematically underestimates the real perimeter of the set, see for example Figure 1.13 (a), where the convex hull of a uniform sample of size $n = 500$ in the disc $B(0, 0.5) \setminus \mathring{B}(0, 0.25)$ is represented. The asymptotic properties of certain statistics of the convex hull of a sample in $\mathbb{R}^2$ were studied by Bräker and Hsing (1998), among others. They obtained the asymptotic normality of the perimeter of H_n as well as its convergence rate in mean. In spite of the fact that the results are really significant, they are established on the assumption that the set of interest is convex, which may be too restrictive in practice as we have already argued. The generalization of the definition of convex hull, leads to new geometric objects that capture the shape of the set of interest, even when the set is not convex. These geometric objects, such as the α-shape, closely related to the α-convex hull, have their origin in the field of computational geometry, and are based on the weakening of the notion of extreme point, see Edelsbrunner et al. (1983). Although the computational geometry framework is deterministic, we can adapt the definitions by substituting a sample $\mathcal{X}_n$ for a finite point set. Thus, given a

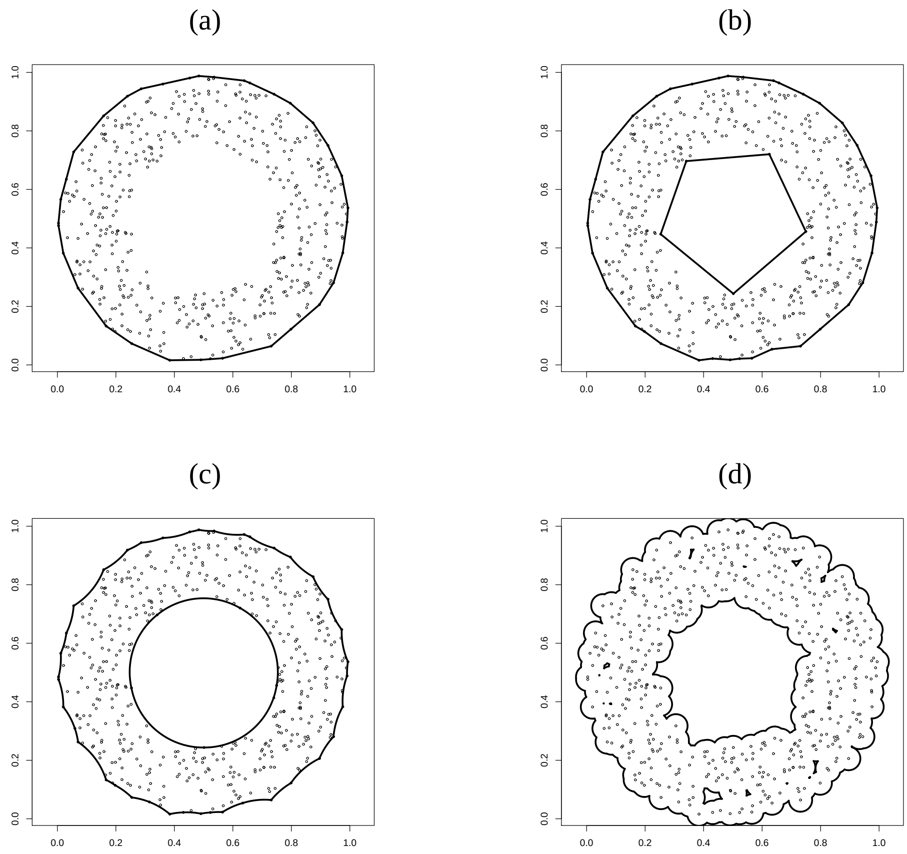


Figure 1.13: *Estimation of the boundary of the disc* $B(0, 0.5) \setminus \mathring{B}(0, 0.25)$ *from a uniform sample of size* $n = 500$. *(a) Convex hull estimator. For* $\alpha = 0.25$, *(b)* α*-shape and (c)* α*-convex hull. (d) Devroye-Wise estimator for* $\varepsilon = 0.04$.

sample $\mathcal{X}_n$ and $\alpha > 0$, the α-shape of $\mathcal{X}_n$ is a polytope which is neither necessarily convex nor necessarily connected. The precise definition of α-shape relies on the notions of α-extreme and α-neighbours.

Definition 1.5.1. *A sample point* X_i *is termed* α*-extreme if there exists a closed ball of radius* α, $B(x, \alpha)$, *such that* X_i *lies on its boundary and* $\mathring{B}(x, \alpha)$ *does not intersect the sample.*

Definition 1.5.2. *If for two* α*-extreme points* X_i *and* X_j *there exists a closed ball of radius* α *such that both points lie on its boundary and the interior of the ball do not enclose any of the points of the sample, then* X_i *and* X_j *are said to be* α*-neighbours.*

Definition 1.5.3. *The* α*-shape is the straight line graph whose vertexes are the* α*-extreme points and whose edges connect the respective* α*-neighbours.*

Figure 1.13 (b) shows the α-shape of the sample in the disc $B(0, 0.5) \setminus \mathring{B}(0, 0.25)$, for $\alpha = 0.25$. The value of the parameter α controls the shape of the estimator. For sufficiently large α, the α-shape is identical to the convex hull of the sample. As α decreases, the shape shrinks until that, for sufficiently small α, the α-shape is the empty set. In Figure 1.14, it is shown the influence of the value of α over the α-shape. Even though the α-shape seems to achieve good results for adequate values of α, it also presents some difficulties. First, the α-shape is a subgraph of the Delaunay triangulation and, therefore, its implementation is based on the construction of the Voronoi diagram and the triangulation of Delaunay of the sample. This implementation is not straightforward and, for the moment, we have programmed this estimator for the bidimensional case in R. Anyway, the main difficulty is related to the manner in which the problem can be tackled from the theoretical point of view.

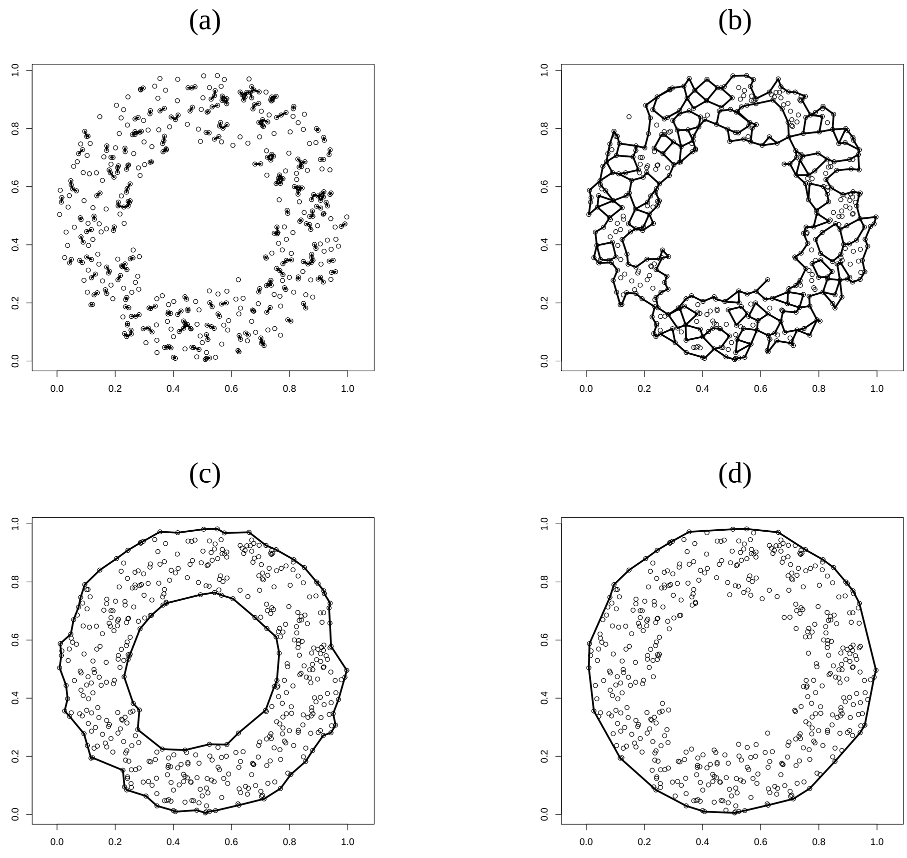


Figure 1.14: *Influence of the value of α over the α-shape of a uniform sample of size $n = 500$ on the disc $B(0, 0.5) \setminus \mathring{B}(0, 0.25)$. (a) $\alpha = 0.01$, (b) $\alpha = 0.03$, (c) $\alpha = 0.07$, (d) $\alpha = 1$.*

Another approach to the estimation of the surface area of a set S consists of making use of the known support estimators. Intuitively, we can think that if a given estimator works well as an estimator of the boundary ∂S, its surface area will also work well as an estimator of the surface area of S. However, this is not always true. See, for example, Figure 1.13 (d), where the Devroye-Wise estimator for the uniform sample of size $n = 500$ on the disc $B(0, 0.5) \setminus \mathring{B}(0, 0.25)$ is represented, with $\varepsilon_n = 0.04$. Recall (1.5) for the definition of the Devroye-Wise estimator. In spite of the fact that the Devroye-Wise estimator works well as an estimator of the support and of the boundary, when computing the surface area of the estimator the results are not so good. The irregularity of the boundary of the estimator does not affect to the Hausdorff distance between the original set and the estimator but it contributes to increase the surface area, as it is shown in a very simple example in Figure 1.15.

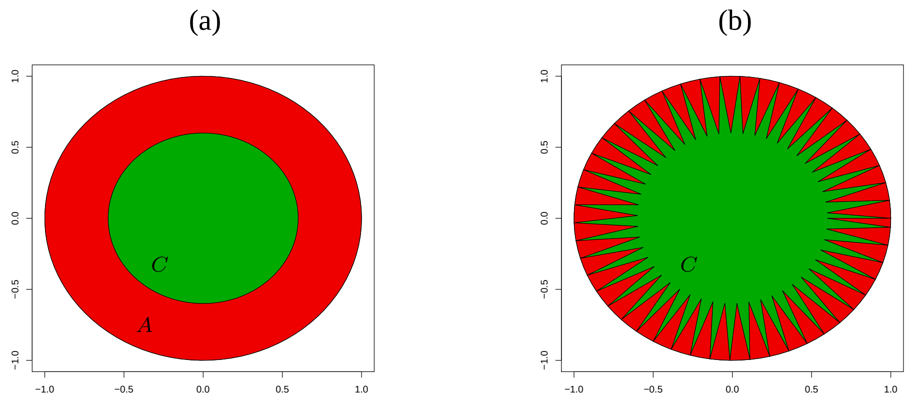


Figure 1.15: *The Hausdorff distance $d_H(A, C)$ is the same in both cases (a) and (b).*

And what about the surface area of the α-convex hull of the sample, $C_\alpha(\mathcal{X}_n)$? Recall (1.10) for the definition of the estimator. In Figure 1.13 (c) we represent the α-convex hull for the example considered along this section, with $\alpha = 0.25$. The main obstacle we encounter when we try to determine the surface area of the α-convex hull of the sample is that, although $C_\alpha(\mathcal{X}_n)$ is completely known, it is hard to identify its boundary explicitly and handle it theoretically.

There is another alternative to the estimation of the surface area of a set, based on the notion of Minkowski content, see Mattila (1995) for a complete discussion of this topic. This approach represents the basis of the work by Cuevas et al. (2007) and serves us as pattern and starting point to develop the results in Chapter 3.

Definition 1.5.4. *The surface area of a body $A \subset \mathbb{R}^d$ is given by the Minkowski content,*

$$L_0(A) = \lim_{\varepsilon \to 0} \frac{\mu(B(\partial A, \varepsilon))}{2\varepsilon} = \lim_{\varepsilon \to 0} L(\varepsilon),$$

provided that this limit exists and it is finite, where

$$L(\varepsilon) = \frac{\mu(B(\partial A, \varepsilon))}{2\varepsilon}.$$

When trying to determine the boundary of a set it seems important to know, not only the points that belong to the set, but also the points that do not belong to it. In fact, the boundary is somewhere in between points of the set and points of its complement. In view of Definition 1.5.4, we realize that the problem of estimating the surface area of a set reduces to the problem of estimating the measure of the dilation of its boundary. And this cannot be done correctly unless we have information of both the set and its complement. The sampling model considered by Cuevas et al. (2007) is justified by this idea. Thus, let G denote the set of interest. Assume without lost of generality that $G \subset (0,1)^d$ and define $R = [0,1]^d \setminus \mathrm{int}(G)$. The sampling information is given by i.i.d. observations $(Z_1, \xi_1), \ldots, (Z_n, \xi_n)$ of a random variable (Z, ξ), where Z is uniformly distributed on the unit square $[0,1]^d$ and $\xi = \mathbb{I}_{\{Z \in G\}}$. Let us denote $\mathcal{X}_n = \{Z_i : \xi_i = 1\}$ and $\mathcal{Y}_n = \{Z_i : \xi_i = 0\}$, see Figure 1.16. For simplicity we abbreviate

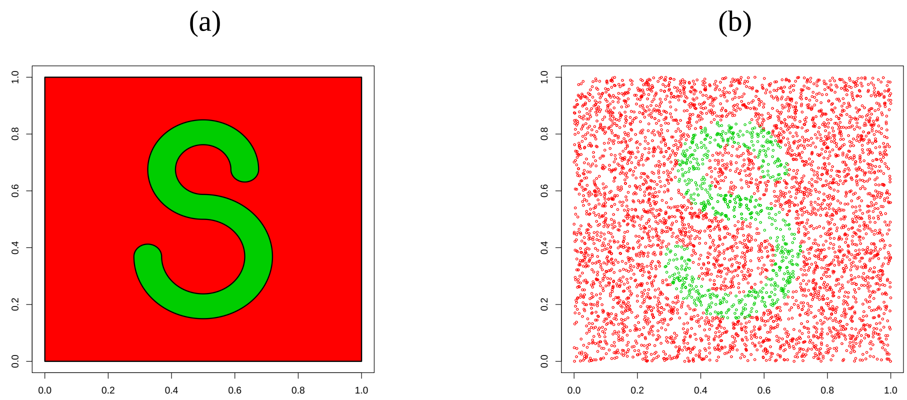


Figure 1.16: *(a) In green the set G. In red $R = [0,1]^d \setminus \mathrm{int}(G)$. (b) Uniform sample of size $n = 5000$ on the unit square. In green $\mathcal{X}_n$ and in red $\mathcal{Y}_n$.*

$L_0(G)$ to L_0 and ∂G to Γ. In view of Definition 1.5.4, a natural estimator of L_0 is given by

$$L_n = \frac{\mu(\Gamma_n)}{2\varepsilon_n}, \tag{1.11}$$

being Γ_n an estimator of $B(\Gamma, \varepsilon_n)$. And how do we estimate the dilation of the boundary $B(\Gamma, \varepsilon_n)$? The key consists of using the following representation of $B(\Gamma, \varepsilon_n)$, valid under mild conditions,

$$B(\Gamma, \varepsilon_n) = B(G, \varepsilon_n) \cap B(R, \varepsilon_n).$$

Therefore, it is possible to construct an estimator of L_0 from estimators of the sets G and R. Thus, if G_n and R_n denote estimators of G and R, respectively, let

$$\Gamma_n = B(G_n, \varepsilon_n) \cap B(R_n, \varepsilon_n). \tag{1.12}$$

The choice of the set estimators G_n and R_n leads us back to Section 1.3 and the comments therein. Cuevas et al. (2007) proposed to estimate G and R empirically by means of the samples $\mathcal{X}_n$ and $\mathcal{Y}_n$, respectively. We will refer to the estimator L_n obtained this way as empirical estimator. Note that if we replace G_n and R_n in (1.12) by $\mathcal{X}_n$ and $\mathcal{Y}_n$, then Γ_n turns out to be the intersection of two Devroye-Wise estimators. Some theoretical properties of the empirical estimator concerning strong consistency, L_1-error and convergence rates can be found in Cuevas et al. (2007). For example, they proved the universal consistency of the estimator, provided that L_0 exists. More precisely, under standardness hypothesis preventing the set G from having too sharp inlets and peaks along Γ, if $\{\varepsilon_n\}$ is a sequence of positive numbers satisfying $\varepsilon_n \to 0$ and $n\varepsilon_n^d / \log n \to \infty$, then $L_n \to L_0$ almost surely. Under stronger assumptions, the L_1-convergence rate for the estimator L_0 is attained, being of order $n^{-1/2d}$.

As occurred with the support estimation problem discussed in Section 1.3, more sophisticated estimators can be considered if we are given some information on the set G. Chapter 3 focuses on the estimation of the surface area of a body G satisfying Assumption (A1), see page 16. It can be proved that, under Assumption (A1), the sets G and R are both α-convex. For this reason we propose to estimate G and R by means of the α-convex hull of $\mathcal{X}_n$ and $\mathcal{Y}_n$, respectively. Thus, let

$$G_n = C_\alpha(\mathcal{X}_n) \text{ and } R_n = C_\alpha(\mathcal{Y}_n).$$

The estimator L_n in (1.11), obtained after substituting $B(C_\alpha(\mathcal{X}_n), \varepsilon_n) \cap B(C_\alpha(\mathcal{Y}_n), \varepsilon_n)$ for Γ_n, is studied in depth in Chapter 3.

1.6 A brief overview of the main results

The aim of this section is to briefly highlight the main results achieved in the course of this research. During this time, our interest has been mainly focused on the support and surface area estimation problems introduced in Sections 1.3, 1.4, and 1.5. As mentioned, the effective estimation of a set is not an easy task and it heavily depends on the assumptions of the model. If no information about the shape of the set is given, then we have no choice but to consider flexible estimators that cover quite different situations. More sophisticated estimators can be considered if we restrict the family of sets to approximate. Traditionally, the support estimation problem has been addressed for the family of convex sets. The convexity assumption, however, may be too restrictive in practice and, for this reason, we concentrate on a more flexible geometrical condition, the α-convexity.

1.6.1 Results on the estimation of α-convex sets

Chapter 2 focuses on the estimation of α-convex sets. Under this assumption, the α-convex hull of a sample of points taken into the set of interest turns out to be the natural estimator. Formally, let $S \subset \mathbb{R}^d$ be a nonempty α-convex compact set with $\alpha > 0$. The goal is to estimate S based on a sample $\mathcal{X}_n$ from a random variable X with absolutely continuous probability distribution P_X and support S. Since the parameter α is typically unknown, we consider the estimator $C_{r_n}(\mathcal{X}_n)$, where r_n is assumed to be lower or equal to α for all n. Is $C_{r_n}(\mathcal{X}_n)$ a consistent estimator of S? Under which conditions? How closely is S approximated by $C_{r_n}(\mathcal{X}_n)$? In Chapter 2 we give answer to these questions. A sufficient and necessary condition for the consistency of the r_n-convex hull estimator is given in Theorem 2.5.1. It is proved that $\mathbb{E}(d_\mu(S, C_{r_n}(\mathcal{X}_n))) \to 0$ if and only if $nr_n^d \to \infty$. It is worth mentioning that the α-convexity assumption is not essential for the consistency of the estimator. In fact, it can be proved that, if $r_n \to 0$ and $nr_n^d \to \infty$, then we still have $\mathbb{E}(d_\mu(S, C_{r_n}(\mathcal{X}_n))) \to 0$, even if S is not α-convex. Note that the assumptions on r_n are identical to those on the smoothing parameter of the Devroye-Wise estimator yielding its consistency in probability, see Devroye and Wise (1980).

Regarding the proximity between S and $C_{r_n}(\mathcal{X}_n)$, we concentrate on the distance in measure between both sets. The almost sure convergence rate for $d_\mu(S, C_{r_n}(\mathcal{X}_n))$ was obtained by Rodríguez-Casal (2007), assuming that S is under the conditions of Theorem 1.4.1. More precisely, it was proved that the order of convergence is $r_n^{-1}(\log n/n)^{2/(d+1)}$. In Theorem 2.5.2 we provide the convergence rate of $\mathbb{E}(d_\mu(S, C_{r_n}(\mathcal{X}_n)))$. As in Rodríguez-Casal (2007), we require an additional condition on S which, in particular, implies the α-convexity. We assume that a ball of radius $\alpha > 0$ rolls freely in S and in $\overline{S^c}$. This free rolling type condition plays a major role in the proofs and it deserves some comments. First, it excludes the presence of sharp peaks in the set. Note that, by merely assuming α-convexity, we cannot ensure that the boundary of the set is smooth. On the other hand, assuming that a ball of radius $\alpha > 0$ rolls freely in S rules sets with isolated points out, for example. Roughly speaking, the free rolling condition in S forces the boundary points to be in direct contact with the interior of the set. At this point one may wonder, in view of the important role of the free rolling condition, why the title of Chapter 2 only refers to the α-convexity. Well, the reason is that the α-convexity is the condition which originally motivated the definition of the estimator. The α-convex hull of a sample makes sense regardless of more restrictive assumptions on S and, for this reason, we have decided to emphasize this property.

Regarding the probability distribution, it is useful to assume that P_X is uniformly bounded on S. Formally, P_X is uniformly bounded on S if there exists $\delta > 0$ such that $P_X(C) \geq \delta\mu(C \cap S)$ for all Borel set $C \subset \mathbb{R}^d$. Is it straightforward to verify that, for example, the uniform distribution on S is uniformly bounded.

Having discussed the assumptions, we are now ready to state the main result of Chapter 2. Then, let S be a nonempty compact subset of $\mathbb{R}^d$ such that a ball of radius $\alpha > 0$ rolls freely in S and in $\overline{S^c}$ and assume that P_X is uniformly bounded on S. Under these conditions, Theorem 2.5.2 states that if the sequence $\{r_n\}$ satisfies

$$\lim_{n\to\infty} \frac{nr_n^d}{\log n} = \infty,$$

then

$$\mathbb{E}(d_\mu(S, C_{r_n}(\mathcal{X}_n))) = O\left(r_n^{-\frac{d-1}{d+1}} n^{-\frac{2}{d+1}}\right).$$

We must not forget to say that the concept of unavoidable family of sets, discussed in detail in Sections 2.3 and 2.4 plays a major role in Chapter 2 and it is essential for proving Theorem 2.5.2. Finally, we prove in Theorem 2.5.3 that the obtained convergence rate of $\mathbb{E}(d_\mu(S, C_{r_n}(\mathcal{X}_n)))$ cannot be improved since there exist sets under the stated conditions for which

$$\liminf_{n\to\infty} r_n^{\frac{d-1}{d+1}} n^{\frac{2}{d+1}} \mathbb{E}(d_\mu(S, C_{r_n}(\mathcal{X}_n))) > 0.$$

These results lead us to compare the convergence rate of $\mathbb{E}(d_\mu(S, C_{r_n}(\mathcal{X}_n)))$, provided by Theorem 2.5.2, to that of $d_\mu(S, C_{r_n}(\mathcal{X}_n))$ (almost sure convergence rate), see Rodríguez-Casal

(2007). The obtained convergence rate for $\mathbb{E}(d_\mu(S, C_{r_n}(\mathcal{X}_n)))$ is faster since the logarithmic term vanishes and the penalty factor $r_n^{-(d-1)/(d+1)}$ is asymptotically smaller than r_n^{-1}.

1.6.2 Results on the surface area estimation

The focus of Chapter 3 is to address the problem of the surface area estimation. When introducing this problem in Section 1.5, we made a distinction between the case where the sampling information comes from points in the set of interest and the case where the sampling information comes from points both in the set of interest $G \subset (0,1)^d$ and in $R = [0,1]^d \setminus \text{int}(G)$. The former case can be described as a further step in support estimation. In spite of the fact that this approach seems more elementary and intuitive, it turns out to be difficult to handle since it is not straightforward to know whether a sample point is close to the boundary. In Chapter 3 we confine ourselves to the case where the sampling information is given by i.i.d. observations $(Z_1, \xi_1), \ldots, (Z_n, \xi_n)$ of a random variable (Z, ξ), where Z is uniformly distributed on the unit square $[0,1]^d$ and $\xi = \mathbb{I}_{\{Z \in G\}}$. Using the notation introduced in Section 1.5, we consider

$$L_n = \frac{\mu(\mathit{\Gamma}_n)}{2\varepsilon_n},$$

being $\mathit{\Gamma}_n$ an estimator of $B(\mathit{\Gamma}, \varepsilon_n)$ and $\varepsilon_n > 0$. Recall that the above expression for L_n is motivated by Definition 1.5.4. Thus, for small values of ε_n, L_n estimates L_0, the Minkowski content of G. Assume that G and R are both α-convex. Then, we propose to estimate $B(\mathit{\Gamma}, \varepsilon_n)$ by

$$\mathit{\Gamma}_n = B(C_\alpha(\mathcal{X}_n), \varepsilon_n) \cap B(C_\alpha(\mathcal{Y}_n), \varepsilon_n)$$

where $\mathcal{X}_n = \{Z_i : \xi_i = 1\}$ and $\mathcal{Y}_n = \{Z_i : \xi_i = 0\}$. A question of theoretical importance is the existence of the Minkowski content L_0. This has to do with the behaviour of the function $\mu(B(\mathit{\Gamma}, \varepsilon))$ and, therefore, with the assumptions on the set G. Regarding the estimator, the natural question is whether or not L_n accurately approximates L_0. Analogous to the support estimation problem, the results in Chapter 3 are obtained under an additional free rolling type condition. Again, it is assumed that a ball of radius $\alpha > 0$ rolls freely in G and in $\overline{G^c}$. This condition ensures that the Minkowski content is well defined. Anyway, L_n makes sense under milder conditions. For example, the α-convexity of G and R is enough to ensure that, with probability one, $\mathit{\Gamma}_n \subset B(\mathit{\Gamma}, \varepsilon_n)$. This fact shows that L_n is biased, as it tends to underestimate L_0. The asymptotic properties of L_n are studied and compared to those of the surface area estimator proposed by Cuevas et al. (2007). Theorems 3.3.1 and 3.3.2 provide, respectively, the almost sure convergence rate and the L_1-convergence rate of L_n to L_0. More precisely, under the stated conditions it is proved that, with probability one,

$$\inf_{\varepsilon_n} |L_n - L_0| = O\left(\frac{\log n}{n}\right)^{\frac{1}{d+1}},$$

where the optimal order is attained for $\varepsilon_n = (\log n/n)^{1/(d+1)}$. Regarding the L_1-convergence rate, we prove that the logarithmic factor can be removed in the above expression and hence,

$$\inf_{\varepsilon_n} \mathbb{E}\,|L_n - L_0| = O\left(n^{-\frac{1}{d+1}}\right).$$

The optimal order in this case is attained for $\varepsilon_n = n^{-1/(d+1)}$. The L_1-convergence of the proposed estimator is, therefore, faster than that of the empirical estimator proposed by Cuevas et al. (2007), which was proved to be of order $n^{-1/2d}$.

1.6.3 Computational issues

Having discussed the theoretical properties of different support and surface area estimators, Chapter 4 focuses on how practical analysis can be carried out. Computing the α-convex hull is not immediate and, for this reason, we devote part of Chapter 4 to the description of the implementation algorithm proposed by Edelsbrunner et al. (1983).

As well as the α-convex hull, we have programmed the surface area estimator discussed in Chapter 3 for the particular case of $\mathbb{R}^2$. We illustrate the surface area estimation problem via a simulation study in which we compare our estimator to that proposed by Cuevas et al. (2007). Since the study did not achieve the expected success, an alternative approach to the surface area estimation problem is discussed. Given the α-convex hull of a sample, we can compute its boundary length by adding the lengths of the arcs that form its boundary. Analogous, we can consider the length of the α-shape. A simulation study on the performance of these kind of surface area estimators is also provided in Chapter 4.

As a conclusion, the obtained results do not suggest that the models based on the Minkowski content are significantly better than those based on the more intuitive idea of measuring the boundary of a support estimator. The promising results that this last simulation study reveals encourage us to find a theoretical justification that explains this good behaviour. Therefore, further research on this topic is needed.

Finally, it is worthwhile to point out that, as a consequence of the implementation in R of the discussed estimators, we have developed a new library named `alphahull`. The complete documentation of the package, including the description of the functions, is available in Appendix C. We would like to highlight here some of the most important features of the library. Apart from the functions that compute the support and surface area estimators used in the simulation studies, the `alphahull` package includes some other functions that can be useful in different contexts. For example, we have programmed the Voronoi diagram and the Delaunay triangulation of a given sample of points. The Voronoi diagram and the Delaunay triangulation are widely used in many fields of research and, as far as we know, there is not a refined code in R providing these geometric structures. Therefore, we aim for the `alphahull` package to be thought not only as collection of functions to carry out the discussed simulation studies, but as a useful tool for further research beyond the context of this work.

Chapter 2

Estimation of α-convex sets

2.1 Introduction

Having reviewed the basics of set estimation, we now turn our attention to the problem of support estimation under the assumption of α-convexity. The α-convexity, defined in Section 1.4, is a condition that affects the shape of the set of interest but which is less restrictive than convexity and therefore, it allows a wider range of applications.

This chapter is organized as follows. We begin with a formal description of the framework and the estimator under study, the α-convex hull of a random sample of points taken in the set of interest. In order to obtain the asymptotic properties of the estimator it will be useful to construct unavoidable families of sets. The precise definition of unavoidable family is given in Section 2.2. Section 2.3 is entirely devoted to the definition of suitable unavoidable families in the bidimensional case. General results on the construction of such families in $\mathbb{R}^d$ are stated in Section 2.4. Finally, the main results on the behaviour of the estimator, regarding its consistency and optimal convergence rate, are proved in Section 2.5.

2.2 Preliminaries

Let S be a nonempty compact subset of $\mathbb{R}^d$ such that S is α-convex for some $\alpha > 0$. Assume that we are given a random sample $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ from X, where X denotes a random variable in $\mathbb{R}^d$ with absolutely continuous probability distribution P_X and support S. Then, $S = C_\alpha(S)$ and the α-convex hull of the sample

$$C_\alpha(\mathcal{X}_n) = (\mathcal{X}_n \oplus \alpha \mathring{B}) \ominus \alpha \mathring{B}$$

turns out to be a natural estimator for the set S. However, the α-convex hull of the sample has the drawback of depending on the unknown parameter α. This difficulty can be overcome by taking a sequence of positive numbers $\{r_n\}$ converging to zero as n tends to infinity. This ensures that $r_n \leq \alpha$ for large enough n. For the sake of simplicity we assume that $r_n \leq \alpha$ for all n and define the estimator

$$S_n = C_{r_n}(\mathcal{X}_n) = (\mathcal{X}_n \oplus r_n \mathring{B}) \ominus r_n \mathring{B}. \tag{2.1}$$

Since with probability one $\mathcal{X}_n \subset S$, we obtain by the properties of the α-convex hull operator that

$$S_n = (\mathcal{X}_n \oplus r_n \mathring{B}) \ominus r_n \mathring{B} \subset (\mathcal{X}_n \oplus \alpha \mathring{B}) \ominus \alpha \mathring{B} \subset (S \oplus \alpha \mathring{B}) \ominus \alpha \mathring{B} = S. \tag{2.2}$$

If we consider the distance in measure to quantify the similarity in content of S and S_n, then (2.2) yields

$$d_\mu(S, S_n) = \mu(S \Delta S_n) = \mu((S \setminus S_n) \cup (S_n \setminus S)) = \mu(S \setminus S_n).$$

Before beginning the systematic study of the random variable $d_\mu(S, S_n)$, it is convenient to make some preliminary comments since the procedure of bounding the expected value of $d_\mu(S, S_n)$ involves a slight change in the estimator which needs to be justified. Although the definition of S_n given in (2.1) arises naturally in connection with the α-convex hull, the derivation of a bound for $\mathbb{E}(d_\mu(S, S_n))$ is a laborious task which can be simplified if, instead of S_n as defined in (2.1), we consider the estimator

$$S_n = (\mathcal{X}_n \oplus r_n B) \ominus r_n B. \tag{2.3}$$

It is important to note that, although we use the same notation S_n for both $(\mathcal{X}_n \oplus r_n B) \ominus r_n B$ and $(\mathcal{X}_n \oplus r_n \mathring{B}) \ominus r_n \mathring{B}$, both estimators are not necessarily equal, see Figure 2.1. However, we prove

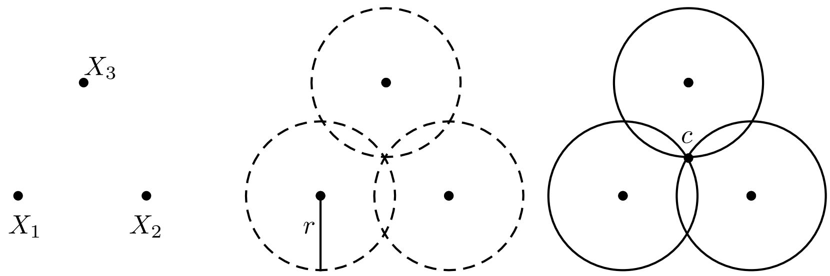


Figure 2.1: *For the point set $\mathcal{X} = \{X_1, X_2, X_3\}$, $(\mathcal{X} \oplus r\mathring{B}) \ominus r\mathring{B} = \mathcal{X}$ and $(\mathcal{X} \oplus rB) \ominus rB = \mathcal{X} \cup \{c\}$.*

in Appendix B that, since P_X is absolutely continuous, with probability one, $(\mathcal{X}_n \oplus r_n B) \ominus r_n B$ coincides with $(\mathcal{X}_n \oplus r_n \mathring{B}) \ominus r_n \mathring{B}$ and hence we can compute $\mathbb{E}(d_\mu(S, S_n))$ by using either (2.1) or (2.3). As we have already commented, the problem of bounding $\mathbb{E}(d_\mu(S, S_n))$ is easier to handle when S_n is defined as in (2.3) and, for this reason, throughout the remainder of this chapter, S_n will refer to the estimator $(\mathcal{X}_n \oplus r_n B) \ominus r_n B$. Then,

$$\begin{aligned}
\mathbb{E}(d_\mu(S, S_n)) &= \mathbb{E}(\mu(S \setminus S_n)) = \mathbb{E}(\mu\{x \in S : x \notin S_n\}) \\
&= \mathbb{E} \int_S \mathbb{I}_{\{x \notin S_n\}} \mu(dx) = \int_S P(x \notin S_n) \mu(dx) \\
&= \int_S P(\exists y \in B(x, r_n) : B(y, r_n) \cap \mathcal{X}_n = \emptyset) \mu(dx).
\end{aligned} \tag{2.4}$$

In order to bound (2.4), we make use of the concept of unavoidable family of sets, defined below.

Definition 2.2.1. *Let $x \in \mathbb{R}^d$, $r > 0$ and $\mathcal{E}_{x,r} = \{B(y,r) : y \in B(x,r)\}$. The family of sets $\mathcal{U}_{x,r}$ is said to be unavoidable for $\mathcal{E}_{x,r}$ if, for all $B(y,r) \in \mathcal{E}_{x,r}$, there exists $U \in \mathcal{U}_{x,r}$ such that $U \subset B(y,r)$.*

As a consequence of Definition 2.2.1, if $\mathcal{U}_{x,r_n}$ is an unavoidable family of sets for $\mathcal{E}_{x,r_n}$, then

$$\{\exists y \in B(x,r_n) : B(y,r_n) \cap \mathcal{X}_n = \emptyset\} \subset \{\exists U \in \mathcal{U}_{x,r_n} : U \cap \mathcal{X}_n = \emptyset\}$$

and then

$$P(\exists y \in B(x,r_n) : B(y,r_n) \cap \mathcal{X}_n = \emptyset) \leq P(\exists U \in \mathcal{U}_{x,r_n} : U \cap \mathcal{X}_n = \emptyset).$$

Moreover, if $\mathcal{U}_{x,r_n}$ is a finite family,

$$\begin{aligned} P(\exists y \in B(x,r_n) : B(y,r_n) \cap \mathcal{X}_n = \emptyset) &\leq P(\exists U \in \mathcal{U}_{x,r_n} : U \cap \mathcal{X}_n = \emptyset) \\ &\leq \sum_{U \in \mathcal{U}_{x,r_n}} P(U \cap \mathcal{X}_n = \emptyset) \\ &= \sum_{U \in \mathcal{U}_{x,r_n}} P(\forall X_j, j = 1,\ldots,n, X_j \notin U) \\ &= \sum_{U \in \mathcal{U}_{x,r_n}} (1 - P_X(U))^n. \end{aligned} \tag{2.5}$$

To sum up, if we define for each $x \in S$ a family $\mathcal{U}_{x,r_n}$ unavoidable and finite for $\mathcal{E}_{x,r_n}$ then, from (2.4) and (2.5), it follows that

$$\begin{aligned} \mathbb{E}(d_\mu(S,S_n)) &= \int_S P(\exists y \in B(x,r_n) : B(y,r_n) \cap \mathcal{X}_n = \emptyset)\mu(dx) \\ &\leq \int_S \sum_{U \in \mathcal{U}_{x,r_n}} (1 - P_X(U))^n \mu(dx). \end{aligned} \tag{2.6}$$

From (2.6) it is apparent that the problem of finding an upper bound for $\mathbb{E}(d_\mu(S,S_n))$ reduces to the problem of finding a lower bound for $P_X(U)$, for all $U \in \mathcal{U}_{x,r_n}$. In view of (2.6) it would be desirable that both the lower bound and the number of elements of the family $\mathcal{U}_{x,r_n}$ depend in the simplest possible way on the point x. How do we define suitable families for $\mathcal{E}_{x,r_n}$? It is clear that, given a point $x \in S$, there is not just one possible unavoidable family $\mathcal{U}_{x,r_n}$ and that the sets $U \subset \mathcal{U}_{x,r_n}$ can substantially change from one family to another. It is important to note that the shape of U determines the value of $P_X(U)$. Therefore, the choice of $\mathcal{U}_{x,r_n}$ is a crucial point in the resolution of (2.6).

As mentioned in the introduction of this chapter, Sections 2.3 and 2.4 are devoted to the definition of suitable unavoidable families in $\mathbb{R}^2$ and $\mathbb{R}^d$, respectively. We wish to emphasize that, although the results in Section 2.3 could have been stated directly in the general framework $\mathbb{R}^d$, we have considered that the proofs in $\mathbb{R}^2$ provide the reader with a more geometric view of the problem. A first approach to the less involved bidimensional case may be helpful since it gives insight into some special features that arise as a consequence of the increase in the dimension of the space. Anyhow, readers may prefer to omit Section 2.3 on their first encounter and return for reference if required.

2.3 Defining unavoidable families in $\mathbb{R}^2$

The main goal of this section is to define unavoidable families of sets $\mathcal{U}_{x,r_n}$ for each $x \in S \subset \mathbb{R}^2$ and find a lower bound for the probability $P_X(U)$, for $U \in \mathcal{U}_{x,r_n}$. But, how can we define the above-mentioned families? How does the point $x \in S$ affect the definition of the family $\mathcal{U}_{x,r_n}$ and the probability $P_X(U)$? In order to find a lower bound for $P_X(U)$ it is useful to assume that the probability distribution P_X is uniformly bounded on S.

Definition 2.3.1. *Let P_X be a probability distribution with support $S \subset \mathbb{R}^2$. It is said that P_X is uniformly bounded on S if*

$$\exists \delta > 0 \text{ such that } P_X(C) \geq \delta \mu(C \cap S) \tag{2.7}$$

for all Borel set $C \subset \mathbb{R}^2$.

Remark 2.3.1. *If the probability distribution P_X is uniform on S, condition (2.7) is satisfied with $\delta = 1/\mu(S)$. In this case we have $P_X(C) = \delta\mu(C \cap S)$. In general, $\delta \leq 1/\mu(S)$.*

Remark 2.3.2. *Let us assume that $\mathcal{U}_{x,r_n}$ is an unavoidable family of sets for $\mathcal{E}_{x,r_n}$. Taking into account condition (2.7), the problem of giving a lower bound for $P_X(U)$, with $U \in \mathcal{U}_{x,r_n}$, reduces to measuring the set $U \cap S$. In particular, if $U \subset S$,*

$$P_X(U) \geq \delta\mu(U). \tag{2.8}$$

Remark 2.3.2 gives us the key to defining suitable unavoidable families of sets. Let us assume that U belongs to an unavoidable family $\mathcal{U}_{x,r_n}$ and that $U \subset S$. Then (2.8) is satisfied. Moreover, by the definition of unavoidable family, $U \subset B(y, r_n)$ for some $y \in B(x, r_n)$ and hence the order of $\mu(U)$ will be r_n^2 at most. In other words, the best lower bound we can obtain for $P_X(U)$ in this context is of order r_n^2. So the question is: can we define unavoidable families $\mathcal{U}_{x,r_n}$ such that $U \subset S$ for all $U \in \mathcal{U}_{x,r_n}$, being the measure of U of order r_n^2? Evidently it will depend on the point $x \in S$ we are considering. If x is not close to the boundary of S, it seems reasonable to think that we will be able to define large sets U totally contained in S. On the contrary, if the point x is close to the boundary of S it does not seem straightforward to find that kind of sets U. For this reason, we will divide the support S into two subsets; the first one, formed by points which are far away from the boundary of S, and the second one, formed by points which are closer to the boundary. Roughly speaking, for those points x which are far away from the boundary, we will be able to define families $\mathcal{U}_{x,r_n}$ such that the sets $U \in \mathcal{U}_{x,r_n}$ are contained in S, $\mu(U)$ does not depend on x and, on top of that, $\mu(U)$ is of order r_n^2. For those points x which are closer to the boundary things are not that simple. In that case we will have to consider different families $\mathcal{U}_{x,r_n}$ and the values of $P_X(U)$ will depend on $d(x, \partial S)$.

Proposition 2.3.1 gives the desired unavoidable families for the points which are far away from the boundary of S. By points which are far away from the boundary we mean those points $x \in S$ such that $d(x, \partial S) > r_n/2$. Taking into account Definition 2.2.1, it will not be difficult to define a suitable family $\mathcal{U}_{x,r_n}$. We need that, given $y \in B(x, r_n)$, there exists $U \in \mathcal{U}_{x,r_n}$ such that $U \subset B(y, r_n)$. In view of the previous comments, it would be also desirable that U was totally contained in S and that $\mu(U)$ was of order r_n^2. Thus, we would ensure the best possible rate for $P_X(U)$. Note that if $x \in S$ and $d(x, \partial S) > r_n/2$, then the ball $B(x, r_n/2)$ is fully contained in S. So, the idea is to divide $B(x, r_n/2)$ into a finite number of subsets. How? In view of the target it seems reasonable to consider a partition of $B(x, r_n/2)$ into circular sectors. Why circular sectors? This choice rests upon two main reasons. First, the measure of a circular sector of $B(x, r_n/2)$ is of order r_n^2. Second, if the central angle of the defined sectors is not too large, then the resulting family $\mathcal{U}_{x,r_n}$ is unavoidable.

Before the statement of Proposition 2.3.1, we give the precise definition of the circular sectors and introduce some basic notation that will be useful later. The definitions given for $\mathbb{R}^2$ can be easily generalized to the d-dimensional case, as it will be shown in Section 2.4. Thus, let $\mathbb{S}_2 = \{u \in \mathbb{R}^2 : \|u\| = 1\}$ denote the unit circle in $\mathbb{R}^2$. Let $\varphi_{u,v}$ denote the angle between the vectors u and v. It is understood that $\varphi_{u,v} \in [0, \pi]$ and $\varphi_{u,v} = \varphi_{v,u}$. Finally, let $e_2 = (0, 1) \in \mathbb{R}^2$.

Definition 2.3.2. *For $u \in \mathbb{S}_2$ and $\theta \in [0, \pi/2]$, we define the sets*

$$C_u^\theta = \{x \in \mathbb{R}^2 : \langle x, u\rangle \geq \|x\| \cos\theta\}$$

and the circular sectors

$$C_{u,r}^\theta = C_u^\theta \cap B(0, r).$$

Remark 2.3.3. *On the basis of Definition 2.3.2, it is straightforward that $C_{u,r}^\theta$ is the circular sector with central angle 2θ enclosed by the radii $v_1 = r\mathcal{R}_\theta(u)$ and $v_2 = r\mathcal{R}_\theta^{-1}(u)$, where $\mathcal{R}_\theta : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ denotes the counter-clockwise rotation of angle θ, whose associated matrix with respect to the canonical basis is*

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

In Figure 2.2 we show an example of $C_{u,r}^\theta$.

Proposition 2.3.1. *Let S be a nonempty compact subset of $\mathbb{R}^2$ such that a ball of radius $\alpha > 0$ rolls freely in S and in $\overline{S^c}$. Let X be a random variable with probability distribution P_X and support S. We assume that the probability distribution P_X satisfies that there exists $\delta > 0$ such that*

$$P_X(C) \geq \delta\mu(C \cap S)$$

for all Borel set $C \subset \mathbb{R}^2$.

Then, for all $x \in S$ such that $d(x, \partial S) > r_n/2$, there exists a finite family $\mathcal{U}_{x,r_n}$ with $m_1 = 6$ elements, unavoidable for $\mathcal{E}_{x,r_n}$ and that satisfies

$$P_X(U) \geq L_1 r_n^2, \quad U \in \mathcal{U}_{x,r_n},$$

where the constant $L_1 > 0$ is independent of x.

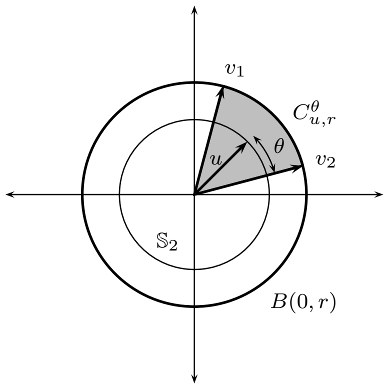


Figure 2.2: *Circular sector* $C^{\theta}_{u,r}$.

Proof. First consider the family

$$\mathcal{U}_{0,r_n} = \{C^{\pi/6}_{u,r_n/2},\ u \in \mathcal{W}\},$$

where $\mathcal{W} \subset \mathbb{R}^2$ denotes a set of unit vectors that divides the unit circle into six circular sectors with central angle $\pi/3$. Figure 2.3 shows one possible choice of $\mathcal{W}$ and the corresponding family $\mathcal{U}_{0,r_n}$. To simplify notation somewhat, we abbreviate $C^{\pi/6}_u$ and $C^{\pi/6}_{u,r_n}$ to C_u and C_{u,r_n}, respectively. Note that the definition of $\mathcal{W}$ implies that

$$B(0, r_n) = \bigcup_{u \in \mathcal{W}} C_{u,r_n}.$$

The fact that $\mathcal{U}_{0,r_n}$ is unavoidable for $\mathcal{E}_{0,r_n}$ easily follows from Lemma 2.3.2, stated below. To see this, note that for $B(y, r_n) \in \mathcal{E}_{0,r_n}$, there exists $u \in \mathcal{W}$ such that $y \in C_{u,r_n}$. Now, by Lemma 2.3.2, $C_{u,r_n} \subset B(y, r_n)$ and therefore $C_{u,r_n/2} \subset B(y, r_n)$. This completes the proof that $\mathcal{U}_{0,r_n}$ is unavoidable. Thus, it remains to prove Lemma 2.3.2. First, however, we need to introduce a preliminary result. Lemma 2.3.1 characterizes the points in C^{θ}_u and will be needed in the proof of 2.3.2.

Lemma 2.3.1. *Let* $x \neq 0$. *Then*

$$x \in C^{\theta}_u \Leftrightarrow \varphi_{x,u} \leq \theta.$$

Proof. Let $x \in C^{\theta}_u$. We have that

$$\|x\| \cos \varphi_{x,u} = \langle x, u \rangle \geq \|x\| \cos \theta. \tag{2.9}$$

The inequality in (2.9) holds if and only if $\varphi_{x,u} \leq \theta$, since the cosine function is monotonically decreasing in $[0, \pi]$. $\square$

We are now ready to state and prove Lemma 2.3.2. This lemma reveals that the partition of $B(0, r_n)$ into circular sectors with central angle $\pi/3$ is indeed a sensible choice, since it guarantees that the constructed sets are unavoidable.

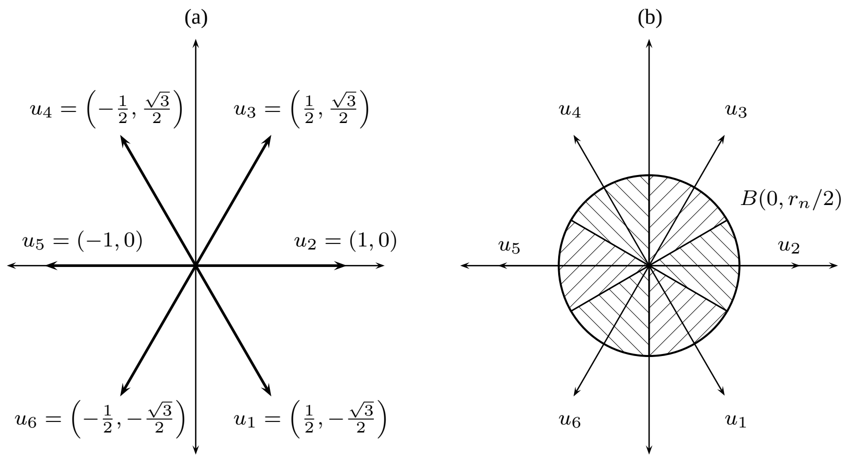


Figure 2.3: *(a) The set* $\mathcal{W} = \{u_i,\ i = 1, \ldots, 6\}$ *divides the unit circle into six circular sectors with central angle* $\pi/3$. *(b) Family* $\mathcal{U}_{0,r_n} = \{C^{\pi/6}_{u,r_n/2}, u \in \mathcal{W}\}$.

Lemma 2.3.2. *For all* $u \in \mathbb{S}_2$ *and* $r > 0$,

$$C_{u,r} \subset \bigcap_{y \in C_{u,r}} B(y, r).$$

Proof. Let $z \in C_{u,r}$. We need to show that, for all $y \in C_{u,r}$, $\|z - y\| \le r$. Assume, without loss of generality, that z and y are both non zero vectors since the result is trivial otherwise. We have that

$$\|z - y\|^2 = \langle z - y, z - y\rangle = \|z\|^2 + \|y\|^2 - 2\langle z, y\rangle = \|z\|^2 + \|y\|^2 - 2\|z\|\|y\| \cos \varphi_{z,y}.$$

By the triangle inequality for angles

$$\varphi_{z,y} \le \varphi_{z,u} + \varphi_{u,y}.$$

Since $z, y \in C_u$, it follows from Lemma 2.3.1 that $\varphi_{z,u} \le \pi/6$ and $\varphi_{y,u} \le \pi/6$. Hence,

$$\varphi_{z,y} \le \varphi_{z,u} + \varphi_{u,y} \le \frac{\pi}{3}$$

and therefore $\cos \varphi_{z,y} \ge \cos(\pi/3) = 1/2$. In short,

$$\|z - y\|^2 \le \|z\|^2 + \|y\|^2 - \|z\|\|y\| \le \max(\|z\|^2, \|y\|^2) \le r^2.$$

$\square$

Once we have proved that $\mathcal{U}_{0,r_n}$ is unavoidable for $\mathcal{E}_{0,r_n}$ consider, for each $x \in S$ such that $d(x, \partial S) > r_n/2$, the family

$$\mathcal{U}_{x,r_n} = \{x\} \oplus \mathcal{U}_{0,r_n} = \{\{x\} \oplus C_{u,r_n/2},\ u \in \mathcal{W}\}.$$

The family $\mathcal{U}_{x,r_n}$, obtained by translating the family $\mathcal{U}_{0,r_n}$ by the vector x, is unavoidable for $\mathcal{E}_{x,r_n}$, as we state in Lemma 2.3.3.

Lemma 2.3.3. *Let $\mathcal{U}_{0,r}$ be an unavoidable family for $\mathcal{E}_{0,r}$. Then $\mathcal{U}_{x,r} = \{x\} \oplus \mathcal{U}_{0,r} = \{\{x\} \oplus U,\ U \in \mathcal{U}_{0,r}\}$ is unavoidable for $\mathcal{E}_{x,r}$.*

Proof. Let $B(y,r) \in \mathcal{E}_{x,r}$. Then $B(y-x,r) \in \mathcal{E}_{0,r}$ and, since $\mathcal{U}_{0,r}$ is unavoidable for $\mathcal{E}_{0,r}$, there exists $U \in \mathcal{U}_{0,r}$ such that $U \subset B(y-x,r)$. The proof is now complete as

$$\{x\} \oplus U \subset \{x\} \oplus B(y-x,r) \equiv B(y,r).$$

□

To complete the proof of Proposition 2.3.1 it remains to give a lower bound for the probability of the sets of the unavoidable family we have just defined. For each $u \in \mathcal{W}$ we have that

$$P_X\left(\{x\} \oplus C_{u,r_n/2}\right) \geq \delta\mu\left(\{x\} \oplus C_{u,r_n/2} \cap S\right) = \delta\mu\left(\{x\} \oplus C_{u,r_n/2}\right) = \delta\mu\left(C_{u,r_n/2}\right).$$

This follows simply because $\{x\} \oplus C_{u,r_n/2} \subset B(x, r_n/2) \subset S$ since $d(x, \partial S) > r_n/2$ and the Lebesgue measure is invariant under translations, see Figure 2.4. Moreover,

$$\mu\left(C_{u,r_n/2}\right) = \frac{1}{6}\mu(B(0, r_n/2)) = \frac{1}{6}\pi\left(\frac{r_n}{2}\right)^2$$

and then

$$P_X\left(\{x\} \oplus C_{u,r_n/2}\right) \geq \delta\frac{1}{6}\pi\left(\frac{r_n}{2}\right)^2.$$

To summarize, we have shown that

$$P_X(U) \geq L_1 r_n^2,\ \ U \in \mathcal{U}_{x,r_n},$$

for $L_1 = \delta\pi/24 > 0$ and the proof of Proposition 2.3.1 is complete.

□

Therefore, given $x \in S$ with $d(x, \partial S) > r_n/2$, Proposition 2.3.1 provides, independently of x, a lower bound for the probability of all the sets in an unavoidable family for $\mathcal{E}_{x,r_n}$. The given family consists of circular sectors with radius $r_n/2$ and central angle $\pi/3$. It is important to note that the collection of unit vectors $\mathcal{W}$ from which the circular sectors are defined is not unique. In particular, any rotation of $\mathcal{W}$ results in a new collection of unit vectors that could also be used to define a new unavoidable family with the same properties as the one considered in Proposition 2.3.1.

Before proceeding to the definition of unavoidable families of sets for points $x \in S$ with $d(x, \partial S) \leq r_n/2$, we wish to emphasize some aspects of this kind of families. So far we have considered circular sectors with central angle $\pi/3$. Which is the role of this angle? Could we have chosen circular sectors with a larger amplitude? And another kind of sets? Of course, we could have defined larger sets provided that they are unavoidable. For points which lie far away from the boundary we have proved that it is enough to consider circular sectors with radius $r_n/2$ and central angle $\pi/3$. Using the same argument for points $x \in S$ such that $\rho = d(x, \partial S) \leq r_n/2$ we only could infer that $B(x, \rho) \subset S$ and hence the lower bound for the probability of

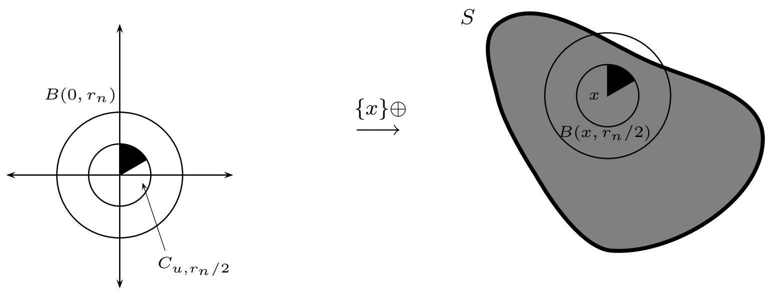


Figure 2.4: *For $x \in S$ under the conditions stated in Proposition 2.3.1, we have that $\{x\} \oplus C_{u,r_n/2} \subset B(x, r_n/2) \subset S$.*

these circular sectors would be of order ρ^2. Can we improve this bound? The answer is yes. We can find larger unavoidable sets. To see this, assume without loss of generality that $x = 0$ and divide $B(0, r)$ into a finite number of sectors $C^{\theta}_{u,r}$ with $\theta > 0$. Then for fixed u,

$$U = \bigcap_{y \in C^{\theta}_{u,r}} B(y, r) \tag{2.10}$$

is the largest set contained in $B(y, r)$ for all $y \in C^{\theta}_{u,r}$. But, what is its measure? Obviously it depends on θ. For example, if $\theta = \pi/2$ then we divide $B(0, r)$ into two circular sectors with central angle π. In that case, it can be easily proved that $U = \{0\}$. Smaller values of θ result in larger sets U. In particular, Lemma 2.3.2 shows that, fixed $\theta = \pi/6$, the set in (2.10) contains at least one circular sector with central angle $\pi/3$. In Proposition 2.3.2 we show that for points $x \in S$ with $\rho = d(x, \partial S) \leq r_n/2$ and $\theta = \pi/6$ we can give a lower bound for $P_X(U)$ of order $r_n^{1/2}\rho^{3/2}$. Note that this bound is better than the one we can obtain for circular sectors of $B(x, \rho)$. Hence, Proposition 2.3.2 provides the second key result in this section. At this point it is worth discussing some of the properties of the sets

$$\bigcap_{y \in C_{u,r}} B(y, r), \text{ with } u \in \mathbb{S}_2, \text{ and } r > 0. \tag{2.11}$$

As will be emphasized later, these sets are known in the literature as Reuleaux triangle, see Figure 2.5. They solve the problem of finding unavoidable families for the bidimensional case. Our first goal was to generalize this concept to the d-dimensional case. However, as will be seen in Section 2.4, the argument in $\mathbb{R}^d$ is somewhat different since it becomes tough to handle with the intersection in (2.11) when $d > 2$. Note that it is fundamental not only to define large unavoidable sets but also to measure them. This causes technical difficulties as the dimension increases. Anyway, an in-depth study of the set (2.11) in $\mathbb{R}^2$ may be helpful since it offers a comprehensive overview of the problem. The following result tells us that these sets are quite simple for the bidimensional case.

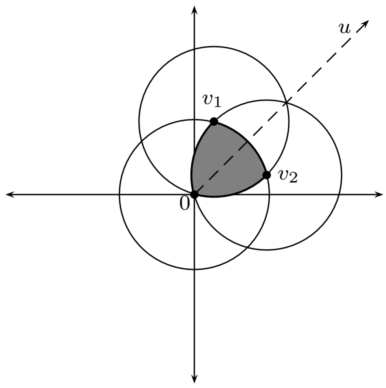


Figure 2.5: *Reuleaux triangle.*

Lemma 2.3.4. *Given $u \in \mathbb{S}_2$, we have*

$$\bigcap_{y \in C_{u,r}} B(y,r) = B(0,r) \cap B(v_1,r) \cap B(v_2,r),$$

where $v_1 = r\mathcal{R}(u)$ and $v_2 = r\mathcal{R}^{-1}(u)$, $\mathcal{R} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ being the counter-clockwise rotation of angle $\pi/6$.

Remark 2.3.4. *As previously discussed, the set $B(0,r) \cap B(v_1,r) \cap B(v_2,r)$ in $\mathbb{R}^2$ is the so-called Reuleaux triangle. Formally, the Reuleaux triangle is defined from an equilateral triangle with sides of length l. It is constructed by drawing the arcs from each polygon vertex of the equilateral triangle between the other two vertices. Thus, the Reuleaux triangle is the set bounded by these three arcs. An important property is that it is a set of constant width l, see Figure 2.6. It is known that the diameter of a set of constant width l is precisely l. See Benson (1966), Croft et al. (1991), Eggleston (1958), and the references cited therein for a detailed development of these concepts.*

Proof. It is straightforward to verify

$$\bigcap_{y \in C_{u,r}} B(y,r) \subset B(0,r) \cap B(v_1,r) \cap B(v_2,r) \tag{2.12}$$

since, by definition, $0 \in C_{u,r}$ and $v_1, v_2 \in C_{u,r}$, as it can be deduced from Remark 2.3.3. Let us now consider the second statement. Let $x \in B(0,r) \cap B(v_1,r) \cap B(v_2,r)$ and $y \in C_{u,r}$. We need to show that

$$\|x - y\| \leq r. \tag{2.13}$$

It follows from (2.12) that $y \in B(0,r) \cap B(v_1,r) \cap B(v_2,r)$ and hence, since the diameter of the Reuleaux triangle is r, (2.13) holds.

□

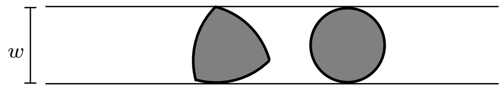


Figure 2.6: *Sets of constant width.*

We now concentrate on the points x which are close to the boundary of S. Recall that by points which are close to the boundary of S we mean those $x \in S$ such that $d(x, \partial S) \leq r_n/2$. As previously described, we shall consider in this context unavoidable sets which are larger than the circular sectors used for points away from ∂S. The unavoidable sets U we shortly define guarantee a lower bound for $P_X(U)$ of order $r_n^{1/2} d(x, \partial S)^{3/2}$. Proposition 2.3.2 makes this ideas precise.

Proposition 2.3.2. *Let S be a nonempty compact subset of $\mathbb{R}^2$ such that a ball of radius $\alpha > 0$ rolls freely in S and in $\overline{S^c}$. Let X be a random variable with probability distribution P_X and support S. We assume that the probability distribution P_X satisfies that there exists $\delta > 0$ such that*

$$P_X(C) \geq \delta \mu(C \cap S)$$

for all Borel set $C \subset \mathbb{R}^2$.

Then, for all $x \in S$ such that $d(x, \partial S) \leq r_n/2$, there exists a finite family $\mathcal{U}_{x,r_n}$ with $m_2 = 6$ elements, unavoidable for $\mathcal{E}_{x,r_n}$ and that satisfies

$$P_X(U) \geq L_2 r_n^{\frac{1}{2}} d(x, \partial S)^{\frac{3}{2}}, \;\; U \in \mathcal{U}_{x,r_n},$$

where the constant $L_2 > 0$ is independent of x.

Proof. Let $x \in S$ such that $d(x, \partial S) \leq r_n/2 < \alpha$. We denote $\rho = d(x, \partial S)$. By Lemmas A.0.7 and A.0.5 there exists a unique point $P_\Gamma x \in \partial S$ and a unique unit vector $\eta \equiv \eta(P_\Gamma x)$ such that

$$B(P_\Gamma x - \alpha\eta, \alpha) \subset S$$

and therefore, given an unavoidable family $\mathcal{U}_{x,r_n}$,

$$P_X(U) \geq \delta\mu(U \cap S) \geq \delta\mu(U \cap B(P_\Gamma x - \alpha\eta, \alpha)), \;\; U \in \mathcal{U}_{x,r_n}. \tag{2.14}$$

Note that this simplifies the proof since by (2.14) it follows that we just need to define a suitable family $\mathcal{U}_{x,r_n}$ and bound $\mu(U \cap B(P_\Gamma x - \alpha\eta, \alpha))$ for $U \in \mathcal{U}_{x,r_n}$. Let us consider a composite function T formed by first applying the orthogonal transformation $\mathcal{O} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ such that $\mathcal{O}(e_2) = -\eta$ and then applying the translation by the vector x, see Figure 2.7. In particular $T(0) = x$, $T((\alpha - \rho)e_2) = x - (\alpha - \rho)\eta = P_\Gamma x - \alpha\eta$, and

$$T(B((\alpha - \rho)e_2, \alpha)) = B(P_\Gamma x - \alpha\eta, \alpha).$$

Then, let $\mathcal{U}_{0,r_n}$ be an unavoidable family for $\mathcal{E}_{0,r_n}$. The following result holds.

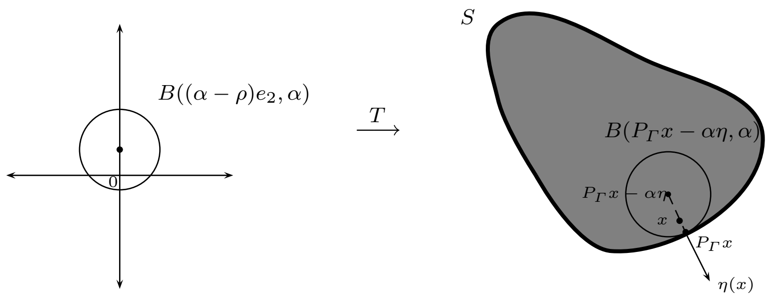


Figure 2.7: *For the function T, $T(B((\alpha-\rho)e_2,\alpha)) = B(P_\Gamma x - \alpha\eta, \alpha)$.*

Lemma 2.3.5. *Let $\mathcal{U}_{0,r}$ be an unavoidable family for $\mathcal{E}_{0,r}$ and let $\mathcal{O} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be an orthogonal transformation. Then $\{\mathcal{O}(U),\ U \in \mathcal{U}_{0,r}\}$ is also an unavoidable family for $\mathcal{E}_{0,r}$.*

Proof. Let $B(y,r) \in \mathcal{E}_{0,r}$. Then $y \in B(0,r)$ and using that $\mathcal{O}$ is an orthogonal transformation, we have that $\mathcal{O}^{-1}(y) \in B(0,r)$. As $\mathcal{U}_{0,r}$ is an unavoidable family for $\mathcal{E}_{0,r}$, there exists $U \in \mathcal{U}_{0,r}$ such that $U \subset B(\mathcal{O}^{-1}(y), r)$. The result is now immediate since

$$\mathcal{O}(U) \subset \mathcal{O}(B(\mathcal{O}^{-1}(y), r)) = B(y,r).$$

□

What Lemma 2.3.5 asserts is that the orthogonal transformation of an unavoidable family for $\mathcal{E}_{0,r_n}$ results in another unavoidable family for $\mathcal{E}_{0,r_n}$. On the other hand, Lemma 2.3.3 established that the result of the translation of an unavoidable family for $\mathcal{E}_{0,r_n}$ by the vector x is an unavoidable family for $\mathcal{E}_{x,r_n}$. As an immediate consequence, we obtain that

$$\mathcal{U}_{x,r_n} = \{T(U), U \in \mathcal{U}_{0,r_n}\}$$

is unavoidable for $\mathcal{E}_{x,r_n}$. Furthermore,

$$\mu(T(U) \cap B(P_\Gamma x - \alpha\eta, \alpha)) = \mu(U \cap B((\alpha-\rho)e_2, \alpha)),$$

as the Lebesgue measure is invariant under translations and orthogonal transformations. Thus, the problem reduces to defining an unavoidable family $\mathcal{U}_{0,r_n}$ for $\mathcal{E}_{0,r_n}$ and finding a lower bound for $\mu(U \cap B((\alpha-\rho)e_2,\alpha))$ for all $U \in \mathcal{U}_{0,r_n}$.

Before continuing the proof of Proposition 2.3.2, it may be useful to make some comments concerning the measure of the sets $U \cap B((\alpha-\rho)e_2,\alpha)$. Note that when defining unavoidable sets for $\mathcal{E}_{0,r_n}$, the main difficulty in giving a lower bound for $\mu(U \cap B((\alpha-\rho)e_2,\alpha))$ arises with those points which lie far away in the direction of the vector $-e_2$. In fact,

$$\min_{y\in B(0,r_n)} \mu\left(B(y,r_n) \cap B((\alpha-\rho)e_2,\alpha)\right) = \mu(B(-r_n e_2, r_n) \cap B((\alpha-\rho)e_2,\alpha))$$

since $y = -r_n e_2$ represents the point where the distance between the centres of both balls attains its maximum and, as a direct consequence, the intersection its minimum. Recall that, by the definition of unavoidable family, for each $y \in B(0, r_n)$ there exists $U \in \mathcal{U}_{0,r_n}$ such that $U \subset B(y, r_n)$. So, it is more involved to find unavoidable sets U with large enough $\mu(U \cap B((\alpha - \rho)e_2, \alpha))$ for points close to $-r_n e_2$. This motivates dividing $B(0, r_n)$ into two subsets as follows

$$B(0, r_n) = \mathcal{G}_{r_n} \cup \mathcal{F}_{r_n},$$

where

$$\mathcal{G}_{r_n} = \left\{ y \in B(0, r_n) : \langle y, e_2 \rangle \geq -\frac{1}{2} \|y\| \right\}$$

and

$$\mathcal{F}_{r_n} = \left\{ y \in B(0, r_n) : \langle y, e_2 \rangle < -\frac{1}{2} \|y\| \right\}.$$

Figure 2.8 shows the sets $\mathcal{G}_{r_n}$ and $\mathcal{F}_{r_n}$. Roughly speaking, $\mathcal{F}_{r_n}$ contains the points $y \in B(0, r_n)$ for which $B(y, r_n) \cap B((\alpha - \rho)e_2, \alpha)$ is small. Therefore, the unavoidable sets U in this case should be carefully selected. On the contrary, $\mathcal{G}_{r_n}$ contains the points $y \in B(0, r_n)$ for which $B(y, r_n) \cap B((\alpha - \rho)e_2, \alpha)$ is larger. For these points the sets U can be circular sectors. Proposition 2.3.3 shows that $\mu(U \cap B((\alpha - \rho)e_2, \alpha))$ is then large enough.

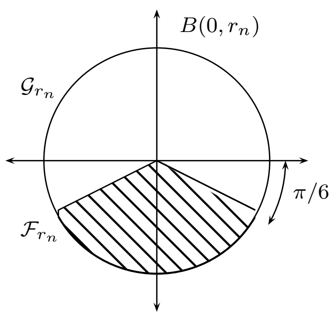


Figure 2.8: $\mathcal{G}_{r_n}$ *and* $\mathcal{F}_{r_n}$.

Proposition 2.3.3. *There exists a finite set of unit vectors* $\mathcal{W}^{\mathcal{G}} \subset \mathbb{S}_2$ *such that, for all* $y \in \mathcal{G}_{r_n}$, *there exists* $u \in \mathcal{W}^{\mathcal{G}}$ *such that* $y \in C_{u,r_n} \subset B(y, r_n)$ *and*

$$\mu(C_{u,r_n} \cap B((\alpha - \rho)e_2, \alpha)) \geq L^{\mathcal{G}} r_n^{\frac{1}{2}} \rho^{\frac{3}{2}},$$

where $L^{\mathcal{G}} > 0$ *is a constant.*

Proof. Let us consider the set

$$\mathcal{W}^{\mathcal{G}} = \{(1, 0), (-1, 0), (1/2, \sqrt{3}/2), (-1/2, \sqrt{3}/2)\}. \tag{2.15}$$

It is straightforward to verify, see Figure 2.9, that

$$\mathcal{G}_{r_n} = \bigcup_{u \in \mathcal{W}^{\mathcal{G}}} C_{u,r_n}.$$

Therefore, for all $y \in \mathcal{G}_{r_n}$ there exists $u \in \mathcal{W}^{\mathcal{G}}$ such that $y \in C_{u,r_n}$. By Lemma 2.3.2 it follows that $C_{u,r_n} \subset B(y, r_n)$. We need to measure $C_{u,r_n} \cap B((\alpha - \rho)e_2, \alpha)$ for $u \in \mathcal{W}^{\mathcal{G}}$. Note that at least half of the set C_{u,r_n} is contained in the halfplane $H_0 = \{x = (x_1, x_2) \in \mathbb{R}^2 : \ x_2 \geq 0\}$ and hence it is sufficient for our purposes to concentrate on $C_{u,r_n} \cap H_0$.

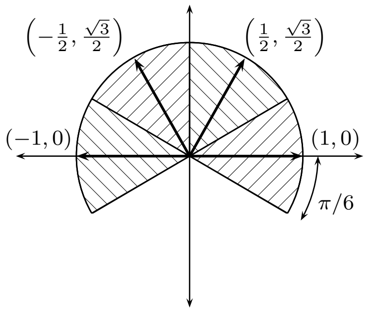


Figure 2.9: *Unit vectors* $\mathcal{W}^{\mathcal{G}} = \{(1,0), (-1,0), (1/2, \sqrt{3}/2), (-1/2, \sqrt{3}/2)\}$ *and* C_{u,r_n}*, for* $u \in \mathcal{W}^{\mathcal{G}}$.

Let $\nu = \sqrt{\rho(2\alpha - \rho)}$. By the Pythagorean theorem, it is straightforward to see that ν represents the distance to the origin from the points such that $\partial B((\alpha - \rho)e_2, \alpha)$ intersects the axis OX, see Figure 2.10.

Lemma 2.3.6.

$$B(0, \nu) \cap H_0 \subset B((\alpha - \rho)e_2, \alpha).$$

Proof. Let $x \in B(0, \nu) \cap H_0$. We have that

$$\begin{aligned} \|x - (\alpha - \rho)e_2\|^2 &= \|x\|^2 + \|(\alpha - \rho)e_2\|^2 - 2(\alpha - \rho)\langle x, e_2 \rangle \\ &\leq \nu^2 + (\alpha - \rho)^2 \\ &= \alpha^2. \end{aligned}$$

The first inequality follows from $x \in H_0$ which implies that $\langle x, e_2 \rangle \geq 0$. The second equality follows immediately from the definition of ν.

$\square$

Lemma 2.3.6 yields the following result. For $u \in \mathcal{W}^{\mathcal{G}}$

$$\begin{aligned} C_{u,r_n} \cap B((\alpha - \rho)e_2, \alpha) &\supset C_{u,r_n} \cap B(0, \nu) \cap H_0 \\ &= C_{u,\tau_n} \cap H_0, \end{aligned}$$

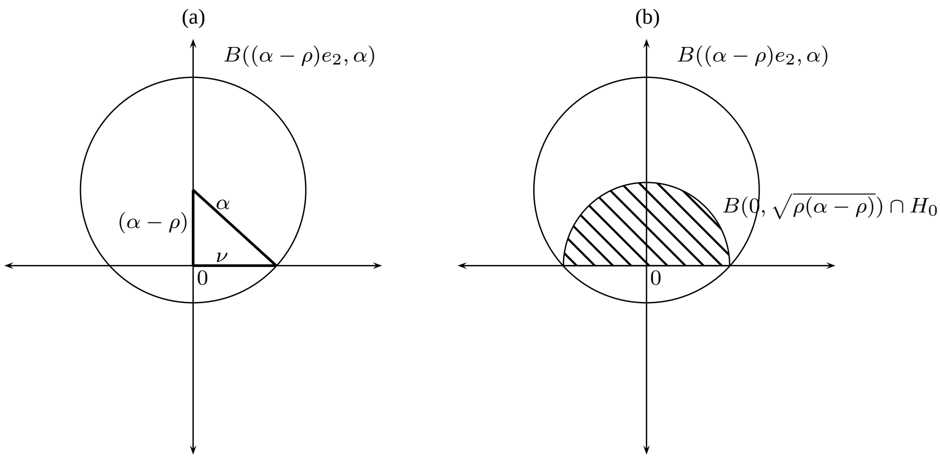


Figure 2.10: *(a)* $\nu=\sqrt{\rho(2\alpha-\rho)}$. *(b)* $B(0,\nu)\cap H_0\subset B((\alpha-\rho)e_2,\alpha)$.

where $\tau_n=\min(\nu,r_n)$. Therefore,

$$\mu(C_{u,r_n}\cap B((\alpha-\rho)e_2,\alpha))\geq\mu(C_{u,\tau_n}\cap H_0)\geq\frac{1}{2}\mu(C_{u,\tau_n})=\frac{\pi}{12}\tau_n^2\geq\frac{\pi}{12}r_n^{1/2}\rho^{3/2}.$$

The second inequality is a direct consequence of the definition of $\mathcal{W}^{\mathcal{G}}$, see Figure 2.11, whereas the last one follows from the fact that $\rho<r_n\leq\alpha$. This completes the proof of Proposition 2.3.3, with $L^{\mathcal{G}}=\pi/12>0$ constant.

□

In view of Proposition 2.3.3 we define, for $x\in\mathcal{G}_{r_n}$, the family

$$\mathcal{U}^{\mathcal{G}}_{0,r_n}=\{C_{u,r_n},u\in\mathcal{W}^{\mathcal{G}}\},$$

formed by $m^{\mathcal{G}}=4$ elements. We now turn to the points in $\mathcal{F}_{r_n}$. The aim is to define for those points a finite family $\mathcal{U}^{\mathcal{F}}_{0,r_n}$, such that, for all $y\in\mathcal{F}_{r_n}$, there exists $U\in\mathcal{U}^{\mathcal{F}}_{0,r_n}$ that satisfies $U\subset B(y,r_n)$ and

$$\mu(U\cap B((\alpha-\rho)e_2,\alpha))\geq L^{\mathcal{F}}r_n^{\frac{1}{2}}\rho^{\frac{3}{2}},\quad\forall U\in\mathcal{U}^{\mathcal{F}}_{0,r_n}. \tag{2.16}$$

At this point, it may be useful to make some comments concerning the main differences between $\mathcal{G}_{r_n}$ and $\mathcal{F}_{r_n}$. One might be tempted to proceed as before for $\mathcal{F}_{r_n}$ and define the set of unit vectors

$$\mathcal{W}^{\mathcal{F}}=\{(-1/2,-\sqrt{3}/2),(1/2,-\sqrt{3}/2)\}.$$

Again we would have that, see Figure 2.12 (a)

$$\mathcal{F}_{r_n}=\bigcup_{u\in\mathcal{W}^{\mathcal{F}}}C_{u,r_n}.$$

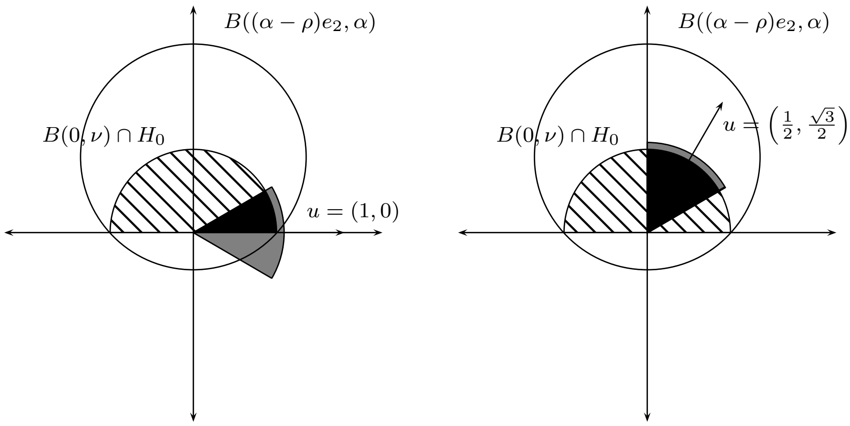


Figure 2.11: *In gray* C_{u,r_n} *for* $u=(1,0)$ *and* $u=(1/2,\sqrt{3}/2)$. *In black* $C_{u,r_n}\cap B(0,\nu)\cap H_0$. *We have that* $\mu(C_{u,r_n}\cap B((\alpha-\rho)e_2,\alpha))\geq\frac{\pi\tau_n^2}{12}$, *for* $\tau_n=\min(\nu,r_n)$.

If we repeat the sketch of the proof for $\mathcal{G}_{r_n}$ and define U to be the circular sectors C_{u,r_n} for $u\in\mathcal{W}^{\mathcal{F}}$, we could no longer guarantee the lower bound in (2.16). Note that the intersection $C_{u,r_n}\cap B((\alpha-\rho)e_2,\alpha)$ for $u\in\mathcal{W}^{\mathcal{F}}$ is considerably smaller than for $u\in\mathcal{W}^{\mathcal{G}}$. In fact, it can be easily proved that, for $u\in\mathcal{W}^{\mathcal{F}}$,

$$\mu(C_{u,r_n}\cap B((\alpha-\rho)e_2,\alpha))\leq\sqrt{3}\rho^2,$$

as it is shown in Figure 2.12 (b). Therefore, we need to consider different sets U. We have previously discussed the possibility of defining unavoidable sets, larger than circular sectors. For a fixed unit vector u,

$$U=\bigcap_{y\in C_{u,r_n}}B(y,r_n) \tag{2.17}$$

is the largest set such that $U\subset B(y,r_n)$ for all $y\in C_{u,r_n}$. Figure 2.13 shows C_{u,r_n}, for an $u\in\mathcal{W}^{\mathcal{F}}$ and the corresponding set U defined in (2.17). Observe that $U\cap B((\alpha-\rho)e_2,\alpha)$ is clearly larger than $C_{u,r_n}\cap B((\alpha-\rho)e_2,\alpha)$. The difference between both intersections will play a fundamental role in obtaining the lower bound in (2.16). In fact, it is not necessary to consider the whole U as defined in (2.17). For our purposes it is sufficient to measure a portion of $U\cap B((\alpha-\rho)e_2,\alpha)$. We shall consider sets as the one represented in gray in Figure 2.14. Its measure is large enough to satisfy (2.16). We give the precise definition of this kind of sets in Proposition 2.3.4, that solves the problem for the points in $\mathcal{F}_{r_n}$.

Proposition 2.3.4. *There exists a finite family of sets* $\mathcal{U}^{\mathcal{F}}_{0,r_n}$ *such that, for all* $y\in\mathcal{F}_{r_n}$, *there exists* $U\in\mathcal{U}^{\mathcal{F}}_{0,r_n}$ *such that* $U\subset B(y,r_n)$ *and*

$$\mu(U\cap B((\alpha-\rho)e_2,\alpha))\geq L^{\mathcal{F}}r_n^{\frac{1}{2}}\rho^{\frac{3}{2}},$$

with $L^{\mathcal{F}}>0$ *a constant.*

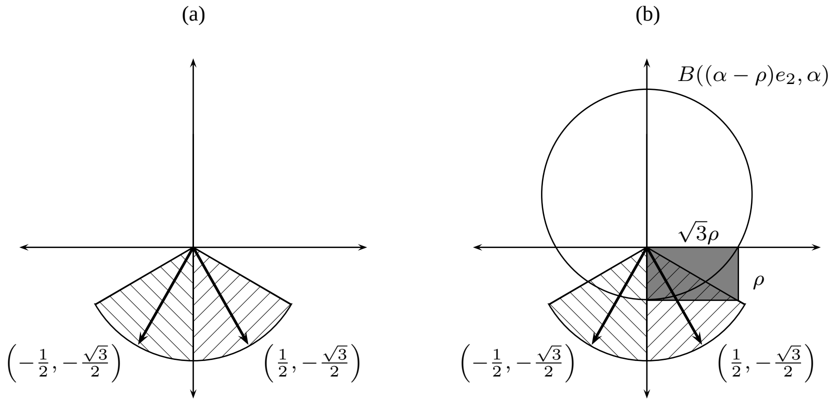


Figure 2.12: *(a)* $\mathcal{W}^{\mathcal{F}} = \{(-1/2,-\sqrt{3}/2),(1/2,-\sqrt{3}/2)\}$ *and* C_{u,r_n}*, for* $u \in \mathcal{W}^{\mathcal{F}}$*. (b) For* $u \in \mathcal{W}^{\mathcal{F}}$*,* $C_{u,r_n} \cap B((\alpha-\rho)e_2,\alpha)$ *is contained in the rectangle of height* ρ *and base* $\sqrt{3}\rho$*.*

Proof. First, let us consider the set

$$B((\alpha-\rho)e_2,\alpha) \cap B(-r_ne_2,r_n),$$

which corresponds to the intersection between two balls of radii α and r_n, respectively, being $\alpha + r_n - \rho$ the distance between their centres, see Figure 2.15 (a). The values of h_1, h_2 and λ in Figure 2.15 (b) can be deduced from the Pythagorean theorem. They satisfy the following equations

$$\begin{cases} (r_n-h_1)^2+\lambda^2 = r_n^2, \\ (\alpha-h_2)^2+\lambda^2 = \alpha^2, \\ h_1+h_2=\rho. \end{cases}$$

By solving the system,

$$h_1 = \frac{\rho(2\alpha-\rho)}{2(\alpha+r_n-\rho)}, \quad h_2 = \frac{\rho(2r_n-\rho)}{2(\alpha+r_n-\rho)} \quad \text{, and} \quad \lambda = \sqrt{2r_nh_1-h_1^2}.$$

We now define the set

$$\mathcal{C}(h_1) = \{x \in \mathbb{R}^2 : \ -h_1 \le \langle x, e_2\rangle \le 0\} \cap B(-r_ne_2,r_n).$$

Lemma 2.3.7 provides a lower bound for the measure of $\mathcal{C}(h_1)$.

Lemma 2.3.7. *Given the previous set* $\mathcal{C}(h_1)$*, then*

$$\mu(\mathcal{C}(h_1)) \ge \frac{\sqrt{2}}{3} r_n^{\frac{1}{2}} \rho^{\frac{3}{2}}.$$

Proof. We have that

$$\mu(\mathcal{C}(h_1)) = \int_0^{h_1} 2\sqrt{2r_ny-y^2}\,dy. \tag{2.18}$$

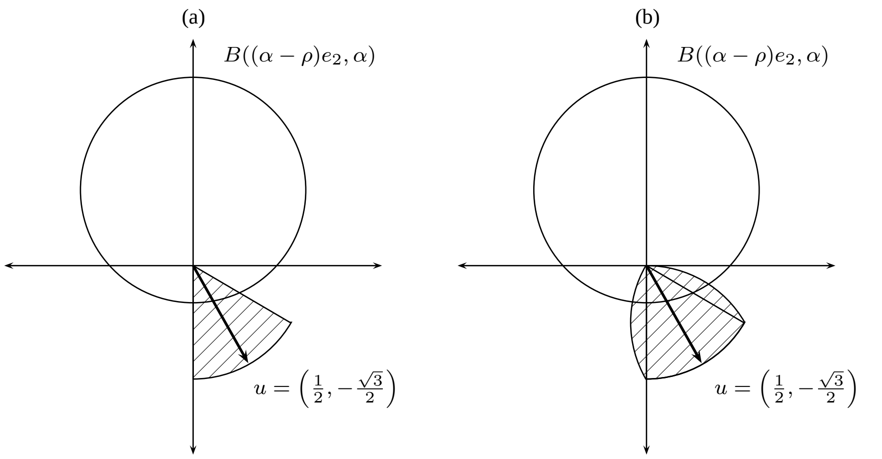


Figure 2.13: *(a)* C_{u,r_n} *with* $u=(1/2,-\sqrt{3}/2)$. *(b)* $\bigcap_{y\in C_{u,r_n}} B(y,r_n)$.

For $y\in[0,h_1]$ we have that $y\leq r_n$, since by construction $h_1\leq\rho$ and by assumption $\rho\leq r_n/2$. Hence, $2r_ny-y^2\geq r_ny$ and

$$\mu(\mathcal{C}(h_1))\geq\int_0^{h_1}2\sqrt{r_ny}dy=\frac{4}{3}r_n^{\frac{1}{2}}h_1^{\frac{3}{2}}.$$

Moreover, $h_1\geq\rho/2$, since $r_n\leq\alpha$ and this completes the proof.

□

Remark 2.3.5. *Note that the exact value of the integral in (2.18) can be explicitly computed since it coincides with the area of the circular segment defined by the chord that joins the intersection points of* $B((\alpha-\rho)e_2,\alpha)\cap B(-r_ne_2,r_n)$*. Thus,*

$$\mu(\mathcal{C}(h_1))=r_n^2\arccos\left(\frac{r_n-h_1}{r_n}\right)-(r_n-h_1)\sqrt{2r_nh_1-h_1^2}.$$

At this point we have defined the set $\mathcal{C}(h_1)$, whose measure verifies the statement of Proposition 2.3.4. Next lemma shows that $\mathcal{C}(h_1)$ is contained in $B((\alpha-\rho)e_2,\alpha)$.

Lemma 2.3.8.

$$\mathcal{C}(h_1)\subset B((\alpha-\rho)e_2,\alpha).$$

Proof. Let $x\in\mathcal{C}(h_1)$.

$$\|x-(\alpha-\rho)e_2\|^2 = \|x\|^2+(\alpha-\rho)^2-2(\alpha-\rho)\langle x,e_2\rangle. \tag{2.19}$$

By definition, $x\in B(-r_ne_2,r_n)$ and therefore

$$\|x\|^2\leq-2r_n\langle x,e_2\rangle.$$

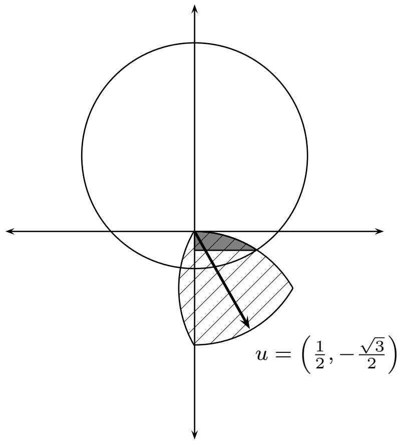


Figure 2.14: *The dashed area corresponds to* $U = \bigcap_{y \in C_{u,r_n}} B(y, r_n)$ *with* $u = (1/2, -\sqrt{3}/2)$.

Furthermore, by definition, $\langle x, e_2 \rangle \geq -h_1$. Turning to (2.19) we get

$$\begin{aligned} \|x - (\alpha - \rho)e_2\|^2 &\leq 2r_n h_1 + (\alpha - \rho)^2 + 2(\alpha - \rho)h_1 \\ &= 2(r_n + \alpha - \rho)h_1 + (\alpha - \rho)^2 \\ &= \rho(2\alpha - \rho) + (\alpha - \rho)^2 \\ &= \alpha^2. \end{aligned}$$

□

It follows from Lemmas 2.3.7 and 2.3.8 that

$$\mu(\mathcal{C}(h_1) \cap B((\alpha - \rho)e_2, \alpha)) \geq L r_n^{\frac{1}{2}} \rho^{\frac{3}{2}}. \tag{2.20}$$

In order to complete the proof, it remains to define the family $\mathcal{U}_{0,r_n}^{\mathcal{F}}$ mentioned in the statement of Proposition 2.3.4. In view of (2.20), it seems natural to divide $\mathcal{C}(h_1)$. Let us first consider the following partition of $\mathbb{R}^2$.

$$\mathbb{R}^2 = \{x = (x_1, x_2) \in \mathbb{R}^2 : \ x_1 \geq 0\} \cup \{x = (x_1, x_2) \in \mathbb{R}^2 : \ x_1 \leq 0\}.$$

We denote $Q_1 = \{x = (x_1, x_2) \in \mathbb{R}^2 : \ x_1 \geq 0\}$ and $Q_2 = \{x = (x_1, x_2) \in \mathbb{R}^2 : \ x_1 \leq 0\}$. Then,

$$\mathcal{F}_{r_n} = (Q_1 \cap \mathcal{F}_{r_n}) \cup (Q_2 \cap \mathcal{F}_{r_n})$$

and, in the same manner,

$$\mathcal{C}(h_1) = (Q_1 \cap \mathcal{C}(h_1)) \cup (Q_2 \cap \mathcal{C}(h_1)).$$

Lemma 2.3.9. *for all* $y \in Q_i \cap \mathcal{F}_{r_n}$ *we have that*

$$Q_i \cap \mathcal{C}(h_1) \subset B(y, r_n), \ \ i = 1, 2.$$

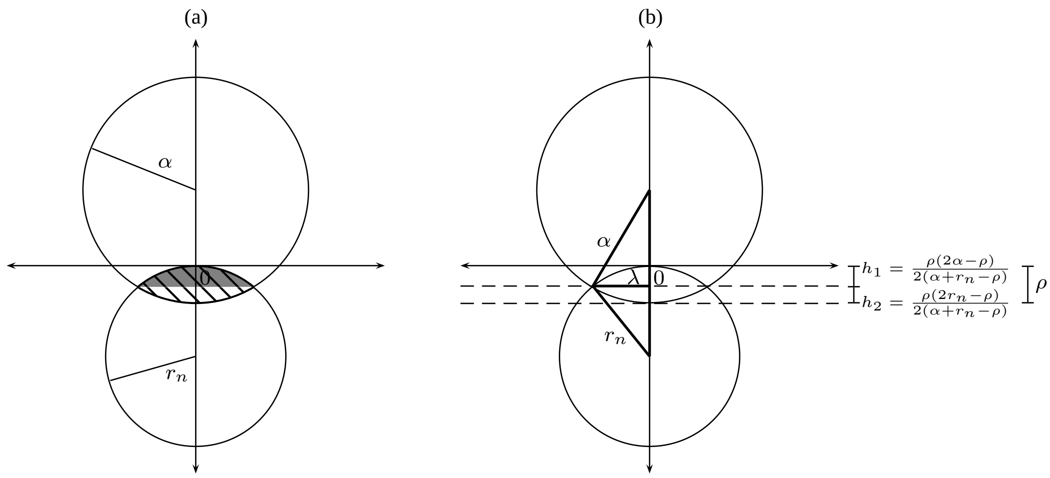


Figure 2.15: *(a) The dashed area corresponds to* $B((\alpha-\rho)e_2,\alpha)\cap B(-r_ne_2,r_n)$*. In gray* $\mathcal{C}(h_1)$*. (b) Values of* h_1*,* h_2 *and* λ*.*

Proof. Let $x\in Q_1\cap\mathcal{C}(h_1)$. First, it can be easily proved that

$$Q_1\cap\mathcal{F}_{r_n}=C_{u,r_n},$$

with $u=(1/2,-\sqrt{3}/2)$. What we need to prove is

$$x\in\bigcap_{y\in C_{u,r_n}}B(y,r_n).$$

It follows from Lemma 2.3.4 that

$$\bigcap_{y\in C_{u,r_n}}B(y,r_n)=B(0,r_n)\cap B(v_1,r_n)\cap B(v_2,r_n),$$

where $v_1=r_n\mathcal{R}(u)=r_n\left(\sqrt{3}/2,-1/2\right)$ and $v_2=r_n\mathcal{R}^{-1}(u)=-r_ne_2$. We have by definition that $x\in B(v_2,r_n)$. Moreover,

$$\|x\|^2\le\lambda^2+h_1^2=2r_nh_1\le r_n^2,$$

since $h_1\le\rho\le r_n/2$. Note that the last inequality justifies the choice of $\rho\le r_n/2$. And,

$$\|x-v_1\|^2=\left(x_1-\frac{\sqrt{3}r_n}{2}\right)^2+\left(x_2+\frac{r_n}{2}\right)^2\le\left(\frac{\sqrt{3}r_n}{2}\right)^2+\left(\frac{r_n}{2}\right)^2=r_n^2,$$

since $0\le x_1\le\lambda\le\sqrt{3}r_n/2$ and $-h_1\le x_2\le 0$, where $h_1\le\rho\le r_n/2$. Thus, we have shown that

$$x\in B(0,r_n)\cap B(v_1,r_n)\cap B(v_2,r_n)$$

and the lemma is proved for $Q_1\cap\mathcal{C}(h_1)$. The proof for $Q_2\cap\mathcal{C}(h_1)$ is analogous.

$\square$

In view of the previous results we define, for $x \in \mathcal{F}_{r_n}$, the family

$$\mathcal{U}^{\mathcal{F}}_{0,r_n} = \{Q_i \cap \mathcal{C}(h_1),\ i = 1, 2\},$$

formed by $m^{\mathcal{F}} = 2$ elements. It follows from Lemma 2.3.9 that, for all $y \in \mathcal{F}_{r_n}$, there exists $i \in \{1, 2\}$ such that $Q_i \cap \mathcal{C}(h_1) \subset B(y, r_n)$. Moreover, by Lemma 2.3.7,

$$Lr_n^{\frac{1}{2}}\rho^{\frac{3}{2}} \le \mu(\mathcal{C}(h_1)) = \sum_{i=1}^{2} \mu(Q_i \cap \mathcal{C}(h_1)).$$

The symmetry of the set $\mathcal{C}(h_1)$ with respect to the axis OY implies that the orthogonal transformation $\mathcal{O} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ such that $\mathcal{O}(x) = \mathcal{O}(x_1, x_2) = (-x_1, x_2)$ transforms $Q_1 \cap \mathcal{C}(h_1)$ into $Q_2 \cap \mathcal{C}(h_1)$ and then both sets measure the same, that is,

$$\mu(Q_1 \cap \mathcal{C}(h_1)) = \mu(Q_2 \cap \mathcal{C}(h_1)) = \frac{1}{2}\mu(\mathcal{C}(h_1)).$$

By Lemma 2.3.8 we further have that, for $i = 1, 2$

$$Q_i \cap \mathcal{C}(h_1) \subset \mathcal{C}(h_1) \subset B((\alpha - \rho)e_2, \alpha)$$

and hence

$$\mu(Q_i \cap \mathcal{C}(h_1) \cap B((\alpha - \rho)e_2, \alpha)) = \mu(Q_i \cap \mathcal{C}(h_1)) \ge L^{\mathcal{F}} r_n^{\frac{1}{2}}\rho^{\frac{3}{2}},$$

where $L^{\mathcal{F}} = \sqrt{2}/6$. This completes the proof of Proposition 2.3.4.

$\square$

Now, we define

$$\mathcal{U}_{0,r_n} = \mathcal{U}^{\mathcal{G}}_{0,r_n} \cup \mathcal{U}^{\mathcal{F}}_{0,r_n}$$

and, as we mentioned at the beginning of Proposition 2.3.2,

$$\mathcal{U}_{x,r_n} = \{T(U),\ U \in \mathcal{U}_{0,r_n}\}$$

is a finite family with $m_2 = m^{\mathcal{G}} + m^{\mathcal{F}} = 6$ elements satisfying that, for each $U \in \mathcal{U}_{0,r_n}$,

$$P_X(T(U)) \ge \delta\mu(T(U) \cap B(P_\Gamma x - \alpha\eta, \alpha)) = \delta\mu(U \cap B((\alpha - \rho)e_d, \alpha)) \ge L_2 r_n^{\frac{1}{2}}\rho^{\frac{3}{2}},$$

where $L_2 = \delta \min(L^{\mathcal{G}}, L^{\mathcal{F}})$. This completes the proof of Proposition 2.3.2.

$\square$

2.4 Defining unavoidable families in $\mathbb{R}^d$

The approach to the general case $\mathbb{R}^d$ will be much more involved than in $\mathbb{R}^2$ even though the sketch of the proofs and the basic ideas still hold. In the same way as for $\mathbb{R}^2$, we shall consider two different situations in order to define unavoidable families and give a lower bound for the probability of the sets in those families. The argument for the points of S which are far away from the boundary in $\mathbb{R}^d$ is analogous to that for $\mathbb{R}^2$. Recall that in $\mathbb{R}^2$ it was sufficient to consider circular sectors of radius r_n and central angle $\pi/3$. By Lemma 2.3.2 the choice of the central angle $\pi/3$ guarantees that the circular sectors are unavoidable. Now we shall consider the generalization to the multidimensional case of the circular sectors in $\mathbb{R}^2$. What, however, is more complicated to handle is the case of the points of S which are closer to the boundary. We have discussed that for points which are close to the boundary in $\mathbb{R}^2$ it was not sufficient for our purposes to consider circular sectors. In the previous section we introduced then the Reuleaux triangles to solve the problem. But, unfortunately, we cannot generalize the concept of Reauleaux triangle to the d-dimensional case so the main difficulty in this section is constructing sets both unavoidable and large enough for points which are close to the boundary.

Thus, the main theorem of this section rests upon two important results. First, Proposition 2.4.1 again defines a finite family of unavoidable sets for $\mathcal{E}_{x,r_n}$ when $x \in S$ and $d(x, \partial S) > r_n/2$. The result also gives a lower bound for the probability of such sets, which is independent of x. In the same manner, Proposition 2.4.2 defines a finite family of unavoidable sets for $\mathcal{E}_{x,r_n}$ and gives a lower bound for the probability of such sets when $x \in S$ and $d(x, \partial S) \leq r_n/2$. In that case the probability depends on the distance from x to the boundary of the set. Moreover, the number of sets that form the unavoidable families is independent of x in both situations.

As we have already comment, we shall work with the d-dimensional generalization of the bidimensional sectors. Hence, before proceeding to the results we introduce the precise definition of these sets and some useful notation. From now on, let $\mathbb{S}_d = \{u \in \mathbb{R}^d : \|u\| = 1\}$ be the unit sphere in $\mathbb{R}^d$. Let $\varphi_{u,v}$ denote the angle between the vectors u and v. As in $\mathbb{R}^2$, it is understood that $\varphi_{u,v} \in [0, \pi]$ and $\varphi_{u,v} = \varphi_{v,u}$. Finally, let $e_d = (0, \ldots, 0, 1) \in \mathbb{R}^d$ and let ω_d be the measure of the unit ball in $\mathbb{R}^d$.

Definition 2.4.1. *For $u \in \mathbb{S}_d$ and $\theta \in [0, \pi/2]$, we define the sets*

$$C_u^\theta = \{x \in \mathbb{R}^d : \langle x, u\rangle \geq \|x\| \cos\theta\}$$

and the generalized circular sectors

$$C_{u,r}^\theta = C_u^\theta \cap B(0, r).$$

Figure 2.16 shows an example of C_u^θ in $\mathbb{R}^3$.

As we said before, Proposition 2.4.1 defines unavoidable families for those points which are far away from the boundary of S and gives a lower bound for the probability of such sets. The proofs of Proposition 2.4.1 and 2.3.1 are essentially the same, apart from some elementary results, which were straightforward in the bidimensional case, but not in the general case. On the other hand, some of the results in $\mathbb{R}^d$ follow either directly from or similarly to the corresponding results in $\mathbb{R}^2$. For this reason we will skip some of the proofs throughout this section.

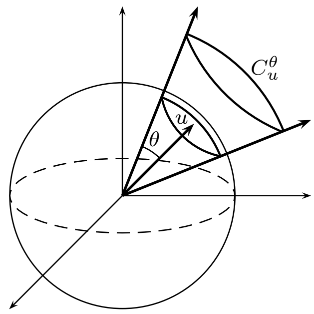


Figure 2.16: C_u^θ *in* $\mathbb{R}^3$.

Proposition 2.4.1. *Let S be a nonempty compact subset of $\mathbb{R}^d$ such that a ball of radius $\alpha > 0$ rolls freely in S and in $\overline{S^c}$. Let X be a random variable with probability distribution P_X and support S. We assume that the probability distribution P_X satisfies that there exists $\delta > 0$ such that*

$$P_X(C) \geq \delta\mu(C \cap S)$$

for all Borel set $C \subset \mathbb{R}^d$.

Then, for all $x \in S$ such that $d(x, \partial S) > r_n/2$, there exists a finite family $\mathcal{U}_{x,r_n}$ with m_1 elements, unavoidable for $\mathcal{E}_{x,r_n}$ and that satisfies

$$P_X(U) \geq L_1 r_n^d, \;\; U \in \mathcal{U}_{x,r_n},$$

where the constants m_1 and $L_1 > 0$ are independent of x.

Proof. The case $d = 1$ is handled separately as it is simpler. For $x \in \mathbb{R}$ under the stated conditions let us consider the unavoidable family

$$\{[x - r_n/2, x], [x, x + r_n/2]\}.$$

The result holds for $L_1 = \delta/2$. The case $d = 2$ was proved in Proposition 2.3.1. Let us then assume that $d \geq 3$. It may be noted that the proof remains valid for the bidimensional case, although Proposition 2.3.1 is simpler. First, we need to define a finite set of unit vectors $\mathcal{W}$ that enables us to divide the ball $B(0, r_n)$ into generalized circular sectors of central angle $\pi/3$. Recall that in $\mathbb{R}^2$, the family $\mathcal{W}$ was explicitly defined. However, for the general case $\mathbb{R}^d$ the family is found by an indirect method. Lemma 2.4.2 states that, since $\mathbb{S}_d$ is compact, we can cover $B(0, r_n)$ by finitely many generalized circular sector with positive central angle. First, we state Lemma 2.4.1, whose proof is identical to that of Lemma 2.3.1 in $\mathbb{R}^2$.

Lemma 2.4.1. *Let $x \neq 0$. Then,*

$$x \in C_u^\theta \Leftrightarrow \varphi_{x,u} \leq \theta.$$

Lemma 2.4.2. *Let $\theta > 0$. There exists a finite family of unit vectors $\mathcal{W}_\theta$ such that*

$$B(0,r) = \bigcup_{u\in\mathcal{W}_\theta} C^\theta_{u,r},$$

for all $r > 0$.

Proof. Let us first consider the sphere $\mathbb{S}_d$. It can be easily proved that $\{\text{int}(C^\theta_u), u \in \mathbb{S}_d\}$ is an open cover of $\mathbb{S}_d$, since for each $u \in \mathbb{S}_d$ we have that $u \in \text{int}(C^\theta_u)$. By the compactness of $\mathbb{S}_d$, there exists a finite family $\mathcal{W}_\theta \subset \mathbb{S}_d$ such that

$$\mathbb{S}_d \subset \bigcup_{u\in\mathcal{W}_\theta} \text{int}(C^\theta_u). \tag{2.21}$$

Now, let $x \in B(0,r)$. For $x = 0$ it is clear that $x \in C^\theta_u$ for all $u \in \mathcal{W}_\theta$. For $x \neq 0$ let us consider $v = x/\|x\| \in \mathbb{S}^d$. By (2.21) there exists $u \in \mathcal{W}_\theta$ such that $v \in C^\theta_u$. Since $\varphi_{x,u} = \varphi_{v,u}$ it follows from Lemma 2.4.1 that $x \in C^\theta_u$. Therefore,

$$B(0,r) \subset \bigcup_{u\in\mathcal{W}_\theta} C^\theta_u \cap B(0,r) = \bigcup_{u\in\mathcal{W}_\theta} C^\theta_{u,r}.$$

By definition of $C^\theta_{u,r}$ it immediate follows that $C^\theta_{u,r} \subset B(0,r)$ for all $u \in \mathcal{W}_\theta$ and the proof is complete.

□

Fix $\theta = \pi/6$ and then, by Lemma 2.4.2, there exists a finite family of unit vectors $\mathcal{W}_{\pi/6}$ such that

$$B(0,r_n) = \bigcup_{u\in\mathcal{W}_{\pi/6}} C^{\pi/6}_{u,r_n}.$$

Recall that, in order to simplify the notation, we write C_u, C_{u,r_n} and $\mathcal{W}$ to refer to $C^{\pi/6}_u$, $C^{\pi/6}_{u,r_n}$ and $\mathcal{W}_{\pi/6}$, respectively. Given $B(y,r_n) \in \mathcal{E}_{0,r_n}$, there exists $u \in \mathcal{W}$ such that $y \in C_{u,r_n}$. In the same manner as Lemma 2.3.2 in $\mathbb{R}^2$, Lemma 2.4.3 shows that $C_{u,r_n} \subset B(y,r_n)$. Since the proof of Lemma 2.4.3 is based on the triangle inequality

$$\varphi_{z,y} \leq \varphi_{z,u} + \varphi_{u,y} \leq \frac{\pi}{3},$$

which remains true for arbitrary dimension, see Figure 2.17, we skip the details.

Lemma 2.4.3. *For all $u \in \mathbb{S}_d$ and $r > 0$,*

$$C_{u,r} \subset \bigcap_{y\in C_{u,r}} B(y,r).$$

Therefore, we define the family

$$\mathcal{U}_{0,r_n} = \{C_{u,r_n/2},\ u \in \mathcal{W}\}.$$

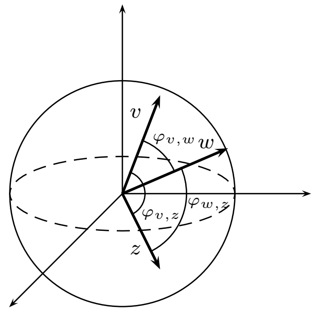


Figure 2.17: *Triangle inequality in* $\mathbb{R}^3$. *We have that* $\varphi_{v,z} \leq \varphi_{v,w} + \varphi_{w,z}$.

Lemma 2.4.3 shows that $\mathcal{U}_{0,r_n}$ is an unavoidable family for $\mathcal{E}_{0,r_n}$. Now, for $x \in S$ such that $d(x, \partial S) > r_n/2$ we define

$$\mathcal{U}_{x,r_n} = \{x\} \oplus \mathcal{U}_{0,r_n} = \{\{x\} \oplus C_{u,r_n/2},\ u \in \mathcal{W}\}.$$

It follows from Lemma 2.4.4 that the family $\mathcal{U}_{x,r_n}$ is unavoidable for $\mathcal{E}_{x,r_n}$. As in the bidimensional case, the translation of unavoidable families gives unavoidable families.

Lemma 2.4.4. *Let $\mathcal{U}_{0,r}$ be an unavoidable family for $\mathcal{E}_{0,r}$. Then the family $\mathcal{U}_{x,r} = \{x\} \oplus \mathcal{U}_{0,r} = \{\{x\} \oplus U,\ U \in \mathcal{U}_{0,r}\}$ is unavoidable for $\mathcal{E}_{x,r}$.*

Proof. Analogous to the proof of Lemma 2.3.3.

□

Finally, for each $u \in \mathcal{W}$

$$P_X\left(\{x\} \oplus C_{u,r_n/2}\right) \geq \delta\mu\left(\{x\} \oplus C_{u,r_n/2} \cap S\right) = \delta\mu\left(\{x\} \oplus C_{u,r_n/2}\right) = \delta\mu\left(C_{u,r_n/2}\right).$$

The last inequality is obtained by using that $d(x, \partial S) > r_n/2$ and that the Lebesgue measure is invariant under translations. Moreover, it follows from Lemma 2.4.2 that we can cover $B(0, r_n/2)$ by a finite number m_1 of generalized circular sectors $C_{u,r_n/2}$, all of them with the same measure. Therefore,

$$\mu\left(C_{u,r_n/2}\right) \geq \frac{1}{m_1}\mu(B(0, r_n/2)) = \frac{1}{m_1}\omega_d\left(\frac{r_n}{2}\right)^d \tag{2.22}$$

and,

$$P_X\left(\{x\} \oplus C_{u,r_n/2}\right) \geq \delta\frac{1}{m_1}\omega_d\left(\frac{r_n}{2}\right)^d.$$

To sum up,

$$P_X(U) \geq L_1 r_n^d,\ \ U \in \mathcal{U}_{x,r_n},$$

being $L_1 = \delta\omega_d/(2^d m_1)$ and the proof of Proposition 2.4.2 is now complete.

□

Remark 2.4.1. *In Proposition 2.3.1 in $\mathbb{R}^2$ we covered $B(0, r_n/2)$ by six nonoverlapping circular sectors $C_{u,r_n/2}$ with central angle $\pi/3$. Thus,*

$$\mu(C_{u,r_n/2}) = \frac{1}{6}\mu(B(0, r_n/2)).$$

However, we cannot guarantee a similar result in general dimension. By Lemma 2.4.2 we can cover $B(0, r_n/2)$ by a not easy to compute finite number of generalized circular sectors but we cannot ensure that they are nonoverlapping sectors. This explains why (2.22) is an enequality.

Once we have proved Proposition 2.4.1 it remains to explain what happens when $x \in S$ and $d(x, \partial S) \leq r_n/2$. As we mentioned when dealing with the problem in $\mathbb{R}^2$, it is not sufficient to consider circular sectors as unavoidable sets. Proposition 2.4.2 below defines unavoidable families for those points which are close to the boundary of S and gives a lower bound for the probability of the sets that form the family. Although the sketch of the proof is almost identical to that of Proposition 2.3.2 we need some extra auxiliary results to handle the more general case of $\mathbb{R}^d$.

Proposition 2.4.2. *Let S be a nonempty compact subset of $\mathbb{R}^d$ such that a ball of radius $\alpha > 0$ rolls freely in S and in $\overline{S^c}$. Let X be a random variable with probability distribution P_X and support S. We assume that the probability distribution P_X satisfies that there exists $\delta > 0$ such that*

$$P_X(C) \geq \delta\mu(C \cap S)$$

for all Borel set $C \subset \mathbb{R}^d$.

Then, for all $x \in S$ such that $d(x, \partial S) \leq r_n/2$, there exists a finite family $\mathcal{U}_{x,r_n}$ with m_2 elements, unavoidable for $\mathcal{E}_{x,r_n}$ and that satisfies

$$P_X(U) \geq L_2 r_n^{\frac{d-1}{2}} d(x, \partial S)^{\frac{d+1}{2}}, \;\; U \in \mathcal{U}_{x,r_n},$$

where the constants m_2 and $L_2 > 0$ are independent of x.

Proof. Let $x \in S$ such that $d(x, \partial S) \leq r_n/2$. We denote $\rho = d(x, \partial S)$. The proof for $d = 1$ is immediate. Consider the unavoidable family

$$\{[x - \rho, x], [x, x + \rho]\}$$

and the result holds for $L_2 = \delta$. The case $d = 2$ was proved in Proposition 2.3.2. Let us assume that $d \geq 3$. Again, the proof remains valid for the bidimensional case. As in $\mathbb{R}^2$, the rolling condition in S simplifies the proof. Using the same notation, let $P_\Gamma x$ be the metric projection of x onto $\Gamma = \partial S$ and η the outward pointing unit normal vector at $P_\Gamma x$. It is enough to define an unavoidable family $\mathcal{U}_{x,r_n}$ and find a lower bound for $\mu(U \cap B(P_\Gamma x - \alpha\eta, \alpha))$ for all $U \in \mathcal{U}_{x,r_n}$. Now, suppose that we are able to define a suitable unavoidable family $\mathcal{U}_{0,r_n}$. Consider the composite function T formed by first applying an orthogonal transformation $\mathcal{O} : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ such that $\mathcal{O}(e_d) = -\eta$ and then applying the translation by the vector x. In particular $T(0) = x$, $T((\alpha - \rho)e_d) = x - (\alpha - \rho)\eta = P_\Gamma x - \alpha\eta$ and

$$T(B((\alpha - \rho)e_d, \alpha)) = B(P_\Gamma x - \alpha\eta, \alpha).$$

Then

$$\mu(T(U) \cap B(P_\Gamma x - \alpha\eta, \alpha)) = \mu(U \cap B((\alpha - \rho)e_d, \alpha)),$$

since the Lebesgue measure is invariant under translations and orthogonal transformations. It suffices to give a lower bound for $\mu(U \cap B((\alpha-\rho)e_d, \alpha))$ for $U \in \mathcal{U}_{0,r_n}$, since as in Lemma 2.3.5 we can prove in the d-dimensional case that the orthogonal transformation of an unavoidable family for $\mathcal{E}_{0,r_n}$ results in another unavoidable family for $\mathcal{E}_{0,r_n}$. This argument is made rigorous in the following lemma. We skip the proof since it is identical to that of Lemma 2.3.5 in $\mathbb{R}^2$.

Lemma 2.4.5. *Let $\mathcal{U}_{0,r}$ be an unavoidable family for $\mathcal{E}_{0,r}$ and $\mathcal{O} : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ an orthogonal transformation. Then $\{\mathcal{O}(U),\ U \in \mathcal{U}_{0,r}\}$ is an unavoidable family for $\mathcal{E}_{0,r}$.*

Therefore, it suffices to consider

$$\mathcal{U}_{x,r_n} = \{T(U),\ U \in \mathcal{U}_{0,r_n}\}$$

which, by Lemmas 2.4.4 and 2.4.5, is also unavoidable for $\mathcal{E}_{x,r_n}$. We will, thus, concentrate on the definition of a family $\mathcal{U}_{0,r_n}$, unavoidable for $\mathcal{E}_{0,r_n}$. We need to bound $\mu(U \cap B((\alpha - \rho)e_d, \alpha))$ for $U \in \mathcal{U}_{0,r_n}$. Recall from Proposition 2.3.2 the comment on the measure of the sets $U \cap B((\alpha - \rho)e_d, \alpha)$. Once again, when defining unavoidable sets for $\mathcal{E}_{0,r_n}$ and giving a lower bound for $\mu(U \cap B((\alpha - \rho)e_d, \alpha))$, one must be careful with those points that lie far away in the direction of $-e_d$. For this reason we divide $B(0, r_n)$ as follows:

$$B(0, r_n) = \mathcal{G}_{r_n} \cup \mathcal{F}_{r_n},$$

where

$$\mathcal{G}_{r_n} = \left\{ y \in B(0, r_n) :\ \langle y, e_d \rangle \geq -\frac{1}{2}\, \|y\| \right\}$$

and

$$\mathcal{F}_{r_n} = \left\{ y \in B(0, r_n) :\ \langle y, e_d \rangle < -\frac{1}{2}\, \|y\| \right\}.$$

Figure 2.18 represents the sets $\mathcal{G}_{r_n}$ and $\mathcal{F}_{r_n}$ in $\mathbb{R}^3$.

Proposition 2.4.3 solves the problem for the points $y \in \mathcal{G}_{r_n}$. This result shows that we can construct a finite unavoidable family $\mathcal{U}^{\mathcal{G}}_{0,r_n}$, such that for all $y \in \mathcal{G}_{r_n}$ there exists $U \in \mathcal{U}^{\mathcal{G}}_{0,r_n}$ such that $U \subset B(y, r_n)$ and

$$\mu(U \cap B((\alpha - \rho)e_d, \alpha)) \geq L^{\mathcal{G}} r_n^{\frac{d-1}{2}} \rho^{\frac{d+1}{2}},\ \ \forall U \in \mathcal{U}^{\mathcal{G}}_{0,r_n}.$$

Before presenting the proof of Proposition 2.4.3 we would like to briefly comment on the main differences between the general case and the bidimensional one, see Proposition 2.3.3. The first step in the proof of Proposition 2.3.3 consisted of covering $\mathcal{G}_{r_n}$ by four circular sectors with central angle $\pi/3$. These circular sectors were determined by the family of unit vectors $\mathcal{W}^{\mathcal{G}}$ given in (2.15). Moreover, the position in the plane of the circular sectors C_{u,r_n}, with $u \in \mathcal{W}^{\mathcal{G}}$ (see Figure 2.9) guaranteed that

$$\mu(C_{u,r_n} \cap H_0) \geq \frac{1}{2}\mu(C_{u,r_n}) \geq \frac{1}{12}\pi r_n^2. \tag{2.23}$$

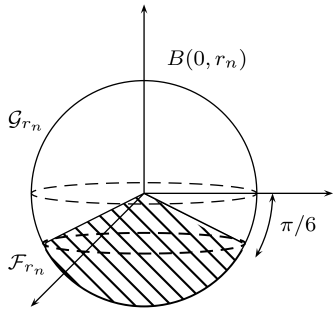


Figure 2.18: $\mathcal{G}_{r_n}$ *and* $\mathcal{F}_{r_n}$ *in* $\mathbb{R}^3$.

Note that the key to obtaining (2.23) is that $\langle u, e_2\rangle \geq 0$ for all $u \in \mathcal{W}^{\mathcal{G}}$. However, this is not true in general dimension since it is not possible to cover $\mathcal{G}_{r_n}$ by generalized circular sectors C_{u,r_n} such that $\langle u, e_d\rangle \geq 0$. For this reason we have to relax this condition. Thus, several extra auxiliary results are needed for the proof of Proposition 2.4.3.

Proposition 2.4.3. *There exists a finite family of unit vectors* $\mathcal{W}^{\mathcal{G}}$ *such that for all* $y \in \mathcal{G}_{r_n}$ *there exists* $u \in \mathcal{W}^{\mathcal{G}}$ *such that* $y \in C_{u,r_n} \subset B(y, r_n)$ *and*

$$\mu(C_{u,r_n} \cap B((\alpha - \rho)e_d, \alpha)) \geq L^{\mathcal{G}} r_n^{\frac{d-1}{2}} \rho^{\frac{d+1}{2}},$$

with $L^{\mathcal{G}} > 0$ *constant.*

Proof. First let us prove that $\mathcal{G}_{r_n}$ can be covered by a finite number of generalized circular sectors C_{u,r_n}. This result is an immediate consequence of Lemma 2.4.2, since $\mathcal{G}_{r_n} \subset B(0, r_n)$. However, in order to guarantee that the measure of $C_{u,r_n} \cap B((\alpha - \rho)e_d, \alpha)$ is large enough, the family of unit vectors $\mathcal{W}^{\mathcal{G}}$ cannot be chosen arbitrarily. Lemma 2.4.6 states that, for fixed $\gamma \in (0, \pi/2]$, we can cover $\mathcal{G}_{r_n}$ by a finite number of generalized circular sectors C_{u,r_n} such that $\langle u, e_d\rangle \geq -\sin\gamma$. This additional property refers to the position in the space of the sets C_{u,r_n} that form the covering. For small values of ρ, that is, for points x which are close to the boundary of S, the ball $B((\alpha - \rho)e_d, \alpha)$ is practically totally contained in the halfspace $H_0 = \{x = (x_1, \ldots, x_d) \in \mathbb{R}^d : \ x_d \geq 0\}$. For this reason, in order to obtain large values of $\mu(C_{u,r_n} \cap B((\alpha - \rho)e_d, \alpha))$, we also need the sets C_{u,r_n} to be contained in H_0, or at least a considerable portion of each set C_{u,r_n}. Hence, the goal is to cover $\mathcal{G}_{r_n}$ by sets C_{u,r_n} with the smallest possible φ_{u,e_d}. In order to ensure this, we restrict ourselves to those $u \in \mathbb{S}^d$ such that $\langle u, e_d\rangle \geq -\sin\gamma$. We shall see after Lemma 2.4.6 that any $\gamma \in (0, \pi/6)$ can be used to construct the desired covering.

Lemma 2.4.6. *Let* $0 < \gamma \leq \pi/2$*. There exists a finite set of unit vectors* $\mathcal{W}^{\mathcal{G}}(\gamma) \subset \mathbb{S}^d$*, such that for all* $u \in \mathcal{W}^{\mathcal{G}}(\gamma)$ *we have that* $\langle u, e_d\rangle \geq -\sin\gamma$ *and*

$$\mathcal{G}_{r_n} \subset \bigcup_{u \in \mathcal{W}^{\mathcal{G}}(\gamma)} C_{u,r_n}.$$

Proof. Let us first consider the set

$$\mathcal{D}_{\mathcal{G}} = \{y \in \mathbb{R}^d : \ \|y\| = 1, \ \langle y, e_d \rangle \geq -1/2\}.$$

Figure 2.19 shows the set $\mathcal{D}_{\mathcal{G}}$ in $\mathbb{R}^3$. We shall prove that

$$\{\text{int}(C_u), u \in \mathcal{D}_\gamma\}$$

is an open cover of $\mathcal{D}_{\mathcal{G}}$, where

$$\mathcal{D}_\gamma = \{u \in \mathbb{R}^d : \ \|u\| = 1, \langle u, e_d \rangle \geq -\sin\gamma\}.$$

Then, let $y \in \mathcal{D}_{\mathcal{G}}$. First, if $\langle y, e_d \rangle \geq -\sin\gamma$, then $y \in \text{int}(C_u)$ for $u = y \in \mathcal{D}_\gamma$. Let us now suppose that $\langle y, e_d \rangle < -\sin\gamma$. Since $y \in \mathcal{D}_{\mathcal{G}}$ we have that

$$\frac{\pi}{2} + \gamma < \varphi_{y,e_d} \leq \frac{2\pi}{3}. \tag{2.24}$$

Let u be a unit vector in the plane passing through the origin and determined by the vectors e_d and y, such that $\langle u, e_d \rangle = -\sin\gamma$. That is,

$$u = ay + be_d,$$

where $a, b \in \mathbb{R}$ satisfy the following system of equations

$$\begin{cases} \langle u, e_d \rangle = a \langle y, e_d \rangle + b = -\sin\gamma, \\ \|u\|^2 = a^2 + b^2 + 2ab \langle y, e_d \rangle = 1. \end{cases}$$

We obtain, by solving the system and using that $\langle y, e_d \rangle = \cos\varphi_{y,e_d} = y_d$,

$$b = -\sin\gamma - a\cos\varphi_{y,e_d} \ \text{ and } \ a = \pm\sqrt{\frac{1-\sin^2\gamma}{1-y_d^2}} = \pm\frac{\cos\gamma}{\sin\varphi_{y,e_d}}.$$

Note that (2.24) guarantees that the obtained solutions are well defined. If we choose

$$a = \frac{\cos\gamma}{\sin\varphi_{y,e_d}},$$

then

$$u = ay + be_d = \frac{\cos\gamma}{\sin\varphi_{y,e_d}} y + \left(-\sin\gamma - \frac{\cos\gamma}{\sin\varphi_{y,e_d}}\cos\varphi_{y,e_d}\right) e_d.$$

Figure 2.19 (c) shows the vector u defined from $y \in \mathcal{D}_{\mathcal{G}}$ with $\langle y, e_d \rangle < -\sin\gamma$. By construction,

$u \in \mathcal{D}_\gamma$. Moreover, $y \in \text{int}(C_u)$, since

$$\begin{aligned}
\langle y, u\rangle &= \langle y, ay + be_d\rangle \\
&= a + b\cos\varphi_{y,e_d} \\
&= \frac{\cos\gamma}{\sin\varphi_{y,e_d}} - \sin\gamma\cos\varphi_{y,e_d} - \frac{\cos^2\varphi_{y,e_d}\cos\gamma}{\sin\varphi_{y,e_d}} \\
&= \frac{\cos\gamma - \sin\gamma\sin\varphi_{y,e_d}\cos\varphi_{y,e_d} - \cos^2\varphi_{y,e_d}\cos\gamma}{\sin\varphi_{y,e_d}} \\
&= \frac{\cos\gamma\sin^2\varphi_{y,e_d} - \sin\gamma\sin\varphi_{y,e_d}\cos\varphi_{y,e_d}}{\sin\varphi_{y,e_d}} \\
&= \cos\gamma\sin\varphi_{y,e_d} - \sin\gamma\cos\varphi_{y,e_d} \\
&= \sin(\varphi_{y,e_d} - \gamma) \\
&> \cos\pi/6.
\end{aligned}$$

The last inequality is a direct consequence of

$$\frac{\pi}{2} < \varphi_{y,e_d} - \gamma \leq \frac{2\pi}{3} - \gamma < \frac{2\pi}{3}.$$

Therefore, $\{\text{int}(C_u), u \in \mathcal{D}_\gamma\}$ is an open cover of $\mathcal{D}_\mathcal{G}$. Since $\mathcal{D}_\mathcal{G}$ is compact, there exists a finite family of unit vectors $\mathcal{W}^\mathcal{G}(\gamma) \subset \mathcal{D}_\gamma$ such that

$$\mathcal{D}_\mathcal{G} \subset \bigcup_{u\in\mathcal{W}^\mathcal{G}(\gamma)} \text{int}(C_u). \tag{2.25}$$

Now, let $y \in \mathcal{G}_{r_n}$. If $y = 0$ then $y \in C_u$ for all $u \in \mathcal{W}^\mathcal{G}(\gamma)$. For $y \neq 0$ we have that $v = y/\|y\| \in \mathcal{D}_\mathcal{G}$ and by (2.25) there exists $u \in \mathcal{W}^\mathcal{G}(\gamma)$ such that $v \in C_u$. This immediately yields that $y \in C_u$, since $\varphi_{y,u} = \varphi_{v,u}$. Hence,

$$\mathcal{G}_{r_n} = \mathcal{G}_{r_n} \cap B(0, r_n) \subset \bigcup_{u\in\mathcal{W}^\mathcal{G}(\gamma)} C_u \cap B(0, r_n) = \bigcup_{u\in\mathcal{W}^\mathcal{G}(\gamma)} C_{u,r_n}$$

and the proof is complete.

□

As we have already mentioned, Lemma 2.4.6 plays an important role in the proof of Proposition 2.4.3. First, because it follows from this result that it is possible to cover $\mathcal{G}_{r_n}$ by a finite number of sets C_{u,r_n}. Second, Lemma 2.4.6 gives us the key to defining the unit vectors from which the sets C_{u,r_n} are constructed and those unit vectors satisfy that $\langle u, e_d\rangle \geq -\sin\gamma$. It is worth commenting at this point the important role of γ. In Section 2.3 we saw that for the bidimensional case, Lemma 2.4.6 remains valid even for $\gamma = 0$. In fact, it suffices to consider $\mathcal{W}^\mathcal{G}(0) \equiv \mathcal{W}^\mathcal{G}$, being $\mathcal{W}^\mathcal{G}$ the family defined in (2.15). It can be easily proved that for all $u \in \mathcal{W}^\mathcal{G}(0)$ we have that $\langle u, e_d\rangle \geq -\sin 0 = 0$ and that $\mathcal{G}_{r_n}$ coincides with the union of sets C_{u,r_n} with $u \in \mathcal{W}^\mathcal{G}(0)$ (see Figure 2.9). However, when $d > 2$ Lemma 2.4.6 is not true for

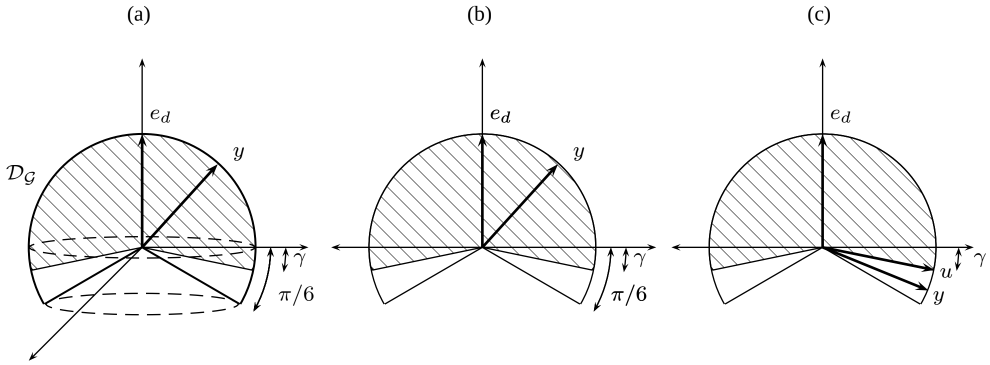


Figure 2.19: *(a) $\mathcal{D}_\mathcal{G}$ in $\mathbb{R}^3$. In the dashed area $\langle y, e_d\rangle \geq -\sin\gamma$. (b) Intersection of $\mathcal{D}_\mathcal{G}$ with the plane defined by e_d and y. (c) For $y \in \mathcal{D}_\mathcal{G}$ with $\langle y, e_d\rangle < -\sin\gamma$ consider u in the plane defined for e_d and y such that $\langle u, e_d\rangle = -\sin\gamma$.*

$\gamma = 0$. For instance, in $\mathbb{R}^3$ it is not possible to cover the points $\{y \in \mathcal{G}_{r_n} : \langle y, e_d\rangle = -\|y\|/2\}$ by a finite number of generalized circular sectors C_{u,r_n} such that $\langle u, e_d\rangle \geq 0$. We would need an infinite number of sectors to cover the set $\mathcal{G}_{r_n}$. To avoid this difficulty, we choose $\gamma > 0$.

If γ is small enough, a considerable portion of C_{u,r_n} will be contained in H_0. As in the bidimensional case, it suffices to consider that portion in order to give a lower bound for $\mu(U \cap B((\alpha-\rho)e_d, \alpha))$ for $U = C_{u,r_n},\ u \in \mathcal{W}^\mathcal{G}(\gamma)$. Let $\nu = \sqrt{\rho(2\alpha-\rho)}$. Then, ν represents the distance to the origin from the points $x = (x_1, \ldots, x_d)$ such that $x \in \partial B((\alpha-\rho)e_d, \alpha)$ and $x_d = 0$, see Figure 2.20.

Lemma 2.4.7.

$$B(0,\nu) \cap H_0 \subset B((\alpha-\rho)e_d, \alpha).$$

Proof. The result is proved analogously to Lemma 2.3.6, using that for $x \in B(0,\nu) \cap H_0$, $\langle x, e_d\rangle \geq 0$.

$\square$

Lemma 2.4.7 establishes that

$$C_{u,r_n} \cap B((\alpha-\rho)e_d, \alpha) \supset C_{u,r_n} \cap H_0 \cap B(0,\nu) = C_{u,\tau_n} \cap H_0,$$

where $\tau_n = \min(\nu, r_n)$. Recall that in $\mathbb{R}^2$ the intersection $C_u \cap H_0$ contains at least one circular sector with central angle $\pi/6$ and hence

$$\mu(C_{u,\tau_n} \cap H_0) \geq \frac{1}{2}\mu(C_{u,\tau_n}).$$

This was enough for our purposes in the bidimensional case. The drawback of working in $\mathbb{R}^d$ is that one cannot immediately infer the position in the space of the sets C_{u,τ_n} and the measure of $C_{u,\tau_n} \cap H_0$. We shall see in Lemma 2.4.8 that if u satisfies $\langle u, e_d\rangle \geq -\sin(\gamma)$, with

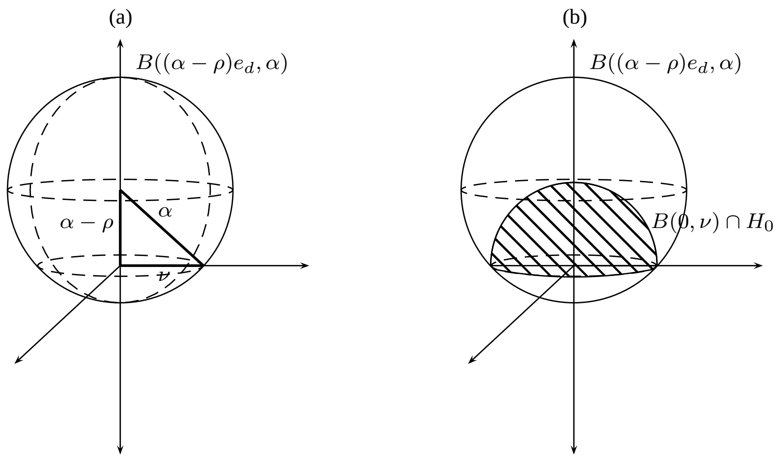


Figure 2.20: *(a) Value of ν. (b) Elements in Lemma 2.4.7 in* $\mathbb{R}^3$.

$0 \le \gamma < \pi/6$, then $C_u \cap H_0$ contains a generalized circular sector $C_{\tilde{u}}^{\theta}$, where $\theta = \theta(\gamma) > 0$. Therefore,

$$\mu(C_{u,\tau_n} \cap H_0) \ge C(\gamma)\mu(C_{u,\tau_n}),\ u \in \mathcal{W}^{\mathcal{G}}(\gamma).$$

Lemma 2.4.8. *Let $0 < \gamma < \pi/6$. For each $u \in \mathbb{S}_d$ such that $\langle u, e_d\rangle \ge -\sin\gamma$ there exists a unit vector $\tilde{u}$ such that*

$$C_{\tilde{u}}^{\theta} \subset C_u \cap H_0,$$

where

$$\theta = \frac{\frac{\pi}{6}-\gamma}{2}.$$

Proof. Let $u \in \mathbb{S}_d$ such that $\langle u, e_d\rangle \ge -\sin\gamma$. If $u = e_d$ the result follows easily by taking $\tilde{u} = u$. If $u \ne e_d$, choose $\tilde{u} \in \mathbb{S}_d$ in the plane passing through the origin and determined by the vectors u and e_d such that the angle between $\tilde{u}$ and u is

$$\varphi_{u,\tilde{u}} = \frac{\frac{\pi}{6}+\gamma}{2}.$$

That is,

$$\tilde{u} = au + be_d,$$

where $a, b \in \mathbb{R}$ are the solutions of the following system of equations.

$$\begin{cases} \langle \tilde{u}, u\rangle = a + b\langle u, e_d\rangle = \cos\varphi_{u,\tilde{u}}, \\ \|\tilde{u}\|^2 = a^2 + b^2 + 2ab\langle u, e_d\rangle = 1. \end{cases}$$

We solve the system and, by using that $\langle u, e_d\rangle = \cos\varphi_{u,e_d}$, we obtain

$$b = \pm\sqrt{\frac{1-\cos^2\varphi_{u,\tilde{u}}}{1-\langle u, e_d\rangle^2}} = \pm\frac{\sin\varphi_{u,\tilde{u}}}{\sin\varphi_{u,e_d}} \quad \text{and} \quad a = \cos\varphi_{u,\tilde{u}} - b\cos\varphi_{u,e_d}.$$

If we choose

$$b = \frac{\sin \varphi_{u,\tilde{u}}}{\sin \varphi_{u,e_d}},$$

then

$$\tilde{u} \quad = \quad \left(\cos \varphi_{u,\tilde{u}} - \frac{\sin \varphi_{u,\tilde{u}}}{\sin \varphi_{u,e_d}} \cos \varphi_{u,e_d}\right) u + \frac{\sin \varphi_{u,\tilde{u}}}{\sin \varphi_{u,e_d}} e_d. \tag{2.26}$$

Figure 2.21 shows the vector $\tilde{u}$. Let us prove that $C_{\tilde{u}}^{\theta} \subset C_u$, where $\theta = (\pi/6 - \gamma)/2$. Then, let $x \in C_{\tilde{u}}^{\theta}$. It follows from Lemma 2.4.1 that $\varphi_{x,\tilde{u}} \leq \theta$ and by construction $\varphi_{\tilde{u},u} = (\pi/6+\gamma)/2$. Then, by the triangle inequality for angles, we have

$$\varphi_{x,u} \leq \varphi_{x,\tilde{u}} + \varphi_{\tilde{u},u} \leq \frac{\frac{\pi}{6} - \gamma}{2} + \frac{\frac{\pi}{6} + \gamma}{2} = \frac{\pi}{6}$$

and then $x \in C_u$. Let us now prove that $C_{\tilde{u}}^{\theta} \subset H_0$. Again, let $x \in C_{\tilde{u}}^{\theta}$. By the triangle inequality for angles,

$$\varphi_{x,e_d} \leq \varphi_{x,\tilde{u}} + \varphi_{\tilde{u},ed} \leq \frac{\frac{\pi}{6} - \gamma}{2} + \varphi_{\tilde{u},ed}. \tag{2.27}$$

Moreover, $\cos \varphi_{\tilde{u},ed} = \langle \tilde{u}, ed \rangle$ and by (2.26) we have

$$\begin{aligned}
\langle \tilde{u}, e_d \rangle &= \langle au + be_d, e_d \rangle \\
&= a \cos \varphi_{u,e_d} + b \\
&= \cos \varphi_{u,\tilde{u}} \cos \varphi_{u,e_d} - \frac{\sin \varphi_{u,\tilde{u}} \cos^2 \varphi_{u,e_d}}{\sin \varphi_{u,e_d}} + \frac{\sin \varphi_{u,\tilde{u}}}{\sin \varphi_{u,e_d}} \\
&= \cos \varphi_{u,\tilde{u}} \cos \varphi_{u,e_d} + \frac{\sin \varphi_{u,\tilde{u}}(1 - \cos^2 \varphi_{u,e_d})}{\sin \varphi_{u,e_d}} \\
&= \cos \varphi_{u,\tilde{u}} \cos \varphi_{u,e_d} + \sin \varphi_{u,\tilde{u}} \sin \varphi_{u,e_d} \\
&= \cos \left(\varphi_{u,\tilde{u}} - \varphi_{u,e_d}\right).
\end{aligned}$$

If $\varphi_{u,\tilde{u}} \geq \varphi_{u,e_d}$,

$$|\varphi_{u,\tilde{u}} - \varphi_{u,e_d}| = \varphi_{u,\tilde{u}} - \varphi_{u,e_d} \leq \varphi_{u,\tilde{u}} = \frac{\frac{\pi}{6} + \gamma}{2} = \frac{\gamma}{2} + \frac{\pi}{12}.$$

If $\varphi_{u,\tilde{u}} < \varphi_{u,e_d}$,

$$|\varphi_{u,\tilde{u}} - \varphi_{u,e_d}| = \varphi_{u,e_d} - \varphi_{u,\tilde{u}} = \varphi_{u,e_d} - \frac{\frac{\pi}{6} + \gamma}{2} \leq \frac{\pi}{2} + \gamma - \frac{\frac{\pi}{6} + \gamma}{2} = \frac{\gamma}{2} + \frac{5\pi}{12},$$

where the last inequality is a consequence of $\langle u, e_d \rangle \geq -\sin \gamma$. Therefore,

$$|\varphi_{u,\tilde{u}} - \varphi_{u,e_d}| \leq \frac{\gamma}{2} + \frac{5\pi}{12}$$

and, turning to (2.27), we obtain that

$$\varphi_{x,e_d} \leq \frac{\frac{\pi}{6} - \gamma}{2} + \frac{\gamma}{2} + \frac{5\pi}{12} = \frac{\pi}{2}.$$

That is,

$$\langle x, e_d \rangle = \|x\| \cos \varphi_{x,e_d} \geq 0$$

and then $C_{\tilde{u}}^{\theta} \subset H_0$.

$\square$

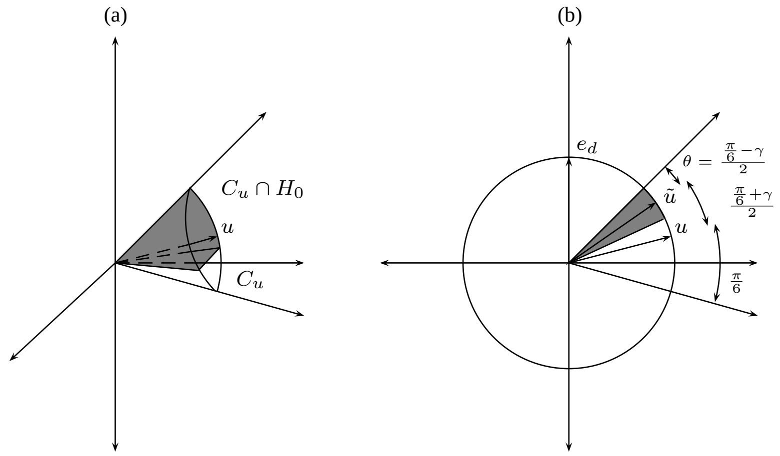


Figure 2.21: *(a) Set $C_u \cap H_0$ for $u \in \mathbb{S}^3$ satisfying $\langle u, e_d \rangle \geq -\sin\gamma$ with $0 < \gamma < \pi/6$. (b) In Lemma 2.4.8 we define the vector $\tilde{u}$ in the plane passing through the origin and determined by the vectors u and e_d such that $\varphi_{u,\tilde{u}} = (\pi/6 + \gamma)/2$ and $C_{\tilde{u}}^{\theta} \subset (C_u \cap H_0)$.*

Note that the choice of γ in the interval $(0, \pi/6)$ is the key to guaranteeing that $C_{\tilde{u}}^{\theta}$ has positive central angle. Fix $\gamma \in (0, \pi/6)$. It follows from Lemma 2.4.6 that we can cover $\mathcal{G}_{r_n}$ by a finite number of generalized circular sectors C_{u,r_n}, such that for all $y \in \mathcal{G}_{r_n}$ there exists $u \in \mathcal{W}^{\mathcal{G}} \equiv \mathcal{W}^{\mathcal{G}}(\gamma)$ with $\langle u, e_d \rangle \geq -\sin\gamma$ such that $y \in C_{u,r_n}$. Moreover, Lemma 2.4.3 yields that $C_{u,r_n} \subset B(y, r_n)$. By Lemmas 2.4.7 and 2.4.8 we have that, for each C_{u,r_n} with $u \in \mathcal{W}^{\mathcal{G}}$,

$$\begin{aligned} C_{u,r_n} \cap B((\alpha-\rho)e_d, \alpha) &\supset C_{u,r_n} \cap H_0 \cap B(0,\nu) \\ &\supset C_{\tilde{u},r_n}^{\theta} \cap B(0,\nu) \\ &= C_{\tilde{u},\tau_n}^{\theta}, \end{aligned}$$

where $\tau_n = \min(\nu, r_n)$. Then, by Lemma 2.4.2,

$$\mu(C_{u,r_n} \cap B((\alpha-\rho)e_d, \alpha)) \geq \mu(C_{\tilde{u},\tau_n}^{\theta}) \geq \frac{1}{m}\mu(B(0,\tau_n)). \tag{2.28}$$

Recall that according to Lemma 2.4.2 the ball $B(0, \tau_n)$ can be covered by a finite number m of circular sectors. It remains to find a lower bound for the measure of $B(0, \tau_n)$. By using that $r_n \leq \alpha$ and $\rho \leq r_n/2$, we have that

$$\mu\left(B(0,\nu)\right) = w_d \rho^{\frac{d}{2}} (2\alpha - \rho)^{\frac{d}{2}} \geq w_d \rho^{\frac{d}{2}} r_n^{\frac{d}{2}} \geq w_d \rho^{\frac{d+1}{2}} r_n^{\frac{d-1}{2}}.$$

On the other hand, since $\rho \leq r_n/2$, we also have that

$$\mu\left(B(0, r_n)\right) = w_d r_n^d = w_d r_n^{\frac{d+1}{2}} r_n^{\frac{d-1}{2}} \geq w_d \rho^{\frac{d+1}{2}} r_n^{\frac{d-1}{2}}.$$

Turning to (2.28) we deduce that

$$\mu(C_{u,r_n} \cap B((\alpha - \rho)e_2, \alpha)) \geq L^{\mathcal{G}} \rho^{\frac{d+1}{2}} r_n^{\frac{d-1}{2}},$$

where $L^{\mathcal{G}} = w_d/m > 0$ is constant. □

Therefore, we have solved the problem for those points $y \in \mathcal{G}_{r_n}$. The unavoidable family of sets $\mathcal{U}^{\mathcal{G}}_{0,r_n}$ we shall consider is

$$\mathcal{U}^{\mathcal{G}}_{0,r_n} = \{C_{u,r_n}, u \in \mathcal{W}^{\mathcal{G}}\},$$

being $m^{\mathcal{G}}$ the number of elements of that family, determined by the number of unit vectors in $\mathcal{W}^{\mathcal{G}}$. We now concentrate on the points $y \in \mathcal{F}_{r_n}$. Recall that

$$\mathcal{F}_{r_n} = \left\{ y \in B(0, r_n) : \langle y, e_d \rangle < -\frac{1}{2} \|y\| \right\}.$$

The aim is to define a finite family of sets $\mathcal{U}^{\mathcal{F}}_{0,r_n}$, such that for all $y \in \mathcal{F}_{r_n}$, there exists $U \in \mathcal{U}^{\mathcal{F}}_{0,r_n}$ such that $U \subset B(y, r_n)$ and

$$\mu(U \cap B((\alpha - \rho)e_d, \alpha)) \geq L^{\mathcal{F}} r_n^{\frac{d-1}{2}} \rho^{\frac{d+1}{2}}, \quad \forall U \in \mathcal{U}^{\mathcal{F}}_{0,r_n}.$$

As in the bidimensional case, the generalized circular sectors C_{u,r_n} are no longer appropriate to form the unavoidable family $\mathcal{U}^{\mathcal{F}}_{0,r_n}$. For example, consider the point $y = -r_n e_d \in \mathcal{F}_{r_n}$ and the generalized circular sector C_{-e_d,r_n}. Then $y \in C_{-e_d,r_n} \subset B(y, r_n)$. The intersection $C_{-e_d,r_n} \cap B((\alpha - \rho)e_d, \alpha)$ is small, as it is shown in Figure 2.22 in $\mathbb{R}^3$. In fact, it can be easily proved that

$$\mu(C_{-e_d,r_n} \cap B((\alpha - \rho)e_d, \alpha)) = O(\rho^d).$$

Even though we could have considered different circular sectors for $y = -r_n e_d$, we need that $C_{u,r_n} \subset B(y, r_n)$. This fact determines the position in the space of C_{u,r_n} in such a way that the measure of $C_{u,r_n} \cap B((\alpha - \rho)e_d, \alpha)$ would not be much larger that $C_{-e_d,r_n} \cap B((\alpha - \rho)e_d, \alpha)$ and hence not large enough for our purposes. We need to define another kind of sets for $y \in \mathcal{F}_{r_n}$. Proposition 2.4.4 provides a solution to this problem.

Proposition 2.4.4. *There exists a finite family of sets $\mathcal{U}^{\mathcal{F}}_{0,r_n}$ such that, for all $y \in \mathcal{F}_{r_n}$, there exists $U \in \mathcal{U}^{\mathcal{F}}_{0,r_n}$ such that $U \subset B(y, r_n)$ and*

$$\mu(U \cap B((\alpha - \rho)e_d, \alpha)) \geq L^{\mathcal{F}} r_n^{\frac{d-1}{2}} \rho^{\frac{d+1}{2}},$$

with $L^{\mathcal{F}} > 0$ a constant.

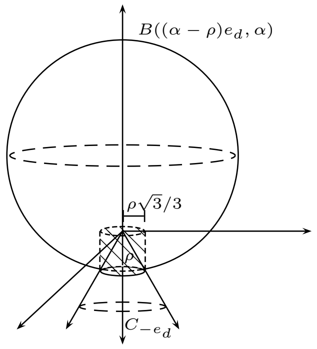


Figure 2.22: *Set C_{-e_d} in $\mathbb{R}^3$. The volume of the represented cylinder of height ρ is $\rho\pi(\rho\sqrt{3}/3)^2$. The cylinder contains the set $C_{-e_d,r_n}\cap B((\alpha-\rho)e_d,\alpha)$. The radius $\rho\sqrt{3}/3$ is derived from the Pythagorean theorem.*

Proof. The sketch of the proof is the same as that of Proposition 2.3.4. First, we shall define a set whose measure is large enough for our purposes. Then we shall construct a partition generated by a finite number of subsets, all of them with the same measure, and satisfying the conditions to form an unavoidable family $\mathcal{U}^{\mathcal{F}}_{0,r_n}$. Let us consider as reference set $B((\alpha-\rho)e_d,\alpha)\cap B(-r_ne_d,r_n)$. We define

$$\mathcal{C}(h_1)=\{x\in\mathbb{R}^d:\ -h_1\le\langle x,e_d\rangle\le 0\}\cap B(-r_ne_d,r_n),\tag{2.29}$$

where

$$h_1=\frac{\rho(2\alpha-\rho)}{2(\alpha+r_n-\rho)}$$

is the distance from the hyperplane $\{x=(x_1,\dots,x_d)\in\mathbb{R}^d:\ x_d=0\}$ to any point of $\partial B((\alpha-\rho)e_d,\alpha)\cap\partial B(-r_ne_d,r_n)$. Its value is computed from the Pythagorean theorem (recall Proposition 2.3.4). In Figure 2.23 the set $\mathcal{C}(h_1)$ is represented for the particular case of $\mathbb{R}^3$. Lemma 2.4.9 gives a lower found for the measure of $\mathcal{C}(h_1)$.

Lemma 2.4.9. *We have that*

$$\mu(\mathcal{C}(h_1))\ge Lr_n^{\frac{d-1}{2}}\rho^{\frac{d+1}{2}}.$$

Proof. Let us consider the translation by the vector h_1e_d. It is straightforward to see that $\mathcal{C}(h_1)\oplus\{h_1e_d\}=\mathcal{C}_0(h_1)$, where

$$\mathcal{C}_0(h_1)=\{x=(x_1,\dots,x_d)\in\mathbb{R}^d:\ 0\le\langle x,e_d\rangle\le h_1\}\cap B(-(r_n-h_1)e_d,r_n).$$

Moreover, since the Lebesgue measure is invariant under translations, we have that $\mu(\mathcal{C}(h_1))=\mu(\mathcal{C}_0(h_1))$. For $0\le l\le h_1$ we define the set

$$\mathcal{C}_0(h_1,l)=\{x=(x_1,\dots,x_{d-1},l):\ x\in\mathcal{C}_0(h_1)\}.$$

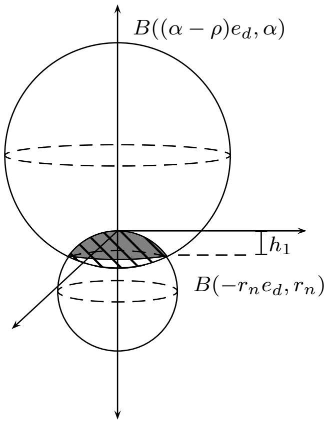


Figure 2.23: *The dashed area corresponds to* $B((\alpha-\rho)e_d,\alpha)\cap B(-r_ne_d,r_n)$ *in* $\mathbb{R}^3$. *In gray it is represented* $\mathcal{C}(h_1)$.

It can be easily seen that $\mathcal{C}_0(h_1,l)$ is the $(d-1)$-dimensional sphere with centre le_d and radius $r(l)$, where

$$r(l)=\sqrt{r_n^2-(r_n-h_1+l)^2}=\sqrt{2r_n(h_1-l)-(h_1-l)^2}.$$

Then, by using that the Lebesgue measure is a product measure,

$$\mu(\mathcal{C}(h_1))=\int_0^{h_1}\mu_{d-1}(\mathcal{C}_0(h_1,l))dl=\omega_{d-1}\int_0^{h_1}r(l)^{d-1}dl,$$

where μ_{d-1} denotes the $(d-1)$-dimensional Lebesgue measure. Therefore,

$$\begin{aligned}\mu(\mathcal{C}(h_1)) &= \omega_{d-1}\int_0^{h_1}\left(2r_n(h_1-l)-(h_1-l)^2\right)^{\frac{d-1}{2}}dl\\ &= \omega_{d-1}\int_0^{h_1}\left(2r_nt-t^2\right)^{\frac{d-1}{2}}dt\\ &\geq \omega_{d-1}\int_0^{h_1}(r_nt)^{\frac{d-1}{2}}dt\\ &= \omega_{d-1}r_n^{\frac{d-1}{2}}h_1^{\frac{d+1}{2}}\frac{2}{d+1}.\end{aligned}\tag{2.30}$$

We have used the change of variables formula with $t=h_1-l$. For $t\in[0,h_1]$ we get that $t\leq r_n$, since by construction $h_1\leq\rho\leq r_n/2$. Moreover, since $r_n\leq\alpha$ we have that

$$h_1=\frac{\rho(2\alpha-\rho)}{2(\alpha+r_n-\rho)}\geq\frac{\rho}{2}$$

and then

$$\mu(\mathcal{C}(h_1))\geq\frac{\omega_{d-1}}{(d+1)2^{\frac{d-1}{2}}}r_n^{\frac{d-1}{2}}\rho^{\frac{d+1}{2}}.$$

□

Lemma 2.4.9 asserts that the measure of $\mathcal{C}(h_1)$ is large enough for our purposes. The following lemma generalizes Lemma 2.3.8.

Lemma 2.4.10.

$$\mathcal{C}(h_1) \subset B((\alpha - \rho)e_d, \alpha).$$

Proof. The proof is analogous to that of Lemma 2.3.8. Again, the result can be proved by using that, if $x \in \mathcal{C}(h_1)$ then $x \in B(-r_n e_d, r_n)$ and hence, $\|x\|^2 \leq -2r_n \langle x, e_d \rangle$, where $\langle x, e_d \rangle \geq -h_1$. □

Lemma 2.4.10 plays an important role since it guarantees that $\mathcal{C}(h_1)$ intersects $B((\alpha - \rho)e_d, \alpha)$ in a large set. To sum up, we have proved that, by Lemmas 2.4.9 and 2.4.10,

$$\mu(\mathcal{C}(h_1) \cap B((\alpha - \rho)e_d, \alpha)) \geq L r_n^{\frac{d-1}{2}} \rho^{\frac{d+1}{2}}.$$

However, it still remains to determine which sets form the finite family $\mathcal{U}^{\mathcal{F}}_{0,r_n}$ mentioned in the statement of Proposition 2.4.4. Using the same arguments as in Proposition 2.3.4, we divide the set $\mathcal{C}(h_1)$ into a finite number of components, all of them with the same measure. They should also fulfill the conditions to form the family $\mathcal{U}^{\mathcal{F}}_{0,r_n}$. Recall that in the bidimensional case we considered the partition $\mathbb{R}^2 = Q_1 \cup Q_2$, where $Q_1 = \{x = (x_1, x_2) \in \mathbb{R}^2 : \ x_1 \geq 0\}$ and $Q_2 = \{x = (x_1, x_2) \in \mathbb{R}^2 : \ x_1 \leq 0\}$. Based on this partition we divided $\mathcal{C}(h_1)$ into two subsets. These two subsets were proved to measure the same since $\mathcal{C}(h_1)$ is symmetric with respect to the axis OX. How can we divide $\mathcal{C}(h_1)$ in $\mathbb{R}^d$? How do we construct a partition in a finite number of sets, all of them with the same measure? We next state an important and general result that provides a finite partition of $\mathbb{R}^d$. An immediate consequence of this general result is that it gives us the key step toward the definition of finite partitions of any subset of $\mathbb{R}^d$.

Lemma 2.4.11. *Let $\theta > 0$. There exists a finite family of unit vectors $\mathcal{W}_\theta \subset \mathbb{S}_{d-1}$ such that*

$$\mathbb{R}^d = \bigcup_{u \in \mathcal{W}_\theta} Q_u^\theta,$$

where, for each $u \in \mathcal{W}_\theta \subset \mathbb{S}_{d-1}$,

$$Q_u^\theta = \left\{ x = (x_1, \ldots, x_{d-1}, x_d) \in \mathbb{R}^d : \ (x_1, \ldots, x_{d-1}) \in C_u^\theta \subset \mathbb{R}^{d-1} \right\}.$$

Proof. Since the unit sphere in $\mathbb{R}^{d-1}$ is compact we get, by the same arguments as in Lemma 2.4.2, that $\mathbb{R}^{d-1}$ can be covered by a finite number of sets C_u^θ, with $\theta > 0$. Let $\mathcal{W}_\theta \subset \mathbb{S}_{d-1}$ be the finite family of unit vectors that determine those sets C_u^θ. We have that

$$\mathbb{R}^{d-1} = \bigcup_{u \in \mathcal{W}_\theta} C_u^\theta.$$

Now, $\mathbb{R}^d = \mathbb{R}^{d-1} \times \mathbb{R}$ and

$$\mathbb{R}^d = \bigcup_{u \in \mathcal{W}_\theta} \left\{ x = (x_1, \ldots, x_{d-1}, x_d) \in \mathbb{R}^d : \ (x_1, \ldots, x_{d-1}) \in C_u^\theta \subset \mathbb{R}^{d-1} \right\} = \bigcup_{u \in \mathcal{W}_\theta} Q_u^\theta.$$

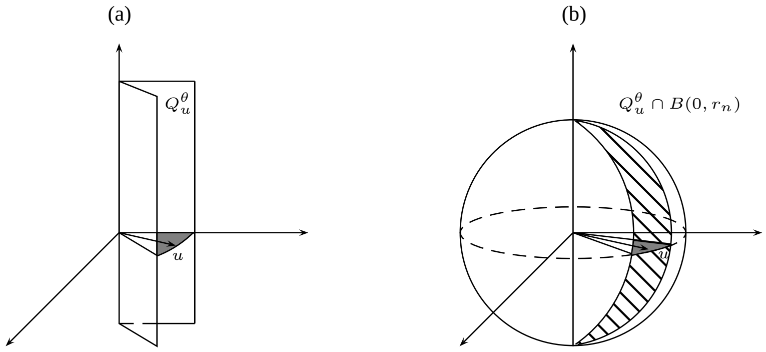


Figure 2.24: *(a) Example of Q_u^θ in $\mathbb{R}^3$. (b) $Q_u^\theta \cap B(0, r_n)$ in $\mathbb{R}^3$.*

□

Figure 2.24 represents a set Q_u^θ in $\mathbb{R}^3$. As mentioned, Lemma 2.4.11 gives us the key to constructing partitions of subsets of $\mathbb{R}^d$. In particular,

$$\mathcal{F}_{r_n} = \bigcup_{u \in \mathcal{W}_\theta} Q_u^\theta \cap \mathcal{F}_{r_n}.$$

Fix $\theta = \pi/6$. As we will see, such choice of θ is arbitrary in some sense. In fact, the following results remain valid for different values of θ. He have chosen the value $\theta = \pi/6$ because it allows us to continue with the same notation. Thus, we write $\mathcal{W}$ and Q_u to refer to $\mathcal{W}_\theta$ and Q_u^θ for $\theta = \pi/6$ as defined in Lemma 2.4.11. Note that the sets Q_1 and Q_2 in Proposition 2.3.4 coincide, returning to the notation of Lemma 2.4.11 for $d = 2$, with Q_1 and Q_{-1}, respectively. Let us consider the partition

$$\mathcal{C}(h_1) = \bigcup_{u \in \mathcal{W}} Q_u \cap \mathcal{C}(h_1). \tag{2.31}$$

Lemma 2.4.16, corresponding to Lemma 2.3.9 in $\mathbb{R}^2$, states that the partition given in (2.31) provides an unavoidable family of sets. Lemma 2.4.17 proves that the sets $Q_u \cap \mathcal{C}(h_1)$ with $u \in \mathcal{W}$ measure the same. First, however, we require several preliminary results, needed in the proof of Lemma 2.4.16. Lemmas 2.4.12 and 2.4.13 prove that, both $\mathcal{C}(h_1)$ and $\mathcal{F}_{r_n}$ are contained in $B(0, r_n) \cap B(-r_n e_d, r_n)$. On the other hand, Lemma 2.4.14 establishes that the distance between $x, y \in B(0, r_n) \cap B(-r_n e_d, r_n)$ such that x lies on the boundary of $B(-r_n e_d, r_n)$ and y lies on the boundary of $B(0, r_n)$ is lower or equal to r_n, whenever x and y fulfill

$$x_1 y_1 + \ldots + x_{d-1} y_{d-1} \geq \beta \sqrt{\|x\|^2 - x_d^2} \sqrt{\|y\|^2 - y_d^2}, \tag{2.32}$$

for some $\beta \geq 1/3$. In spite of the fact that it seems to be an artifitial condition, Lemma 2.4.15 shows that in particular, $x, y \in Q_u$ fulfill the restriction (2.32) for $\beta = 1/2$.

Lemma 2.4.12.

$$\mathcal{C}(h_1) \subset B(0, r_n) \cap B(-r_n e_d, r_n).$$

Proof. The lemma will be proved if we can show that $\mathcal{C}(h_1) \subset B(0, r_n)$. Let $x \in \mathcal{C}(h_1)$. Since $x \in B(-r_n e_d, r_n)$, we have that

$$\|x + r_n e_d\|^2 = \|x\|^2 + r_n^2 + 2r_n \langle x, e_d \rangle \leq r_n^2.$$

Then, by the definition of $\mathcal{C}(h_1)$ and the fact that $h_1 \leq \rho \leq r_n/2$,

$$\|x\|^2 \leq -2r_n \langle x, e_d \rangle \leq 2r_n h_1 \leq r_n^2.$$

$\square$

Lemma 2.4.13.

$$\mathcal{F}_{r_n} \subset B(0, r_n) \cap B(-r_n e_d, r_n).$$

Proof. We have to show that $\mathcal{F}_{r_n} \subset B(-r_n e_d, r_n)$. Let $y \in \mathcal{F}_{r_n}$. We have that

$$\|y + r_n e_d\|^2 = \|y\|^2 + r_n^2 + 2r_n \langle y, e_d \rangle < \|y\|^2 + r_n^2 - r_n \|y\| \leq \max(\|y\|^2, r_n^2) \leq r_n^2.$$

$\square$

Observe that if $x \in \mathcal{C}(h_1)$, then $\|x - (\alpha - \rho)e_d\|^2 \leq \alpha^2$ by Lemma 2.4.10. Moreover, it follows from Lemma 2.4.12 that $\|x\|^2 \leq r_n^2$ and $\|x + r_n e_d\|^2 \leq r_n^2$. If $y \in \mathcal{F}_{r_n}$, Lemma 2.4.13 yields that $\|y\|^2 \leq r_n^2$ and $\|y + r_n e_d\|^2 \leq r_n^2$.

Lemma 2.4.14. *Let $x, y \in \mathbb{R}^d$ such that*

i) $\|x\|^2 \leq r_n^2$, $\|x\|^2 = -2r_n x_d$,

ii) $\|y\|^2 = r_n^2$, $\|y\|^2 \leq -2r_n y_d$,

iii) $x_1 y_1 + \ldots + x_{d-1} y_{d-1} \geq \beta \sqrt{\|x\|^2 - x_d^2} \sqrt{\|y\|^2 - y_d^2}$, *for any* $\beta \geq 1/3$.

Then,

$$\|x - y\|^2 \leq r_n^2.$$

Proof. Let x and y be under the stated conditions.

$$\|x - y\|^2 = \|x\|^2 + \|y\|^2 - 2\langle x, y \rangle = -2r_n x_d + r_n^2 - 2\langle x, y \rangle = r_n^2 - 2\left(\langle x, y \rangle + r_n x_d\right).$$

We denote $E(x, y) = \langle x, y \rangle + r_n x_d$. Then,

$$\|x - y\|^2 \leq r_n^2 \Leftrightarrow E(x, y) \geq 0.$$

By iii), we obtain

$$\begin{aligned}
E(x,y) &= x_1y_1 + \ldots + x_{d-1}y_{d-1} + x_dy_d + r_nx_d \\
&\geq \beta\sqrt{\|x\|^2 - x_d^2}\sqrt{\|y\|^2 - y_d^2} + x_d(y_d + r_n) \\
&= \beta\sqrt{-2r_nx_d - x_d^2}\sqrt{r_n^2 - y_d^2} + x_d(y_d + r_n) \\
&= \beta\sqrt{2r_nu - u^2}\sqrt{2r_nv - v^2} - uv.
\end{aligned}$$

The last equality follows from the change of variables formula with $u = -x_d$ and $v = y_d + r_n$. Now, i) yields

$$0 \leq \|x\|^2 = -2r_nx_d \leq r_n^2$$

and hence $0 \leq u \leq r_n/2$. Similarly, by ii)

$$\|y\|^2 = r_n^2 \leq -2r_ny_d$$

and, therefore, $y_d \leq -r_n/2$. Moreover, $\|y\|^2 = r_n^2$ yields $y_d^2 \leq r_n^2$ and, in particular, $y_d \geq -r_n$. Finally,

$$-r_n \leq y_d \leq -r_n/2.$$

Then, $0 \leq u \leq r_n/2$ and $0 \leq v \leq r_n/2$. By using that $\beta \geq 1/3$ we complete the proof since

$$\begin{aligned}
E(x,y) &\geq \beta\sqrt{2r_nu - u^2}\sqrt{2r_nv - v^2} - uv \\
&= \beta\sqrt{u(2r_n - u)}\sqrt{v(2r_n - v)} - uv \\
&\geq \beta\sqrt{u\frac{3r_n}{2}}\sqrt{v\frac{3r_n}{2}} - uv \\
&= \beta\frac{3r_n}{2}\sqrt{uv} - uv \\
&\geq \frac{r_n}{2}\sqrt{uv} - uv \\
&\geq \sqrt{uv}\sqrt{uv} - uv \\
&= 0.
\end{aligned}$$

□

Lemma 2.4.15. *Let $u \in \mathcal{W} \subset \mathbb{S}_{d-1}$. For all $x, y \in Q_u$ we have that*

$$x_1y_1 + \ldots + x_{d-1}y_{d-1} \geq \frac{1}{2}\sqrt{\|x\|^2 - x_d^2}\sqrt{\|y\|^2 - y_d^2}.$$

Proof. Let $x = (x_1, \ldots, x_d)$, $y = (y_1, \ldots, y_d) \in Q_u$. We denote $x_{-d} = (x_1, \ldots, x_{d-1})$ and $y_{-d} = (y_1, \ldots, y_{d-1})$ the $(d-1)$-dimensional vectors obtained after removing the last component of the original vectors. Then

$$x_1y_1 + \ldots + x_{d-1}y_{d-1} = \langle x_{-d}, y_{-d}\rangle = \|x_{-d}\|\,\|y_{-d}\|\cos\varphi_{x_{-d},y_{-d}}.$$

Since $x, y \in Q_u$, we have that $x_{-d}, y_{-d} \in C_u \subset \mathbb{R}^{d-1}$ and hence

$$\cos \varphi_{x_{-d},y_{-d}} \geq \cos(\varphi_{x_{-d},u} + \varphi_{y_{-d},u}) \geq \cos\frac{\pi}{3} = \frac{1}{2}.$$

The result is a straightforward consequence of $\|x_{-d}\| = \sqrt{\|x\|^2 - x_d^2}$. We complete the proof of the lemma by applying the same arguments to y.

□

Note that the proof of Lemma 2.4.15 makes clear that the choice of θ defining the sets Q_u^θ is, in some sense, arbitrary. In fact, by proceeding in an analogous manner we can deduce a more general result. For all $x, y \in Q_u^\theta$, with $\theta \leq \frac{\arccos(1/3)}{2}$ we have that

$$x_1y_1 + \ldots + x_{d-1}y_{d-1} \geq \frac{1}{3}\sqrt{\|x\|^2 - x_d^2}\sqrt{\|y\|^2 - y_d^2}. \tag{2.33}$$

Note that (2.33) is exactly the same as condition iii) of Lemma 2.4.14. Even so, the choice $\theta = \pi/6$ is enough for our purposes.

Now we are ready to prove that the sets $Q_u \cap \mathcal{C}(h_1)$, with $u \in \mathcal{W}$, satisfy the conditions to form the family $\mathcal{U}_{0,r_n}^{\mathcal{F}}$ we are trying to define for two reasons. First, the $\theta = \pi/6$ is small enough to guarantee that $(Q_u \cap \mathcal{C}(h_1)) \subset B(y, r_n)$ for all $y \in Q_u \cap \mathcal{F}_{r_n}$, see Lemma 2.4.16. Second, the partition in (2.31) is such that all sets $Q_u \cap \mathcal{C}(h_1)$ with $u \in \mathcal{W}$ measure the same, see Lemma 2.4.17.

Lemma 2.4.16. *Let $u \in \mathcal{W} \subset \mathbb{S}_{d-1}$. For all $y \in Q_u \cap \mathcal{F}_{r_n}$,*

$$Q_u \cap \mathcal{C}(h_1) \subset B(y, r_n).$$

Proof. Let $y \in Q_u \cap \mathcal{F}_{r_n}$. We define

$$y^* = r_n \frac{y}{\|y\|}.$$

Thereby $y^* = (y_1^*, \ldots, y_d^*)$ satisfies:

i) $\|y^*\|^2 = r_n^2$.

ii) $y^* \in \mathcal{F}_{r_n}$.

iii) $y^* \in Q_u$.

Then, by i), ii) and Lemma 2.4.13 we deduce that $\|y^*\|^2 = r_n^2$ and $\|y^*\|^2 \leq -2r_n y_d^*$.

Let $x \in Q_u \cap \mathcal{C}(h_1)$, with $\|x\|^2 = -2r_n x_d$. Figure 2.25 represents a set $Q_u \cap \mathcal{C}(h_1)$ in $\mathbb{R}^3$. It follows from Lemma 2.4.12 that $\|x\|^2 \leq r_n^2$. By the definition of x and by iii) we have that $x, y^* \in Q_u$ and hence, by Lemma 2.4.15, we can conclude that all the hypothesis of Lemma 2.4.14 are fulfilled. Therefore,

$$\|x - y^*\| \leq r_n.$$

That is, $y^* \in B(x, r_n)$ for all $x \in Q_u \cap \mathcal{C}(h_1)$, with $\|x\|^2 = -2r_n x_d$. Moreover, it follows from Lemma 2.4.12 that $0 \in B(x, r_n)$ for all $x \in Q_u \cap \mathcal{C}(h_1)$, with $\|x\|^2 = -2r_n x_d$. By using that $B(x, r_n)$ is convex and that the point y lies on the segment that joins 0 with y^*, we have that

$$y \in B(x, r_n) \text{ for all } x \in Q_u \cap \mathcal{C}(h_1) \text{ with } \|x\|^2 = -2r_n x_d. \tag{2.34}$$

Now, let $x \in Q_u \cap \mathcal{C}(h_1)$ arbitrary. Then

$$x^* = \frac{x + r_n e_d}{\|x + r_n e_d\|} r_n - r_n e_d$$

satisfies:

i) $\|x^*\|^2 = -2r_n x_d^*$.

ii) $x^* \in \mathcal{C}(h_1)$. To prove this note first that i) yields $x^* \in B(-r_n e_d, r_n)$. Moreover, $\langle x^*, e_d \rangle \leq 0$. The remainder of the proof consists of showing that $\langle x^*, e_d \rangle \geq -h_1$. Thus,

$$\langle x^*, e_d \rangle = \langle \frac{x + r_n e_d}{\|x + r_n e_d\|} r_n - r_n e_d, e_d \rangle = \frac{r_n}{\|x + r_n e_d\|} (\langle x, e_d \rangle + r_n) - r_n.$$

Since $x \in \mathcal{C}(h_1)$, we have that $\|x + r_n e_d\| \leq r_n$ and hence $\frac{r_n}{\|x + r_n e_d\|} \geq 1$. Moreover, $\langle x, e_d \rangle + r_n \geq -h_1 + r_n \geq 0$. That is,

$$\langle x^*, e_d \rangle \geq \langle x, e_d \rangle + r_n - r_n = \langle x, e_d \rangle \geq -h_1.$$

iii) $x^* \in Q_u$. We need to show that $x_{-d}^* = (x_1^*, \ldots, x_{d-1}^*) \in C_u$. By using that

$$x^* = \frac{x + r_n e_d}{\|x + r_n e_d\|} r_n - r_n e_d = \frac{r_n}{\|x + r_n e_d\|} x + \left(\frac{r_n^2}{\|x + r_n e_d\|} - r_n \right) e_d,$$

we have that

$$x_{-d}^* = \frac{r_n}{\|x + r_n e_d\|} x_{-d}.$$

Since $x \in Q_u$,

$$\begin{aligned} \langle x_{-d}^*, u \rangle &= \frac{r_n}{\|x + r_n e_d\|} \langle x_{-d}, u \rangle \\ &\geq \frac{r_n}{\|x + r_n e_d\|} \|x_{-d}\| \cos \frac{\pi}{6} \\ &= \|x_{-d}^*\| \cos \frac{\pi}{6}. \end{aligned}$$

Then $x^* \in Q_u \cap \mathcal{C}(h_1)$ with $\|x^*\|^2 = -2r_n x_d$ and by (2.34) we have that $x^* \in B(y, r_n)$ for all $y \in Q_u \cap \mathcal{F}_{r_n}$. It follows from Lemma 2.4.13 that $-r_n e_d \in B(y, r_n)$ for all $y \in Q_u \cap \mathcal{F}_{r_n}$. Moreover, x lies on the segment that joins x^* with $-r_n e_d$. In fact, we shall see that $x = ax^* - (1 - a) r_n e_d$ with $a \in [0, 1]$. By the definition of x^*, it follows that

$$x = \frac{\|x + r_n e_d\|}{r_n} x^* - \left(1 - \frac{\|x + r_n e_d\|}{r_n}\right) r_n e_d,$$

where

$$0 \leq \frac{\|x + r_n e_d\|}{r_n} \leq 1$$

as $x \in \mathcal{C}(h_1)$. Finally, since $B(y, r_n)$ is convex, we have that

$$x \in B(y, r_n) \text{ for all } y \in Q_u \cap \mathcal{F}_{r_n}.$$

To sum up, we have proved that

$$Q_u \cap \mathcal{C}(h_1) \subset B(y, r_n) \text{ for all } y \in Q_u \cap \mathcal{F}_{r_n}.$$

This completes the proof of Lemma 2.4.16. □

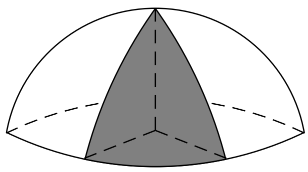


Figure 2.25: *Example of set* $Q_u \cap \mathcal{C}(h_1)$ *in* $\mathbb{R}^3$.

We are now ready to complete the proof of Proposition 2.4.4. In view of the previous results we define the finite family

$$\mathcal{U}^{\mathcal{F}}_{0,r_n} = \{Q_u \cap \mathcal{C}(h_1), u \in \mathcal{W} \subset \mathbb{S}_{d-1}\}.$$

Since $y \in \mathcal{F}_{r_n}$, there exists $u \in \mathcal{W}$ such that $y \in Q_u \cap \mathcal{F}_{r_n}$. Lemma 2.4.16 yields $Q_u \cap \mathcal{C}(h_1) \subset B(y, r_n)$. Moreover,

$$\mathcal{C}(h_1) = \bigcup_{u \in \mathcal{W}} Q_u \cap \mathcal{C}(h_1)$$

and, by appealing to Lemma 2.4.9, we get

$$L r_n^{\frac{d-1}{2}} \rho^{\frac{d+1}{2}} \leq \mu(\mathcal{C}(h_1)) \leq \sum_{u \in \mathcal{W}} \mu(\mathcal{C}(h_1) \cap Q_u). \tag{2.35}$$

Next we prove that the sets $Q_u \cap \mathcal{C}(h_1)$ measure all the same, independently of $u \in \mathcal{W}$.

Lemma 2.4.17. *For all* $u, v \in \mathcal{W}$,

$$\mu(Q_u \cap \mathcal{C}(h_1)) = \mu(Q_v \cap \mathcal{C}(h_1)).$$

Proof. Let $u, v \in \mathcal{W}$. Consider the orthogonal transformation

$$\mathcal{O}_{-d} : \mathbb{R}^{d-1} \longrightarrow \mathbb{R}^{d-1}$$

such that $\mathcal{O}_{-d}(u) = v$. Consider the function

$$\mathcal{O} : \mathbb{R}^d \longrightarrow \mathbb{R}^d,$$

where $\mathcal{O}(x) = \mathcal{O}(x_1, \dots, x_{d-1}, x_d) := (\mathcal{O}_{-d}(x_{-d}), x_d)$. The function $\mathcal{O}$ is also an orthogonal transformation. We have that

$$\mathcal{O}(\mathcal{C}(h_1)) = \mathcal{C}(h_1). \tag{2.36}$$

Let us first prove that $\mathcal{O}(\mathcal{C}(h_1)) \subset \mathcal{C}(h_1)$. Consider the vector $\mathcal{O}(x)$ with $x \in \mathcal{C}(h_1)$. Then $\mathcal{O}(x) \in B(-r_n e_d, r_n)$, since

$$\|\mathcal{O}(x) + r_n e_d\|^2 = \|\mathcal{O}_{-d}(x_{-d})\|^2 + (x_d + r_n)^2 = \|x_{-d}\|^2 + (x_d + r_n)^2 = \|x + r_n e_d\|^2 \leq r_n^2.$$

Moreover, $-h_1 \leq \langle \mathcal{O}(x), e_d \rangle \leq 0$, since $\langle \mathcal{O}(x), e_d \rangle = x_d$ and $x \in \mathcal{C}(h_1)$ and hence $\mathcal{O}(x) \in \mathcal{C}(h_1)$. Next we prove that $\mathcal{C}(h_1) \subset \mathcal{O}(\mathcal{C}(h_1))$. Let $x \in \mathcal{C}(h_1)$. We have that $x = \mathcal{O}(\mathcal{O}^{-1}(x))$, where $\mathcal{O}^{-1}$ denotes the inverse orthogonal transformation. We shall see that $\mathcal{O}^{-1}(x) \in \mathcal{C}(h_1)$. First, $\mathcal{O}^{-1}(x) \in B(-r_n e_d, r_n)$ as

$$\left\|\mathcal{O}^{-1}(x) + r_n e_d\right\|^2 = \left\|\mathcal{O}_{-d}^{-1}(x_{-d})\right\|^2 + (x_d + r_n)^2 = \|x_{-d}\|^2 + (x_d + r_n)^2 = \|x + r_n e_d\|^2 \leq r_n^2.$$

We have used that $\mathcal{O}_{-d}^{-1}$, inverse of $\mathcal{O}_{-d}$, is also an orthogonal transformation. Moreover $-h_1 \leq \left\langle \mathcal{O}^{-1}(x), e_d \right\rangle \leq 0$ since $\left\langle \mathcal{O}^{-1}(x), e_d \right\rangle = x_d$ and $x \in \mathcal{C}(h_1)$. Then $\mathcal{O}^{-1}(x) \in \mathcal{C}(h_1)$ and the proof of (2.36) is complete.

Now, we see that

$$\mathcal{O}(Q_u) = Q_v. \tag{2.37}$$

Consider the vector $\mathcal{O}(x)$ with $x \in Q_u$. Then $\mathcal{O}(x) = (\mathcal{O}_{-d}(x_{-d}), x_d)$, where $x_{-d} = (x_1, \dots, x_{d-1}) \in C_u$. We have that

$$\langle \mathcal{O}_{-d}(x_{-d}), v \rangle = \langle \mathcal{O}_{-d}(x_{-d}), \mathcal{O}_{-d}(u) \rangle = \langle x_{-d}, u \rangle \geq \|x_{-d}\| \cos\frac{\pi}{6} = \|\mathcal{O}_{-d}(x_{-d})\| \cos\frac{\pi}{6},$$

and hence $\mathcal{O}(x) \in Q_v$. Let $x \in Q_v$. We can write $x = \mathcal{O}(\mathcal{O}^{-1}(x)) = (\mathcal{O}_{-d}^{-1}(x_{-d}), x_d)$ and

$$\left\langle \mathcal{O}_{-d}^{-1}(x_{-d}), u \right\rangle = \left\langle \mathcal{O}_{-d}^{-1}(x_{-d}), \mathcal{O}_{-d}^{-1}(v) \right\rangle = \langle x_{-d}, v \rangle \geq \|x_{-d}\| \cos\frac{\pi}{6} = \left\|\mathcal{O}_{-d}^{-1}(x_{-d})\right\| \cos\frac{\pi}{6}.$$

We have used that $\mathcal{O}_{-d}^{-1}$ is also an orthogonal transformation. Then $\mathcal{O}^{-1}(x) \in Q_u$ and hence $Q_v \subset \mathcal{O}(Q_u)$. This completes the proof of (2.37).

Finally, given $u, v \in \mathcal{W}$, we have shown that there is an orthogonal transformation $\mathcal{O}$ such that $\mathcal{O}(Q_u \cap \mathcal{C}(h_1)) = Q_v \cap \mathcal{C}(h_1)$. Since the measure remains invariant under orthogonal transformations we get that

$$\mu(Q_u \cap \mathcal{C}(h_1)) = \mu\left(\mathcal{O}\left(Q_u \cap \mathcal{C}(h_1)\right)\right) = \mu\left(Q_v \cap \mathcal{C}(h_1)\right).$$

□

Turning to equation (2.35) we can conclude that, for all $u \in \mathcal{W}$,

$$\mu\left(Q_u \cap \mathcal{C}(h_1)\right) \geq \frac{1}{m^{\mathcal{F}}} L r_n^{\frac{d-1}{2}} \rho^{\frac{d+1}{2}},$$

where $m^{\mathcal{F}}$ represents the number of elements that form $\mathcal{W}$. Lastly, by Lemma 2.4.10 it follows that

$$Q_u \cap \mathcal{C}(h_1) \subset \mathcal{C}(h_1) \subset B((\alpha-\rho)e_d, \alpha)$$

and

$$\mu\left(Q_u \cap \mathcal{C}(h_1) \cap B((\alpha-\rho)e_d,\alpha)\right) = \mu(Q_u \cap \mathcal{C}(h_1)) \geq L^{\mathcal{F}} r_n^{\frac{d-1}{2}} \rho^{\frac{d+1}{2}},$$

completing the proof of Proposition 2.4.4.

□

We finish the proof of Proposition 2.4.2 by defining the family

$$\mathcal{U}_{0,r_n} = \mathcal{U}^{\mathcal{G}}_{0,r_n} \cup \mathcal{U}^{\mathcal{F}}_{0,r_n}.$$

Then,

$$\mathcal{U}_{x,r_n} = \{T(U),\ U \in \mathcal{U}_{0,r_n}\}$$

is a finite family with $m_2 = m^{\mathcal{G}} + m^{\mathcal{F}}$ elements satisfying that, for each $U \in \mathcal{U}_{0,r_n}$,

$$P_X(T(U)) \geq \delta\mu(T(U) \cap B(P_\Gamma x - \alpha\eta, \alpha)) = \delta\mu(U \cap B((\alpha-\rho)e_d, \alpha)) \geq L_2 r_n^{\frac{d-1}{2}} \rho^{\frac{d+1}{2}},$$

where $L_2 = \delta \min(L^{\mathcal{G}}, L^{\mathcal{F}})$. □

2.5 Main results

The aim of this section is to present the achieved results on the consistency and convergence rate of the estimator S_n defined in (2.3). The concept of unavoidable family, discussed in Sections 2.3 and 2.4 for the particular case of $\mathbb{R}^2$ and the general case of $\mathbb{R}^d$, respectively, plays a major role in the development of this section. Propositions 2.4.1 and 2.4.2 and their counterparts in the bidimensional case are the key results in deriving the convergence rate of $\mathbb{E}(d_\mu(S, S_n))$, which is given in Theorem 2.5.2, the main result of this chapter. In Theorem 2.5.3 we show that the obtained convergence rate cannot be improved. Finally, some general ideas about unavoidable families will be particularly useful for proving the consistency of the estimator, established in Theorem 2.5.1, below.

Theorem 2.5.1. *Let $S \subset \mathbb{R}^d$ be a nonempty α-convex compact set with $\alpha > 0$. Let X be a random variable with probability distribution P_X and density f whose support is S. Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a random sample from X and let $\{r_n\}$ be a sequence of positive terms which do not depend on the sample such that $r_n \leq \alpha$. Then,*

$$\lim_{n\to\infty} \mathbb{E}(d_\mu(S, S_n)) = 0$$

if and only if $\lim_{n\to\infty} n r_n^d = \infty$.

Proof. Recall that, according to the definition of the estimator S_n in (2.3) and by (2.4),

$$\mathbb{E}(d_\mu(S,S_n)) = \mathbb{E}(\mu(S\setminus S_n)) = \int_S P(\exists y \in B(x,r_n) : B(y,r_n)\cap \mathcal{X}_n = \emptyset)\mu(dx).$$

Let us first assume that $\lim_{n\to\infty} nr_n^d = \infty$. We shall see that, for almost all $x\in\mathbb{R}$,

$$\lim_{n\to\infty} P(\exists y \in B(x,r_n) : B(y,r_n)\cap \mathcal{X}_n = \emptyset) = 0. \tag{2.38}$$

Note that if (2.38) holds, then by the dominated convergence theorem

$$\begin{aligned}\lim_{n\to\infty}\mathbb{E}(d_\mu(S,S_n)) &= \lim_{n\to\infty}\int_S P(\exists y \in B(x,r_n) : B(y,r_n)\cap \mathcal{X}_n = \emptyset)\mu(dx)\\ &= \int_S \lim_{n\to\infty} P(\exists y \in B(x,r_n) : B(y,r_n)\cap \mathcal{X}_n = \emptyset)\mu(dx)\\ &= 0. \end{aligned}\tag{2.39}$$

For each $x\in S$ let us consider the family $\mathcal{U}_{x,r_n} = \{U^u_{x,r_n}, u\in\mathcal{W}\}$, where $\mathcal{W}$ is the finite family of unit vectors given by Lemma 2.4.2 for $\theta=\pi/6$ and for each $u\in\mathcal{W}$, $U^u_{x,r_n} = \{x\}\oplus C_{u,r_n}$ is the translation of the set C_{u,r_n} by x. Then $\mathcal{U}_{x,r_n}$ is a finite unavoidable family for $\mathcal{E}_{x,r_n}$ as can be deduced from Lemmas 2.4.2, 2.4.3 and 2.4.4. Denote by m the number of sets of $\mathcal{U}_{x,r_n}$, which coincides with the number of unit vectors of $\mathcal{W}$. Then, using the same argument as in (2.6) we have that

$$P(\exists y \in B(x,r_n) : B(y,r_n)\cap \mathcal{X}_n = \emptyset) \le \sum_{u\in\mathcal{W}}(1-P_X(U^u_{x,r_n}))^n. \tag{2.40}$$

In order to give a lower bound for $P_X(U^u_{x,r_n})$ in (2.40) it will be useful following general version of the Lebesgue density theorem. See Devroye (1983) for the proof of the lemma.

Lemma 2.5.1 (Lebesgue density theorem, Devroye (1983))**.** *If f is a density in $\mathbb{R}^d$ and A is a compact set of $\mathbb{R}^d$ with $\mu(A)>0$, then*

$$\lim_{h\to 0}\frac{1}{\mu(hA)}\int_{\{x\}\oplus hA} f(y)dy = f(x), \text{ almost all } x.$$

Lemma 2.5.1 gives us the key to bounding the probability of small compact sets in a neighbourhood of the point x, from the value of the density in x and the Lebesgue measure of the set. Thus, let us consider the compact set $C_{u,1}$ and $h>0$. We have that

$$\{x\}\oplus hC_{u,1} = \{x\}\oplus C_{u,h} = U^u_{x,h}.$$

It follows from Lemma 2.5.1 that for almost all x, there exists h_x such that for all $h\le h_x$ we have

$$P_X(U^u_{x,h}) = \int_{U^u_{x,h}} f(y)dy \ge \frac{f(x)}{2}\mu(C_{u,h}). \tag{2.41}$$

For each $n\in\mathbb{N}$ let $h_n\equiv h_{n,x} = \min(r_n,h_x)$. Then $U^u_{x,h_n}\subset U^u_{x,r_n}$ and we can apply (2.41) to conclude that

$$P_X(U^u_{x,r_n}) \ge P_X(U^u_{x,h_n}) \ge \frac{f(x)}{2}\mu(C_{u,h_n}) \ge \frac{f(x)}{2}\frac{\mu(B(0,h_n))}{m} = \frac{f(x)}{2}\frac{w_d h_n^d}{m} = L_x h_n^d,$$

where, if $x \in S$, $L_x = \frac{f(x)w_d}{2m} > 0$. Returning to (2.40) we have

$$P(\exists y \in B(x,r_n) : B(y,r_n) \cap \mathcal{X}_n = \emptyset) \leq m(1 - L_x h_n^d)^n \leq m e^{-nL_x h_n^d}.$$

The last inequality follows from the fact that $(1-z)^n \leq \mathrm{e}^{-nz}$, for $z \in [0,1]$. Note that we can guarantee that $L_x h_n^d \leq 1$ since $L_x h_n^d \leq P_X(U^u_{x,r_n})$. Then

$$\lim_{n\to\infty} P(\exists y \in B(x,r_n) : B(y,r_n) \cap \mathcal{X}_n = \emptyset) \leq \lim_{n\to\infty} m e^{-nL_x h_n^d}.$$

Finally, the definition of h_n and the assumption $\lim_{n\to\infty} n r_n^d = \infty$ yield $\lim_{n\to\infty} n L_x h_n^d = \infty$. As a consequence,

$$\lim_{n\to\infty} P(\exists y \in B(x,r_n) : B(y,r_n) \cap \mathcal{X}_n = \emptyset) = 0, \text{ for almost all } x \in S,$$

which yields (2.39).

We now prove the converse assertion. Thus, let us assume that $\lim_{n\to\infty} \mathbb{E}(d_\mu(S,S_n)) = 0$. Note that

$$\begin{aligned} P(\exists y \in B(x,r_n) : B(y,r_n) \cap \mathcal{X}_n = \emptyset) &\geq P(B(x,r_n) \cap \mathcal{X}_n = \emptyset) \\ &= (1 - P_X(B(x,r_n)))^n. \end{aligned} \tag{2.42}$$

If the sequence $\{nr_n^d\}$ does not converge to infinity as $n \to \infty$, then we may find a bounded subsequence $\{n_k r_{n_k}^d\}$. Therefore, there exists $M > 0$ such that $n_k r_{n_k}^d \leq M$ for all n_k and as an immediate consequence $\lim_{k\to\infty} r_{n_k}^d = 0$. In this case Lemma 2.5.1 ensures that, for almost all x, for large enough k,

$$P_X(B(x,r_{n_k})) = \int_{B(x,r_{n_k})} f(y)dy \leq 2f(x)\mu(B(0,r_{n_k})) = 2f(x)w_d r_{n_k}^d = L_x r_{n_k}^d, \tag{2.43}$$

where now $L_x = 2f(x)w_d$. In order to simplify the notation let

$$\Psi_n(x) = P(\exists y \in B(x,r_n) : B(y,r_n) \cap \mathcal{X}_n = \emptyset)$$

and consider the subsequence $\{\Psi_{n_k}(x)\}$. We now combine (2.42) and (2.43) to get

$$\begin{aligned} \liminf_{k\to\infty} \Psi_{n_k}(x) &\geq \liminf_{k\to\infty} (1 - P_X(B(x,r_{n_k})))^{n_k} \\ &\geq \liminf_{k\to\infty} (1 - L_x r_{n_k}^d)^{n_k} \\ &\geq \liminf_{k\to\infty} \exp\left(\frac{-n_k L_x r_{n_k}^d}{1 - L_x r_{n_k}^d}\right) \\ &\geq e^{-L_x M}. \end{aligned} \tag{2.44}$$

We have used that $(1-z)^n \geq \exp(-nz/(1-z))$ for $z \in [0,1)$. The case when $z = 0$ is straightforward and for $z \in (0,1)$ write $(1-z)^n = \exp(n\log(1-z))$ and use the fact that

$\log(1-z) > -z/(1-z)$. The last inequality holds since $\lim_{k\to\infty} r_{n_k}^d = 0$ and $\{n_k r_{n_k}^d\}$ is bounded by M. By the Fatou's Lemma and (2.44) we obtain

$$\lim_{k\to\infty} \mathbb{E}(d_\mu(S, S_{n_k})) = \lim_{k\to\infty} \int_S \Psi_{n_k}(x)\mu(dx)$$
$$= \liminf_{k\to\infty} \int_S \Psi_{n_k}(x)\mu(dx) \geq \int_S \liminf_{k\to\infty} \Psi_{n_k}(x)\mu(dx) > 0,$$

which is a contradiction since we are assuming that $\lim_{n\to\infty} \mathbb{E}(d_\mu(S, S_n)) = 0$ and hence every subsequence of $\mathbb{E}(d_\mu(S, S_n))$ must also converge to zero. So, the sequence $\{nr_n^d\}$ must converge to infinity and this concludes the proof of the theorem.

□

Remark 2.5.1. *By definition, $d_\mu(S, S_n) = \mu(S \setminus S_n) + \mu(S_n \setminus S)$. The α-convexity assumption of Theorem 2.5.1 ensures that $S_n \subset S$ and, therefore, $\mu(S_n \setminus S) = 0$. Anyway, if the set S is not assumed to be α-convex, a similar consistency result can be stated under an extra condition on the parameter r_n. It can be proved that, if $\{r_n\}$ is a sequence of positive terms such that $\lim_{n\to\infty} r_n = 0$ and $\lim_{n\to\infty} nr_n^d = \infty$, then $\lim_{n\to\infty} \mathbb{E}(d_\mu(S, S_n)) = 0$. Without going into details, the proof follows easily from*

$$\mathbb{E}(d_\mu(S, S_n)) = \mathbb{E}(\mu(S \setminus S_n)) + \mathbb{E}(\mu(S_n \setminus S)). \tag{2.45}$$

Note that the first term in the right-hand side of (2.45) was studied in Theorem 2.5.1 and that the α-convexity assumption is not needed to guarantee that $\lim_{n\to\infty} \mathbb{E}(\mu(S \setminus S_n)) = 0$ for a compact set S. For the second term in the right-hand side of (2.45) we have $\mathbb{E}(\mu(S_n \setminus S)) \leq \mu(S \oplus r_n B) - \mu(S)$ since, with probability one, $S_n \subset (S \oplus r_n B)$. The Lebesgue dominated convergence theorem ensures that $\lim_{n\to\infty} \mu(S \oplus r_n B) = \mu(S)$ if $\lim_{n\to\infty} r_n = 0$.

Having obtained the consistency of the estimator, we now focus on the convergence rate of $\mathbb{E}(d_\mu(S, S_n))$. As mentioned in Chapter 1, Rodríguez-Casal (2007) obtains, under similar conditions on S, the almost sure convergence rate of $d_\mu(S, S_n)$. A more detail comparison of these results is given in Remark 2.5.2, after the statement Theorem 2.5.2, below.

Theorem 2.5.2. *Let S be a nonempty compact subset of $\mathbb{R}^d$ such that a ball of radius $\alpha > 0$ rolls freely in S and in $\overline{S^c}$. Let X be a random variable with probability distribution P_X and support S. We assume that the probability distribution P_X satisfies that there exists $\delta > 0$ such that $P_X(C) \geq \delta\mu(C \cap S)$ for all Borel subset $C \subset \mathbb{R}^d$. Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a random sample from X and let $\{r_n\}$ be a sequence of positive numbers which do not depend on the sample such that $r_n \leq \alpha$. If the sequence $\{r_n\}$ satisfies*

$$\lim_{n\to\infty} \frac{nr_n^d}{\log n} = \infty, \tag{2.46}$$

then

$$\mathbb{E}(d_\mu(S, S_n)) = O\left(r_n^{-\frac{d-1}{d+1}} n^{-\frac{2}{d+1}}\right). \tag{2.47}$$

Remark 2.5.2. *Rodríguez-Casal (2007) proves that, if S is under the conditions of Theorem 1.4.1 and $\{r_n\}$ is a sequence of positive numbers satisfying (2.46), then $d_\mu(S, S_n) = O(r_n^{-1}(\log n/n)^{2/(d+1)})$, almost surely. The convergence rate of $\mathbb{E}(d_\mu(S, S_n))$ obtained in Theorem 2.5.2 is, therefore, faster than the almost sure convergence rate of $d_\mu(S, S_n)$. Note that the logarithmic term vanishes in (2.47). Moreover, the penalty factor $r_n^{-(d-1)/(d+1)}$ is asymptotically smaller than r_n^{-1}.*

Proof. Recall that, if we define for each $x \in S$ a family $\mathcal{U}_{x,r_n}$ unavoidable and finite for $\mathcal{E}_{x,r_n}$, then

$$\begin{aligned}
\mathbb{E}(d_\mu(S, S_n)) &= \int_S P(\exists y \in B(x, r_n) : B(y, r_n) \cap \mathcal{X}_n = \emptyset)\mu(dx) \\
&\leq \int_S \sum_{U \in \mathcal{U}_{x,r_n}} (1 - P_X(U))^n \mu(dx) \\
&\leq \int_S \sum_{U \in \mathcal{U}_{x,r_n}} \exp(-nP_X(U))\mu(dx).
\end{aligned}$$

The last inequality follows by applying that $(1-z)^n \leq \mathrm{e}^{-nz}$, for $z \in [0, 1]$. We divide S into two subsets

$$S = \left\{x \in S : \ d(x, \partial S) > \frac{r_n}{2}\right\} \cup \left\{x \in S : \ d(x, \partial S) \leq \frac{r_n}{2}\right\}$$

and then

$$\begin{aligned}
\mathbb{E}(d_\mu(S, S_n)) &\leq \int_S \sum_{U \in \mathcal{U}_{x,r_n}} \exp(-nP_X(U))\mu(dx) \\
&= \int_{\left\{x\in S:\ d(x,\partial S) > \frac{r_n}{2}\right\}} \sum_{U \in \mathcal{U}_{x,r_n}} \exp(-nP_X(U))\mu(dx) \\
&+ \int_{\left\{x\in S:\ d(x,\partial S) \leq \frac{r_n}{2}\right\}} \sum_{U \in \mathcal{U}_{x,r_n}} \exp(-nP_X(U))\mu(dx). \qquad (2.48)
\end{aligned}$$

For those $x \in S$ such that $d(x, \partial S) > r_n/2$ we make use of the families $\mathcal{U}_{x,r_n}$ given in Proposition 2.4.1. Recall that Proposition 2.4.1 ensures the existence of suitable finite families $\mathcal{U}_{x,r_n}$ and provides a lower bound on the probability of the sets U, independent of x. Thus,

$$\begin{aligned}
&\int_{\left\{x\in S:\ d(x,\partial S) > \frac{r_n}{2}\right\}} \sum_{U \in \mathcal{U}_{x,r_n}} \exp(-nP_X(U))\mu(dx) \\
&\leq \int_{\left\{x\in S:\ d(x,\partial S) > \frac{r_n}{2}\right\}} m_1 \exp(-nL_1 r_n^d)\mu(dx) \\
&= O\left(\mathrm{e}^{-L_1 n r_n^d}\right), \qquad (2.49)
\end{aligned}$$

where m_1 denotes the finite number of elements of $\mathcal{U}_{x,r_n}$. Note that m_1 is also independent of x. Now, for those $x \in S$ such that $d(x, \partial S) \leq r_n/2$, we may consider the unavoidable families $\mathcal{U}_{x,r_n}$ given in Proposition 2.4.2. Let m_2 be the number of elements of $\mathcal{U}_{x,r_n}$. We have that

$$\begin{aligned}
&\int_{\left\{x\in S:\ d(x,\partial S)\leq \frac{r_n}{2}\right\}} \sum_{U\in\mathcal{U}_{x,r_n}} \exp(-nP_X(U))\mu(dx) \\
&\leq \int_{\left\{x\in S:\ d(x,\partial S)\leq \frac{r_n}{2}\right\}} m_2 \exp\left(-L_2 n r_n^{\frac{d-1}{2}} d(x,\partial S)^{\frac{d+1}{2}}\right)\mu(dx) \\
&= \int_{\mathcal{T}^{-1}([0,r_n/2])} g(\mathcal{T}(x))\mu(dx),
\end{aligned}$$

where $\mathcal{T} : S \to \mathbb{R}$ is defined as $\mathcal{T}(x) = d(x, \partial S)$ and $g(z) = m_2 \exp(-L_2 n r_n^{\frac{d-1}{2}} z^{\frac{d+1}{2}})$. It follows from the change of variables formula (see Theorem 16.12 of Billingsley (1995)) that

$$\int_{\mathcal{T}^{-1}([0,r_n/2])} g(\mathcal{T}(x))\mu(dx) = \int_{[0,r_n/2]} g(\rho)\mu\mathcal{T}^{-1}(d\rho)$$

where $\rho = \mathcal{T}(x)$ and $\mu\mathcal{T}^{-1}$ is the measure on $\mathbb{R}$ defined by

$$\mu\mathcal{T}^{-1}(A) = \mu(\mathcal{T}^{-1}(A)),$$

for $A \subset \mathbb{R}$. The measure $\mu\mathcal{T}^{-1}$ is characterized by

$$F(z) = \mu\{x \in S :\ d(x, \partial S) \leq z\}.$$

Under the stated conditions it can be proved that, for $0 \leq z < \alpha$, $F(z)$ is a polynomial of degree at most d in z, see Federer (1959). Therefore, it is a differentiable function and $F'(z)$ is bounded on compact sets. In short, we obtain

$$\begin{aligned}
&\int_{[0,r_n/2]} g(\rho)\mu\mathcal{T}^{-1}(d\rho) \\
&= \int_{[0,r_n/2]} m_2 \exp\left(-L_2 n r_n^{\frac{d-1}{2}} \rho^{\frac{d+1}{2}}\right) F'(\rho)d\rho \\
&\leq K \int_0^{\frac{r_n}{2}} m_2 \exp\left(-L_2 n r_n^{\frac{d-1}{2}} \rho^{\frac{d+1}{2}}\right) d\rho \\
&= K \int_0^{\frac{L_2 n}{2^{(d+1)/2}} r_n^d} m_2 \frac{1}{\frac{d+1}{2} L_2^{2/(d+1)}} r_n^{-\frac{d-1}{d+1}} n^{-\frac{2}{d+1}} \mathrm{e}^{-v} v^{\frac{1-d}{d+1}} dv \\
&= O\left(r_n^{-\frac{d-1}{d+1}} n^{-\frac{2}{d+1}}\right),
\end{aligned} \tag{2.50}$$

where we have used the change of variables formula $v = L_2 n r_n^{\frac{d-1}{2}} \rho^{\frac{d+1}{2}}$ and also the fact that $\int_0^\infty \mathrm{e}^{-v} v^{\frac{1-d}{d+1}} dv < \infty$. Turning to the computation of $\mathbb{E}(d_\mu(S, S_n))$ in (2.48), it follows from (2.49) and (2.50) that

$$\mathbb{E}(d_\mu(S, S_n)) = O\left(\mathrm{e}^{-L_1 n r_n^d} + r_n^{-\frac{d-1}{d+1}} n^{-\frac{2}{d+1}}\right). \tag{2.51}$$

Now if (2.46) holds, then for all $M > 0$ there exists $N \in \mathbb{N}$ such that

$$nr_n^d \geq M \log n,$$

for all $n \geq N$ and hence

$$\mathrm{e}^{-L_1 n r_n^d} \leq \mathrm{e}^{-L_1 M \log n} = n^{-L_1 M}.$$

As a consequence

$$\limsup_{n\to\infty} \frac{\mathrm{e}^{-L_1 n r_n^d}}{r_n^{-\frac{d-1}{d+1}} n^{-\frac{2}{d+1}}} \leq \limsup_{n\to\infty} \frac{n^{-L_1 M}}{r_n^{-\frac{d-1}{d+1}} n^{-\frac{2}{d+1}}} = \limsup_{n\to\infty} r_n^{\frac{d-1}{d+1}} n^{\left(\frac{2}{d+1} - L_1 M\right)} = 0, \tag{2.52}$$

for large enough M. Remember that r_n is bounded ($r_n \leq \alpha$ by assumption). We now combine (2.51) and (2.52) to obtain

$$\mathbb{E}(d_\mu(S, S_n)) = O\left(\mathrm{e}^{-L_1 n r_n^d} + r_n^{-\frac{d-1}{d+1}} n^{-\frac{2}{d+1}}\right) = O\left(r_n^{-\frac{d-1}{d+1}} n^{-\frac{2}{d+1}}\right),$$

which completes the proof. □

Finally, next lemma shows that the obtained rate in Theorem 2.5.2 cannot be improved since there exist sets under the stated conditions for which

$$\liminf_{n\to\infty} r_n^{\frac{d-1}{d+1}} n^{\frac{2}{d+1}} \mathbb{E}(d_\mu(S, S_n)) > 0.$$

Theorem 2.5.3. *Under the conditions of Theorem 2.5.2, there exist sets S for which*

$$\liminf_{n\to\infty} r_n^{\frac{d-1}{d+1}} n^{\frac{2}{d+1}} \mathbb{E}(d_\mu(S, S_n)) > 0.$$

Proof. Let $S = B(0, \alpha)$ and assume that the distribution P_X is uniform on S. Our aim is to find a lower bound for $\mathbb{E}(d_\mu(S, S_n))$. Thus,

$$\begin{aligned}
\mathbb{E}(d_\mu(S, S_n)) &= \int_S P(\exists y \in B(x, r_n) : B(y, r_n) \cap \mathcal{X}_n = \emptyset) \mu(dx) \\
&\geq \int_{\left\{x \in S:\ d(x, \partial S) \leq \frac{r_n}{2}\right\}} P(\exists y \in B(x, r_n) : B(y, r_n) \cap \mathcal{X}_n = \emptyset) \mu(dx).
\end{aligned}$$

For each $x \in S$ such that $d(x, \partial S) \leq r_n/2$ let $\eta = x/\|x\|$ and

$$\tilde{x} = (\alpha + r_n - d(x, \partial S))\eta = (\|x\| + r_n)\eta. \tag{2.53}$$

In Figure 2.26 we show an example of the definition of $\tilde{x}$ in the particular case of $\mathbb{R}^2$. Note that $\tilde{x} \in B(x, r_n)$ and hence

$$P(\exists y \in B(x, r_n) : B(y, r_n) \cap \mathcal{X}_n = \emptyset) \geq P(B(\tilde{x}, r_n) \cap \mathcal{X}_n = \emptyset) = (1 - P_X(B(\tilde{x}, r_n)))^n.$$

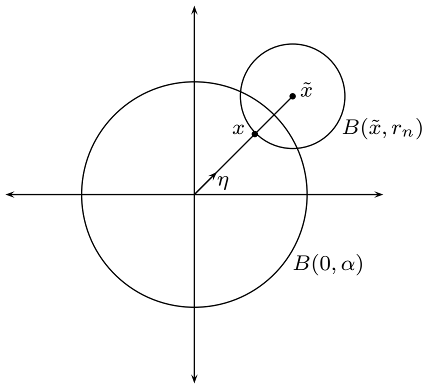


Figure 2.26: *Given* $x \in B(0,\alpha)$ *such that* $d(x,\partial S) \le r_n/2$*, we define* $\tilde{x} = (\alpha + r_n - d(x,\partial S))\eta$*.*

In short,

$$\mathbb{E}(d_\mu(S,S_n)) \ge \int_{\{x\in S:\ d(x,\partial S)\le \frac{r_n}{2}\}} (1 - P_X(B(\tilde{x},r_n)))^n \mu(dx), \tag{2.54}$$

where $\tilde{x}$ is given by (2.53). First we shall see that $P_X(B(\tilde{x},r_n)) \le 1/2$. Remember that, under the assumption of the uniform distribution on S, we have

$$P_X(B(\tilde{x},r_n)) = \frac{\mu(B(\tilde{x},r_n)\cap S)}{\mu(S)}. \tag{2.55}$$

Let us consider an orthogonal transformation $\mathcal{O} : \mathbb{R}^d \to \mathbb{R}^d$ such that $\mathcal{O}(\eta) = -e_d$. Then

$$\mathcal{O}(B(\tilde{x},r_n)\cap S) = B(-(\alpha + r_n - d(x,\partial S))e_d, r_n) \cap B(0,\alpha).$$

It is easy to see that

$$B(-(\alpha + r_n - d(x,\partial S))e_d, r_n) \subset \{z \in \mathbb{R}^d :\ \langle z, e_d\rangle \le 0\}$$

and, since the Lebesgue measure is invariant under orthogonal transformations, we have

$$\begin{aligned} \mu(B(\tilde{x},r_n)\cap S) &= \mu(B(-(\alpha + r_n - d(x,\partial S))e_d, r_n) \cap B(0,\alpha)) \\ &\le \mu(\{z \in \mathbb{R}^d :\ \langle z, e_d\rangle \le 0\} \cap B(0,\alpha)) \\ &= \frac{1}{2}\mu(B(0,\alpha)). \end{aligned} \tag{2.56}$$

Combine (2.55) and (2.56) to get

$$P_X(B(\tilde{x},r_n)) \le \frac{1}{2}. \tag{2.57}$$

We return to (2.54) to obtain that

$$\begin{aligned}\mathbb{E}(d_\mu(S,S_n)) &\geq \int_{\{x\in S:\ d(x,\partial S)\leq \frac{r_n}{2}\}} (1-P_X(B(\tilde{x},r_n)))^n\mu(dx)\\ &\geq \int_{\{x\in S: d(x,\partial S)\leq r_n/2\}} \exp\left(\frac{-nP_X(B(\tilde{x},r_n))}{1-P_X(B(\tilde{x},r_n))}\right)\mu(dx)\\ &\geq \int_{\{x\in S: d(x,\partial S)\leq r_n/2\}} \exp\left(-2nP_X(B(\tilde{x},r_n))\right)\mu(dx). \qquad (2.58)\end{aligned}$$

We have used again the fact that $(1-z)^n \geq \exp(-nz/(1-z))$ for $z\in[0,1)$ together with (2.57). In view of (2.58) we need again an upper bound for $P_X(B(\tilde{x},r_n))$. The bound in (2.57) will be now too rough for our purposes and so we shall see that it can be sharpened. Let us now consider the composed function formed by first applying the previous orthogonal transformation $\mathcal{O}:\mathbb{R}^d\rightarrow\mathbb{R}^d$ such that $\mathcal{O}(\eta)=-e_d$ and then applying the translation by the vector $(\alpha-d(x,\partial S))e_d$, see Figure 2.27. Using again that the Lebesgue measure is invariant under orthogonal transformations and translations we have that

$$\mu(B(B(\tilde{x},r_n)\cap S)) = \mu(B(-r_ne_d,r_n)\cap B((\alpha-d(x,\partial S))e_d,\alpha)).$$

The set $B(-r_ne_d,r_n)\cap B((\alpha-d(x,\partial S))e_d,\alpha)$ is the intersection of two balls with radius r_n

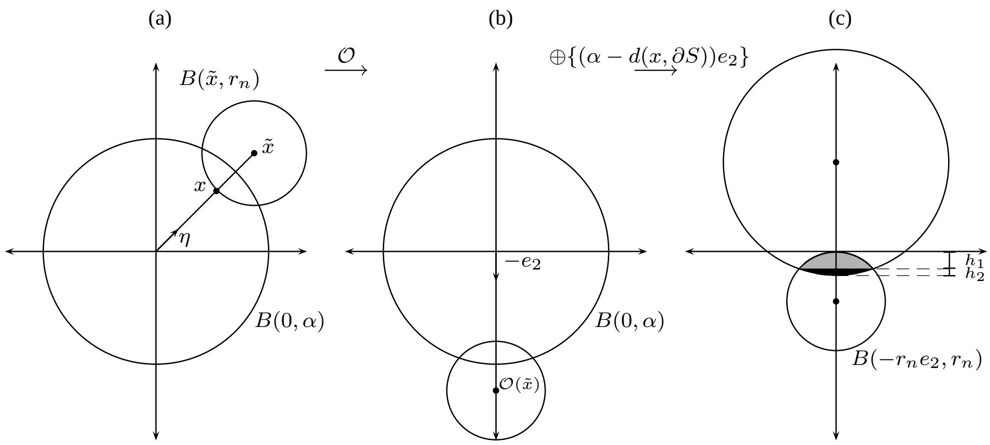


Figure 2.27: *(a)* $B(\tilde{x},r_n)\cap S$. *(b) Result of applying an orthogonal transformation* $\mathcal{O}:\mathbb{R}^2\rightarrow\mathbb{R}^2$ *such that* $\mathcal{O}(\eta)=-e_2$. *(c) Translation by the vector* $(\alpha-d(x,\partial S))e_2$. *In black* $\mathcal{A}(h_2)$ *and in gray* $\mathcal{C}(h_1)$.

and α such that the distance between their centres is equal to $\alpha+r_n-d(x,\partial S)$. Recall that this set appeared for the first time in Proposition 2.4.4. Following the notation used previously,

$$B(-r_ne_d,r_n)\cap B((\alpha-d(x,\partial S))e_d,\alpha) = \mathcal{C}(h_1)\cup\mathcal{A}(h_2),$$

where $\mathcal{C}(h_1)$ is given by (2.29) and

$$\mathcal{A}(h_2) = \{z\in\mathbb{R}^d: -(h_1+h_2)\leq\langle z,e_d\rangle\leq -h_1\}\cap B((\alpha-d(x,\partial S))e_d,\alpha).$$

Recall that the values of h_1 and h_2 were easily deduced from the Pythagorean theorem by solving the system

$$\begin{cases} (r_n - h_1)^2 + \lambda^2 = r_n^2, \\ (\alpha - h_2)^2 + \lambda^2 = \alpha^2, \\ h_1 + h_2 = d(x, \partial S). \end{cases}$$

Thus,

$$h_1 = \frac{d(x,\partial S)(2\alpha - d(x,\partial S))}{2(\alpha + r_n - d(x,\partial S))}, \quad h_2 = \frac{d(x,\partial S)(2r_n - d(x,\partial S))}{2(\alpha + r_n - d(x,\partial S))}.$$

Since $\mathcal{C}(h_1)$ and $\mathcal{A}(h_2)$ are disjoint, up to a zero measure set, we have

$$\mu(B(-r_n e_d, r_n) \cap B((\alpha - d(x,\partial S))e_d, \alpha)) = \mu(\mathcal{C}(h_1)) + \mu(\mathcal{A}(h_2)). \tag{2.59}$$

First, in order to find an upper bound in (2.59), we shall see that $\mu(\mathcal{A}(h_2)) \leq \mu(\mathcal{C}(h_1))$. It can be easily proved that $\mu(\mathcal{A}(h_2)) = \mu(\mathcal{A}_0(h_2))$, where

$$\mathcal{A}_0(h_2) = \{z = (z_1, \ldots, z_d) \in \mathbb{R}^d : \ 0 \leq \langle z, e_d \rangle \leq h_2\} \cap B(-(\alpha - h_2)e_d, \alpha).$$

Note that $\mathcal{A}_0(h_2)$ is obtained after applying an orthogonal transformation and a translation to $\mathcal{A}(h_2)$. Using a similar argument as in the proof of Lemma 2.4.9 let $0 \leq l \leq h_2$ and define the set

$$\mathcal{A}_0(h_2, l) = \{z = (z_1, \ldots, l) \in \mathbb{R}^d : \ z \in \mathcal{A}_0(h_2)\}.$$

Then

$$\mu(\mathcal{A}_0(h_2)) = \int_0^{h_2} \mu_{d-1}(\mathcal{A}_0(h_2, l))dl$$

where μ_{d-1} denotes the $(d-1)$-dimensional Lebesgue measure and $\mathcal{A}_0(h_2, l)$ refers to the $(d-1)$-dimensional sphere with centre le_d and radius $s(l)$, being

$$s(l) = \sqrt{\alpha^2 - (\alpha - h_2 + l)^2}.$$

Therefore,

$$\mu(\mathcal{A}(h_2)) = \omega_{d-1} \int_0^{h_2} s(l)^{d-1} dl. \tag{2.60}$$

Recall from Lemma 2.4.9 that

$$\mu(\mathcal{C}(h_1)) = \omega_{d-1} \int_0^{h_1} r(l)^{d-1} dl, \tag{2.61}$$

where $r(l) = \sqrt{r_n^2 - (r_n - h_1 + l)^2}$, for $0 \leq l \leq h_1$. In view of (2.60) and (2.61) and since $h_2 \leq h_1$, if we are able to prove that $s(l) \leq r(l)$ for $0 \leq l \leq h_2$, then

$$\mu(\mathcal{A}(h_2)) = \omega_{d-1} \int_0^{h_2} s(l)^{d-1} dl \leq \omega_{d-1} \int_0^{h_2} r(l)^{d-1} dl \leq \omega_{d-1} \int_0^{h_1} r(l)^{d-1} dl = \mu(\mathcal{C}(h_1)).$$

As $r(l) \geq 0$ and $s(l) \geq 0$ it suffices to show that $s(l)^2 \leq r(l)^2$ or, equivalently, $r(l)^2 - s(l)^2 \geq 0$. By construction $r(0)^2 = s(0)^2 = \lambda^2$. and an easy computation shows that $r(l)^2 - s(l)^2$ is an increasing function. Indeed,

$$r(l)^2 - s(l)^2 = 2l(\alpha - r_n + h_1 - h_2) + (h_2^2 - h_1^2 + 2r_n h_1 - 2\alpha h_2) \tag{2.62}$$

and the derivative of (2.62) with respect to l satisfies

$$2(\alpha - r_n + h_1 - h_2) \geq 0,$$

since $r_n \leq \alpha$ and $h_2 \leq h_1$. Therefore $s(l) \leq r(l)$ for $0 \leq l \leq h_2$ and $\mu(\mathcal{A}(h_2)) \leq \mu(\mathcal{C}(h_1))$. Now. if we return to the equation (2.59), we get

$$\mu(B(\tilde{x}, r_n) \cap S) \leq 2\mu(\mathcal{C}(h_1)). \tag{2.63}$$

We will thus concentrate on $\mathcal{C}(h_1)$. Lemma 2.4.9 provided a lower bound for the measure of $\mathcal{C}(h_1)$. However, we now need an upper bound for $\mu(\mathcal{C}(h_1))$. Proceed as in the proof of Lemma 2.4.9 to get

$$\mu(\mathcal{C}(h_1)) \quad = \quad \omega_{d-1} \int_0^{h_1} \left(2r_n t - t^2\right)^{\frac{d-1}{2}} dt,$$

see (2.30). It is immediate that $2r_n t - t^2 \leq 2r_n t$, for $0 \leq t \leq h_1$ and hence

$$\mu(\mathcal{C}(h_1)) \leq \omega_{d-1} \int_0^{h_1} (2r_n t)^{\frac{d-1}{2}} dt = \frac{\omega_{d-1}}{d+1} 2^{\frac{d+1}{2}} r_n^{\frac{d-1}{2}} h_1^{\frac{d+1}{2}}.$$

Since $h_1 \leq d(x, \partial S)$, we have

$$\mu(\mathcal{C}(h_1)) \leq \frac{\omega_{d-1}}{d+1} 2^{\frac{d+1}{2}} r_n^{\frac{d-1}{2}} d(x, \partial S)^{\frac{d+1}{2}}. \tag{2.64}$$

Combine (2.63) and (2.64) to obtain

$$\mu(B(\tilde{x}, r_n) \cap S) \leq \frac{\omega_{d-1}}{d+1} 2^{\frac{d+3}{2}} r_n^{\frac{d-1}{2}} d(x, \partial S)^{\frac{d+1}{2}}.$$

As a consequence,

$$P_X(B(\tilde{x}, r_n)) \leq \frac{1}{\mu(S)} \frac{\omega_{d-1}}{d+1} 2^{\frac{d+3}{2}} r_n^{\frac{d-1}{2}} d(x, \partial S)^{\frac{d+1}{2}} = L r_n^{\frac{d-1}{2}} d(x, \partial S)^{\frac{d+1}{2}}.$$

Finally, if we apply the latter bound to (2.58), then we have that

$$\begin{aligned} \mathbb{E}(d_\mu(S, S_n)) \quad &\geq \quad \int_{\{x \in S : d(x, \partial S) \leq r_n/2\}} \exp\left(-2nL r_n^{\frac{d-1}{2}} d(x, \partial S)^{\frac{d+1}{2}}\right) \mu(dx) \\ &= \quad \int_{\mathcal{T}^{-1}([0, r_n/2])} g(\mathcal{T}(x)) \mu(dx), \end{aligned}$$

where $\mathcal{T}: S \to \mathbb{R}$ is defined as $\mathcal{T}(x) = d(x, \partial S)$ and $g(z) = \exp(-2nLr_n^{\frac{d-1}{2}} z^{\frac{d+1}{2}})$. By the change of variables formula (see Theorem 16.12 of Billingsley (1995))

$$\int_{\mathcal{T}^{-1}([0,r_n/2])} g(\mathcal{T}(x))\mu(dx) = \int_{[0,r_n/2]} g(\rho)\mu\mathcal{T}^{-1}(d\rho)$$

where $\rho = \mathcal{T}(x)$ and $\mu\mathcal{T}^{-1}$ is the measure on $\mathbb{R}$ defined by

$$\mu\mathcal{T}^{-1}(A) = \mu(\mathcal{T}^{-1}(A)),$$

for $A \subset \mathbb{R}$. The measure $\mu\mathcal{T}^{-1}$ is characterized by

$$F(z) = \mu\{x \in S : \ d(x, \partial S) \leq z\}.$$

We know from Federer (1959) that $F(z)$ is a polynomial of degree at most d in z. In fact, in this particular case, for $z < \alpha$, $F(z) = \omega_d(\alpha^d - (\alpha - z)^d)$. Therefore F is differentiable and

$$\begin{aligned}
\mathbb{E}(d_\mu(S, S_n)) &\geq \int_{[0,r_n/2]} g(\rho)\mu\mathcal{T}^{-1}(d\rho) \\
&= \int_0^{r_n/2} \exp\left(-2nLr_n^{\frac{d-1}{2}}\rho^{\frac{d+1}{2}}\right) F'(\rho)d\rho \\
&= \int_0^{r_n/2} \exp\left(-2nLr_n^{\frac{d-1}{2}}\rho^{\frac{d+1}{2}}\right) \omega_d d(\alpha - \rho)^{d-1}d\rho.
\end{aligned}$$

It is immediate to show that for $0 \leq \rho \leq r_n/2$ the function $F'(\rho) = \omega_d d(\alpha-\rho)^{d-1}$ is decreasing with $F'(\rho) \geq F'(r_n/2) = \omega_d d(\alpha - r_n/2)^{d-1} \geq \omega_d d(\alpha/2)^{d-1}$. Therefore

$$\begin{aligned}
\mathbb{E}(d_\mu(S, S_n)) &\geq \int_0^{r_n/2} \exp\left(-2nLr_n^{\frac{d-1}{2}}\rho^{\frac{d+1}{2}}\right) \omega_d d\left(\frac{\alpha}{2}\right)^{d-1} d\rho \\
&= \omega_d d\left(\frac{\alpha}{2}\right)^{d-1} \int_0^{\frac{2nL}{2^{(d+1)/2}}r_n^d} \frac{1}{\frac{d+1}{2}(2L)^{2/(d+1)}} r_n^{-\frac{d-1}{d+1}} n^{-\frac{2}{d+1}} e^{-v} v^{\frac{1-d}{d+1}} dv \\
&= \omega_d d\left(\frac{\alpha}{2}\right)^{d-1} \frac{1}{\frac{d+1}{2}(2L)^{2/(d+1)}} r_n^{-\frac{d-1}{d+1}} n^{-\frac{2}{d+1}} \int_0^{\frac{2nL}{2^{(d+1)/2}}r_n^d} e^{-v} v^{\frac{1-d}{d+1}} dv.
\end{aligned}$$

We have used the change of variables formula with $v = 2nLr_n^{\frac{d-1}{2}}\rho^{\frac{d+1}{2}}$. Therefore

$$\liminf_{n\to\infty} r_n^{\frac{d-1}{d+1}} n^{\frac{2}{d+1}} \mathbb{E}(d_\mu(S, S_n)) \geq \liminf_{n\to\infty} \frac{\omega_d d\,(\alpha/2)^{d-1}}{\frac{d+1}{2}(2L)^{2/(d+1)}} \int_0^{\frac{2nL}{2^{(d+1)/2}}r_n^d} e^{-v} v^{\frac{1-d}{d+1}} dv.$$

Since $nr_n^d \to \infty$, we have

$$\liminf_{n\to\infty} r_n^{\frac{d-1}{d+1}} n^{\frac{2}{d+1}} \mathbb{E}(d_\mu(S, S_n)) \geq \frac{\omega_d d\,(\alpha/2)^{d-1}}{\frac{d+1}{2}(2L)^{2/(d+1)}} \int_0^{\infty} e^{-v} v^{\frac{1-d}{d+1}} dv > 0.$$

This completes the proof of the theorem. □

Remark 2.5.3. *We conjecture that*

$$\liminf_{n\to\infty} r_n^{\frac{d-1}{d+1}} n^{\frac{2}{d+1}} \mathbb{E}(d_\mu(S, S_n)) > 0$$

for any set S under the conditions of Theorem 2.5.2. The proof relies on the following "local convexity" property, which we think S fulfills. We say that S is "locally convex" in $B(s,\tau)\cap\partial S$ for $s\in\partial S$ and $\tau>0$ if there exists $\varepsilon>0$ such that for all $t\in B(s,\tau)\cap\partial S$, the set $B(t,\varepsilon)\cap S$ is contained in the halfspace $\{x\in\mathbb{R}^d : \langle x-t,\eta(t)\rangle \le 0\}$. Note that this local convexity property holds for any $s\in\partial S$, $\tau>0$, and $\varepsilon>0$ when S is a ball of radius α as in Theorem 2.5.3.

Chapter 3

Surface area estimation

3.1 Introduction

The surface area estimation problem was briefly introduced in Chapter 1. In this chapter we propose an in-depth study of a new estimator for the surface area, based on the notion of Minkowski content and on the α-convexity assumption. We have structured this chapter as follows. In Section 3.2 we introduce the estimator L_n along with a brief discussion of the sampling model and the assumptions. In Section 3.3 the asymptotic behaviour of the proposed estimator is analysed. More precisely, the almost sure convergence rate and the L_1-convergence rate are provided in Subsections 3.3.1 and 3.3.2, respectively. The results in Chapter 2 will be useful in order to derive the L_1-convergence rate. The results in Subsection 3.3.1 can be also found in Pateiro-López and Rodríguez-Casal (2008), accepted for its publication in Advances in Applied Probability.

3.2 The sampling model and the estimator

As has been argued, the notion of Minkowski content serves us as starting point for defining a suitable surface area estimator. The assumptions of the model are motivated by its definition. Thus, let G be a nonempty compact set in $\mathbb{R}^d$ and assume, without lost of generality, that $G \subset (0,1)^d$. The Minkowski content of G, recall Definition 1.5.4, is given by

$$L_0 \equiv L_0(G) = \lim_{\varepsilon \to 0} \frac{\mu(B(\partial G, \varepsilon))}{2\varepsilon} = \lim_{\varepsilon \to 0} L(\varepsilon), \tag{3.1}$$

provided that this limit exists and is finite, being

$$L(\varepsilon) = \frac{\mu(B(\partial G, \varepsilon))}{2\varepsilon}. \tag{3.2}$$

Note that $B(\partial G, \varepsilon)$ in (3.2) represents the closed ε-neighbourhood of the boundary ∂G and that, in order to estimate $B(\partial G, \varepsilon)$, it would be desirable to have information from both G and $R = [0,1]^d \setminus \text{int}(G)$ since ∂G is somewhere in between points of the set and points of its complement. For this reason the sampling information is assumed to be given by i.i.d. observations

$(Z_1, \xi_1), \ldots, (Z_n, \xi_n)$ of a random variable (Z, ξ), where Z is uniformly distributed on the unit square $[0,1]^d$ and $\xi = \mathbb{I}_{\{Z \in G\}}$. Let us denote by P_X and P_Y the conditional distributions of the observations in G and in R, that is, the distributions of $X = Z|\{\xi = 1\}$ and $Y = Z|\{\xi = 0\}$, respectively. It is not difficult to prove that P_X and P_Y are both uniform on G and R, respectively. Let $\{\varepsilon_n\}$ be a deterministic sequence of positive numbers which converges to zero as n tends to infinity. We propose to estimate L_0 by means of

$$L_n = \frac{\mu(\Gamma_n)}{2\varepsilon_n}, \tag{3.3}$$

being Γ_n an estimator of $B(\partial G, \varepsilon_n)$. We saw in Chapter 1 that the problem of estimating L_0 can be tackled as a problem of set estimation, since, assuming the mild condition $\overline{\mathrm{int}(G)} = G$, $B(\partial G, \varepsilon_n)$ can be written as the intersection $B(G, \varepsilon_n) \cap B(R, \varepsilon_n)$. Thus, if G_n and R_n estimate G and R, respectively, then

$$\Gamma_n = B(G_n, \varepsilon_n) \cap B(R_n, \varepsilon_n) \tag{3.4}$$

estimates $B(\partial G, \varepsilon_n)$. Continuing with the theme of α-convexity discussed in Chapter 2, this chapter deals with the case where G and R are both α-convex. In this situation we propose to estimate G and R by means of the α-convex hull of the samples $\mathcal{X}_n = \{Z_i : \xi_i = 1\}$ and $\mathcal{Y}_n = \{Z_i : \xi_i = 0\}$, respectively. Therefore, let

$$G_n = C_\alpha(\mathcal{X}_n) = (\mathcal{X}_n \oplus \alpha \mathring{B}) \ominus \alpha \mathring{B}, \tag{3.5}$$

$$R_n = C_\alpha(\mathcal{Y}_n) = (\mathcal{Y}_n \oplus \alpha \mathring{B}) \ominus \alpha \mathring{B}, \tag{3.6}$$

and let Γ_n be the estimator obtained after replacing (3.5) and (3.6) in (3.4). Thus, the estimator L_n in (3.3) is now completely defined. Before proceeding to the analysis of the properties of L_n, it is convenient to make some comments on its definition. First, G_n and R_n in (3.5) and (3.6) do not coincide exactly with the set estimators studied in Chapter 2, see (2.3) where the estimator with closed balls was defined. Anyway, remember that in Appendix B we prove that, with probability one, both definitions are equivalent and hence, it makes no difference whether we consider the estimator defined with open or closed balls. Considering (2.3) helped us to obtain the theoretical properties of the α-convex hull estimator and for that reason we used it in Chapter 2. However, in the case of L_n the definition in (2.3) does not facilitate the proofs and we have decided to work with (3.5) and (3.6) since they reliably reproduce the definition of α-convex hull. Anyway, we recall here Lemma B.0.9 since we will refer to it when computing $\mathbb{E}(L_n)$.

Second, in view of (3.2) we would like to make a remark on the behaviour of the function $\mu(B(\partial G, \varepsilon))$. In Chapter 1 we commented that Federer (1959) provides a generalization of the Steiner's formula for sets with positive reach. There, it is established that the d-dimensional measure of the closed ε-neighbourhood of a set with positive reach in $\mathbb{R}^d$ can be expressed as a polynomial of degree at most d in ε. The positive reach of a set is closely related to the free rolling condition. In Appendix B we prove that if G is a nonempty closed set in $\mathbb{R}^d$ such that a ball of radius $\alpha > 0$ rolls freely in G and in $\overline{G^c}$, then ∂G has positive reach. Therefore, under this rolling condition we may use Federer's theorem to conclude that $\mu(B(\partial G, \varepsilon))$ coincides locally with a polynomial of degree at most d in ε. From the Lebesgue density theorem and the rolling condition it can be proved that $\mu(\partial G) = 0$. It follows from this fact and from the polynomial representation of $\mu(B(\partial G, \varepsilon))$ that the limit in (3.2) exists and, as a consequence, the coefficient of ε in the polynomial must coincide with $2L_0$. This property has useful implications. For example, we can conclude that $|L(\varepsilon_n) - L_0| = O(\varepsilon_n)$.

Finally, we would like to note that G and R do not play completely interchangeable roles even though they are both α-convex. In Appendix A we list some useful results under the assumption that G is a nonempty closed set such that a ball of radius α rolls freely in G and in $\overline{G^c}$. In Lemma A.0.4 we prove that in those results G can be replaced by $\overline{G^c}$. However, it is important to emphasize that R is not equal to $\overline{G^c}$. For example, we cannot ensure that a ball of radius α rolls freely in R, see Figure 3.1. Note that ∂R does not coincide with ∂G, since it also includes the boundary of $[0, 1]^d$. Even so, Assumption (A1) in page 16 applied to the set G will be enough for our purposes.

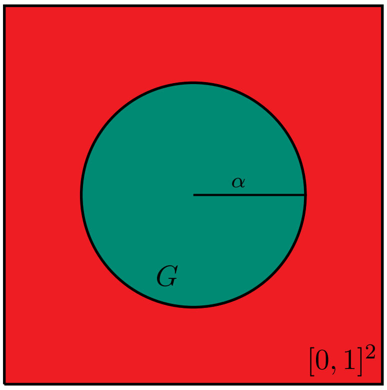


Figure 3.1: *G in green and R in red are both α-convex. A ball of radius α rolls freely in G and in $\overline{G^c}$. A ball of radius α does not roll freely in R.*

3.3 Asymptotic behaviour of L_n

In this section we present the main results regarding the behaviour of the estimator L_n defined in (3.2) with G_n and R_n as given in (3.5) and (3.6). First, we proof Theorem 3.3.1, which gives the almost sure rate of convergence of L_n to L_0. Under the same conditions, Theorem 3.3.2 gives us the L_1-convergence rate of L_n. From now on and for the sake of simplicity, we use the notation $\Gamma = \partial G$.

3.3.1 Almost sure convergence rate

Theorem 3.3.1. *Let $G \subset (0, 1)^d$ be a nonempty compact set. Assume that a ball of radius $\alpha > 0$ rolls freely in G and in $\overline{G^c}$. Then, with probability one,*

$$\inf_{\varepsilon_n} |L_n - L_0| = O\left(\frac{\log n}{n}\right)^{\frac{1}{d+1}},$$

and the optimal order is attained for $\varepsilon_n = (\log n/n)^{1/(d+1)}$.

Remark 3.3.1. *In Cuevas et al. (2007) a similar estimator to the one studied here is proposed. There, it is considered the estimator* L_n *defined in (3.2) with* $G_n = \mathcal{X}_n$ *and* $R_n = \mathcal{Y}_n$. *Its almost sure consistency and* L_1*-convergence rate are provided, but not the almost sure convergence rate. In order to compare both estimators we refer to Theorem 3.3.2 where the* L_1*-convergence rate for the estimator proposed in this chapter is provided.*

Proof. We follow the ideas of the proof of Theorem 3 in Rodríguez-Casal (2007). The proof is based on Propositions 3.3.1, 3.3.2 and 3.3.3. Proposition 3.3.1 establishes that if $\Gamma \subset B(Z_n^{\mathcal{X}}, 2\varrho_n) \cap B(Z_n^{\mathcal{Y}}, 2\varrho_n)$, where $Z_n^{\mathcal{X}} = \{Z_i \in \mathcal{X}_n : d(Z_i, \Gamma) \leq \varrho_n^2\}$ and $Z_n^{\mathcal{Y}} = \{Z_i \in \mathcal{Y}_n : d(Z_i, \Gamma) \leq \varrho_n^2\}$, then $B(\Gamma, \varepsilon_n) \backslash \Gamma_n$ is contained in the disc $D_n = B(\Gamma, \varepsilon_n) \setminus B(\Gamma, \varepsilon_n - K\varrho_n^2)$ for large enough K. Proposition 3.3.2 relies on $\mu(D_n) = O(\varrho_n^2)$ to find a bound for $|L_n - L_0|$ depending only on ε_n and ϱ_n. Finally, in Proposition 3.3.3 we determine the order of ϱ_n for which, with probability one, we have that $\Gamma \subset B(Z_n^{\mathcal{X}}, 2\varrho_n) \cap B(Z_n^{\mathcal{Y}}, 2\varrho_n)$ for large enough n, that is, ϱ_n satisfies

$$P\left(\Gamma \subset B(Z_n^{\mathcal{X}}, 2\varrho_n) \cap B(Z_n^{\mathcal{Y}}, 2\varrho_n) \text{ eventually }\right) = 1.$$

Theorem 3.3.1 is a straightforward consequence of these three results.

Proposition 3.3.1. *Let* G *be a set under the conditions of Theorem 3.3.1. Then the following results hold.*

i) With probability one, $\Gamma_n \subset B(\Gamma, \varepsilon_n)$.

ii) Let us assume that $\varrho_n \to 0$ *satisfies* $\varrho_n^2 \varepsilon_n^{-1} \to 0$ *and that*

$$P\left(\Gamma \subset B(Z_n^{\mathcal{X}}, 2\varrho_n) \cap B(Z_n^{\mathcal{Y}}, 2\varrho_n) \textit{ eventually }\right) = 1,$$

where $Z_n^{\mathcal{X}} = \{Z_i \in \mathcal{X}_n : d(Z_i, \Gamma) \leq \varrho_n^2\}$ *and* $Z_n^{\mathcal{Y}} = \{Z_i \in \mathcal{Y}_n : d(Z_i, \Gamma) \leq \varrho_n^2\}$. *Then, if* $K \geq \max(2, 8/\alpha)$, *we have that*

$$P\left(B(\Gamma, \varepsilon_n - K\varrho_n^2) \subset \Gamma_n \textit{ eventually }\right) = 1.$$

Remark 3.3.2. *The proof of i) remains true under milder conditions. It is only needed that the sets* G *and* R *are both* α*-convex. The* α*-convexity of* G *and* R *follows easily from Assumption (A1) as we mention in the proof below.*

Proof. Under the conditions of the proposition, G and $\overline{G^c}$ are both α-convex, see Lemma A.0.8. It can be easily seen that, as a consequence, R is also α-convex. Since, with probability one, $\mathcal{X}_n \subset G$ and $\mathcal{Y}_n \subset R$,

$$G_n = C_\alpha(\mathcal{X}_n) \subset C_\alpha(G) = G \text{ and } R_n = C_\alpha(\mathcal{Y}_n) \subset C_\alpha(R) = R.$$

Thus, with probability one,

$$\Gamma_n = B(G_n, \varepsilon_n) \cap B(R_n, \varepsilon_n) \subset B(G, \varepsilon_n) \cap B(R, \varepsilon_n) = B(\Gamma, \varepsilon_n),$$

which concludes the proof of i). On the other hand, the proof of ii) will be finished if we show that

$$P\left(\Gamma \subset B(G_n, K\varrho_n^2) \cap B(R_n, K\varrho_n^2) \text{ eventually }\right) = 1, \tag{3.7}$$

since if $\Gamma \subset B(G_n, K\varrho_n^2) \cap B(R_n, K\varrho_n^2)$ and $\varepsilon_n > K\varrho_n^2$, then

$$\begin{aligned} B(\Gamma, \varepsilon_n - K\varrho_n^2) &\subset B\left(B\left(G_n, K\varrho_n^2\right) \cap B\left(R_n, K\varrho_n^2\right), \varepsilon_n - K\varrho_n^2\right) \\ &\subset B(G_n, \varepsilon_n) \cap B(R_n, \varepsilon_n) = \Gamma_n. \end{aligned}$$

In order to prove (3.7) it is suffices to show that with probability one, for large enough n,

$$x_G = x - K\varrho_n^2 \eta(x) \in G_n \text{ and } x_R = x + K\varrho_n^2 \eta(x) \in R_n \tag{3.8}$$

for all $x \in \Gamma$, where $\eta(x)$ is the outward pointing unit normal vector at x, see Lemma A.0.5.

To prove (3.8) we need to show that x_G cannot be contained in an open ball of radius α which does not meet the sample $\mathcal{X}_n$. In the same manner, we need to prove that x_R cannot be contained in an open ball of radius α which does not meet the sample $\mathcal{Y}_n$. The situation in which the centre of the ball is close to Γ is analysed in Lemma 3.3.2. This lemma yields the result for x_G. For x_R we have also to analyse the situation in which the centre of the ball is far from Γ. This case is studied in Lemma 3.3.3. Finally, in Lemma 3.3.4 both results are used to establish the precise conditions under which (3.8) is satisfied. Proposition 3.3.1 is a consequence of this result. We begin with a geometric lemma, needed to prove Lemma 3.3.2.

Lemma 3.3.1. *Let G be a set under the conditions of Theorem 3.3.1 and $y \in \mathbb{R}^d$ such that $d(y, \Gamma) = \alpha - \kappa$ where $0 \leq \kappa \leq \alpha$. Then, for all $x \in \mathbb{R}^d$ with $d(x, \Gamma) \leq \kappa/2$ and $\|x - y\| \geq \alpha$ we have that*

$$\|x - P_\Gamma y\| \geq \sqrt{\frac{\alpha\kappa}{2}},$$

where $P_\Gamma y$ is the metric projection of y onto Γ.

Proof. This lemma is similar to Lemma 1 in Rodríguez-Casal (2007) and its proof is almost identical. Let $y \in \mathbb{R}^d$ be a point such that $d(y, \Gamma) = \alpha - \kappa$, where $0 \leq \kappa \leq \alpha$. The result is trivial for $\kappa = 0$. Hence, let us assume that $\kappa > 0$. Since reach$(\Gamma) \geq \alpha$, we denote by $P_\Gamma y$ the unique metric projection of y onto Γ, see Lemma A.0.6. Let η be the outward pointing unit normal vector at $P_\Gamma y$, see Figure 3.2.

First, we assume that $y \in G$. Then, Lemma A.0.1 ensures that $y = P_\Gamma y - (\alpha - \kappa)\eta$. Let $t = P_\Gamma y + \alpha\eta$. Then, for $x \in \mathbb{R}^d$ with $d(x, \Gamma) \leq \kappa/2$ and $\|x - y\| \geq \alpha$,

$$\alpha^2 \leq \|x - y\|^2 = \|x - P_\Gamma y + (\alpha - \kappa)\eta\|^2 = \|x - P_\Gamma y\|^2 + (\alpha - \kappa)^2 + 2(\alpha - \kappa)\langle x - P_\Gamma y, \eta\rangle,$$

$$\left(\alpha - \frac{\kappa}{2}\right)^2 \leq \|x - t\|^2 = \|x - P_\Gamma y - \alpha\eta\|^2 = \|x - P_\Gamma y\|^2 + \alpha^2 - 2\alpha\langle x - P_\Gamma y, \eta\rangle.$$

The second inequality is consequence of $d(x, \Gamma) \leq \kappa/2$ and $d(t, \Gamma) = \alpha$. Then

$$\|x - P_\Gamma y\|^2 + 2(\alpha - \kappa)\langle x - P_\Gamma y, \eta\rangle \geq 2\alpha\kappa - \kappa^2,$$

$$\|x - P_\Gamma y\|^2 - 2\alpha \langle x - P_\Gamma y, \eta\rangle \geq -\alpha\kappa + \frac{\kappa^2}{4}.$$

Multiplying the first inequality by α and the second by $(\alpha - \kappa)$ and adding, we have that

$$\|x - P_\Gamma y\|^2 \geq \frac{2\alpha^2\kappa - \alpha\kappa^2 - (\alpha-\kappa)\alpha\kappa + (\alpha-\kappa)\frac{\kappa^2}{4}}{2\alpha - \kappa} = \frac{\alpha^2\kappa + (\alpha-\kappa)\frac{\kappa^2}{4}}{2\alpha-\kappa} \geq \frac{\alpha\kappa}{2},$$

where the last inequality is a consequence of $0 < \kappa \leq \alpha$.

For $y \in G^c$ we can apply the previous result to $\overline{G^c}$. In this case $y \in \overline{G^c}$ and Γ is also the boundary of $\overline{G^c}$, see Lemma A.0.2.

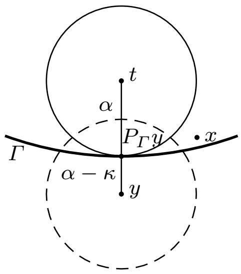


Figure 3.2: *Main elements considered in the proof of Lemma 3.3.1.*

□

Lemma 3.3.2. *Let G be a set under the conditions of Theorem 3.3.1 and let us assume that*

$$\Gamma \subset B(Z_n^{\mathcal{X}}, 2\varrho_n) \cap B(Z_n^{\mathcal{Y}}, 2\varrho_n),$$

where $Z_n^{\mathcal{X}} = \{Z_i \in \mathcal{X}_n : d(Z_i, \Gamma) \leq \varrho_n^2\}$ and $Z_n^{\mathcal{Y}} = \{Z_i \in \mathcal{Y}_n : d(Z_i, \Gamma) \leq \varrho_n^2\}$. Then, for all $y \in \mathbb{R}^d$ such that $d(y, \Gamma) = \alpha - \kappa$ with $\max(2, 8/\alpha)\varrho_n^2 < \kappa \leq \alpha$,

$$\mathring{B}(y,\alpha) \cap \mathcal{X}_n \neq \emptyset \ \text{ and } \ \mathring{B}(y,\alpha) \cap \mathcal{Y}_n \neq \emptyset.$$

Proof. Let $y \in \mathbb{R}^d$ be a point such that $d(y, \Gamma) = \alpha - \kappa$ with $\max(2, 8/\alpha)\varrho_n^2 < \kappa \leq \alpha$. We denote by $P_\Gamma y$ the metric projection of y onto Γ. Since $\Gamma \subset B(Z_n^{\mathcal{X}}, 2\varrho_n)$, there exists $z_x \in Z_n^{\mathcal{X}}$ such that $\|z_x - P_\Gamma y\| \leq 2\varrho_n$. Furthermore, $d(z_x, \Gamma) \leq \varrho_n^2 < \kappa/2$. If $\|z_x - y\| \geq \alpha$, then Lemma 3.3.1 yields that

$$\|z_x - P_\Gamma y\| \geq \sqrt{\frac{\alpha\kappa}{2}} > 2\varrho_n,$$

which leads to a contradiction. The last inequality is a consequence of $\kappa > 8\varrho_n^2/\alpha$. Therefore $\|z_x - y\| < \alpha$ and $\mathring{B}(y,\alpha) \cap \mathcal{X}_n \neq \emptyset$. Analogously, it can be proved that $\mathring{B}(y,\alpha) \cap \mathcal{Y}_n \neq \emptyset$.

□

Before stating Lemma 3.3.3 it is necessary to introduce some notation. Since we are assuming $G \subset (0,1)^d$, for all $x \in G$

$$d(x, \mathbb{R}^d \setminus (0,1)^d) > 0.$$

The function $d(\cdot, \mathbb{R}^d \setminus (0,1)^d)$ is continuous and therefore it reaches its minimum in the compact set G. Let us denote by e this minimum, that is,

$$e = \min_{x \in G} d(x, \mathbb{R}^d \setminus (0,1)^d) > 0. \tag{3.9}$$

Note that $B(G, e) \subset [0,1]^d$.

Lemma 3.3.3. *Let $x \in \mathbb{R}^d$ be a point such that $0 \leq d(x, G) \leq e/2$ and let $y \notin [0,1]^d$ such that $x \in \mathring{B}(y, \alpha)$. Then there exists $z_0 \in R$ for which $B(z_0, e/4) \subset \mathring{B}(y, \alpha)$.*

Proof. The function

$$d(\lambda) = d(\lambda x + (1-\lambda)y, G), \ \ 0 \leq \lambda \leq 1,$$

is continuous. Since $y \notin [0,1]^d$, we have that $d(0) = d(y, G) > e$. Furthermore

$$d(1) = d(x, G) \leq e/2.$$

Bolzano's Theorem ensures that there exists z_0 on the segment with endpoints x and y such that $d(z_0, G) = 3e/4$. Moreover, $z_0 \in R$ since $z_0 \in B(G, e) \subset [0,1]^d$ and $z_0 \notin G$, see Figure 3.3. Now, let us prove that $B(z_0, e/4) \subset \mathring{B}(y, \alpha)$. Let $z \in B(z_0, e/4)$. We have that

$$\|z - y\| \leq \|z - z_0\| + \|z_0 - y\| \leq \frac{e}{4} + \|z_0 - y\|.$$

Since z_0 lies on the segment with endpoints x and y,

$$\|z_0 - y\| = \|x - y\| - \|x - z_0\|.$$

From $d(z_0, G) = 3e/4$ and $d(x, G) \leq e/2$ it follows that $\|x - z_0\| \geq e/4$ and, therefore,

$$\|z_0 - y\| = \|x - y\| - \|x - z_0\| < \alpha - \frac{e}{4}.$$

Thus,

$$\|z - y\| < \frac{e}{4} + \alpha - \frac{e}{4} = \alpha.$$

□

Lemma 3.3.4. *Let us assume that $\Gamma \subset B(Z_n^{\mathcal{X}}, 2\varrho_n) \cap B(Z_n^{\mathcal{Y}}, 2\varrho_n)$ and $K\varrho_n^2 < \min(e/2, \alpha)$, where $K \geq \max(2, 8/\alpha)$. Let us also assume that $d_H(\mathcal{X}_n, G) < \alpha$ and $d_H(\mathcal{Y}_n, R) < \min(e/4, \alpha)$. Then, for all $x \in \Gamma$,*

$$x - K\varrho_n^2 \eta(x) \in G_n \ \ \textit{and} \ \ x + K\varrho_n^2 \eta(x) \in R_n,$$

where $\eta(x)$ is the outward pointing unit normal vector at x.

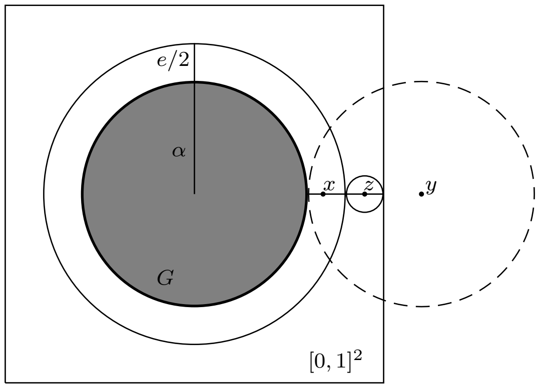


Figure 3.3: *Main elements considered in the proof of Lemma 3.3.3.*

Proof. Let $x \in \Gamma$ and $x_G = x - K\varrho_n^2\eta(x)$. The point x_G belongs to G_n if any open ball of radius α that contains the point x_G meets the sample $\mathcal{X}_n$. Thus, let $y \in \mathbb{R}^d$ be a point such that $x_G \in \mathring{B}(y,\alpha)$. We want to show that $\mathring{B}(y,\alpha) \cap \mathcal{X}_n$ is not empty. This is straightforward when $y \in G$, since by assumption $d_H(\mathcal{X}_n, G) < \alpha$. Now, let us suppose that $y \in G^c$. Since $x_G \in \mathring{B}(y,\alpha) \cap G$ ($K\varrho_n^2 < \alpha$), then $d(y,\Gamma) = \alpha - \kappa$, where $\kappa > K\varrho_n^2 \geq \max(2, 8/\alpha)\,\varrho_n^2$. By Lemma 3.3.2, we have that $\mathring{B}(y,\alpha) \cap \mathcal{X}_n \neq \emptyset$.

Now, let $x_R = x + K\varrho_n^2\eta(x)$. As before, in order to prove that x_R belongs to R_n, we need to show that $\mathring{B}(y,\alpha) \cap \mathcal{Y}_n$ is not empty, for any $y \in \mathbb{R}^d$ such that $x_R \in \mathring{B}(y,\alpha)$. Again, this is straightforward when $y \in R$, since $d_H(\mathcal{Y}_n, R) < \alpha$ by assumption. Now, let us assume that $y \notin R$. There are two possibilities: $y \in G$ or $y \notin [0,1]^d$. For the first one, as $x_R \in \mathring{B}(y,\alpha) \cap G^c$ ($K\varrho_n^2 < \alpha$), we have that $d(y,\Gamma) = \alpha - \kappa$, with $\kappa > K\varrho_n^2 \geq \max(2, 8/\alpha)\,\varrho_n^2$. Lemma 3.3.2 implies that $\mathring{B}(y,\alpha) \cap \mathcal{Y}_n \neq \emptyset$. Finally, if $y \notin [0,1]^d$, by the definition of x_R, we have that $d(x_R, G) = K\varrho_n^2 < e/2$. Then Lemma 3.3.3 establishes that there exists $z_0 \in R$ such that $B(z_0, e/4) \subset \mathring{B}(y,\alpha)$. Since $d_H(\mathcal{Y}_n, R) < e/4$, we have that $B(z_0, e/4) \cap \mathcal{Y}_n \neq \emptyset$. Thus, we have that $\mathring{B}(y,\alpha) \cap \mathcal{Y}_n \neq \emptyset$. □

The proof of Proposition 3.3.1 is now complete since the conditions of Lemma 3.3.4 are satisfied with probability one for large enough n.

□

Now, we are ready to prove Proposition 3.3.2, which gives a bound for the distance $|L_n - L_0|$ by splitting it into a *bias* term $|L(\varepsilon_n) - L_0|$ and a *variance* term $|L_n - L(\varepsilon_n)|$. Recall that L_0, $L(\varepsilon_n)$ and L_n are defined in equations (3.1), (3.2) and (3.3), respectively.

Proposition 3.3.2. *Under the conditions of Proposition 3.3.1 we have that, with probability one,*

$$|L_n - L_0| \leq |L(\varepsilon_n) - L_0| + O\left(\frac{\varrho_n^2}{\varepsilon_n}\right) = O(\varepsilon_n) + O\left(\frac{\varrho_n^2}{\varepsilon_n}\right)$$

and

$$\inf_{\varepsilon_n} |L_n - L_0| = O(\varrho_n). \tag{3.10}$$

Proof. We have that

$$|L_n - L_0| \leq |L_n - L(\varepsilon_n)| + |L(\varepsilon_n) - L_0|.$$

In Section 3.2 we discussed that $|L(\varepsilon_n) - L_0| = O(\varepsilon_n)$. On the other hand, Proposition 3.3.1 yields

$$P\left(B(\Gamma, \varepsilon_n - K\varrho_n^2) \subset \Gamma_n \subset B(\Gamma, \varepsilon_n) \text{ eventually }\right) = 1.$$

Then with probability one, for large enough n,

$$|L_n - L(\varepsilon_n)| = \frac{\mu(B(\Gamma, \varepsilon_n))}{2\varepsilon_n} - \frac{\mu(\Gamma_n)}{2\varepsilon_n} \leq \frac{\mu(B(\Gamma, \varepsilon_n)) - \mu(B(\Gamma, \varepsilon_n - K\varrho_n^2))}{2\varepsilon_n}.$$

In the following lemma the convergence rate of the last term in the previous inequality is determined.

Lemma 3.3.5. *Assume that $F(\varepsilon) = \mu(B(\Gamma, \varepsilon))$ is differentiable in a neighbourhood of zero and that the derivative F' is continuous at zero. Then*

$$\lim_{n \to \infty} \frac{\mu(D_n)}{2K\varrho_n^2} = L_0,$$

where $D_n = B(\Gamma, \varepsilon_n) \setminus B(\Gamma, \varepsilon_n - K\varrho_n^2)$.

Remark 3.3.3. *In Lemma 3.3.5 we assume that the function $F(\varepsilon) = \mu(B(\Gamma, \varepsilon))$ is smooth in a neighbourhood of zero. This holds in particular for the boundary Γ of a set G such that a ball of radius α rolls freely in G and in $\overline{G^c}$. In that case, it can be proved that Γ satisfies the conditions of Theorem 5.6 in Federer (1959), see Lemma A.0.6. As we have already observed this result ensures that $F(\varepsilon)$ coincides locally, for $\varepsilon \in (0, \alpha)$, with a polynomial of degree at most d. Therefore Lemma 3.3.5 is enough for our purposes.*

Proof. For large enough n (since $\varepsilon_n, \varrho_n \to 0$ and $\varrho_n^2 \varepsilon_n^{-1} \to 0$) we have that

$$\begin{aligned} \frac{\mu(D_n)}{2K\varrho_n^2} &= \frac{\mu\left(B(\Gamma, \varepsilon_n)\right) - \mu\left(B(\Gamma, \varepsilon_n - K\varrho_n^2)\right)}{2K\varrho_n^2} \\ &= \frac{F(\varepsilon_n) - F(\varepsilon_n - K\varrho_n^2)}{2K\varrho_n^2} = \frac{F'(\varsigma_n)K\varrho_n^2}{2K\varrho_n^2} = \frac{F'(\varsigma_n)}{2}, \end{aligned}$$

where we have applied the Mean Value Theorem, being ς_n a point in the interval $(\varepsilon_n - K\varrho_n^2, \varepsilon_n)$. Since F' is continuous at zero,

$$\lim_{n \to \infty} \frac{\mu(D_n)}{2K\varrho_n^2} = \frac{F'(0)}{2} = L_0,$$

where the last equality is a consequence of (3.1).

$\square$

By Lemma 3.3.5, with probability one,

$$|L_n - L(\varepsilon_n)| = O\left(\frac{\varrho_n^2}{\varepsilon_n}\right).$$

Therefore, with probability one,

$$|L_n - L_0| \leq |L(\varepsilon_n) - L_0| + O\left(\frac{\varrho_n^2}{\varepsilon_n}\right) = O(\varepsilon_n) + O\left(\frac{\varrho_n^2}{\varepsilon_n}\right).$$

Now, if we make equal the convergence orders of both terms on the right-hand side, then (3.10) holds for $\varepsilon_n = \varrho_n$. This completes the proof of Proposition 3.3.2. □

As we mentioned at the beginning of the proof of Theorem 3.3.1, in the following proposition we determine the rate for ϱ_n which guarantees that, with probability one, $\Gamma \subset B(Z_n^{\mathcal{X}}, 2\varrho_n) \cap B(Z_n^{\mathcal{Y}}, 2\varrho_n)$ for large enough n.

Proposition 3.3.3. *If $c > 0$ is large enough then*

$$P\left(\Gamma \subset B(Z_n^{\mathcal{X}}, 2\varrho_n) \cap B(Z_n^{\mathcal{Y}}, 2\varrho_n) \textit{ eventually}\right) = 1,$$

where

$$\varrho_n = \left(\frac{c \log n}{n}\right)^{\frac{1}{d+1}},$$

$Z_n^{\mathcal{X}} = \{Z_i \in \mathcal{X}_n : d(Z_i, \Gamma) \leq \varrho_n^2\}$, *and* $Z_n^{\mathcal{Y}} = \{Z_i \in \mathcal{Y}_n : d(Z_i, \Gamma) \leq \varrho_n^2\}$.

Proof. Theorem 1 of Dümbgen and Walther (1996) establishes that, for $\varrho_n > 0$,

$$P(\Gamma \nsubseteq B(Z_n^{\mathcal{U}}, 2\varrho_n)) \leq \varrho_n^{-d} \Pi(G, Z_n^{\mathcal{U}}, \varrho_n), \quad \mathcal{U} = \mathcal{X}, \mathcal{Y},$$

where $\Pi(G, Z_n^{\mathcal{U}}, \varrho_n) = \sup_{x \in \Gamma} P(B(x, \varrho_n) \cap Z_n^{\mathcal{U}} = \emptyset)$. By the Borel-Cantelli lemma, it is enough to prove prove that

$$\sum_{n=1}^{\infty} \varrho_n^{-d} \Pi(G, Z_n^{\mathcal{U}}, \varrho_n) < \infty, \quad \mathcal{U} = \mathcal{X}, \mathcal{Y}. \tag{3.11}$$

Let $x \in \Gamma$. Since Z is uniformly distributed on $[0,1]^d$ we have that, for $\varrho_n^2 < e$ (recall equation (3.9)),

$$\begin{aligned} P(B(x, \varrho_n) \cap Z_n^{\mathcal{X}} = \emptyset) &= P(Z_i \notin B(x, \varrho_n) \cap B(\Gamma, \varrho_n^2) \cap G, i = 1, \ldots, n) \\ &= \left(1 - \mu(B(x, \varrho_n) \cap B(\Gamma, \varrho_n^2) \cap G)\right)^n \\ &\leq \exp\left(-n\mu(B(x, \varrho_n) \cap B(\Gamma, \varrho_n^2) \cap G)\right). \end{aligned}$$

Likewise,

$$P(B(x, \varrho_n) \cap Z_n^{\mathcal{Y}} = \emptyset) \leq \exp\left(-n\mu(B(x, \varrho_n) \cap B(\Gamma, \varrho_n^2) \cap R)\right).$$

Lemma 3.3.6 stated below is proved in Rodríguez-Casal (2007). It gives a lower bound for $\mu(B(x, \varrho_n) \cap B(\Gamma, \varrho_n^2) \cap G)$ and $\mu(B(x, \varrho_n) \cap B(\Gamma, \varrho_n^2) \cap R)$ for large enough n.

Lemma 3.3.6. *If G is under the conditions of Theorem 3.3.1, then there exist constants $\gamma, \beta > 0$ such that, $\forall \varepsilon \in [0, \beta]$, $\forall x \in \Gamma$,*

$$\mu(B(x,\varepsilon)\cap B(\Gamma,\varepsilon^2)\cap G) \geq \gamma\varepsilon^{d+1} \text{ and } \mu(B(x,\varepsilon)\cap B(\Gamma,\varepsilon^2)\cap \overline{G^c}) \geq \gamma\varepsilon^{d+1}.$$

It is easy to prove that if ϱ_n is small enough, for example $\varrho_n^2 < e$, then

$$\mu(B(x,\varrho_n)\cap B(\Gamma,\varrho_n^2)\cap R) = \mu(B(x,\varrho_n)\cap B(\Gamma,\varrho_n^2)\cap \overline{G^c})$$

and, therefore, by Lemma 3.3.6

$$P(B(x,\varrho_n)\cap Z_n^{\mathscr{U}} = \emptyset) \leq \exp\left(-n\gamma\varrho_n^{d+1}\right), \quad \mathscr{U} = \mathcal{X}, \mathcal{Y}.$$

It is not hard to prove that for large enough c, (3.11) is satisfied. Thus, the proof of Proposition 3.3.3 is complete. □

Theorem 3.3.1 is a straightforward consequence of Propositions 3.3.1, 3.3.2 and 3.3.3. □

3.3.2 L_1- convergence rate

Theorem 3.3.2. *Let $G \subset (0,1)^d$ be a nonempty compact set. Assume that a ball of radius $\alpha > 0$ rolls freely in G and in $\overline{G^c}$. Then,*

$$\inf_{\varepsilon_n} \mathbb{E}\left|L_n - L_0\right| = O\left(n^{-\frac{1}{d+1}}\right)$$

and the optimal order is attained for $\varepsilon_n = n^{-1/(d+1)}$.

Remark 3.3.4. *The L_1-convergence rate for the estimator of L_0 based on the empirical approximation of $B(\Gamma,\varepsilon_n)$ proposed by Cuevas et al. (2007) is $n^{-1/(2d)}$, which is worse than the L_1-convergence rate $n^{-1/(d+1)}$ attained by the estimator proposed in this chapter. The main reason for this improvement is that smoothing the samples $\mathcal{X}_n$ and $\mathcal{Y}_n$ allows us to choose smaller radius ε_n of order $n^{-1/(d+1)}$. In Cuevas et al. (2007), the order of the optimal ε_n was $n^{-1/(2d)}$.*

Proof. We have that

$$\left|L_n - L_0\right| \leq \left|L_n - L(\varepsilon_n)\right| + \left|L(\varepsilon_n) - L_0\right|. \tag{3.12}$$

It follows from Proposition 3.3.1 that, with probability one, $\Gamma_n \subset B(\Gamma,\varepsilon_n)$. As a consequence $L_n \leq L(\varepsilon_n)$ and hence $\left|L_n - L(\varepsilon_n)\right| = L(\varepsilon_n) - L_n$. Taking expectations in (3.12) we get

$$\mathbb{E}\left|L_n - L_0\right| \leq L(\varepsilon_n) - \mathbb{E}(L_n) + \left|L(\varepsilon_n) - L_0\right|. \tag{3.13}$$

Under the stated conditions $\left|L(\varepsilon_n) - L_0\right| = O(\varepsilon_n)$. Now, by the definition of L_n,

$$\mathbb{E}(L_n) = \mathbb{E}\left(\frac{\mu(\Gamma_n)}{2\varepsilon_n}\right) = \frac{1}{2\varepsilon_n}\mathbb{E}\left(\int \mathbb{I}_{\{x\in\Gamma_n\}}\mu(dx)\right)$$
$$= \frac{1}{2\varepsilon_n}\int \mathbb{E}\left(\mathbb{I}_{\{x\in\Gamma_n\}}\right)\mu(dx) = \frac{1}{2\varepsilon_n}\int_{B(\Gamma,\varepsilon_n)} P(x\in\Gamma_n)\mu(dx),$$

where we have used again that, with probability one, $\Gamma_n \subset B(\Gamma, \varepsilon_n)$. Then,

$$\begin{aligned}
\mathbb{E}(L_n) &= \frac{1}{2\varepsilon_n} \int_{B(\Gamma,\varepsilon_n)} (1 - P(x \notin \Gamma_n))\, \mu(dx) \\
&= \frac{\mu(B(\Gamma,\varepsilon_n))}{2\varepsilon_n} - \frac{1}{2\varepsilon_n} \int_{B(\Gamma,\varepsilon_n)} P(x \notin \Gamma_n) \mu(dx) \\
&= L(\varepsilon_n) - \frac{1}{2\varepsilon_n} \int_{B(\Gamma,\varepsilon_n)} P(x \notin \Gamma_n) \mu(dx).
\end{aligned}$$

Therefore,

$$|\mathbb{E}(L_n) - L(\varepsilon_n)| = \frac{1}{2\varepsilon_n} \int_{B(\Gamma,\varepsilon_n)} P(x \notin \Gamma_n) \mu(dx). \tag{3.14}$$

It follows from (3.4) that

$$\begin{aligned}
P(x \notin \Gamma_n) &= P(x \notin B(G_n, \varepsilon_n) \cap B(R_n, \varepsilon_n)) \\
&= P(x \in B(G_n, \varepsilon_n)^c \cup B(R_n, \varepsilon_n)^c) \\
&\leq P(x \notin B(G_n, \varepsilon_n)) + P(x \notin B(R_n, \varepsilon_n))
\end{aligned}$$

and hence the left-hand side in (3.14) can be bounded above by two integrals. To be precise,

$$\begin{aligned}
|\mathbb{E}(L_n) - L(\varepsilon_n)| &\leq \frac{1}{2\varepsilon_n} \int_{B(\Gamma,\varepsilon_n)} P(x \notin B(G_n, \varepsilon_n)) \mu(dx) \\
&+ \frac{1}{2\varepsilon_n} \int_{B(\Gamma,\varepsilon_n)} P(x \notin B(R_n, \varepsilon_n)) \mu(dx) \\
&= (I) + (II).
\end{aligned} \tag{3.15}$$

First, let us consider the term

$$(I) = \frac{1}{2\varepsilon_n} \int_{B(\Gamma,\varepsilon_n)} P(x \notin B(G_n, \varepsilon_n)) \mu(dx). \tag{3.16}$$

Let $x \in B(\Gamma, \varepsilon_n)$ and $s \in \Gamma$ such that $\|x - s\| = d(x, \Gamma)$. Define

$$x_G = x - \varepsilon_n \eta(s),$$

being $\eta(s)$ the outward pointing unit normal vector at s, see Lemma A.0.5. Under the stated conditions the vector $\eta(s)$ is unique. Moreover, if $d(x, \Gamma) < \alpha$ the point $s \in \Gamma$ such that $\|x - s\| = d(x, \Gamma)$ is also unique, see Lemma A.0.7. Lemma A.0.5 ensures that $B(s - \alpha\eta(s), \alpha) \subset G$ and hence $x_G \in G$ for $\varepsilon_n \leq \alpha$, see Figure 3.4. Furthermore,

$$x \notin B(G_n, \varepsilon_n) \Rightarrow x_G = x - \varepsilon_n \eta(s) \notin G_n$$

which yields

$$P(x \notin B(G_n, \varepsilon_n)) \leq P(x_G \notin G_n).$$

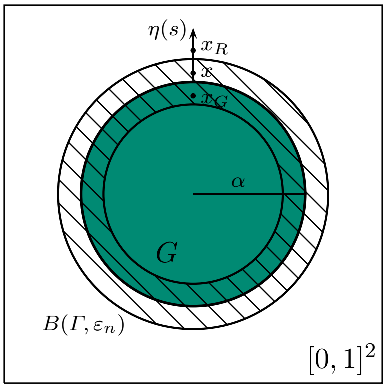


Figure 3.4: *The dashed area corresponds with $B(\Gamma, \varepsilon_n)$. In Theorem 3.3.2 we define for $x \in B(\Gamma, \varepsilon_n)$ the points $x_G = x - \varepsilon_n \eta(s)$ and $x_R = x + \varepsilon_n \eta(s)$. For large enough n, $x_G \in G$ and $x_R \in R$.*

Replace into (3.16) to get

$$(I) \leq \frac{1}{2\varepsilon_n} \int_{B(\Gamma,\varepsilon_n)} P(x_G \notin G_n)\mu(dx). \tag{3.17}$$

In Chapter 2 we bounded probabilities like the one that appears in (3.17), see (2.4) and (2.5). However, the estimator was not exactly the same as G_n since it was defined with closed balls, recall (2.3). We prove in Appendix B that it makes no difference whether we consider the estimator defined with open or closed balls, since with probability one both estimator are equal. Therefore, the results in Chapter 2 can also be applied here. Following the same steps as in (2.5), if we define a finite family $\mathcal{U}_{x_G,\alpha}$ unavoidable for $\mathcal{E}_{x_G,\alpha}$, we have that

$$P(x_G \notin G_n) \leq \sum_{U \in \mathcal{U}_{x_G,\alpha}} P(U \cap \mathcal{X}_n = \emptyset). \tag{3.18}$$

Consider, for $A \subset \mathbb{R}^d$, the random variable $\xi(A) = \sum_{i=1}^n \mathbb{I}_{\{Z_i \in A, \xi_i = 1\}}$. It is easy to see that $\xi(A)$ has a binomial distribution $B(n, p_A)$ where

$$p_A = P(Z \in A, \xi = 1) = P(\xi = 1)P(Z \in A|\xi = 1) = \mu(G)P_X(A).$$

Then,

$$P(U \cap \mathcal{X}_n = \emptyset) = P(\xi(U) = 0) = (1 - \mu(G)P_X(U))^n$$

and replacing $P(U \cap \mathcal{X}_n = \emptyset)$ in (3.18) we get

$$P(x_G \notin G_n) \leq \sum_{U \in \mathcal{U}_{x_G,\alpha}} (1 - \mu(G)P_X(U))^n \leq \sum_{U \in \mathcal{U}_{x_G,\alpha}} \exp(-n\mu(G)P_X(U)). \tag{3.19}$$

In short, we need to define a suitable unavoidable family for $\mathcal{E}_{x_G,\alpha}$ and give a lower bound for $P_X(U)$. Remember that the definition of a suitable unavoidable family for $\mathcal{E}_{x_G,\alpha}$ depends on the distance from the point x_G to the boundary Γ. This distance is easy to compute since, if $2\varepsilon_n < \alpha$, then

$$d(x_G,\Gamma) = d(x-\varepsilon_n\eta(s),\Gamma) = \begin{cases} \varepsilon_n + d(x,\Gamma) & \text{if } x \in G, \\ \varepsilon_n - d(x,\Gamma) & \text{if } x \notin G. \end{cases}$$

From the latter,

$$\varepsilon_n - d(x,\Gamma) \le d(x_G,\Gamma) \le 2\varepsilon_n.$$

Therefore, we can make $d(x_G,\Gamma)$ as small as desired for large enough n. Let $\varepsilon_n \le \alpha/4$. Note that G , P_X, and x_G satisfy the conditions of Proposition 2.4.2, since we are assuming that a ball of radius α rolls freely in G and in $\overline{G^c}$, P_X is the uniform distribution on G, and $x_G \in G$ with $d(x_G,\Gamma) \le \alpha/2$[1]. By Proposition 2.4.2, there exists a finite family $\mathcal{U}_{x_G,\alpha/2}$ with m_2 elements, unavoidable for $\mathcal{E}_{x_G,\alpha/2}$ and that satisfies

$$P_X(U) \ge L_2\left(\frac{\alpha}{2}\right)^{\frac{d-1}{2}} d(x_G,\Gamma)^{\frac{d+1}{2}}, \quad U \in \mathcal{U}_{x_G,\alpha/2}$$

where $L_2 > 0$ is a constant. Turning to (3.17) and (3.19),

$$\begin{aligned}
(I) &\le \frac{1}{2\varepsilon_n}\int_{B(\Gamma,\varepsilon_n)} m_2 \exp\left(-n\mu(G)L_2\left(\frac{\alpha}{2}\right)^{\frac{d-1}{2}} d(x_G,\Gamma)^{\frac{d+1}{2}}\right)\mu(dx) \\
&\le \frac{1}{2\varepsilon_n}\int_{B(\Gamma,\varepsilon_n)} m_2 \exp\left(-n\mu(G)L_2\left(\frac{\alpha}{2}\right)^{\frac{d-1}{2}} (\varepsilon_n - d(x,\Gamma))^{\frac{d+1}{2}}\right)\mu(dx) \\
&= \frac{1}{2\varepsilon_n}\int_{B(\Gamma,\varepsilon_n)} g_n(d(x,\Gamma))\mu(dx),
\end{aligned}$$

being $g_n(z) = m_2\exp(-n\mu(G)L_2\left(\frac{\alpha}{2}\right)^{\frac{d-1}{2}}(\varepsilon_n - z)^{\frac{d+1}{2}})$. It follows from the change of variables formula that

$$\int_{B(\Gamma,\varepsilon_n)} g_n(d(x,\Gamma))\mu(dx) = \int_0^{\varepsilon_n} g_n(\rho)\mu\mathcal{T}^{-1}(d\rho),$$

being $\rho = d(x,\Gamma)$ and $\mu\mathcal{T}^{-1}$ the measure on $\mathbb{R}$ characterized by

$$F(z) = \mu\{x \in \mathbb{R}^d : \ d(x,\Gamma) \le z\} = \mu(B(\Gamma,z)).$$

As we have seen, for $0 \le z < \alpha$, $F(z)$ is a polynomial of degree at most d in z. Therefore, it is a differentiable function and $F'(z)$ is bounded on compact sets. In short, we obtain

$$(I) \le \frac{1}{2\varepsilon_n}\int_0^{\varepsilon_n} g_n(\rho)F'(\rho)d\rho \le \frac{1}{2\varepsilon_n}K\int_0^{\varepsilon_n} g_n(\rho)d\rho. \tag{3.20}$$

[1]We apply Proposition 2.4.2 with $r_n = \alpha$.

By the change of variables $t = \varepsilon_n - \rho$ we obtain

$$\begin{aligned}\int_0^{\varepsilon_n} g_n(\rho)d\rho &= \int_0^{\varepsilon_n} m_2 \exp\left(-n\mu(G)L_2\left(\frac{\alpha}{2}\right)^{\frac{d-1}{2}}(\varepsilon_n-\rho)^{\frac{d+1}{2}}\right)d\rho\\ &= \int_0^{\varepsilon_n} m_2 \exp\left(-n\mu(G)L_2\left(\frac{\alpha}{2}\right)^{\frac{d-1}{2}} t^{\frac{d+1}{2}}\right)dt.\end{aligned}$$

Let $u = n\mu(G)L_2\left(\frac{\alpha}{2}\right)^{\frac{d-1}{2}} t^{\frac{d+1}{2}}$. Then,

$$\begin{aligned}&\int_0^{\varepsilon_n} m_2 \exp\left(-n\mu(G)L_2\left(\frac{\alpha}{2}\right)^{\frac{d-1}{2}} t^{\frac{d+1}{2}}\right)dt\\ &= \int_0^{n\mu(G)L_2\left(\frac{\alpha}{2}\right)^{\frac{d-1}{2}}\varepsilon_n^{\frac{d+1}{2}}} m_2 \frac{1}{\frac{d+1}{2}(\mu(G)L_2)^{2/(d+1)}}\left(\frac{\alpha}{2}\right)^{-\frac{d-1}{d+1}} n^{-\frac{2}{d+1}}\mathrm{e}^{-u}u^{\frac{1-d}{d+1}}du\\ &= O\left(n^{-\frac{2}{d+1}}\right).\end{aligned}$$

Finally, replace in (3.20) to get

$$(I) = O\left(\varepsilon_n^{-1} n^{-\frac{2}{d+1}}\right). \tag{3.21}$$

It remains to study the term

$$(II) = \frac{1}{2\varepsilon_n}\int_{B(\Gamma,\varepsilon_n)} P(x \notin B(R_n,\varepsilon_n))\mu(dx).$$

Again, let $x \in B(\Gamma,\varepsilon_n)$ and $s \in \Gamma$ such that $\|x-s\| = d(x,\Gamma)$. We define

$$x_R = x + \varepsilon_n\eta(s).$$

For large enough n we have that $x_R \in \overline{G^c}$, see Figure 3.4. Note that Lemma A.0.5 ensures that $B(s+\alpha\eta(s),\alpha) \subset \overline{G^c}$. Therefore, it suffices to consider $\varepsilon_n \leq \alpha$ in order to guarantee that $x_R \in \overline{G^c}$. Moreover,

$$x \notin B(R_n,\varepsilon_n) \Rightarrow x_R = x + \varepsilon_n\eta(s) \notin R_n$$

and hence

$$P(x \notin B(R_n,\varepsilon_n)) \leq P(x_R \notin R_n).$$

As we explained before, the definition of unavoidable families helps us to find an upper bound for $P(x_R \notin R_n)$. Thus, if $\mathcal{U}_{x_R,\alpha}$ is a finite and unavoidable family for $\mathcal{E}_{x_R,\alpha}$, then

$$\begin{aligned}(II) &\leq \frac{1}{2\varepsilon_n}\int_{B(\Gamma,\varepsilon_n)} P(x_R \notin R_n)\mu(dx)\\ &\leq \frac{1}{2\varepsilon_n}\int_{B(\Gamma,\varepsilon_n)} \sum_{U\in\mathcal{U}_{x_R,\alpha}} (1-\mu(R)P_Y(U))^n\mu(dx)\\ &\leq \frac{1}{2\varepsilon_n}\int_{B(\Gamma,\varepsilon_n)} \sum_{U\in\mathcal{U}_{x_R,\alpha}} \exp(-n\mu(R)P_Y(U))\mu(dx). \end{aligned}\tag{3.22}$$

We obtain (3.22) using the same arguments as in (3.18) but defining $\xi(A) = \sum_{i=1}^{n} \mathbb{I}_{\{Z_i \in A, \xi_i = 0\}}$. Now, if $2\varepsilon_n < \alpha$,

$$d(x_R, \Gamma) = d(x + \varepsilon_n \eta(s), \Gamma) = \begin{cases} \varepsilon_n + d(x, \Gamma) & \text{if } x \notin G, \\ \varepsilon_n - d(x, \Gamma) & \text{if } x \in G. \end{cases}$$

Therefore,

$$\varepsilon_n - d(x, \Gamma) \leq d(x_R, \Gamma) \leq 2\varepsilon_n \tag{3.23}$$

and $d(x_R, \Gamma)$ can be as small as desired for large enough n. However, in this situation we cannot directly apply Proposition 2.4.2 in order to define unavoidable families. As we have already discussed in Section 3.2, the set R does not satisfy Assumption (A1). We need the following auxiliary result.

Lemma 3.3.7. *Let $G \subset (0,1)^d$ be a nonempty compact set. Assume that a ball of radius $\alpha > 0$ rolls freely in G and in $\overline{G^c}$. Let $0 < \alpha_0 < \alpha$ and*

$$S = B(\Gamma, \alpha_0) \cap \overline{G^c}.$$

Then S is a nonempty compact set such that a ball of radius $\alpha^* > 0$ rolls freely in S and in $\overline{S^c}$, being $\alpha^* = \min(\alpha_0/2, \alpha - \alpha_0)$.

Proof. It can be easily seen that S is a nonempty compact set. Note that S is defined as the intersection of compacts sets. In order to prove the result we shall see that a ball of radius $\alpha_0/2$ rolls freely in S and a ball of radius $\alpha - \alpha_0$ rolls freely in $\overline{S^c}$. In Figure 3.5 we show the main elements considered in this result. First, we need to determine ∂S. We have that

$$\partial S = \Gamma \cup (\partial B(\Gamma, \alpha_0) \cap \overline{G^c}). \tag{3.24}$$

Equation (3.24) can be easily deduced using the definition of boundary of a set, Lemma A.0.3, and basic properties on the behaviour of the closure of the finite union of sets.

$$\begin{aligned} \partial S &= B(\Gamma, \alpha_0) \cap \overline{G^c} \cap \overline{(B(\Gamma, \alpha_0) \cap \overline{G^c})^c} \\ &= (B(\Gamma, \alpha_0) \cap \overline{G^c}) \cap (\overline{B(\Gamma, \alpha_0)^c} \cup \overline{\overline{G^c}^c}) \\ &= (B(\Gamma, \alpha_0) \cap \overline{G^c}) \cap (\overline{B(\Gamma, \alpha_0)^c} \cup G) \\ &= (B(\Gamma, \alpha_0) \cap \overline{B(\Gamma, \alpha_0)^c} \cap \overline{G^c}) \cup (B(\Gamma, \alpha_0) \cap \overline{G^c} \cap G) \\ &= (\partial B(\Gamma, \alpha_0) \cap \overline{G^c}) \cup \Gamma. \end{aligned}$$

Moreover, we shall see that

$$\partial B(\Gamma, \alpha_0) = \{x \in \mathbb{R}^d : \ d(x, \Gamma) = \alpha_0\}.$$

Since $B(\Gamma, \alpha_0) = \{x \in \mathbb{R}^d : \ d(x, \Gamma) \leq \alpha_0\}$ and the function $d(\cdot, \Gamma)$ is continuous, it follows easily that $\partial B(\Gamma, \alpha_0) \subset \{x \in \mathbb{R}^d : \ d(x, \Gamma) = \alpha_0\}$. On the other hand, it is not difficult to prove that, by the free rolling condition in G and in $\overline{G^c}$ and since $0 < \alpha_0 < \alpha$, $\{x \in \mathbb{R}^d : d(x, \Gamma) = \alpha_0\} \subset \partial B(\Gamma, \alpha_0)$. Therefore, we have

$$\partial B(\Gamma, \alpha_0) \cap \overline{G^c} = \{x \in \overline{G^c} : \ d(x, \Gamma) = \alpha_0\}, \tag{3.25}$$

To summarize, (3.24) and (3.25) characterize ∂S. Now we are ready to complete the proof. Let $s \in \partial S$. We shall see that there exists $x \in S$ such that $s \in B(x, \alpha_0/2) \subset S$. By (3.24) and (3.25), we must consider two different situations.

i) Suppose that $s \in \Gamma$. The free rolling condition in G and in $\overline{G^c}$ guarantees that there exists $\eta(s)$, such that $s \in B(s + \alpha\eta(s), \alpha) \subset \overline{G^c}$. Define

$$x = s + \frac{\alpha_0}{2}\eta(s)$$

and then

$$s \in B(x, \alpha_0/2) \subset B(s + \alpha\eta(s), \alpha) \subset \overline{G^c}.$$

Moreover, $B(x, \alpha_0/2) \subset B(\Gamma, \alpha_0)$ since for all $y \in B(x, \alpha_0/2)$ we have

$$d(y, \Gamma) \leq \|y - s\| \leq \|y - x\| + \|x - s\| \leq \frac{\alpha_0}{2} + \frac{\alpha_0}{2} = \alpha_0.$$

In short, $s \in B(x, \alpha_0/2) \subset S$.

ii) Now suppose that $s \in \partial B(\Gamma, \alpha_0) \cap \overline{G^c}$. By (3.25), $d(s, \Gamma) = \alpha_0$. Lemma A.0.7 establishes that there exists a unique point $t \in \Gamma$ such that $\|s - t\| = \alpha_0$. Moreover, by Lemma A.0.1, $s = t + \alpha_0\eta(t)$, being $\eta(t)$ the unique unit vector such that $B(t + \alpha\eta(t), \alpha) \subset \overline{G^c}$. Again, if we define

$$x = s - \frac{\alpha_0}{2}\eta(t) = t + \frac{\alpha_0}{2}\eta(t),$$

then we get $s \in B(x, \alpha_0/2) \subset S$.

It remains to prove that a ball of radius $\alpha - \alpha_0$ rolls freely in $\overline{S^c}$. Note that, since a ball of radius $\alpha_0/2$ rolls freely in S, it follows from Lemma A.0.2 that $\partial\overline{S^c} = \partial S$. Moreover

$$\overline{S^c} = \overline{(B(\Gamma, \alpha_0) \cap \overline{G^c})^c} = \overline{B(\Gamma, \alpha_0)^c \cup \overline{G^c}^c} = \overline{B(\Gamma, \alpha_0)^c} \cup \overline{\overline{G^c}^c} = \overline{B(\Gamma, \alpha_0)^c} \cup G.$$

Let $x \in \partial\overline{S^c}$. Again, we must consider two different situations.

i) If $s \in \Gamma$, then

$$s \in B(s - \alpha\eta(s), \alpha - \alpha_0) \subset B(s - \alpha\eta(s), \alpha) \subset G \subset \overline{S^c}.$$

ii) If $s \in \partial B(\Gamma, \alpha_0) \cap \overline{G^c}$, we have proved that $d(s, \Gamma) = \alpha_0$ and there exists a unique $t \in \Gamma$ such that $\|s - t\| = \alpha_0$. Let

$$x = s + (\alpha - \alpha_0)\eta(t) = t + \alpha\eta(t).$$

Then $s \in B(x, \alpha - \alpha_0)$ and it is straightforward to verify that $B(x, \alpha - \alpha_0) \subset \overline{S^c}$.

The proof is complete. $\square$

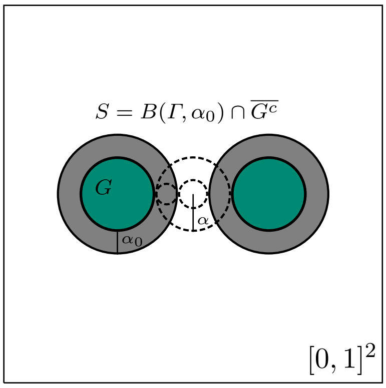


Figure 3.5: *In green G. A ball of radius α rolls freely in G and in $\overline{G^c}$. In gray $S = B(\Gamma, \alpha_0) \cap \overline{G^c}$, where $\alpha_0 < \alpha$. A ball of radius $\alpha_0/2$ rolls freely in S. A ball of radius $\alpha - \alpha_0$ rolls freely in $\overline{S^c}$.*

Turning back to the proof of Theorem 3.3.2, recall the definition of e given in (3.9) and let $\alpha_0 < \min(e, \alpha)$. By Lemma 3.3.7, $S = B(\Gamma, \alpha_0) \cap \overline{G^c}$ is a nonempty compact set such that a ball of radius $\alpha^* > 0$ rolls freely in S and in $\overline{S^c}$, being $\alpha^* = \min(\alpha_0/2, \alpha - \alpha_0)$. Let $\varepsilon_n \leq \alpha^*/4$ and then the following results hold.

i) $S \subset R$, since by assumption $\alpha_0 < e$.

ii) $x_R \in S$. Note that $x_R \in \overline{G^c}$ since $\varepsilon_n \leq \alpha$ and $d(x_R, \Gamma) \leq 2\varepsilon_n \leq \alpha_0$.

iii) $d(x_R, \partial S) = d(x_R, \Gamma) \leq \alpha^*/2$. We prove this equality by using that

$$d(x_R, \partial S) = \min(d(x_R, \Gamma), d(x_R, \partial B(\Gamma, \alpha_0) \cap \overline{G^c})).$$

Suppose that $d(x_R, \Gamma) \geq d(x_R, \partial B(\Gamma, \alpha_0) \cap \overline{G^c}) = \|x_R - z\|$, with $z \in \partial B(\Gamma, \alpha_0) \cap \overline{G^c}$. We shall see that this yields a contradiction. Since $x_R = x + \varepsilon_n \eta(s)$ with $s \in \Gamma$, we have that

$$d(z, \Gamma) \leq \|z - s\| \leq \|z - x_R\| + \|x_R - s\| \leq 2d(x_R, \Gamma) \leq 4\varepsilon_n \leq \alpha^* < \alpha_0,$$

where we have used (3.23). However, by (3.25), $d(z, \Gamma) = \alpha_0$. Therefore $d(x_R, \partial S) = d(x_R, \Gamma) \leq 2\varepsilon_n \leq \alpha^*/2$.

To summarize, we have defined a nonempty compact set $S \subset R$ such that a ball of radius $\alpha^* > 0$ rolls freely in S and in $\overline{S^c}$. Moreover, $x_R \in S$ with $d(x_R, \partial S) = d(x_R, \Gamma) \leq \alpha^*/2$, for large enough n. Let us consider the random variable X_S with uniform distribution on S. Then, by Proposition 2.4.2 there exists a finite family $\mathcal{U}_{x_R,\alpha^*}$ with m_2 elements, unavoidable for $\mathcal{E}_{x_R,\alpha^*}$ and that satisfies

$$P_{X_s}(U) = \frac{\mu(U \cap S)}{\mu(S)} \geq L_2 \alpha^{*\frac{d-1}{2}} d(x_R, \Gamma)^{\frac{d+1}{2}},$$

for all $U \in \mathcal{U}_{x_R,\alpha^*}$, being $L_2 > 0$ a constant. We prove in Lemma 3.3.8 that $\mathcal{U}_{x_R,\alpha^*}$ is also unavoidable for $\mathcal{E}_{x_R,\alpha}$ since $\alpha^* < \alpha$. Use that Y is the uniform distribution on R and that $S \subset R$ to obtain that, for all $U \in \mathcal{U}_{x_R,\alpha^*}$,

$$P_Y(U) = \frac{\mu(U \cap R)}{\mu(R)} \geq \frac{\mu(U \cap S)}{\mu(R)} = \frac{\mu(S)}{\mu(R)} P_{X_S}(U) \geq \frac{\mu(S)}{\mu(R)} L_2 {\alpha^*}^{\frac{d-1}{2}} d(x_R, \Gamma)^{\frac{d+1}{2}}.$$

Turning back to (3.22) we obtain

$$(II) \leq \frac{1}{2\varepsilon_n} \int_{B(\Gamma,\varepsilon_n)} m_2 \exp\left(-n\mu(S) L_2 {\alpha^*}^{\frac{d-1}{2}} d(x_R, \Gamma)^{\frac{d+1}{2}}\right) \mu(dx).$$

Analogous to the term (I), we get

$$(II) = O(\varepsilon_n^{-1} n^{-\frac{2}{d+1}}). \tag{3.26}$$

Replace (3.21) and (3.26) into (3.15) and then

$$|\mathbb{E}(L_n) - L(\varepsilon_n)| = O(\varepsilon_n^{-1} n^{-\frac{2}{d+1}}). \tag{3.27}$$

Now, going back to (3.13) and using (3.27) and the fact that $|L(\varepsilon_n) - L_0| = O(\varepsilon_n)$, we get that

$$\mathbb{E}\,|L_n - L_0| = O(\varepsilon_n^{-1} n^{-\frac{2}{d+1}}) + O(\varepsilon_n).$$

Making equal the convergence orders of both terms in the right-hand side we obtain the optimal convergence order for $\mathbb{E}\,|L_n - L_0|$. Therefore, for $\varepsilon_n = n^{-1/(d+1)}$

$$\mathbb{E}\,|L_n - L_0| = O\left(n^{-\frac{1}{d+1}}\right).$$

The proof of Theorem 3.3.2 is complete. □

Lemma 3.3.8. *Let $\mathcal{U}_{x,r_0}$ be an unavoidable family for $\mathcal{E}_{x,r_0}$, with $r_0 > 0$. Then $\mathcal{U}_{x,r_0}$ is unavoidable for $\mathcal{E}_{x,r}$, for $r \geq r_0$.*

Proof. Let $y \in B(x,r)$. If $\|y - x\| \leq r_0$, then by definition of unavoidable family for $\mathcal{E}_{x,r_0}$, there exists $U \in \mathcal{U}_{x,r_0}$ such that

$$U \subset B(y, r_0) \subset B(y, r).$$

Suppose that $r_0 < \|y - x\| \leq r$ and define

$$y* = x + r_0 \frac{y - x}{\|y - x\|} \in B(x, r_0).$$

Then, there exists $U \in \mathcal{U}_{x,r_0}$ such that $U \subset B(y^*, r_0) \subset B(y, r)$ since, for all $z \in B(y^*, r_0)$, we have that

$$\begin{aligned}\|z - y\| &\leq \|z - y^*\| + \|y^* - y\| \leq r_0 + \left\| x + r_0 \frac{y - x}{\|y - x\|} - y \right\| \\ &= r_0 + \left(1 - \frac{r_0}{\|y - x\|}\right) \|y - x\| = \|y - x\| \leq r.\end{aligned}$$

This completes the proof of the lemma. □

Chapter 4

Implementation issues and simulation results

4.1 Introduction

In the previous chapters we have discussed the support and surface area estimation problems from a theoretical point of view. We now turn our attention to how practical analysis can be carried out in the R computing environment.

Some of the estimators defined in Chapter 1, such as the Devroye-Wise support estimator, are easy to understand and implement. The implementation of the α-convex hull, however, is not so immediate and some effort is required in order to compute it efficiently. Edelsbrunner et al. (1983) proposed an algorithm to construct the α-convex hull of a finite set of points in $\mathbb{R}^2$. The algorithm is based on the closed relationship that exists between this construct and Delaunay triangulations. Following the methodology described by Edelsbrunner et al. (1983), we have programmed the α-convex hull of a sample. To be precise, we have programmed the complement of the α-convex hull of a sample, which can be written as a union of open balls and halfplanes. Many times, however, our interest lies in the surface area of a set (boundary length in $\mathbb{R}^2$) rather than in its support. Given the α-convex hull of a sample we can compute its boundary length by adding the lengths of the arcs that form its boundary. Another way to estimate the boundary length of a set is by using the α-shape of a sample of points taken in it. The notion of α-shape, briefly discussed in Chapter 1, is derived from a generalization of the convex hull definition. This construct is also closely related to the Delaunay triangulations and an algorithm for determining the α-shape of a finite set of points is given in Edelsbrunner et al. (1983). We have also implemented this algorithm in R. An alternative perspective to the boundary length estimation problem relies on the notion of Minkowski content. Based on this notion and on the α-convex hull implementation, we can compute the boundary length estimator L_n discussed in Chapter 3. Note that L_n is defined from the α-convex hull of two given samples. Therefore, we consider in this chapter two different approaches to the boundary length estimation problem, depending on the available information. If the information comes from a sample of points in the set of interest, then we estimate the boundary length of the set via the perimeter of a support estimator such as the α-convex hull estimator or the α-shape estimator. If we are provided with a sample of points in the set and in its complement, then we estimate the boundary length of the set via the Minkowski content. We shall refer to these two different situations as one sample approach and two sample approach, respectively.

This chapter is organized as follows. Section 4.2 starts by discussing some computational issues in the α-convex hull estimator. A brief overview of the Voronoi and Delaunay geometric structures is included along with the description of the implementation algorithm of the α-convex hull. Section 4.3 is devoted to the boundary length estimation problem, more precisely to the two sample approach. We present the results of a simulation study on the estimator L_n defined in Chapter 3. Section 4.4 is devoted to the one sample approach, stressing the α-shape estimator. All the programmed functions have been put together in R package format. The resulting library of functions, named `alphahull`, is intended to provide a means of better understanding the different estimators discussed throughout this dissertation. Details on how to use the package and short scripts to execute some basic examples are inserted along the chapter in typewriter font. More extended documentation of the package is available in Appendix C.

4.2 Programming the α-convex hull

Let S be a nonempty compact subset of $\mathbb{R}^2$ and let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a random sample from X, where X denotes a random variable with support S. Remember that, for $\alpha > 0$, the α-convex hull of $\mathcal{X}_n$ is given by

$$C_\alpha(\mathcal{X}_n) = \bigcap_{\{\mathring{B}(x,\alpha):\ \mathring{B}(x,\alpha)\cap\mathcal{X}_n=\emptyset\}} \left(\mathring{B}(x,\alpha)\right)^c. \tag{4.1}$$

Edelsbrunner et al. (1983) defined a similar construct, the α-hull of a finite set of points, for an arbitrary $\alpha \in \mathbb{R}$. According to the terminology used by Edelsbrunner et al. (1983), the α-convex hull in (4.1) coincides with the $-1/\alpha$-hull. By DeMorgan's law, the complement of $C_\alpha(\mathcal{X}_n)$ can be written as the union of all open balls of radius α which contain no point of $\mathcal{X}_n$, that is,

$$C_\alpha(\mathcal{X}_n)^c = \bigcup_{\{\mathring{B}(x,\alpha):\ \mathring{B}(x,\alpha)\cap\mathcal{X}_n=\emptyset\}} \mathring{B}(x,\alpha). \tag{4.2}$$

This representation of $C_\alpha(\mathcal{X}_n)^c$ provides a means of computing the α-convex hull. Thus, the problem is to determine the union of the open balls in (4.2). The solution to this problem is closely related to the Voronoi diagrams and Delaunay triangulations. We provide a brief introduction to these geometric structures which are further discussed in Aurenhammer (1991), Aurenhammer and Klein (2000), and Møller (1994), among others. A tessellation or mosaic of the d-dimensional Euclidean space $\mathbb{R}^d$ is a subdivision of $\mathbb{R}^d$ into d-dimensional non-overlapping sets. Depending on the situation, these sets are called cells, crystals, regions, *etc*. The Voronoi diagram is one of the most attracting tessellations since it provides models for many natural phenomena and has numerous mathematical and statistical applications. We define the Voronoi diagram of a sample $\mathcal{X}_n$ in $\mathbb{R}^2$. This definition can be straightforwardly generalized to the d-dimensional Euclidean space and to a deterministic finite set of points.

Definition 4.2.1. *The Voronoi diagram of $\mathcal{X}_n$ is a covering of the plane by n regions V_i, where for $i = 1, \ldots, n$,*

$$V_i = \{x \in \mathbb{R}^2 : \ \|x - X_i\| \leq \|x - X_j\| \text{ for all } X_j \in \mathcal{X}_n\}.$$

Remark 4.2.1. *Definition 4.2.1 corresponds with the closest point Voronoi diagram defined by Edelsbrunner et al. (1983), who distinguishes between closest and furthest point Voronoi diagram.*

For each sample point X_i, the set V_i consists of all points in $\mathbb{R}^2$ which have X_i as nearest sample point. The cells V_i are closed and convex and they can be proved to have disjoint topological interiors. The cell V_i is unbounded if and only if X_i is a point which belongs to the boundary of the convex hull of $\mathcal{X}_n$. Otherwise V_i is a nonempty convex polygon. Figure 4.1 (a) shows the Voronoi diagram of a uniform random sample of size $n = 30$ on the unit square. Each region V_i contains the point X_i. The dashed lines represent the semi-infinite line segments of the unbounded Voronoi cells. Two sample points X_i and X_j are said to be Voronoi neighbours if the cells V_i and V_j share a common point.

Another interesting structure, closely related to the Voronoi diagram, is the Delaunay triangulation.

Definition 4.2.2. *The Delaunay triangulation of $\mathcal{X}_n$ is defined as the straight line dual to the Voronoi diagram of $\mathcal{X}_n$, that is, there exists a straight line edge between X_i and X_j if and only if V_i and V_j are Voronoi neighbours.*

Figure 4.1 (b) shows the Delaunay triangulation of the above-mentioned uniform random sample. Each Delaunay cell is a triangle whose vertices are sample points and whose circumcentre coincides with a Voronoi cell vertex, see Figure 4.2. Observe that the Delaunay triangulation constitutes a tessellation of the convex hull of $\mathcal{X}_n$.

There is a one-to-one correspondence between the Voronoi diagram and the Delaunay triangulation. The algorithm for the construction of the Voronoi diagram is based on this duality. Edelsbrunner et al. (1983) underlines the fact that both the Voronoi diagram and the Delaunay triangulation can be constructed in $O(n \log n)$ time. Moreover, the Voronoi diagram can be constructed from the respective Delaunay triangulation in $O(n)$ time and vice versa. Lemma 4.2.1 stated below is the key result that relates the problem of computing $C_\alpha(\mathcal{X}_n)^c$ in (4.2) with the Voronoi diagram and Delaunay triangulation. See Edelsbrunner et al. (1983) for the proof.

Lemma 4.2.1. *Let $\mathring{B}(x, r)$ be an open ball which does not contain any point of a sample $\mathcal{X}_n$. Either $\mathring{B}(x, r)$ lies entirely outside the convex hull of $\mathcal{X}_n$ or there is an open ball which contains $\mathring{B}(x, r)$ but no points of $\mathcal{X}_n$ and which has its centre on an edge of the Voronoi diagram of $\mathcal{X}_n$.*

Note that by Lemma 4.2.1, the union of open balls in (4.2) can be reduced to the union of open balls with centres on the edges of the Voronoi diagram which do not contain any point of the sample and whose radii are greater or equal to α. This fact considerably simplifies the problem of computing (4.2). Anyway, the union of open balls with centres on the edges of the Voronoi diagram which do not contain any point of the sample and whose radii are greater or equal to α is still too complex. How can we compute these sets? Let us consider the following

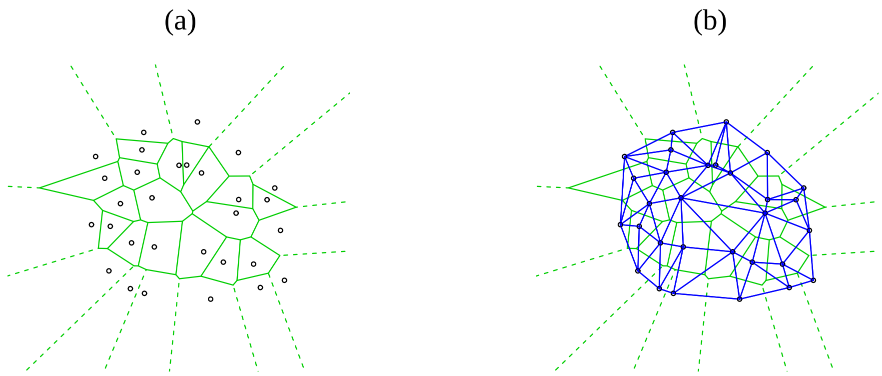


Figure 4.1: *(a) Voronoi diagram of a uniform random sample of size* $n = 30$ *on the unit square. The dashed lines represent the semi-infinite line segments of the unbounded Voronoi cells. (b) In blue the Delaunay triangulation dual to the Voronoi diagram.*

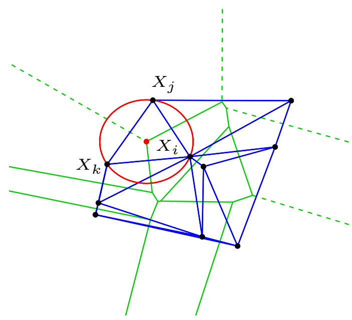


Figure 4.2: *Uniform random sample* $\mathcal{X}_n$ *of size* $n = 10$ *on the unit square. In green Voronoi diagram of* $\mathcal{X}_n$*. In blue Delaunay triangulation. Each Delaunay cell is a triangle whose vertices are sample points. The circumcentre of each Delaunay triangle* $\widehat{X_i X_j X_k}$ *coincides with a Voronoi cell vertex.*

two possible situations. First, let X_i and X_j be two Voronoi neighbours such that the cells V_i and V_j share a common closed line segment $[a, b]$, see Figure 4.2. It follows from the duality between the Voronoi diagram and the Delaunay triangulation that the union of open balls with centres on the edge $[a, b]$ which do not contain any point of a sample is equal to

$$\mathring{B}(a, \|a - X_i\|) \cup \mathring{B}(b, \|b - X_i\|),$$

see Figure 4.3. Therefore, the existence of an open ball with centre x on $[a, b]$ and radius α such that $\mathring{B}(x, \alpha) \cap \mathcal{X}_n = \emptyset$ will depend on the values $\|a - X_i\|$ and $\|b - X_i\|$.

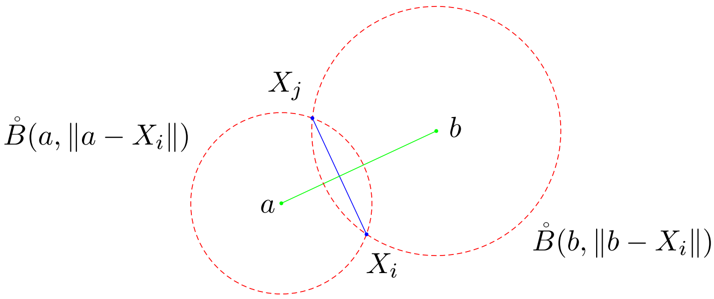


Figure 4.3: *Consider the sample $\mathcal{X}_n$ in Figure 4.2. The points X_i and X_j are Voronoi neighbours such that the cells V_i and V_j share a common closed line segment $[a, b]$. The union of open balls with centres on the edge $[a, b]$ which do not contain any point of $\mathcal{X}_n$ is equal to $\mathring{B}(a, \|a - X_i\|) \cup \mathring{B}(b, \|b - X_i\|)$.*

Second, let X_j and X_k be two Voronoi neighbours such that the cells V_j and V_k share a common semi-infinite line segment $[a, +\infty)$, see Figure 4.2. Now, the union of open balls with centres on the edge $[a, +\infty)$ which do not contain any point of a sample can be written as

$$\mathring{B}(a, \|a - X_j\|) \cup H(X_j, X_k),$$

where $H(X_j, X_k)$ denotes the open halfplane defined by the straight line through X_j and X_k, see Figure 4.4. The existence of an open ball with centre x on $[a, +\infty)$ and radius α such that $\mathring{B}(x, \alpha) \cap \mathcal{X}_n = \emptyset$ will depend on the value $\|a - X_j\|$.

All these considerations have been taken into account in the programming of the complement of the α-convex hull. As the number on edges in the Voronoi diagram is linear in n, see Edelsbrunner et al. (1983), is follows that $C_\alpha(\mathcal{X}_n)^c$ can be written as union of $O(n)$ open balls and halfplanes. Note that for each edge of the Voronoi diagram, the union of open balls with centres on it and radius α which do not contain any point of $\mathcal{X}_n$ can be written as the union of at most four open balls or halfplanes. To summarize, the algorithm for the construction of $C_\alpha(\mathcal{X}_n)^c$ is as follows.

1. Construct the Voronoi diagram and Delaunay triangulation of $\mathcal{X}_n$.
2. For each edge of the Voronoi diagram determine the union of open balls with centres on it and radius α which do not contain any point of $\mathcal{X}_n$.
3. Output $C_\alpha(\mathcal{X}_n)^c$.

Once $C_\alpha(\mathcal{X}_n)^c$ is constructed we can decide whether a given point of $\mathbb{R}^2$ belongs to the α-convex hull or not, by checking if it belongs to any of the open balls or halfplanes that form the complement of the α-convex hull. Moreover, the boundary of $C_\alpha(\mathcal{X}_n)$ can be completely determined and hence, we can also measure the perimeter of $C_\alpha(\mathcal{X}_n)$.

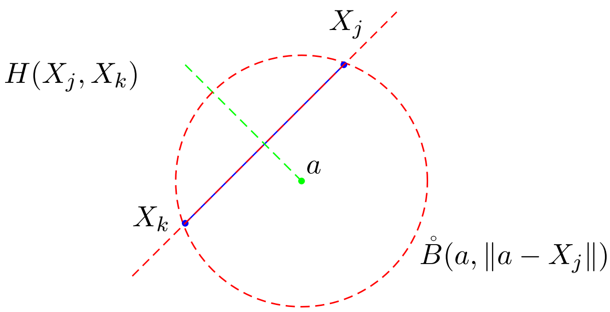


Figure 4.4: *Consider the sample $\mathcal{X}_n$ in Figure 4.2. The points X_j and X_k are Voronoi neighbours such that the cells V_j and V_k share a common semi-infinite line segment $[a, +\infty)$. The union of open balls with centres on the edge $[a, +\infty)$ which do not contain any point of $\mathcal{X}_n$ can be written as $\mathring{B}(a, \|a - X_j\|) \cup H(X_j, X_k)$.*

We next show an example on how to use the `alphahull` library to compute the α-convex hull of a uniform random sample of size $n = 30$ on the unit square. The following commands generate the sample and return the corresponding Voronoi diagram and Delaunay triangulation. We also produce a plot of the resulting geometric structures.

```
> library(alphahull)
> sample<-matrix(runif(60),nc=2)
> info<-inform.vor.tri(sample)
> plot(sample[,1],sample[,2])
> add.voronoi(info$mat.info,col=3)
> plot(info$tri.obj,add = T,col=4)
```

The function `inform.vor.tri` invokes internal functions from the `TRIPACK` package. The `TRIPACK` is a Fortran 77 software package that employs an incremental algorithm to construct a constrained Delaunay triangulation of a set of points in the plane, see Renka (1996). We have programmed the Voronoi diagram from the Delaunay triangulation returned by `TRIPACK`. The matrix `info$mat.info` contains all the information relating to the Voronoi diagram and Delaunay triangulation of the sample.

```
> info$mat.info[1:5,]
           l11       l12       l21       l22       m11
[1,] 0.7348463 0.5473715 0.7333052 0.7953018 0.7869268
[2,] 0.9059224 0.6084090 0.7348463 0.5473715 0.7869268
[3,] 0.8693576 0.5464288 0.7348463 0.5473715 0.8026671

......................................................
```

```
           m12       m21       m22 dum1 dum2 l1 l2 m1 m2
[1,] 0.6716652 0.6892330 0.6710579    0    0 15 28  1 39
[2,] 0.6716652 0.8026671 0.6275482    0    0  1 15  1  2
[3,] 0.6275482 0.8017665 0.4990455    0    0  4 15  2 13
```

For each row of `info$mat.info`, the columns `l1`, `l2`, `m1`, and `m2` store the indexes of two Voronoi neighbour sample points and the corresponding vertices of the shared Voronoi edge, respectively. Their coordinates are given by the first eight columns. Finally, `dum1` and `dum2` indicate whether the vertices of the Voronoi edge are infinite or not, that is, whether the shared edge is a closed line segment (both `dum1` and `dum2` equal to zero) or an semi-infinite line segment.

```
> alpha<-0.15
> compl<-complement(alpha,info$mat.info)
> shape<-alpha.shape(info,alpha)
> ahull<-alpha.hull(shape,compl)
> plot.ahull(ahull,pvor=T,pdel=T,pshape=F,new=F,col=2)
```

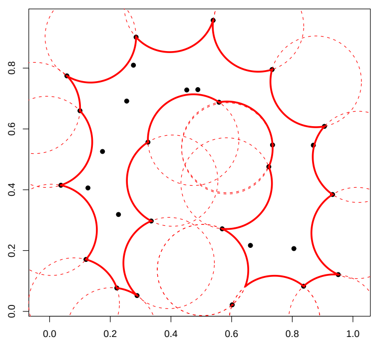


Figure 4.5: *Uniform random sample $\mathcal{X}_n$ of size $n = 30$ on the unit square. In red the boundary of $C_\alpha(\mathcal{X}_n)$ for $\alpha = 0.15$. The boundary is determined by arcs of balls of radius α that hit the sample (dashed balls) and isolated sample points.*

The function `complement` returns the open balls and halfplanes determining the complement of the α-convex hull for a given value $\alpha > 0$. In order to plot the boundary of the α-convex hull, we need to determine which of the balls with radius α in the complement hit two Voronoi neighbour points. This is closely related to the concept of α-neighbours, recall Definition 1.5.2. The function `alpha.shape` returns, among other arguments, the α-extremes and α-neighbours of the sample. The function `alpha.hull` returns the boundary of the α-convex hull and its length. By using the information provided by the function `alpha.shape`, we characterize the boundary of the α-convex hull by arcs of balls of radius α that hit the sample and isolated sample points when this is the case, see Figure 4.5. The perimeter coincides with the sum of the lengths of the arcs that form the boundary. For example, for the discussed sample and $\alpha = 0.15$, we have

```
> ahull$length
[1] 4.900739.
```

We can also determine whether a given point in $\mathbb{R}^2$ belongs to the α-convex hull or not by means of the function `in.alpha.hull`.

```
> in.alpha.hull(ahull,c(0.5,0.5))
[1] FALSE
> in.alpha.hull(ahull,c(0.2,0.5))
[1] TRUE
```

Finally, the function `plot.ahull` produces a plot of the boundary of the α-convex hull. If desired it also adds the Voronoi diagram, the Delaunay triangulation or the α-shape of $\mathcal{X}_n$ to the plot, see Appendix C for more details on the use of these functions. An in-depth study of the α-shape is given in Section 4.4. In Figure 4.6 we show the α-convex hull of $\mathcal{X}_n$ for different values of α. It is clear from the plots that $C_\alpha(\mathcal{X}_n)$ changes considerably depending on the value of α. We discuss the choice of α in Section 4.4.

(a) (b) (c)

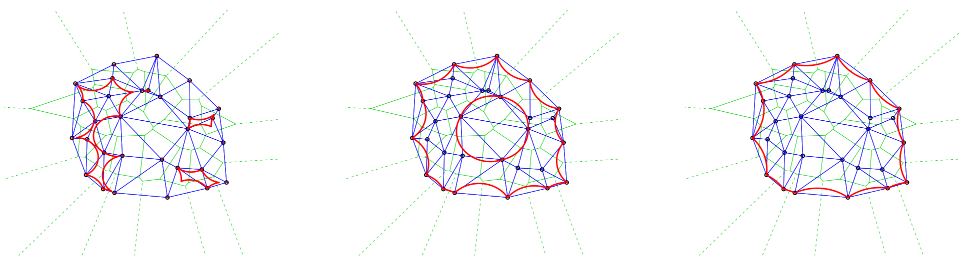


Figure 4.6: *We plot in red the boundary of the α-convex hull of a uniform random sample $\mathcal{X}_n$ of size $n = 30$ on the unit square. (a) $\alpha = 0.13$. (b) $\alpha = 0.2$. (c) $\alpha = 0.3$.*

4.3 Boundary length estimation: the two samples approach

Let $G \subset (0,1)^2$ be a nonempty compact set in $\mathbb{R}^2$. In this section we present the results of a simulation study comparing the behaviour of two boundary length estimators of the form

$$L_n = \frac{\mu(\Gamma_n)}{2\varepsilon_n},$$

where Γ_n is an estimator of $B(\partial G, \varepsilon_n)$. In this situation the sampling information is assumed to be given by i.i.d. observations $(Z_1, \xi_1), \ldots, (Z_n, \xi_n)$ of a random variable (Z, ξ), where Z is uniformly distributed on the unit square $[0,1]^2$ and $\xi = \mathbb{I}_{\{Z \in G\}}$. Let $\mathcal{X}_n = \{Z_i : \xi_i = 1\}$ and $\mathcal{Y}_n = \{Z_i : \xi_i = 0\}$. First, we consider the estimator L_n proposed by Cuevas et al. (2007), that is, with

$$\Gamma_n = B(\mathcal{X}_n, \varepsilon_n) \cap B(\mathcal{Y}_n, \varepsilon_n).$$

We shall denote this estimator by L_n^{DW} since Γ_n is defined as the intersection of two Devroye-Wise estimators. Second, we consider the estimator L_n with

$$\Gamma_n = B(C_\alpha(\mathcal{X}_n), \varepsilon_n) \cap B(C_\alpha(\mathcal{Y}_n), \varepsilon_n).$$

We shall denote this estimator by L_n^{α} since Γ_n is defined from the α-convex hull of $\mathcal{X}_n$ and $\mathcal{Y}_n$. Before presenting the simulation study, we briefly comment on some aspects of the implementation of both boundary length estimators. The main difference between L_n^{DW} and L_n^{α} resides in the construction of Γ_n. Once Γ_n is obtained, the procedure for computing $\mu(\Gamma_n)$ is similar. Cuevas et al. (2007) comment on the difficulty to give the exact value of $\mu(\Gamma_n)$ and suggest the possibility of approximating $\mu(\Gamma_n)$ by using the Monte Carlo method. Thus, following the notation in Cuevas et al. (2007), let $Z_1^*, \ldots, Z_B^*$ be a random sample, independent of $Z_1, \ldots, Z_n$, from the uniform distribution on $[0,1]^2$. For both L_n^{DW} and L_n^{α}, we have that $\Gamma_n \subset [0,1]^2$, with probability one for large enough n, and $\mu(\Gamma_n) = P(Z_1^* \in \Gamma_n)$. For large B,

$$\mu_B(\Gamma_n) = \frac{\sum_{i=1}^{B} \mathbb{I}_{\{Z_i^* \in \Gamma_n\}}}{B}$$

approximates $\mu(\Gamma_n)$. There are alternatives to this design, see Cruz-Orive (2001/02) for a tutorial on geometric sampling. For example, instead of generating B independent points $Z_1^*, \ldots, Z_B^*$ from the uniform distribution on $[0,1]^2$, we can proceed as follows. We divide the unit square into b by b cells and generate a uniform random point $x = (x_1, x_2)$ on $[0, 1/b]^2$. Then,

$$\left\{ \left(x_1 + \frac{i}{b}, x_2 + \frac{j}{b} \right),\ i, j = 0, \ldots, b-1 \right\}$$

constitutes a so-called systematic sequence of b^2 points on $[0,1]^2$, $Z_1^*, \ldots, Z_{b^2}^*$, see Figure 4.7. As before,

$$\mu_{b^2}(\Gamma_n) = \frac{\sum_{i=1}^{b^2} \mathbb{I}_{\{Z_i^* \in \Gamma_n\}}}{b^2} \tag{4.3}$$

approximates $\mu(\Gamma_n)$. One of the advantages of the systematic sampling design is that we only need to generate one sample point x. Furthermore, we have noticed in practice that the estimations we obtain with this procedure are more stable than with the Monte Carlo sampling method. Thus, the systematic sampling design allows us to efficiently estimate $\mu(\Gamma_n)$ with fewer points, reducing the computational cost.

Once we have explained the method to estimate $\mu(\Gamma_n)$, the problem reduces to determining whether a given point Z_i^* belongs to Γ_n or not. Let us first consider the estimator L_n^{DW}. It is easy to see that $Z_i^* \in \Gamma_n$ for all ε_n such that

$$\varepsilon_n \geq \max_i \{ \min_j \{ \|Z_i^* - Z_j\|,\ \xi_j = 1 \}, \min_j \{ \|Z_i^* - Z_j\|,\ \xi_j = 0 \} \}.$$

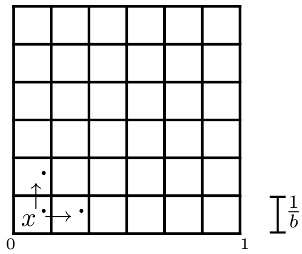


Figure 4.7: *Systematic design.*

Let us now consider the estimator L_n^{α}. Then, Γ_n is defined as the intersection of the dilations of radius ε_n of both $C_\alpha(\mathcal{X}_n)$ and $C_\alpha(\mathcal{Y}_n)$. In Section 4.3 we explained how to compute the α-convex hull of a given sample. In order to determine if Z_i^* belongs to Γ_n we have programmed the function `in.BTnEn`. The arguments of this function are the α-convex hull of $\mathcal{X}_n$ and $\mathcal{Y}_n$, the point Z_i^*, and the radius ε_n. The procedure is as follows. The auxiliary function `dilation` returns the distances $d_{\mathcal{X}}$ and $d_{\mathcal{Y}}$ from Z_i^* to $C_\alpha(\mathcal{X}_n)$ and $C_\alpha(\mathcal{Y}_n)$, respectively. Then, $Z_i^* \in \Gamma_n$ for all ε_n such that

$$\varepsilon_n \geq \max\{d_{\mathcal{X}}, d_{\mathcal{Y}}\}.$$

Figure 4.8 gives insight into the procedure for determining Γ_n for both L_n^{DW} and L_n^{α}. In Figure 4.9 we represent the set Γ_n for the estimator L_n^{DW}.

(a) (b)

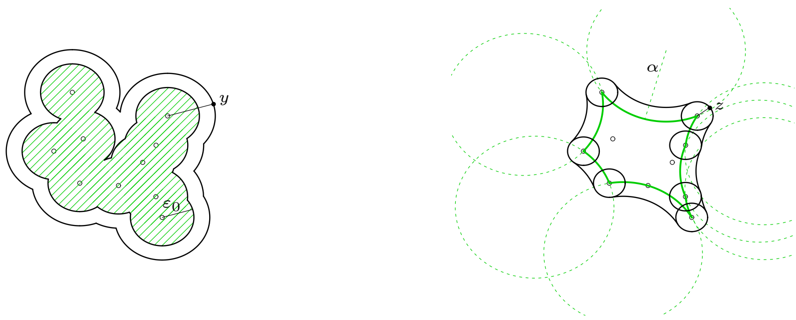


Figure 4.8: *Sample $\mathcal{X}_k = \{X_1, \ldots, X_k\}$ (black dots). (a) The dashed green area corresponds to $\bigcup_{i=1}^k B(X_i, \varepsilon_0)$. The point y belongs to $\bigcup_{i=1}^k B(X_i, \varepsilon)$ for all $\varepsilon \geq \min\{\|y - X_i\|,\ i = 1, \ldots, k\}$. (b) In green boundary of $C_\alpha(\mathcal{X}_k)$. The point z belongs to $B(C_\alpha(\mathcal{X}_k), \varepsilon)$ for all $\varepsilon \geq d(z, C_\alpha(\mathcal{X}_k))$.*

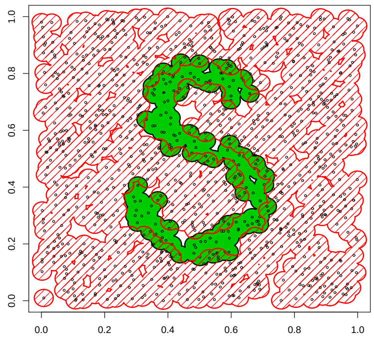


Figure 4.9: *Uniform random sample* (Z_i, ξ_i), $i = 1\ldots, 1000$, *on the unit square. In green* $B(\mathcal{X}_n, \varepsilon_n)$ *and in red (dashed)* $B(\mathcal{Y}_n, \varepsilon_n)$, *for* $\varepsilon_n = 0.03$. *The intersection of both regions,* Γ_n, *estimates* $B(\partial G, \varepsilon_n)$. *The original set is represented in Figure 4.10 (a), below.*

4.3.1 Simulation study

In order to evaluate and compare the behaviour of L_n^α and L_n^{DW}, we have considered two different sets, see Figure 4.10 (a), (b). We denote them by S and C, respectively, referring to their shape. We have chosen these sets for several reasons. First, a ball of radius α_S rolls freely in S and in $\overline{S^c}$ for some $\alpha_S > 0$. The same property holds for C, being $\alpha_C > \alpha_S$ (note that C is wider than S). In fact, $\alpha_S = 0.035$ whereas $\alpha_C = 0.1$ Therefore, S and C fulfill the conditions under which the theoretical properties of the estimators have been obtained. Second, we know for both sets the exact value of the boundary length, see Table 4.1, since they are constructed from a union of arcs. We shall use this information to evaluate the performance of the estimators.

	S	C
Boundary length	3.16	2.51

Table 4.1: *Boundary length of the sets* S *and* C *in Figure 4.10.*

Proceeding as in the simulation study carried out by Cuevas et al. (2007), both estimators L_n^α and L_n^{DW} have been evaluated for 500 samples of sizes $n = 1000$ and $n = 5000$. Regarding the parameters, we have chosen $\alpha_S = 0.03$ and have slightly modified the estimator L_n^α for the set C. Note that, by definition, we should estimate both C and $[0,1]^2 \setminus \mathrm{int}(C)$ with the same parameter α_C. However, it is clear that both sets are α-convex for different values of α. For this reason we estimate $B(\partial C, \varepsilon_n)$ by means of

$$\Gamma_n = B(C_{0.2}(\mathcal{X}_n), \varepsilon_n) \cap B(C_{0.1}(\mathcal{Y}_n), \varepsilon_n).$$

We have considered 250 equidistant values of ε_n from $\varepsilon_n = 0.001$ to $\varepsilon_n = 0.250$. The resampling parameter b, used in the systematic design described at the beginning of this section, is

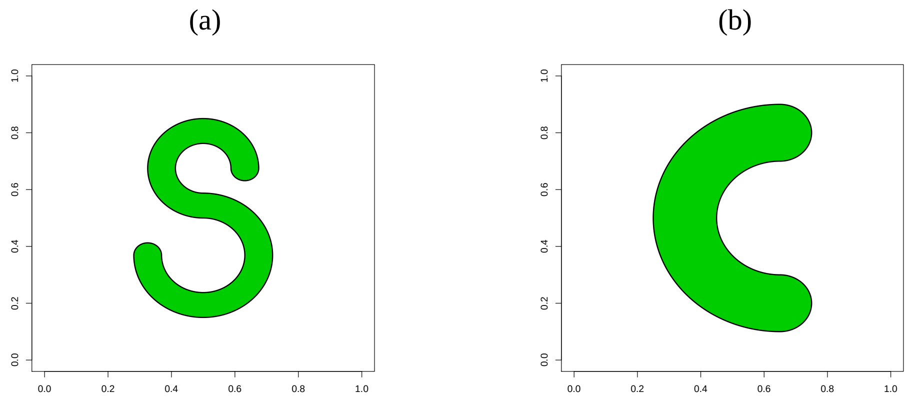


Figure 4.10: *(a) Set S. (b) Set C.*

$b = 40$. Therefore, we approximate $\mu(\Gamma_n)$ by evaluating 1600 points of the unit square, recall the definition of $\mu_{b^2}(\Gamma_n)$ in (4.3). First, we show the results for the set C. Tables 4.2 and 4.3 provide, for $n = 1000$ and some values of ε_n, the average, standard deviation and median of both L_n^α and L_n^{DW} computed from the 500 replications. The same results, for $n = 5000$, are shown in Tables 4.4, 4.5. Finally, we plot the Mean Square Error (MSE) for L_n^α and L_n^{DW} with $n = 1000$, see Figure 4.11 (a), and $n = 5000$, see Figure 4.11 (b).

ε_n	0.001	0.020	0.039	0.058	0.077	0.096	0.115
Average	0.00000	0.81684	1.60635	1.91578	2.07226	2.16685	2.19261
Std. deviation	0.00000	0.12498	0.08362	0.05696	0.04453	0.03623	0.02738
Median	0.00000	0.81250	1.60256	1.91810	2.07386	2.16634	2.19293
ε_n	0.135	0.154	0.173	0.192	0.211	0.230	0.250
Average	2.09326	2.00396	1.93119	1.86917	1.81670	1.76831	1.72203
Std. deviation	0.02214	0.01916	0.01772	0.01607	0.01479	0.01375	0.01250
Median	2.09491	2.00487	1.93280	1.87174	1.81724	1.76902	1.72313

Table 4.2: *Average, standard deviation and median of L_n^α for the set C, based on 500 uniform samples on the unit square with sample size n=1000.*

Regarding the results for the set S, and for the sake of brevity, we only show the summary results corresponding to the sample size $n = 5000$. Tables 4.6 and 4.7 provide, for $n = 5000$ and some values of ε_n, the average, standard deviation and median of both L_n^α and L_n^{DW} computed from the 500 replications.

At this point, we make some comments on the behaviour of the estimator L_n^α. First of all, we must admit that the results are not so good as expected, according to the theoretical properties

ε_n	0.001	0.020	0.039	0.058	0.077	0.096	0.115
Average	0.00000	0.31344	1.29790	1.81036	2.03090	2.14738	2.18399
Std. deviation	0.00000	0.07044	0.08990	0.06049	0.04526	0.03688	0.02873
Median	0.00000	0.31250	1.29808	1.81304	2.03328	2.14518	2.18750
ε_n	0.135	0.154	0.173	0.192	0.211	0.230	0.250
Average	2.08929	2.00265	1.92996	1.86910	1.81660	1.76830	1.72173
Std. deviation	0.02216	0.01923	0.01753	0.01630	0.01509	0.01385	0.01268
Median	2.09028	2.00487	1.93100	1.87012	1.81724	1.76902	1.72250

Table 4.3: *Average, standard deviation and median of L_n^{DW} for the set C, based on 500 uniform samples on the unit square with sample size n=1000.*

ε_n	0.001	0.020	0.039	0.058	0.077	0.096	0.115
Average	0.00625	1.92284	2.21537	2.31475	2.36476	2.39559	2.32049
Std. deviation	0.04379	0.07450	0.04173	0.02872	0.01985	0.01697	0.00994
Median	0.00000	1.93750	2.21955	2.31681	2.36607	2.39583	2.32065
ε_n	0.135	0.154	0.173	0.192	0.211	0.230	0.250
Average	2.18389	2.08127	1.99714	1.92794	1.86857	1.81365	1.76147
Std. deviation	0.00811	0.00698	0.00647	0.00575	0.00585	0.00488	0.00503
Median	2.18519	2.08198	1.99783	1.92708	1.86908	1.81386	1.76125

Table 4.4: *Average, standard deviation and median of L_n^{α} for the set C, based on 500 uniform samples on the unit square with sample size n=5000.*

of the estimator L_n^{α}. A larger sample size might produce more remarkable features although the computational cost substantially increases as n is larger. In spite of this first discouraging impression, we have noticed some interesting peculiarities we next discuss.

1. As happened with the estimator L_n^{DW}, see Cuevas et al. (2007), L_n^{α} underestimates systematically the true value L_0. The bias decreases as the sample size increases.

2. The variance also decreases by increasing the sample size.

3. The bias is the main source of error in L_n^{α}.

4. We have also observed that the values of the mean and the median are very close. This suggests that the sampling distributions are almost symmetric.

Apart from the previous conclusions, we consider that it is interesting to compare the behavior of L_n^{α} and L_n^{DW}. While neither of theses alternatives feature an accurate approximation to the target, we observe that:

ε_n	0.001	0.020	0.039	0.058	0.077	0.096	0.115
Average	0.00000	1.51294	2.11728	2.27876	2.35119	2.39009	2.31759
Std. deviation	0.00000	0.08438	0.04194	0.02651	0.01956	0.01627	0.00988
Median	0.00000	1.51562	2.11538	2.27909	2.34984	2.38932	2.31793

ε_n	0.135	0.154	0.173	0.192	0.211	0.230	0.250
Average	2.18235	2.08006	1.99712	1.92790	1.86846	1.81317	1.76156
Std. deviation	0.00800	0.00741	0.00675	0.00600	0.00569	0.00512	0.00496
Median	2.18287	2.07995	1.99603	1.92871	1.86759	1.81386	1.76125

Table 4.5: *Average, standard deviation and median of L_n^{DW} for the set C, based on 500 uniform samples on the unit square with sample size n=5000.*

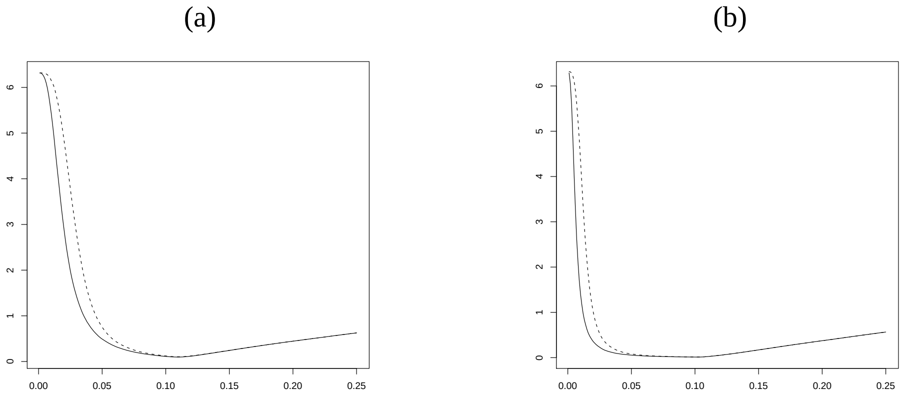


Figure 4.11: *Mean Square Error (MSE) of L_n^{α} (solid line) and L_n^{DW} (dashed line) for the set C. (a) $n = 1000$. (b) $n = 5000$.*

1. In terms of the Mean Square Error (MSE), the estimator L_n^{α} improves the results obtained with L_n^{DW}, see Figure 4.11. These differences are more obvious for small values of the smoothing parameter ε_n and they are mainly due to the bias term, which is smaller for the estimator L_n^{α}.

2. Observe that the differences between both estimators is less obvious as ε_n increases. The reason is that when we dilate both estimators with large values of ϵ_n, the influence of those points of G_n and R_n which are not in $\mathcal{X}_n$ and $\mathcal{Y}_n$ is not noticeable any more.

ε_n	0.001	0.020	0.039	0.058	0.077	0.096	0.115
Average	0.00000	1.94110	2.64593	2.60890	2.36584	2.18777	2.05832
Std. deviation	0.00000	0.12758	0.10116	0.07361	0.05891	0.04732	0.03954
Median	0.00000	1.93750	2.64530	2.61351	2.36797	2.18924	2.05797

ε_n	0.135	0.154	0.173	0.192	0.211	0.230	0.250
Average	1.95846	1.88848	1.82453	1.76424	1.71131	1.66469	1.62216
Std. deviation	0.03427	0.02990	0.02552	0.02259	0.01992	0.01740	0.01478
Median	1.95802	1.88853	1.82466	1.76389	1.71011	1.66377	1.62200

Table 4.6: *Average, standard deviation and median of L_n^α for the set S, based on 500 uniform samples on the unit square with sample size n=5000.*

ε_n	0.001	0.020	0.039	0.058	0.077	0.096	0.115
Average	0.00033	1.90640	2.66134	2.61699	2.36965	2.18995	2.05963
Std. deviation	0.00745	0.13115	0.09951	0.07545	0.05887	0.04949	0.04076
Median	0.00000	1.90000	2.66239	2.61782	2.36797	2.18924	2.06087

ε_n	0.135	0.154	0.173	0.192	0.211	0.230	0.250
Average	1.95762	1.88830	1.82507	1.76487	1.71191	1.66497	1.62230
Std. deviation	0.03527	0.03120	0.02620	0.02288	0.02006	0.01737	0.01498
Median	1.95802	1.88745	1.82370	1.76476	1.71248	1.66522	1.62267

Table 4.7: *Average, standard deviation and median of L_n^{DW} for the set S, based on 500 uniform samples on the unit square with sample size n=5000.*

4.4 Boundary length estimation: the one sample approach

It may be the case that we only have information on the set of interest $S \subset \mathbb{R}^2$, that is, we are provided with a random sample $\mathcal{X}_n$ from a random variable X with support S. In Chapter 1 we have discussed different support estimators, such as the convex hull estimator and the Devroye-Wise estimator. Suppose that we are interested in a geometric characteristic of the set S, for example the boundary length. It seems natural to estimate the set S by means of a support estimator S_n and then compute the boundary length of S_n. The main difference with respect to the problem in Section 4.3 is that now we have to estimate the boundary length from inside the set since we do not have any kind of information about the complement. In this section we compare the results obtained when we estimate the perimeter of a set S by using the α-convex hull estimator and α-shape estimator of a given sample in S. In Section 4.2 we have commented on the implementation of the α-convex hull estimator. The α-shape estimator, however, was briefly introduced in Chapter 1 but nothing was said about its structure or the algorithm to compute it. Before presenting the results of the simulation study we would like to give some insight into the construction of the α-shape. First, recall Definitions 1.5.1, 1.5.2, and 1.5.3. The α-shape of a sample $\mathcal{X}_n$ is defined to be a straight line graph. One of the drawbacks of its definition is that the α-shape does not provide us with a criterion to differentiate between the *inside* and the *outside* of the α-shape. In Figure 4.12 (a) the α-shape of a uniform random sample $\mathcal{X}_n$ on the unit square with $\alpha = 0.13$ is represented. It does not seem clear how to define the interior of the α-shape. On the contrary, the α-convex hull is completely characterized and we can determine whether a given point belongs to the α-convex hull or not. As with the α-convex

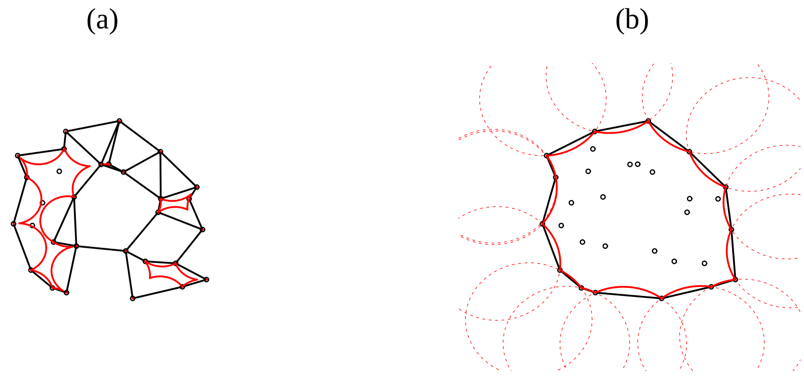


Figure 4.12: *Uniform random sample $\mathcal{X}_n$ of size $n = 30$ on the unit square. The boundary of $C_\alpha(\mathcal{X}_n)$ is represented in solid red. The α-shape is represented by a straight black line graph. (a) $\alpha = 0.13$. (b) $\alpha = 0.3$.*

hull, the value of α affects the α-shape. Thus, the α-shape tends to the convex hull of the sample for large values of α, whereas as α tends to zero the α-shape tends to the empty set. Another drawback of the α-shape is that there exist few theoretical results on its behaviour, mainly due to the difficulty formalizing its definition. In spite of these disadvantages we have decided to study the performance of the length of the α-shape as an estimator of the boundary length, since it is relatively easy to implement and we have observed that it achieves good results.

In Edelsbrunner et al. (1983) it is proved that the α-shape is a subgraph of the Delaunay triangulation defined in Section 4.3. The algorithm for its construction is as follows.

1. Construct the Voronoi diagram and Delaunay triangulation of $\mathcal{X}_n$.

2. Determine the α-extremes of $\mathcal{X}_n$.

   2.1. Determine the convex hull of the sample. The points $X_i \in \mathcal{X}_n$ which lie on the convex hull are α-extremes for all $\alpha > 0$.

   2.2. For each X_i which is not on the convex hull compute the distances from X_i to the vertices v of the Voronoi cell V_i. Then X_i is α-extreme for all α satisfying

$$0 < \alpha \leq \max\{\|X_i - v\|\,,\ v \text{ vertex of } V_i\}.$$

3. Determine the α-neighbours of $\mathcal{X}_n$.

4. Output the α-shape.

We now briefly comment on how to solve step 3 of the algorithm. Given an edge of the Delaunay triangulation $[X_i, X_j]$ and its dual edge of the Voronoi diagram, the extremes X_i and X_j are α-neighbours for all α satisfying

$$\alpha_{\min} \leq \alpha \leq \alpha_{\max}, \tag{4.4}$$

where $\alpha_{\min}$ and $\alpha_{\max}$ are computed from the position of X_i and X_j with respect to the vertices of the dual Voronoi edge. In particular, if the Voronoi edge is a closed line segment $[a, b]$, then $\alpha_{\max} = \max\{\|X_i - a\|, \|X_i - b\|\}$ whereas if the Voronoi edge is a semi-infinite line segment $[a, +\infty)$, then $\alpha_{\max} = \infty$ and equation (4.4) reduces to $\alpha \geq \alpha_{\min}$. Obtaining the value of $\alpha_{\min}$ is a bit trickier. For example, if the Voronoi edge is a closed line segment $[a, b]$, then $\alpha_{\min}$ is not necessarily equal to $\min\{\|X_i - a\|, \|X_i - b\|\}$, see Figure 4.13. Of course, we have programmed all possible values of $\alpha_{\min}$ but we omit here the details since they do not contribute to the understanding of the subject matter.

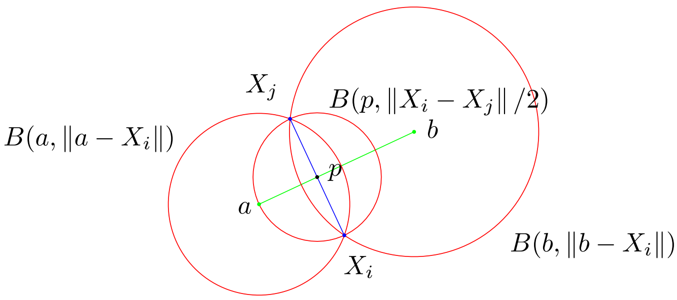


Figure 4.13: *Let $[X_i, X_j]$ be an edge of the Delaunay triangulation and let $[a, b]$ be the corresponding dual Voronoi edge. Then, we can find a ball of radius α such that both X_i and X_j lie on its boundary for all $\alpha_{\min} \leq \alpha \leq \alpha_{\max}$, being $\alpha_{\max} = \max\{\|X_i - a\|, \|X_i - b\|\}$ and for the given example $\alpha_{\min} = \|X_i - X_j\|/2$.*

Regarding the implemented code, the function `alpha.shape` returns the α-shape of a sample. The input arguments are the output of the function `inform.vor.tri` and the value of α. Continuing with the example of Section 4.3, we compute the α-shape of the sample $\mathcal{X}_n$ for $\alpha = 0.3$.

```
> alpha<-0.3
> shape<-alpha.shape(info,alpha)
```

The output object `shape` contains the following components:

```
> names(shape)
[1] "sample"        "info"          "alp.shape"     "alpha"
[5] "alpha.extremes" "possibles"     "length"
```

Among other information, `shape` stores in `shape$alpha.extremes` the indexes of the sample points that are α-extremes. The component `shape$alp.shape` contains the coordinates of each pair of α-neighbours, the corresponding dual Voronoi edges and the values from which $\alpha_{\min}$ and $\alpha_{\max}$ are computed. The length of the α-shape is stored in `shape$length`.

```
> shape$alpha.extremes
[1] 21  7  5  2 30 10  8 22 13 28  1 20 23  9
> shape$length
[1] 3.039860
```

Finally, the function `plot.ashape` produces a plot of the α-shape. If desired it also adds the Voronoi diagram and the Delaunay triangulation of $\mathcal{X}_n$. In Figure 4.14 we show the plot of the α-shape for the discussed example.

```
> plot.ashape(shape,pvor=T,pdel=T,new=T)
```

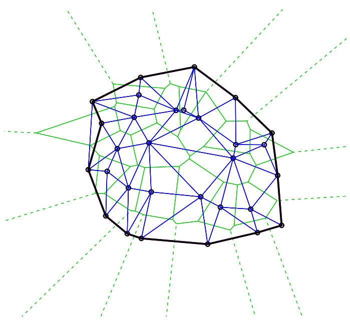


Figure 4.14: *Uniform random sample $\mathcal{X}_n$ of size $n = 30$ on the unit square. In green Voronoi diagram of $\mathcal{X}_n$, in blue Delaunay triangulation and in black α-shape for $\alpha = 0.3$.*

4.4.1 Simulation study

Again, we have considered the same sets S and C defined in Subsection 4.3.1, see Figure 4.10. The setting is, however, different since now we are provided with random samples of points generated into the sets under study but not into their complementaries. We generate 500 uniform samples of size $n = 1000$ on each set and evaluate the estimators for different values of α, see Table 4.8. Summarizing, for each sample $\mathcal{X}_n$ and each value of α we construct $C_\alpha(\mathcal{X}_n)$ and the corresponding α-shape, denoted by $\alpha^{\text{shape}}(\mathcal{X}_n)$. Finally, we compute the boundary length of both estimators. The results corresponding to the set C are summarized in Tables 4.9, 4.10, and

α	0.01	0.03	0.05	0.07	0.09

Table 4.8: *Values of α.*

4.11. Tables 4.12, 4.13, and 4.14 provide the results corresponding to the set S. Recall Table 4.1 for the exact values of the boundary length.

α	0.01	0.03	0.05	0.07	0.09
Average	7.03104	2.70356	2.57810	2.53918	2.52043
Std. deviation	0.22676	0.02905	0.01207	0.00989	0.01036
Median	7.05082	2.69996	2.57740	2.53939	2.52082

Table 4.9: *Average, standard deviation and median of the boundary length of $C_\alpha(\mathcal{X}_n)$, based on 500 uniform samples $\mathcal{X}_n$ on C with sample size n=1000.*

α	0.01	0.03	0.05	0.07	0.09
Average	14.17756	2.56695	2.51315	2.49881	2.49186
Std. deviation	0.38347	0.02445	0.00906	0.00751	0.00673
Median	14.16505	2.56389	2.51289	2.49883	2.49204

Table 4.10: *Average, standard deviation and median of the boundary length of $\alpha^{\text{shape}}(\mathcal{X}_n)$, based on 500 uniform samples $\mathcal{X}_n$ on C with sample size n=1000.*

α	0.01	0.03	0.05	0.07	0.09
$C_\alpha(\mathcal{X}_n)$	20.46161	0.03705	0.00435	0.00077	0.00016
$\alpha^{\text{shape}}(\mathcal{X}_n)$	136.20260	0.00348	0.00008	0.00027	0.00050

Table 4.11: *Mean square error of the boundary length of $C_\alpha(\mathcal{X}_n)$ and $\alpha^{\text{shape}}(\mathcal{X}_n)$, based on 500 uniform samples $\mathcal{X}_n$ on C with sample size n=1000.*

α	0.01	0.03	0.05	0.07	0.09
Average	7.56055	3.31702	3.27558	3.21966	2.66260
Std. deviation	0.22075	0.01844	0.03622	0.02116	0.07993
Median	7.56726	3.31649	3.28373	3.21991	2.65317

Table 4.12: *Average, standard deviation and median of the boundary length of $C_\alpha(\mathcal{X}_n)$, based on 500 uniform samples $\mathcal{X}_n$ on S with sample size n=1000.*

α	0.01	0.03	0.05	0.07	0.09
Average	10.54785	3.19447	3.21567	3.19925	2.74672
Std. deviation	0.35094	0.01527	0.02962	0.09567	0.06215
Median	10.55026	3.19327	3.21205	3.14840	2.73819

Table 4.13: *Average, standard deviation and median of the boundary length of* $\alpha^{\text{shape}}(\mathcal{X}_n)$*, based on 500 uniform samples* $\mathcal{X}_n$ *on* S *with sample size n=1000.*

α	0.01	0.03	0.05	0.07	0.09
$C_\alpha(\mathcal{X}_n)$	19.40276	0.02461	0.01439	0.00386	0.25502
$\alpha^{\text{shape}}(\mathcal{X}_n)$	54.68538	0.00134	0.00384	0.01060	0.17568

Table 4.14: *Mean square error of the boundary length of* $C_\alpha(\mathcal{X}_n)$ *and* $\alpha^{\text{shape}}(\mathcal{X}_n)$*, based on 500 uniform samples* $\mathcal{X}_n$ *on* S *with sample size n=1000.*

We observe, for some values of the parameter α, a significant improvement compared to the results of the simulation study discussed in Section 4.3. Small values of α, however, provide considerably biased estimations, especially in the case of the α-shape, see the first column of Tables 4.10 and 4.13. This fact can be explained by the definition of the α-shape. Recall that the α-shape was defined as the straight line graph whose edges connect α-neighbours. When α is small, a considerable number of interior points of the set turn out to be α-extremes and the α-shape looks like a mesh connecting many of them, see Figure 4.15. As a consequence, the length of the α-shape is large, as it is the result of the addition of many small segments.

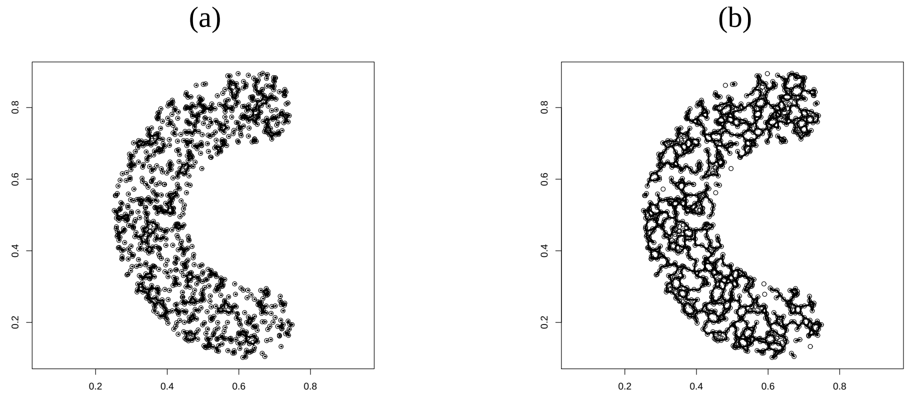


Figure 4.15: *Boundary length estimation with small* α*. (a) The boundary length of* $C_\alpha(\mathcal{X}_n)$ *for* $\alpha = 0.01$ *is* 6.782*. (b) The boundary length of* $\alpha^{\text{shape}}(\mathcal{X}_n)$ *for* $\alpha = 0.01$ *is* 14.053*.*

The estimations are also biased for large values of α (compared to the real values α_C and

α_S). This fact cannot be appreciated in the particular case of the set C since $\alpha_C > 0.09$. Note that a ball of radius α rolls freely in C and in $\overline{C^c}$ for all α in Table 4.8. However, observe the last column of Tables 4.12 and 4.13. It seems that there is an inflection point in the estimations of the boundary length of S. The reason is that $\alpha = 0.09$ is too large and the estimator is not longer able to identify the cavities of the set. For example, the α-shape joins α-extreme points from the upper and lower part of S. The same occurs with $C_\alpha(\mathcal{X}_n)$, see Figure 4.16. As a consequence, the boundary length is underestimated.

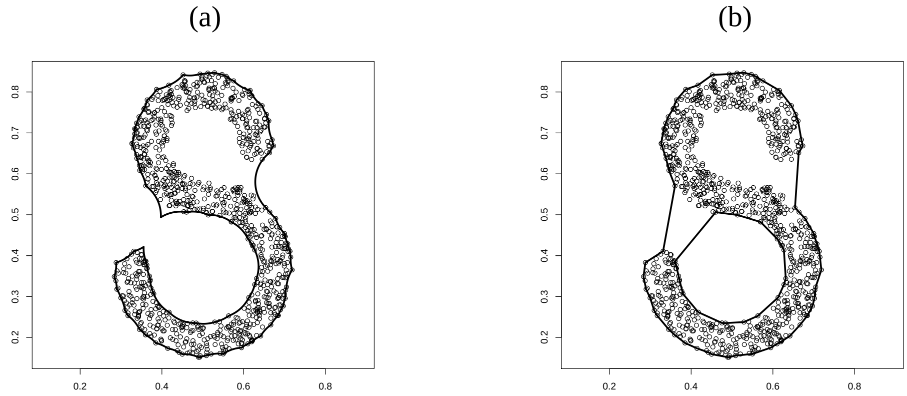


Figure 4.16: *Boundary length estimation with large* α. *(a) The boundary length of* $C_\alpha(\mathcal{X}_n)$ *for* $\alpha = 0.09$ *is* 2.659. *(b) The boundary length of* $\alpha^{\text{shape}}(\mathcal{X}_n)$ *for* $\alpha = 0.09$ *is* 2.724.

Finally, we also include some descriptive graphs. Figures 4.17 and 4.18 show boxplots of the estimates for different values of α. Due to the bias problems explained before, the scale for the case $\alpha = 0.01$ is much higher than for the rest of values of α. For the sake of clarity, we have omitted this case in the plots.

(a) (b)

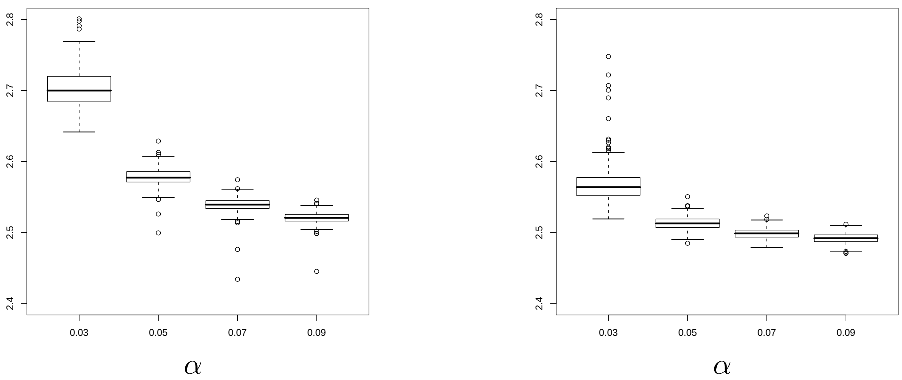


Figure 4.17: *Summary graphs for the set C. (a) Boxplots for the boundary length of $C_\alpha(\mathcal{X}_n)$. (b) Boxplots for the boundary length of $\alpha^{\text{shape}}(\mathcal{X}_n)$.*

(a) (b)

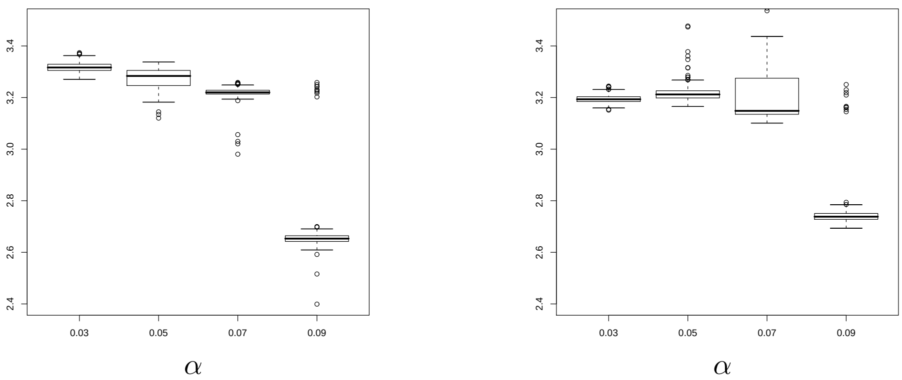


Figure 4.18: *Summary graphs for the set S. (a) Boxplots for the boundary length of $C_\alpha(\mathcal{X}_n)$. (b) Boxplots for the boundary length of $\alpha^{\text{shape}}(\mathcal{X}_n)$.*

Appendix A

Rolling condition, positive reach and α-convexity

The free rolling condition, recall Definition 1.4.5, has useful implications which are worth noting. In this appendix we list some results about the rolling condition that play an important role throughout this work. For example, sufficient conditions relating the rolling condition to the positive reach or the α-convexity of a set are given.

We begin by making some preliminary comments. Assume that a ball of radius $\alpha > 0$ rolls freely in a nonempty closed set $A \subset \mathbb{R}^d$ and let $a \in \partial A$. By definition there exists $x \in A$ such that $a \in B(x, \alpha) \subset A$ and, necessarily, $\|x - a\| = \alpha$. Observe that if $\|x - a\| < \alpha$, then it easily follows that $a \in \mathring{B}(a, \alpha - \|x - a\|) \subset \mathring{B}(x, \alpha) \subset \text{int}(A)$, yielding a contradiction since $a \in \partial A$. Define the unit vector $\eta(a) = (a - x)/\|a - x\|$. Then we can write $B(a - \alpha\eta(a), \alpha) \subset A$ since $x = a - \alpha\eta(a)$. It is important to note that the free rolling condition in A does not imply that the point x and, consequently, the vector $\eta(a)$ are unique, see Figure A.1.

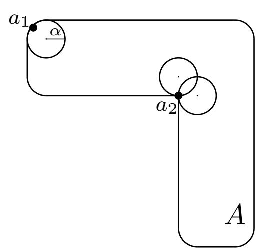


Figure A.1: *A ball of radius α rolls freely in A. For the point $a_1 \in \partial A$ there exists a unique $x \in A$ such that $a_1 \in B(x, \alpha) \subset A$. However, for the point $a_2 \in \partial A$, $a_2 \in B(x, \alpha) \subset A$ for infinite $x \in A$.*

Lemma A.0.1 shows that the uniqueness of the unit vector $\eta(a)$ such that $B(a - \alpha\eta(a), \alpha) \subset A$ is closely related to the existence of some $x \notin A$ such that a coincides with the metric projection of x onto A.

Lemma A.0.1. *Let $A \subset \mathbb{R}^d$ be a nonempty closed set and $a \in \partial A$. Assume that there exists $x \notin A$ such that*

$$\rho = \|x - a\| = d(x, A),$$

that is, a is a metric projection of x onto A. If there exists $\alpha > 0$ and a unit vector $\eta(a)$ such that $B(a - \alpha\eta(a), \alpha) \subset A$, then

$$x = a + \rho\eta(a).$$

Proof. To see this suppose the contrary, that is, suppose that there exists x under the stated conditions such that $x \neq a + \rho\eta(a)$. Then, it can be easily seen that x, a, and $a - \alpha\eta(a)$ cannot lie on the same line and hence,

$$\|a - \alpha\eta(a) - x\| < \|a - \alpha\eta(a) - a\| + \|a - x\| = \alpha + \rho. \tag{A.1}$$

Now, let $z \in \partial B(a - \alpha\eta(a), \alpha) \cap [x, a - \alpha\eta(a)]$, where $[x, a - \alpha\eta(a)]$ denotes the line segment with endpoints x and $a - \alpha\eta(a)$, see Figure A.2. We have

$$\|a - \alpha\eta(a) - x\| = \|a - \alpha\eta(a) - z\| + \|z - x\| = \alpha + \|z - x\|.$$

Therefore, by (A.1)

$$\|z - x\| = \|a - \alpha\eta(a) - x\| - \alpha < \alpha + \rho - \alpha = \rho,$$

which is a contradiction since $z \in A$ and $\rho = d(x, A)$. $\square$

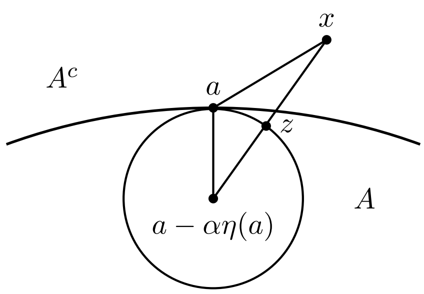


Figure A.2: *Elements of Lemma A.0.1.*

Remark A.0.1. *A direct consequence of Lemma A.0.1 is that the vector $\eta(a)$ is unique, whenever a is the metric projection of some $x \notin A$ onto A. Another interpretation is that if $a \in \partial A$ and there exists more that one ball such that $a \in B(x, \alpha) \subset A$, then a cannot be the metric projection of any point $x \notin A$, see Figure A.3.*

The following lemma shows that the rolling condition guarantees some regularity on the boundary of the set.

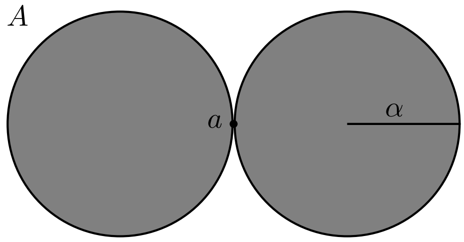


Figure A.3: *For the set A in gray and the point $a \in \partial A$ we can find two unit vectors $\eta(a)$ such that $B(a - \alpha\eta(a), \alpha) \subset A$. It follows from Lemma A.0.1 that a cannot be the metric projection of any $x \notin A$ onto A.*

Lemma A.0.2. *Let $A \subset \mathbb{R}^d$ be a nonempty closed set. Assume that a ball of radius $\alpha > 0$ rolls freely in A. Then,*

$$\text{int}(\overline{A^c}) = A^c \ \textit{and} \ \partial A = \partial\overline{A^c}.$$

Proof. First we prove that $\text{int}(\overline{A^c}) = A^c$. It is straightforward to see that $A^c \subset \text{int}(\overline{A^c})$ by using that A^c is open. Now we prove that $\text{int}(\overline{A^c}) \subset A^c$. Suppose the contrary, that is, suppose that there exists $x \in \text{int}(\overline{A^c})$ such that $x \notin A^c$. Then, $x \in A \cap \overline{A^c} = \partial A$. By the free rolling condition in A, there exists $p \in A$ such that $x \in B(p, \alpha) \subset A$. Moreover, as we have seen $\|x - p\| = \alpha$. Since $x \in \text{int}(\overline{A^c})$, there exists $\varepsilon > 0$ such that $B(x, \varepsilon) \subset \overline{A^c}$. Assume that $\varepsilon < \alpha$ and consider the point

$$y_\lambda = x + \lambda \frac{p - x}{\|p - x\|}, \quad \lambda \in (0, \varepsilon).$$

We have $y_\lambda \in \mathring{B}(p, \alpha) \subset \text{int}(A)$. We get a contradiction since $y_\lambda \in B(x, \varepsilon) \subset \overline{A^c}$. The proof for $\partial A = \partial\overline{A^c}$ is now straightforward if we use that the boundary of a set can be written as the adherence of the set minus its interior. Since A^c is open and $\text{int}(\overline{A^c}) = A^c$, we obtain

$$\partial\overline{A^c} = \overline{\overline{A^c}} \setminus \text{int}(\overline{A^c}) = \overline{A^c} \setminus A^c = \overline{A^c} \setminus \text{int}(A^c) = \partial A^c = \partial A.$$

$\square$

An immediate consequence of Lemma A.0.2 is given in the following lemma.

Lemma A.0.3. *Let $A \subset \mathbb{R}^d$ be a nonempty closed set. Assume that a ball of radius $\alpha > 0$ rolls freely in A. Then,*

$$A = \overline{\overline{A^c}^c}.$$

Proof. The result is a straightforward application of Lemma A.0.2. Use that $\text{int}(\overline{A^c}) = A^c$ to obtain

$$\overline{\overline{A^c}^c} = \text{int}(\overline{A^c})^c = A.$$

$\square$

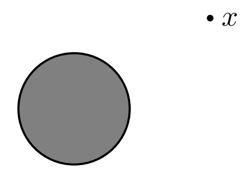


Figure A.4: *For $A = B \cup \{x\}$, we have that $\overline{\overline{A^c}^c} = \overline{\operatorname{int}(A)} = B$. Note that A does not fulfill the free rolling condition in A.*

Remark A.0.2. *The set $\overline{\overline{A^c}^c}$ can also be written as $\overline{\operatorname{int}(A)}$. Since A is closed, it is straightforward to verify that $\overline{\operatorname{int}(A)} \subset A$. We then deduce that the rolling condition in A is essential in order to guarantee that $A \subset \overline{\operatorname{int}(A)}$, since in general this is not true, see Figure A.4.*

From here on, we will assume that $A \subset \mathbb{R}^d$ is a nonempty closed set such that a ball of radius $\alpha > 0$ rolls freely not only in A but also in $\overline{A^c}$. The implications of this assumption are established in Lemmas A.0.5, A.0.6, A.0.7, and A.0.8. First, we would like to comment on the symmetric roles that A and $\overline{A^c}$ play in this assumption. It can be proved that the roles of A and $\overline{A^c}$ are interchangeable in the sense that if a ball of radius $\alpha > 0$ rolls freely in A and in $\overline{A^c}$, then we also have that a ball of radius $\alpha > 0$ rolls freely in $\overline{A^c}$ and in $\overline{\overline{A^c}^c}$. The precise statement is given in Lemma A.0.4, which relies on Lemma A.0.3.

Lemma A.0.4. *Let $A \subset \mathbb{R}^d$ be a nonempty closed set. Assume that a ball of radius $\alpha > 0$ rolls freely in A and in $\overline{A^c}$. Then, a ball of radius $\alpha > 0$ rolls freely in $\overline{A^c}$ and in $\overline{\overline{A^c}^c}$.*

Proof. The result is a direct consequence of Lemma A.0.3 which states that $\overline{\overline{A^c}^c} = A$. □

Lemma A.0.5. *Let $A \subset \mathbb{R}^d$ be a nonempty closed set. Assume that a ball of radius $\alpha > 0$ rolls freely in A and in $\overline{A^c}$. Then, for all $a \in \partial A$ there exists a unique unit vector $\eta(a)$ such that*

$$B(a - \alpha\eta(a), \alpha) \subset A \ \ \textit{and} \ \ B(a + \alpha\eta(a), \alpha) \subset \overline{A^c}.$$

Proof. Let $a \in \partial A$. By the free rolling condition in A, there exists $x \in A$ such that $a \in B(x, \alpha) \subset A$. Moreover, x can be written as $x = a - \alpha\eta(a)$, where $\eta(a) = (a - x)/\left\|a - x\right\|$. By Lemma A.0.2, $\partial A = \partial\overline{A^c}$ and hence $a \in \partial\overline{A^c}$. The free rolling condition in $\overline{A^c}$ yields that there exists $y \in \overline{A^c}$ such that $a \in B(y, \alpha) \subset \overline{A^c}$ and then $\|y - a\| = d(y, A) = \alpha$, that is, a is the metric projection of $y \notin A$ onto A. It follows from Lemma A.0.1 that

$$y = a + \alpha\eta(a),$$

and therefore $B(a + \alpha\eta(a), \alpha) \subset \overline{A^c}$. □

Remark A.0.3. *Note that by Lemma A.0.2 we can conclude that if $B(a + \alpha\eta(a), \alpha) \subset \overline{A^c}$, then $\mathring{B}(a + \alpha\eta(a), \alpha) \subset A^c$, since $\operatorname{int}(\overline{A^c}) = A^c$.*

Next we focus on the relation between the free rolling condition and the positive reach of a set. Recall that the reach of a nonempty set A, $\mathrm{reach}(A)$, is defined as the largest α, possibly infinity, such that if $x \in \mathbb{R}^d$ and $d(x, A) < \alpha$, then the metric projection of x onto A is unique. Lemma A.0.6 states that if A is a nonempty closed subset of $\mathbb{R}^d$ such that a ball of radius α rolls freely in A and in $\overline{A^c}$, then ∂A has positive reach, being $\mathrm{reach}(\partial A) \geq \alpha$. As a consequence every point whose distance to ∂A is lower than α has a unique metric projection onto ∂A.

Lemma A.0.6. *Let $A \subset \mathbb{R}^d$ be a nonempty closed set. Assume that a ball of radius $\alpha > 0$ rolls freely in A and in $\overline{A^c}$. Then, for all $x \in \mathbb{R}^d$ such that $\rho = d(x, \partial A) < \alpha$ there exists a unique point $a \in \partial A$ such that $\|x - a\| = d(x, \partial A)$. That is, the reach of ∂A is greater or equal to α.*

Proof. Let $x \in \mathbb{R}^d$ such that $\rho = d(x, \partial A) < \alpha$. We can assume that $x \notin \partial A$ since the result is trivial otherwise. First, suppose that $x \notin A$. If there exist two metric projections of x onto ∂A, namely a_1 and a_2, then by the free rolling condition in A and by Lemmas A.0.1 and A.0.5, we have that

$$x = a_1 + \rho\eta(a_1) = a_2 + \rho\eta(a_2),$$

where $\eta(a_1)$ and $\eta(a_2)$ are the unique unit vector such that

$$B(a_i - \alpha\eta(a_i), \alpha) \subset A \ \text{ and } \ B(a_i + \alpha\eta(a_i), \alpha) \subset \overline{A^c}, \ \ i = 1, 2.$$

The points $x, a_2 + \alpha\eta(a_2)$, and a_1 cannot lie on the same line. Otherwise

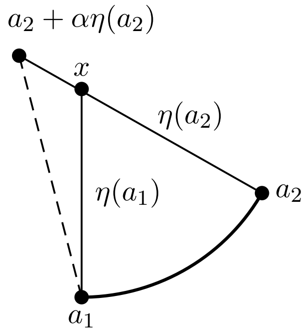
$$a_1 = a_2 + \lambda\eta(a_2)$$

for some $\lambda \in \mathbb{R}$. But by assumption $a_1 = a_2 + \rho\eta(a_2) - \rho\eta(a_1)$ and hence $|\lambda - \rho| = \rho$, that is, $\lambda = 0$ or $\lambda = 2\rho$.

None of these two values is valid. First, $\lambda = 0$ yields $a_1 = a_2$ which is a contradiction since we are assuming that both points are different. Second, $\lambda = 2\rho < 2\alpha$ yields

$$\|a_1 - (a_2 + \alpha\eta(a_2))\| = |2\rho - \alpha| < \alpha,$$

and hence $a_1 \in \mathring{B}(a_2 + \alpha\eta(a_2), \alpha) \subset A^c$, which is another contradiction since $a_1 \in \partial A$. Therefore, x, $a_2 + \alpha\eta(a_2)$, and a_1 do not lie on the same line. Finally, using the strict triangle inequality and $\rho \leq \alpha$ we have that



$$\|a_1 - (a_2 + \alpha\eta(a_2))\| < \|a_1 - x\| + \|x - (a_2 + \alpha\eta(a_2))\| = \rho + (\alpha - \rho) = \alpha.$$

This is again a contradiction since $a_1 \in \partial A$. Therefore, the projection onto ∂A of $x \notin A$ such that $\rho = d(x, \partial A) < \alpha$ is unique. Now suppose that $x \in A$. Since we are assuming that $x \notin \partial A$ it can be easily seen that $x \notin \overline{A^c}$. Moreover, $\partial\overline{A^c} = \partial A$ by Lemma A.0.2 and hence $d(x, \partial\overline{A^c}) < \alpha$. The result is now straightforward if we repeat the same steps as before and use that, by Lemma A.0.4, the roles of A and $\overline{A^c}$ are interchangeable.

□

Therefore, Lemma A.0.6 proves that a sufficient condition for $\partial A \subset \mathbb{R}^d$ to have positive reach is that a ball of radius $\alpha > 0$ rolls freely in A and in $\overline{A^c}$. It is convenient to note, as it is shown in Figure A.5, that it is not enough that a ball of radius α rolls freely in A in order to guarantee that $\mathrm{reach}(\partial A) \geq \alpha$. The same occurs if a ball of radius α only rolls freely in $\overline{A^c}$, see Figure A.6. In Lemma A.0.7 we state a useful application of Lemma A.0.6.

Lemma A.0.7. *Let $A \subset \mathbb{R}^d$ be a nonempty closed set. Assume that a ball of radius $\alpha > 0$ rolls freely in A and in $\overline{A^c}$. Then A and $\overline{A^c}$ are both sets with positive reach, being $\mathrm{reach}(A)$ and $\mathrm{reach}(\overline{A^c})$ greater or equal to α.*

Proof. The result is an immediate consequence of Lemma A.0.6. □

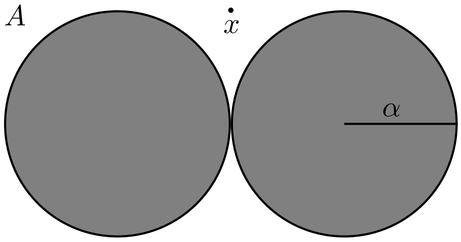


Figure A.5: *A ball of radius α rolls freely in A, $d(x, \partial A) < \alpha$, and the metric projection of x onto ∂A is not unique.*

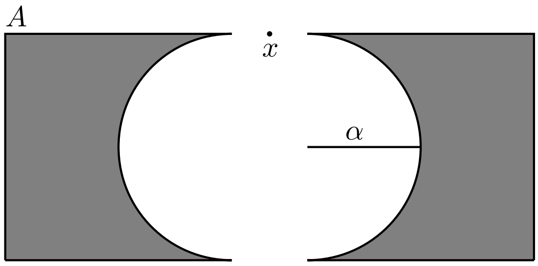


Figure A.6: *A ball of radius α rolls freely in $\overline{A^c}$, $d(x, \partial A) < \alpha$, and the metric projection of x onto ∂A is not unique.*

Finally, it remains to establish the relation between the rolling condition and the α-convexity, recall Definition 1.4.1. Lemma A.0.8 states the result.

Lemma A.0.8. *Let $A \subset \mathbb{R}^d$ be a nonempty closed set. Assume that a ball of radius $\alpha > 0$ rolls freely in A and in $\overline{A^c}$. Then A and $\overline{A^c}$ are both α-convex.*

Proof. First we shall prove that $A = C_\alpha(A)$. Since by definition $A \subset C_\alpha(A)$, it suffices to show that if $x \in A^c$ then $x \notin C_\alpha(A)$. Thus, let $x \in A^c$ and $\rho = d(x, A)$. If $\rho \geq \alpha$, then

$x \in \mathring{B}(x, \alpha) \subset A^c$ and therefore $x \notin C_\alpha(A)$. If $\rho < \alpha$, then by Lemmas A.0.7 and A.0.5 there exists a unique point $a \in \partial A$ and a unique unit vector $\eta(a)$ such that $x = a + \rho\eta(a)$ and

$$x \in \mathring{B}(a + \alpha\eta(a), \alpha) \subset A^c,$$

which yields $x \notin C_\alpha(A)$. It remains to proof that $\overline{A^c}$ is α-convex. The result is an immediate consequence of the latter and Lemma A.0.3. □

Remark A.0.4. *The converse of Lemma A.0.8 may fail, that is, we may find sets A such that A and $\overline{A^c}$ are both α-convex but do not satisfy the rolling condition in A and in $\overline{A^c}$. See for example Figure A.4, where the sets $A = B \cup \{x\}$ and $\overline{A^c} = \mathbb{R}^2 \setminus \mathring{B}$ are both α-convex for $\alpha = 1$. However, a ball of radius 1 does not roll freely in A because of the point x.*

Appendix B

Closing of a sample with respect to open and closed balls

Chapter 2 focused on the study of a set estimator for a compact set $S \subset \mathbb{R}^d$ under the assumption of α-convexity. As has been argued, the α-convex hull serves as basis to define a natural estimator in this context. Thus, given a random sample $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ from a random variable with support S, we defined

$$S_n = (\mathcal{X}_n \oplus r_n \mathring{B}) \ominus r_n \mathring{B}.$$

Note that, according to Definition 1.4.3, S_n is the closing of $\mathcal{X}_n$ with respect to $\mathring{B}(0, r_n)$. We have pointed out that the method to bound $\mathbb{E}(d_\mu(S, S_n))$ can be simplified if instead of S_n as defined above we considered

$$S_n = (\mathcal{X}_n \oplus r_n B) \ominus r_n B.$$

Although we have not introduced the precise definition in Chapter 1, the latter estimator S_n is the closing of $\mathcal{X}_n$ with respect to $B(0, r_n)$. The result of this appendix states that, with probability one, both estimators are equal. More precisely, geometric arguments yield that if the closing of a sample $\mathcal{X}_n$ with respect to the open ball $\mathring{B}(0, r)$ is not the same as the closing of $\mathcal{X}_n$ with respect to $B(0, r)$, then there exist $y \in \mathbb{R}^d$ and $d + 1$ sample points whose distance to y is exactly equal to r. The proof concludes by showing that, under the assumption of an absolutely continuous distribution, this is a zero-probability event. In short, it makes no difference whether we consider S_n defined with open or closed balls.

Lemma B.0.9. *Let $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ be a random sample from X, where X denotes a random variable in $\mathbb{R}^d$ with absolutely continuous distribution P_X. For any $r > 0$, let*

$$C_r(\mathcal{X}_n) = (\mathcal{X}_n \oplus r\mathring{B}) \ominus r\mathring{B} \quad \textit{and} \quad C_r^B(\mathcal{X}_n) = (\mathcal{X}_n \oplus rB) \ominus rB.$$

Then, with probability one,

$$C_r(\mathcal{X}_n) = C_r^B(\mathcal{X}_n).$$

Proof. It is straightforward to verify that $C_r(\mathcal{X}_n) \subset C_r^B(\mathcal{X}_n)$ and hence

$$P(C_r(\mathcal{X}_n) \neq C_r^B(\mathcal{X}_n)) = P(C_r^B(\mathcal{X}_n) \setminus C_r(\mathcal{X}_n) \neq \emptyset).$$

By using the definitions of the Minkowski addition and Minkowski subtraction it can be easily proved that if $C_r^B(\mathcal{X}_n) \setminus C_r(\mathcal{X}_n) \neq \emptyset$, then

$$\operatorname{int}\left(\bigcup_{i=1}^{n} B(X_i, r)\right) \neq \bigcup_{i=1}^{n} \mathring{B}(X_i, r). \tag{B.1}$$

Note that $\bigcup_{i=1}^{n} \mathring{B}(X_i, r) \subset \operatorname{int}\left(\bigcup_{i=1}^{n} B(X_i, r)\right)$ and hence (B.1) implies that there exists $y \in \mathbb{R}^d$ such that

$$y \in \operatorname{int}\left(\bigcup_{i=1}^{n} B(X_i, r)\right) \tag{B.2}$$

and, for all $i \in \{1, \ldots, n\}$,

$$\|y - X_i\| \geq r. \tag{B.3}$$

It follows from (B.2) that there exists $\delta > 0$ such that $B(y, \delta) \subset \bigcup_{i=1}^{n} B(X_i, r)$. In particular, there exists $i_1 \in \{1 \ldots, n\}$ such that $\|y - X_{i_1}\| \leq r$ and by (B.3) we have that

$$\|y - X_{i_1}\| = r.$$

Now, define

$$u_1 = \frac{y - X_{i_1}}{\|y - X_{i_1}\|} \tag{B.4}$$

and let $y_{1,m} = y + \varepsilon_m u_1$ be a sequence of vectors, where ε_m represents a sequence of positives numbers that converges to zero as m tends to infinity. Then, $y_{1,m} \notin B(X_{i_1}, r)$ since

$$\|y_{1,m} - X_{i_1}\| = \|y + \varepsilon_m u_1 - X_{i_1}\| = \|(\varepsilon_m + \|y - X_{i_1}\|)u_1\| = \varepsilon_m + r > r, \tag{B.5}$$

for all $m \in \mathbb{N}$. By construction $\|y - y_{1,m}\| = \varepsilon_m$ and hence, for large enough m, $y_{1,m} \in B(y, \delta) \subset \bigcup_{i=1}^{n} B(X_i, r)$, that is, there exists $i_m \in \{1, \ldots, n\}$ such that $\|y_{1,m} - X_{i_m}\| \leq r$. Note that (B.5) yields $i_m \neq i_1$. Now, since the sequence $\{X_{i_m}\}_m$ has a finite range, then there must be at least one value which is taken on by infinitely many terms of $\{X_{i_m}\}_m$. More precisely, there exists a constant subsequence $\{X_{i_{m_l}}\}_l$. For all l,

$$r \leq \left\|y - X_{i_{m_l}}\right\| \leq \|y - y_{1,m}\| + \left\|y_{1,m} - X_{i_{m_l}}\right\| = \varepsilon_{m_l} + r.$$

Let $X_{i_{m_l}} \equiv X_{i_2}$ and, since ε_{m_l} tends to zero as l tends to infinity, the latter expression yields

$$\|y - X_{i_2}\| = r. \tag{B.6}$$

Now, if $d > 2$ we proceed as follows. Iteratively, for $k \in \{2, \ldots, d\}$ let us consider the $(k-1)$-hyperplane Π_k defined by the points $X_{i_1}, \ldots, X_{i_k}$. Note that, since the distribution is absolutely continuous, Π_k is well defined with probability one. Let $P_{\Pi_k} y$ be the projection of y onto Π_k. Note that, since $\left\|y - X_{i_j}\right\| = r$ for all $j \le k$, it is straightforward from the Pythagorean theorem that $P_{\Pi_k} y$ is the unique point in Π_k equidistant from $X_{i_1}, \ldots, X_{i_k}$. More precisely, for all $j \le k$,

$$\left\|P_{\Pi_k} y - X_{i_j}\right\|^2 = \left\|y - X_{i_j}\right\|^2 - \|y - P_{\Pi_k} y\|^2 = r^2 - \|y - P_{\Pi_k} y\|^2 .$$

For instance, for $k = 2$, $P_{\Pi_2} y$ is the middle point of the segment defined by X_{i_1} and X_{i_2}. Similarly, for $k = 3$, $P_{\Pi_3} y$ is the circumcentre of the triangle defined by X_{i_1}, X_{i_2}, and X_{i_3}. Define

$$u_k = \frac{y - P_{\Pi_k} y}{\|y - P_{\Pi_k} y\|}$$

whenever $y \neq P_{\Pi_k} y$. Otherwise, let u_k be any unit normal vector to Π_k at the point $y \in \Pi_k$. Let $y_{k,m} = y + \varepsilon_m u_k$ be a sequence of vectors, where ε_m represents a sequence of positive numbers that converges to zero as m tends to infinity. Then, $y_{k,m} \notin B(X_{i_j}, r)$ for $j = 1, \ldots, k$, since

$$\begin{aligned}
\left\|y_{k,m} - X_{i_j}\right\|^2 &= \|y_{k,m} - P_{\Pi_k} y\|^2 + \left\|P_{\Pi_k} y - X_{i_j}\right\|^2 \\
&= \|y + \varepsilon_m u_k - P_{\Pi_k} y\|^2 + \left\|P_{\Pi_k} y - X_{i_j}\right\|^2 \\
&= (\varepsilon_m + \|y - P_{\Pi_k} y\|)^2 + \left\|P_{\Pi_k} y - X_{i_j}\right\|^2 \\
&\ge \varepsilon_m^2 + \|y - P_{\Pi_k} y\|^2 + \left\|P_{\Pi_k} y - X_{i_j}\right\|^2 \\
&= \varepsilon_m^2 + r^2 > r, \qquad \text{(B.7)}
\end{aligned}$$

for all $m \in \mathbb{N}$. Furthermore, by construction $\|y - y_{k,m}\| = \varepsilon_m$ and therefore, for large enough m, $y_{k,m} \in B(y, \delta) \subset \bigcup_{i=1}^n B(X_i, r)$, that is, there exists $i_m \in \{1, \ldots, n\}$ such that $\|y_{k,m} - X_{i_m}\| \le r$. Note that by (B.7) $i_m \neq i_j$ for $j = 1, \ldots, k$. We use a similar argument as in the proof of (B.6) to establish that there exists $i_{k+1} \in \{1, \ldots, n\}$ such that $i_{k+1} \neq i_j$ for $j = 1, \ldots, k$ and

$$\left\|y - X_{i_{k+1}}\right\| = r.$$

To summarize, if $C_r^B(\mathcal{X}_n) \setminus C_r(\mathcal{X}_n) \neq \emptyset$, then we have proved that there exist $y \in \mathbb{R}^d$ and $d+1$ sample points $X_{i_1}, \ldots, X_{i_{d+1}}$ such that $\left\|y - X_{i_j}\right\| = r$ for all $j = 1, \ldots, d+1$. This implies that

$$P(C_r^B(\mathcal{X}_n) \setminus C_r(\mathcal{X}_n) \neq \emptyset) \le P(\exists y, X_{i_1}, \ldots, X_{i_{d+1}} : \left\|y - X_{i_j}\right\| = r, j = 1, \ldots, d+1).$$

Note that

$$\begin{aligned}
&\exists y, X_{i_1}, \ldots, X_{i_{d+1}} : \left\|y - X_{i_j}\right\| = r, j = 1, \ldots, d+1 \\
\Leftrightarrow\ &\exists X_{i_1}, \ldots, X_{i_{d+1}} : \left\|c(X_{i_1}, \ldots, X_{i_{d+1}}) - X_{i_j}\right\| = r, j = 1, \ldots, d+1,
\end{aligned}$$

where $c(X_{i_1}, \ldots, X_{i_{d+1}})$ denotes the unique point equidistant from $X_{i_1}, \ldots, X_{i_{d+1}}$. Therefore,

$$P(C_r^B(\mathcal{X}_n) \setminus C_r(\mathcal{X}_n) \neq \emptyset) \le \binom{n}{d+1} P(\|c(X_1, \ldots, X_{d+1}) - X_1\| = r).$$

In order to compute $P(\|c(X_1,\ldots,X_{d+1}) - X_1\| = r)$ we apply Lemma B.0.10, which states that if $\|c(X_1,\ldots,X_{d+1}) - X_1\| = r$, then $c(X_1,\ldots,X_{d+1}) \in \{p_1,p_2\}$, where p_1 and p_2 depend only on $X_2,\ldots,X_{d+1}$. Note that a random set of $d+1$ points from a d-variate continuous distribution are noncoplanar with probability one. Thus, the convex hull of a random set of $d+1$ such points is a simplex with probability one. Therefore we are under the conditions in Lemma B.0.10 and

$$P(\|c(X_1,\ldots,X_{d+1}) - X_1\| = r) \leq \sum_{i=1}^{2} P(\|p_i(X_2,\ldots,X_{d+1}) - X_1\| = r)$$

$$= \sum_{i=1}^{2} \mathbb{E}\left(P_X(\|p_i(X_2,\ldots,X_{d+1}) - X_1\| = r | X_2,\ldots,X_{d+1})\right) = 0$$

since $\mu\{x \in \mathbb{R}^d : \|y - x\| = r\} = 0$ for all $y \in \mathbb{R}^d$ and P_X is absolutely continuous with respect to μ. Therefore, we have proved that

$$P(C_r^B(\mathcal{X}_n) \setminus C_r(\mathcal{X}_n) \neq \emptyset) = 0$$

and the proof is complete. □

Lemma B.0.10. *Let $X_1 \ldots, X_{d+1}$ be the vertices of a d-dimensional simplex in $\mathbb{R}^d$ and let $c(X_1,\ldots,X_{d+1})$ be the unique point in $\mathbb{R}^d$ which is equidistant to $X_1,\ldots,X_{d+1}$. Assume that $\|c(X_1,\ldots,X_{d+1}) - X_1\| = r$ for fixed $r > 0$. Then*

$$c(X_1,\ldots,X_{d+1}) \in \{p_1,p_2\},$$

where $p_i \in \mathbb{R}^d$ do not depend on X_1, that is, $p_i \equiv p_i(X_2,\ldots,X_{d+1})$, $i = 1,2$.

Proof. Let Π_d be the $(d-1)$-hyperplane defined by $X_2 \ldots, X_{d+1}$ and let $c(X_2,\ldots,X_{d+1})$ be the unique point in Π_d equidistant to $X_2 \ldots, X_{d+1}$. If $c(X_1,\ldots,X_{d+1}) \in \Pi_d$, then the result follows immediately since $c(X_1,\ldots,X_{d+1}) = c(X_2,\ldots,X_{d+1})$ and

$$p_1 = p_2 = c(X_2,\ldots,X_{d+1}).$$

Otherwise, define for $i = 1,2$,

$$p_i = c(X_2,\ldots,X_{d+1}) + \sqrt{r^2 - \|c(X_2,\ldots,X_{d+1}) - X_2\|^2}\, u_i, \tag{B.8}$$

where $u_i, i = 1,2$ are the two possible unit normal vectors to Π_d at $c(X_2,\ldots,X_{d+1})$, that is, $u_1 = -u_2$. We shall see now that $c(X_1,\ldots,X_{d+1}) = p_i$ for $i = 1$ or $i = 2$. Let $P_{\Pi_d} c(X_1,\ldots,X_{d+1})$ be the projection of $c(X_1,\ldots,X_{d+1})$ onto Π_d. To simplify notation we abbreviate $P_{\Pi_d} c(X_1,\ldots,X_{d+1})$ to $P_{\Pi_d} c$. Then

$$c(X_1,\ldots,X_{d+1}) = P_{\Pi_d} c + \|c(X_1,\ldots,X_{d+1}) - P_{\Pi_d} c\| \frac{c(X_1,\ldots,X_{d+1}) - P_{\Pi_d} c}{\|c(X_1,\ldots,X_{d+1}) - P_{\Pi_d} c\|} \tag{B.9}$$

It follows from the Pythagorean theorem that, for all $j = 2, \ldots, d+1$,

$$\begin{aligned} \|P_{\Pi_d} c - X_j\|^2 &= \|c(X_1, \ldots, X_{d+1}) - X_j\|^2 - \|c(X_1, \ldots, X_{d+1}) - P_{\Pi_d} c\|^2 \\ &= r^2 - \|c(X_1, \ldots, X_{d+1}) - P_{\Pi_d} c\|^2 \,, \end{aligned} \tag{B.10}$$

that is, $P_{\Pi_d} c \in \Pi_d$ is equidistant to $X_2, \ldots, X_{d+1}$ and hence $P_{\Pi_d} c = c(X_2, \ldots, X_{d+1})$. If we replace $P_{\Pi_d} c$ in (B.9), then

$$c(X_1, \ldots, X_{d+1}) = c(X_2, \ldots, X_{d+1}) + \|c(X_1, \ldots, X_{d+1}) - P_{\Pi_d} c\| \, u,$$

where

$$u = \frac{c(X_1, \ldots, X_{d+1}) - P_{\Pi_d} c}{\|c(X_1, \ldots, X_{d+1}) - P_{\Pi_d} c\|}$$

is a unit normal vector to Π_d at $c(X_2, \ldots, X_{d+1})$ and by (B.10)

$$\|c(X_1, \ldots, X_{d+1}) - P_{\Pi_d} c\|^2 = r^2 - \|P_{\Pi_d} c - X_2\|^2 = r^2 - \|c(X_2, \ldots, X_{d+1}) - X_2\|^2 \,.$$

Therefore, $c(X_1, \ldots, X_{d+1})$ corresponds to one of the p_i defined in (B.8). This completes the proof of the lemma. $\square$

Appendix C

The alphahull Package

The alphahull package is the result of the implementation of the estimators discussed throughout this dissertation. Over the last years, the R computing environment has become a powerful scientific tool that offers a rich collection of classical and modern statistical modeling techniques. Motivated by its flexibility and its widely acceptance among the scientific community, we have chosen R as programming language to develop this library of functions.

Title Generalization of the convex hull of a sample of points in the plane

Version 1.0

Date 2008-03-01

Author Beatriz Pateiro-López, Alberto Rodríguez-Casal

Maintainer Beatriz Pateiro-López <beatriz.pateiro@usc.es>

Depends R, tripack

Description This package computes the alpha-shape and alpha-convex hull of a given sample of points in the plane. The concepts of alpha-shape and alpha-convex hull generalize the definition of the convex hull. The programming is based on the Voronoi diagram and Delaunay triangulation of the sample. The package also includes functions to calculate the dilation of the alpha-convex hull of a given sample and to determine whether a point belongs to it. A function to estimate the Minkowsky content of a compact set is also included.

License R functions: GPL, Fortran code: ACM, free for noncommercial use

R topics documented:

`add.voronoi` *Voronoi diagram*

Description

This function adds the Voronoi diagram to an open plot.

Usage

```
add.voronoi(mat.info, ...)
```

Arguments

`mat.info`	Output matrix from the `inform.vor.del` function, see Details.
`...`	Arguments to be passed to methods, such as graphical parameters (see `par`).

Details

The input matrix `mat.info` is one of the arguments included in the output list that the function `inform.vor.del` returns. It contains all the necessary information of the Delaunay triangulation and Voronoi diagram. For each edge of the Delaunay triangulation `mat.info` contains the indexes and coordinates of the sample points that form the edge, the indexes and coordinates of the extremes of the corresponding segment in the Voronoi diagram, and an indicator that takes the value 1 for those extremes of the Voronoi diagram that represent a boundless extreme. The semi-infinite segments of the Voronoi diagram are represented with dashed lines.

See Also

`inform.vor.tri`.

Examples

```
# Simple example from TRIPACK
data(tritest)
sample<-matrix(c(tritest$x,tritest$y),nc=2,byrow=F)
plot(sample[,1],sample[,2],xlim=c(-1,2),ylim=c(-1,2))
# Delaunay triangulation and Voronoi diagram calculation
info<-inform.vor.tri(sample)
# Add Voronoi diagram
add.voronoi(info$mat.info,col=3)

# Random sample in the unit square
sample<-matrix(runif(20),nc=2)
plot(sample[,1],sample[,2],xlim=c(-1,2),ylim=c(-1,2))
# Delaunay triangulation and Voronoi diagram calculation
info<-inform.vor.tri(sample)
# Add Voronoi diagram
add.voronoi(info$mat.info,col=3)
```

`alpha.hull` *alpha-convex hull calculation*

Description

This function calculates the boundary of the α-convex hull of a given sample, from the complement of the α-convex hull.

Usage

```
alpha.hull(shape, compl)
```

Arguments

`shape`	Output list from the `alpha.shape` function.
`compl`	Output matrix from the `complement` function.

Details

The boundary of the α-convex hull is formed by arcs of the open balls that define the complement of the α-convex hull. The arcs are determined by the intersections of some of these balls. The extremes of an arc are given by $c+rA_\theta v$ and $c+rA_{-\theta}v$ where c and r represent the centre and radius of the arc, repectively and $A_\theta v$ represents the clockwise rotation of angle θ of the unitary vector v.

Value

A list with the following components:

`sample`	A 2-column matrix with the coordinates of the sample points.
`ahull.archs`	For each arc in the boundary of the α-convex hull, `ahull.archs` contains the centre and radius of the arc, the unitary vector v and the angle θ that define the arc, see Details.
`length`	Length of the boundary of the α-convex hull, see `length.ahull`.
`ashape`	Output list from the `alpha.shape` function.
`alpha`	Value of α.
`complement`	Output matrix from the `complement` function.

See Also

`alpha.shape`, `complement`, `rotation.cw`, `inter`, `length.ahull`, `plot.ahull`.

Examples

```
# Random sample in the unit square
sample<-matrix(runif(100),nc=2)
# value of alpha
alpha<-0.2
# Triangulation information
info<-inform.vor.tri(sample)
# alpha-shape
shape<-alpha.shape(info,alpha)
# Complement of the alpha-convex hull and alpha-hull boundary
compl<-complement(alpha,info$mat.info)
ahull<-alpha.hull(shape,compl)
```

`alpha.shape` *alpha-shape calculation*

Description

This function calculates the α-shape of a given sample.

Usage

```
alpha.shape(info, alpha)
```

Arguments

`info`	Output list from the `inform.vor.tri` function.
`alpha`	Value of α.

Details

The α-shape is implemented with the algorithm described in Edelsbrunnner *et al.* (1983).

Value

A list with the following components:

`sample`	A 2-column matrix with the coordinates of the sample points.
`info`	Output list from the `inform.vor.tri` function.
`alp.shape`	A *nseg*-row matrix with the coordinates and indexes of the edges of the Delaunay triangulation that form the α-shape. The number of rows *nseg* coincides with the number of segments of the α-shape. The matrix also includes information of the Voronoi extremes corresponding to each segment.
`alpha`	Value of α.
`alpha.extremes`	Vector with the indexes of the sample points that are α-extremes. See Edelsbrunnner *et al.* (1983).
`possibles`	Matrix with the coordinates and indexes of the edges of the Delaunay triangulation that are candidates to form the α-shape. It includes the edges whose extremes are α-extremes, not necessarily α-neighbours.
`length`	Length of the α-shape.

References

Edelsbrunner, H., Kirkpatrick, D.G. and Seidel, R. (1983) *On the shape of a set of points in the plane*. IEEE Transactions on Information Theory, Vol IT-29, No. 4.

See Also

`inform.vor.tri`, `plot.ashape`.

Examples

```
# Uniform sample of size n=300 on the disc B(c,0.5)\B(c,0.25),
# with c=(0.5,0.5).
n<-300
m<-0
sample<-matrix(0,n,2)
while(m<n){
x<-runif(1)
y<-runif(1)
d<-(x-0.5)^2+(y-0.5)^2
if((d<=(0.5)^2)&(d>=(0.25)^2)){
m<-m+1
sample[m,]<-c(x,y)
}
}
# Value of alpha
alpha<-0.1
# Triangulation information
info<-inform.vor.tri(sample)
# alpha-shape
shape<-alpha.shape(info,alpha)
```

`alphahull-package` *Generalization of the convex hull of a sample of points in the plane*

Description

This package computes the α-shape and α-convex hull of a given sample of points in the plane. The concepts of α-shape and α-convex hull generalize the definition of the convex hull. The programming is based on the Voronoi diagram and Delaunay triangulation of the sample. The package also includes functions to calculate the dilation of the α-convex hull of a given sample and to determine whether a point belongs to it. A function to estimate the Minkowsky content of a compact set is also included.

Details

Package:	alphahull
Type:	Package
Version:	1.0
Date:	2008-03-01
License:	R functions: GPL, Fortran code: ACM, free for noncommercial use

Author(s)

Beatriz Pateiro-López, Alberto Rodríguez-Casal.

Maintainer: Beatriz Pateiro-López <beatriz.pateiro@usc.es>

`angs.arch` *Angles of the extremes of an arc*

Description

Given a vector v and an angle θ, `angs.arch` returns the angles that $A_\theta v$ and $A_{-\theta} v$ form with the axis *OX*, where $A_\theta v$ represents the clockwise rotation of angle θ of the vector v.

Usage

```
angs.arch(v, theta)
```

Arguments

`v`	Vector v in the plane.
`theta`	Angle θ.

Details

The angle that forms the vector v with the axis *OX* takes its value in $[0, 2\pi)$.

Value

`angs`	Numeric vector with two components.

Examples

```
# Let v=c(0,1) and theta=pi/4
# Consider the arc such that v is the internal angle bisector
# that divides the angle 2*theta into two equal angles
# The angles that the arc forms with the OX axis are pi/4
# and 3*pi/4
v<-c(0,1)
theta<-pi/4
angs.arch(v,theta)
```

`arch` *Add an arc to a plot*

Description

This function adds the arc of $B(c, r)$ between angles $\theta 1$ and $\theta 2$ to a plot.

Usage

```
arch(c1, c2, r, theta1, theta2, col=1, lwd=3, lty=1,...)
```

Arguments

`c1`	*X*-coordinate of the centre.
`c2`	*Y*-coordinate of the centre.
`r`	Radius of the ball.
`theta1`	Angle that forms the vector defining one of the extremes of the arc with the axis *OX*.
`theta2`	Angle that forms the vector defining one of the extremes of the arc with the axis *OX*.
`col`	Color parameter, by default black.
`lwd`	Line width, by default 3.
`lty`	Line type, by default solid.
`...`	Arguments to be passed to methods, such as graphical parameters (see `par`).

See Also

`plot.ahull`.

Examples

```
# Plot of the circumference of radius 1
theta<-seq(0,2*pi,length=100)
r<-1
plot(r*cos(theta),r*sin(theta),type="l")
# Add in red the arc between pi/4 and 3*pi/4
arch(0,0,1,pi/4,3*pi/4,col=2,lwd=3)
```

`complement` *Complement of the alpha-convex hull*

Description

This function calculates the complement of the α-convex hull of a given sample.

Usage

```
complement(alpha, mat.coor)
```

Arguments

`alpha`	Value of α.
`mat.coor`	Output matrix from the `inform.vor.tri` function.

Details

The complement of the α-convex hull is calculated as the union of open balls and halfplanes that do not contain any point of the sample. See Edelsbrunnner *et al.* (1983) for a basic description of the algorithm. The construction of the complement is based on the Delaunay triangulation and Voronoi diagram of the sample, provided by the `inform.vor.tri` function. The function `complement` returns a matrix `compl`. For each row, `compl[i,]` contains the information relative to an open ball or halfplane of the complement. The first three columns are assigned to the characterization of the ball or halfplane `i`. The information relative to the edge of the Delaunay triangulation that generates the ball or halfplane `i` is contained in `compl[i,4:17]`. Thus, if the row `i` refers to an open ball, `compl[i,1:3]` contains the centre and radius of the ball. Furthermore, the components `compl[i,19:20]` and `compl[i,21]` refer to the unitary vector v and the angle θ that characterize the arc that joins the two sample points that define the ball `i`. If the row `i` refers to a halfplane, `compl[i,1:3]` determines its equation. For the halfplane $y > a+bx$, `compl[i,1:3]=(a,b,-1)`. In the same way, for the halfplane $y < a+bx$, `compl[i,1:3]=(a,b,-2)`, for the halfplane $x > a$, `compl[i,1:3]=(a,0,-3)` and for the halfplane $x < a$, `compl[i,1:3]=(a,0,-4)`.

Value

`compl`	Output matrix. For each row, `compl[i,]` contains the information relative to an open ball or halfplane of the complement of the α-convex hull, see Details.

References

Edelsbrunner, H., Kirkpatrick, D.G. and Seidel, R. (1983) *On the shape of a set of points in the plane*. IEEE Transactions on Information Theory, Vol IT-29, No. 4.

See Also

`inform.vor.tri`, `alpha.hull`.

Examples

```
# Random sample in the unit square
sample<-matrix(runif(100),nc=2)
# value of alpha
alpha<-0.2
# Triangulation information
info<-inform.vor.tri(sample)
# Complement of the alpha-convex hull
compl<-complement(alpha,info$mat.info)
```

`dilation` *Dilation of the alpha-convex hull*

Description

This function determines if a given point *p* belongs to the dilation of radius ε of the α-convex hull of a sample.

Usage

```
dilation(Shull, p, eps)
```

Arguments

`Shull`	Output list from the `alpha.hull` function.
`p`	Numeric vector with two components describing a point in the plane.
`eps`	Value of ε.

Details

The dilation of radius ε of a set S is given by the points x such that $d(x, S) \leq \varepsilon$, where $d(x, S) = \inf\{d(x, s),\ s \in S\}$. The function `dilation` determines if the given point p belongs to the α-convex hull of the sample by using the function `in.alpha.hull`. If the point does not belong to the α-convex hull the function `dilation` computes the distance to the boundary and establishes if the distance is lower or equal to ε.

Value

A list with the following components:

`in.dilation`	A logical value specifying whether the point p belongs to the dilation.
`eps.max`	Distance from p to the boundary of the α-convex hull. If p belongs to the α-convex hull, `eps.max=0`.

Examples

```
# Random sample in the unit square
sample<-matrix(runif(100),nc=2)
# Value of alpha and epsilon
alpha<-0.2
eps<-0.05
# Triangulation information
info<-inform.vor.tri(sample)
# alpha-shape
shape<-alpha.shape(info,alpha)
# alpha-hull
compl<-complement(alpha,info$mat.info)
ahull<-alpha.hull(shape,compl)
plot.ahull(ahull,pvor=F,pdel=F,pshape=F,new=T,col=1)
# Dilation of radius alpha
# Grid
n=100
x<-seq(0,1,length=n)
y<-numeric()
for (i in 1:n){
 y<-c(y,rep(x[i],n))
}
grid<-matrix(c(rep(x,n),y),nc=2)
for (i in 1:n^2){
in.dilation<-dilation(ahull,grid[i,],eps)$in.dilation
if (in.dilation==1){points(grid[i,1],grid[i,2],pch=19,col=4)}
}
```

`dummy.coor` *Semi-infinite edge of the Voronoi diagram*

Description

This function determines fictitious coordinates for the boundless extreme of a semi-infinite edge of the Voronoi diagram.

Usage

```
dummy.coor(tri.obj, l1, l2, m, away)
```

Arguments

`tri.obj`	Object of class `"tri"`. See `tri.mesh` in package **tripack**.
`l1`	Index of the sample point correponding to one vertex of a triangle of Delaunay that lies on the convex hull, see Details.
`l2`	Index of the sample point correponding to other vertex of a triangle of Delaunay that lies on the convex hull, see Details.
`m`	Index of the circumcentre of the triangle of Delaunay with one edge on the convex hull.
`away`	Constant that determines how far away the fictitious boundless extreme is located.

Details

When a triangle of the Delaunay triangulation has one of its edges (given by the segment that joins the sample points `l1` and `l2`) on the convex hull, the corresponding segment of the Voronoi diagram is semi-infinite. The finite extreme coincides with the circumcentre of the triangle and the direction of the line is given by the prependicular bisector of the edge that lies on the convex hull.

Value

`dum`	Fictitious coordinates of the boundless extreme.

See Also

`inform.vor.tri`.

`in.BTnEn` *Estimated dilation of the boundary of a set*

Description

This function determines whether a point belongs to the estimated dilation of radius ε of the boundary of a set. The dilation of the boundary is expressed as the intersection of the dilation of the set and the dilation of its complement. Both the set and its complement are estimated by means of the α-convex hull.

Usage

```
in.BTnEn(Ghull, Rhull, p, eps)
```

Arguments

`Ghull`	Output list from the `alpha.hull` function applied to a sample of points taken in the set of interest.
`Rhull`	Output list from the `alpha.hull` function applied to a sample of points taken in the complement of the set of interest.
`p`	Numeric vector with two components describing a point in the plane.
`eps`	Value of ε.

Details

Let G be a compact set in $[0,1]^2$ and let R be $\overline{[0,1]^2 \setminus G}$. Let T denote the boundary of G. Based on the fact that $B(T,\varepsilon) = B(G,\varepsilon) \cap B(R,\varepsilon)$ it is possible to construct an estimator for $B(T,\varepsilon)$ from estimators of the sets G and R. The estimators of G and R considered by the function `in.BTnEn` are the α-convex hull of samples taken in both sets.

Value

A list with the following components:

`in.B`	A logical value specifying whether the point p belongs to the estimated dilation of the boundary of the set.
`eps.max`	The point p belongs to the estimated dilation of the boundary of the set for $\varepsilon \geq$ `eps.max`.

References

Pateiro-López, B., Rodríguez-Casal, A. (2008) *Length and surface area estimation under convexity type restrictions*. Advances in Applied Probability, Vol 40.2.

See Also

`alpha.hull`, `in.dilation`.

Examples

```
# Ellipse of centre (0.5,0.5) and radius a=0.45, b=0.25
n<-2000
c<-c(0.5,0.5)
a<-0.45
b<-0.25
alpha<-b^2/a
x<-runif(n)
y<-runif(n)
inside<-ifelse(((x-c[1])/a)^2+((y-c[2])/b)^2<=1,1,0)
sample<-matrix(c(x,y,inside),n,3)
# alpha-convex hull of the sample in G
sample.G<-sample[sample[,3]==1,1:2]
info.G<-inform.vor.tri(sample.G)
shape.G<-alpha.shape(info.G,alpha)
compl.G<-complement(alpha,info.G$mat.info)
Ghull<-alpha.hull(shape.G,compl.G)
# alpha-convex hull of the sample in R
sample.R<-sample[sample[,3]==0,1:2]
info.R<-inform.vor.tri(sample.R)
shape.R<-alpha.shape(info.R,alpha)
compl.R<-complement(alpha,info.R$mat.info)
Rhull<-alpha.hull(shape.R,compl.R)
# Plots
plot.ahull(Rhull,pvor=T,pdel=F,pshape=F,new=T,col=2)
plot.ahull(Ghull,pvor=T,pdel=F,pshape=F,new=F,col=3)
# Grid
n=100
x<-seq(0,1,length=n)
y<-numeric()
for (i in 1:n){
 y<-c(y,rep(x[i],n))
}
grid<-matrix(c(rep(x,n),y),nc=2)
npunt<-n^2
# Plot in green of the dilation of radius 0.05 of the
# alpha-convex hull of the sample in G
# Add in red the dilation of the alpha-convex hull
# of the sample in R
```

```
# Represent in blue the points in the estimated dilation of the
# boundary
eps<-0.05
for (i in 1:npunt){
in.dilation<-dilation(Ghull,grid[i,],eps)$in.dilation
if (in.dilation==1){points(grid[i,1],grid[i,2],pch=19,col=3)}
in.dilation<-dilation(Rhull,grid[i,],eps)$in.dilation
if (in.dilation==1){points(grid[i,1],grid[i,2],pch=19,col=2)}
sal[i]<-in.BTnEn(Ghull,Rhull,grid[i,],eps)$in.B
if (sal[i]==1){points(grid[i,1],grid[i,2],pch=19,col=4)}
}
```

`in.alpha.hull` *Determine whether a point belongs to the alpha-convex hull*

Description

This function determines whether a given point p belongs to the α-convex hull of a sample.

Usage

```
in.alpha.hull(ahull, p)
```

Arguments

`ahull`	Output list from the `alpha.hull` function.
`p`	Numeric vector with two components describing a point in the plane.

Details

The complement of the α-convex hull of a sample is calculated by the `complement` function. The function `in.alpha.hull` checks whether the point p belongs to any of the open balls or halfplanes that define the complement.

Value

`in.alpha.hull`
A logical value specifying whether the point p belongs to the α-convex hull.

See Also

`alpha.hull`, `complement`.

Examples

```
# Random sample in the unit square
sample<-matrix(runif(100),nc=2)
# value of alpha
alpha<-0.2
# Triangulation information
info<-inform.vor.tri(sample)
# alpha-shape
shape<-alpha.shape(info,alpha)
# Complement of the alpha-convex hull and alpha-hull boundary
compl<-complement(alpha,info$mat.info)
ahull<-alpha.hull(shape,compl)
# Check if the point (0.5,0.5) belongs to the alpha-convex hull
in.alpha.hull(ahull,p=c(0.5,0.5))
```

`inform.vor.tri` *Delaunay triangulation and Voronoi diagram*

Description

This function returns a matrix with information of the Delaunay triangulation and Voronoi diagram of a given sample.

Usage

```
inform.vor.tri(sample)
```

Arguments

`sample` Matrix of sample points in the plane. The dimension of `sample` is $n \times 2$, where n is the sample size.

Details

The function `tri.mesh` from package **tripack** calculates the Delaunay triangulation of a finite number of points using Fortran functions from the library TRIPACK. Using the Delaunay triangulation, the function `inform.vor.tri` calculates the corresponding Voronoi diagram. For each edge of the Delaunay triangulation there is a segment in the Voronoi diagram, given by the union of the circumcentres of the two neighbour triangles that share the edge. For those triangles with edges on the convex hull, the corresponding line in the Voronoi diagram is a semi-infinite segment, whose boundless extreme is calculated by the function `dummy.coor`. The function `inform.vor.tri` returns the sample, the output object of class "tri" from the function tri.mesh and a matrix with all the necessary information of the Delaunay triangulation and Voronoi diagram. Thus, for each edge of the Delaunay triangulation the output matrix contains the indexes and coordinates of the sample points that form the edge, the indexes and coordinates of the extremes of the corresponding segment in the Voronoi diagram, and an indicator that takes the value 1 for those extremes of the Voronoi diagram that represent a boundless extreme.

Value

A list with the following components:

`sample`	A 2-column matrix with the coordinates of the sample points.
`mat.info`	Matrix of dimension $n.edges \times 14$, where $n.edges$ is the total number of different edges of the Delaunay triangulation.
`tri.obj`	Object of class "tri". See tri.mesh in package **tripack**.

See Also

add.voronoi, dummy.coor.

Examples

```
# Simple example from TRIPACK
data(tritest)
sample<-matrix(c(tritest$x,tritest$y),nc=2,byrow=F)
# Delaunay triangulation and Voronoi diagram calculation
info<-inform.vor.tri(sample)

# Random sample in the unit square
sample<-matrix(runif(20),nc=2)
# Delaunay triangulation and Voronoi diagram calculation
info<-inform.vor.tri(sample)
```

`inter` *Intersection of two circumferences*

Description

This function calculates the intersection of two circumferences, given their centres and radius c_1, r_1 and c_2, r_2, respectively.

Usage

```
inter(c11, c12, r1, c21, c22, r2)
```

Arguments

`c11`	X-coordinate of the centre c_1.
`c12`	Y-coordinate of the centre c_1.
`r1`	Radius r_1.
`c21`	X-coordinate of the centre c_2.
`c22`	Y-coordinate of the centre c_2.
`r2`	Radius r_2.

Details

The function `inter` is internally called by the function `alpha.hull`.

Value

A list with the following components:

`n.cut`	Number of intersection points.
`v1`	If there are two intersection points, `v1` is the numeric vector whose components are the coordinates of the unitary vector that has its origin in c_1 and it's perpendicular to the chord that joins the intersection points of the two circumferences.
`theta1`	Angle that forms `v1` with the radius that joins the centre c_1 with an intersection point.
`v2`	If there are two intersection points, `v2` is the numeric vector whose components are the coordinates of the unitary vector that has its origin in c_2 and it's perpendicular to the chord that joins the intersection points of the two circumferences.
`theta2`	Angle that forms `v2` with the radius that joins the centre c_2 with an intersection point.

`length.ahull` *Length of the boundary of the alpha-convex hull*

Description

This function calculates the length of the boundary of the α-convex hull of a given sample.

Usage

```
length.ahull(ahull.archs)
```

Arguments

`ahull.archs` Output matrix from the `alpha.hull` function.

Details

The function `length.ahull` is internally called by the function `alpha.hull`.

Value

`length` Length of the boundary of the α-convex hull.

See Also

`alpha.hull`.

Examples

```
# Random sample in the unit square
sample<-matrix(runif(100),nc=2)
# value of alpha
alpha<-0.2
# Triangulation information
info<-inform.vor.tri(sample)
# alpha-shape
shape<-alpha.shape(info,alpha)
# Complement of the alpha-convex hull and alpha-hull boundary
compl<-complement(alpha,info$mat.info)
ahull<-alpha.hull(shape,compl)
# Length of the alpha-convex hull
ahull$length
```

`plot.ahull` *Plot the alpha-convex hull*

Description

This function returns a plot of the α-convex hull. If desired, it also adds the Delaunay triangulation, Voronoi diagram and α-shape of the sample.

Usage

```
plot.ahull(ahull, pvor=F, pdel=F, pshape=F, new=T,...)
```

Arguments

`ahull`	Output list from the `alpha.hull` function.
`pvor`	Logical, indicates if Voronoi diagram should be added to the plot.
`pdel`	Logical, indicates if Delaunay triangulation should be added to the plot.
`pshape`	Logical, indicates if the α-shape should be added to the plot.
`new`	Logical, indicates if a new plot is opened.
`...`	Arguments to be passed to methods, such as graphical parameters (see `par`).

See Also

`alpha.hull`, `angs.arch`, `add.voronoi`, `plot.ashape`.

Examples

```
# Random sample in the unit square
sample<-matrix(runif(100),nc=2)
# value of alpha
alpha<-0.2
# Triangulation information
info<-inform.vor.tri(sample)
# alpha-shape
shape<-alpha.shape(info,alpha)
# Complement of the alpha-convex hull and alpha-hull boundary
compl<-complement(alpha,info$mat.info)
ahull<-alpha.hull(shape,compl)
# Plot including the alpha-convex hull, alpha-shape,
# voronoi diagram and Delaunay triangulation
plot.ahull(ahull,pvor=T,pdel=T,pshape=T,new=T,col=1)
```

`plot.ashape` *Plot the alpha-shape*

Description

This function returns a plot of the α-shape.

Usage

```
plot.ashape(ashape, pvor = F, pdel = F, new = F, ...)
```

Arguments

`ashape`	Output list from the `alpha.shape` function.
`pvor`	Logical, indicates if Voronoi diagram should be added to the plot.
`pdel`	Logical, indicates if Delaunay triangulation should be added to the plot.
`new`	Logical, indicates if a new plot is opened.
`...`	Arguments to be passed to methods, such as graphical parameters (see `par`).

See Also

objects to See Also as `alpha.shape`, `add.voronoi`.

Examples

```
# Uniform sample of size n=300 on the disc B(c,0.5)\B(c,0.25),
# with c=(0.5,0.5).
n<-300
m<-0
sample<-matrix(0,n,2)
while(m<n){
x<-runif(1)
y<-runif(1)
d<-(x-0.5)^2+(y-0.5)^2
if((d<=(0.5)^2)&(d>=(0.25)^2)){
m<-m+1
sample[m,]<-c(x,y)
}
}
# Value of alpha
alpha<-0.1
```

```
# Triangulation information
info<-inform.vor.tri(sample)
# alpha-shape
shape<-alpha.shape(info,alpha)
plot.ashape(shape,pvor=T,pdel=T,new=T)
```

`rotation.cw` *Clockwise rotation*

Description

This function calculates the clockwise rotation of angle θ of a given vector v in the plane.

Usage

```
rotation.cw(v, theta)
```

Arguments

`v`	Vector v in the plane.
`theta`	Angle θ.

Value

`v.rot`	Vector after rotation.

Examples

```
# Rotation of angle pi/4 of the vector (0,1)
rotation.cw(v=c(0,1),theta=pi/4)
```

Resumen en castellano

Nuestro objetivo en este resumen es destacar brevemente los principales resultados que hemos obtenido durante este período de investigación. En este tiempo, nuestro interés se ha centrado, fundamentalmente, en los problemas de estimación del soporte y el área superficial, que se enmarcan dentro de la teoría general de estimación de conjuntos. La reconstrucción de un conjunto S a partir de un conjunto finito de puntos tomados en él es un problema que ha sido abordado en diferentes campos de investigación. Por ejemplo, en geometría computacional, la construcción eficiente de la envoltura convexa tiene importantes aplicaciones en reconocimiento de patrones, procesamiento de imágenes y análisis cluster, entre otros. Véase Preparata y Shamos (1985) para una introducción a la geometría computacional y sus aplicaciones. En determinadas situaciones es razonable suponer que el conjunto de puntos a partir del cual se pretende reconstruir S es no-determinista. Obtener buenas estimaciones de un conjunto a partir de una muestra de puntos no es una tarea fácil y la resolución de este problema depende, en gran medida, de las hipótesis del modelo. Así, si no disponemos de ninguna información sobre el conjunto de interés, no tendremos otra elección más que considerar estimadores flexibles que nos permitan abordar eficientemente la mayor cantidad de situaciones posibles. En cambio, si restringimos la familia de conjuntos a estimar, podremos considerar estimadores más sofisticados, que se adapten mejor a las restricciones de forma establecidas.

Formalmente, el problema de **estimación del soporte** se establece como el problema de aproximar el soporte de una distribución de probabilidad absolutamente continua P_X, a partir de una muestra aleatoria simple $\mathcal{X}_n = \{X_1, \ldots, X_n\}$ de X. Tradicionalmente, el problema de estimación del soporte ha sido abordado para la familia de conjuntos convexos. Así, Korostelëv y Tsybakov (1993) cita los trabajos de Geffroy (1964), Rényi y Sulanke (1963) y Rényi y Sulanke (1964) como las primeras referencias que tratan el problema de la estimación del soporte. En concreto, Rényi y Sulanke (1963) y Rényi y Sulanke (1964) estudiaron el caso en el que el soporte $S \subset \mathbb{R}^2$ es convexo y propusieron un estimador natural, la envoltura convexa de la muestra. Sin embargo, la hipótesis de convexidad puede resultar demasiado restrictiva en la práctica y si S no es convexo, entonces la envoltura convexa de la muestra no resulta ser un estimador apropiado. La pregunta es entonces, ¿cómo podemos estimar S si no tenemos ninguna hipótesis sobre la forma del conjunto? En este sentido, Chevalier (1976) y Devroye y Wise (1980) proponen estimar el soporte de una distribución de probabilidad desconocida mediante una versión suavizada de la muestra $\mathcal{X}_n$. El estimador propuesto, al que nos referiremos como estimador Devroye-Wise, se define como

$$\bigcup_{i=1}^{n} B(X_i, \varepsilon),$$

donde $\varepsilon > 0$ y $B(X_i, \varepsilon)$ denota la bola cerrada de centro X_i y radio ε. El problema de la estimación del soporte se presenta en Devroye y Wise (1980) relacionado con una aplicación práctica, la detección de comportamiento anormal de un sistema, planta o máquina. Los resultados del comportamiento del estimador fueron analizados, entre otros, por Chevalier (1976), Devroye y Wise (1980) y Korostelëv y Tsybakov (1993). Por supuesto, existen situaciones intermedias entre las dos citadas anteriormente, es decir, podemos asumir que el conjunto S satisface una condición de forma más flexible que la convexidad. En Rodríguez-Casal (2006) se estudia la estimación de un soporte α-convexo. Se dice que el conjunto $S \subset \mathbb{R}^d$ es α-convexo, para $\alpha > 0$, si $S = C_\alpha(S)$, siendo

$$C_\alpha(S) = \bigcap_{\{\mathring{B}(x,\alpha):\ \mathring{B}(x,\alpha)\cap S=\emptyset\}} \left(\mathring{B}(x,\alpha)\right)^c.$$

En la ecuación anterior, $\mathring{B}(x, \alpha)$ denota la bola abierta de centro x y radio α y $(\mathring{B}(x, \alpha))^c$ su complementario. El conjunto $C_\alpha(S)$ se denomina envoltura α-convexa de S y es la base para la definición del nuevo estimador de soporte propuesto por Rodríguez-Casal (2006), la envoltura α-convexa de la muestra. Dicho estimador es estudiado en profundidad en el Capítulo 2 de esta tesis. La hipótesis de α-convexidad, que juega un papel fundamental en nuestro trabajo, está íntimamente relacionada con otra interesante condición de forma, la condición de rodamiento libre. Diremos que una bola de radio α rueda libremente en el conjunto S si para cada punto a de la frontera del conjunto, existe un punto $x \in S$ tal que $a \in B(x, \alpha) \subset S$.

Antes de continuar, nos gustaría hacer hincapié sobre algo fundamental que hemos obviado hasta el momento. Estamos hablando de estimadores del soporte y de determinar si dichos estimadores se aproximan al conjunto original S. Sin embargo, no hemos establecido ningún criterio para evaluar dicha proximidad. Existen distintas alternativas para definir la distancia entre conjuntos, como por ejemplo, la distancia de Hausdorff o la distancia en medida. Nos centramos en la definición de esta última. Así, dados dos conjuntos de Borel A y $C \subset \mathbb{R}^d$, se define la distancia en medida entre A y C como

$$d_\mu(A, C) = \mu(A \Delta C),$$

donde μ denota la medida de Lebesgue d-dimensional y $A \Delta C$ denota la diferencia simétrica entre A y C, es decir,

$$A \Delta C = (A \setminus C) \cup (C \setminus A).$$

La distancia en medida nos da una idea de la similitud en el contenido de dos conjuntos. En particular, si S_n es un estimador del soporte S, entonces $d_\mu(S, S_n)$ mide la proximidad entre ambos conjuntos, sirviendo así como criterio para evaluar el comportamiento del estimador del soporte.

En cuanto a la **estimación del área superficial**, abordamos el problema desde dos puntos de vista diferentes. En el primer planteamiento que comentamos, la información muestral viene dada por una muestra de puntos tomada en el conjunto de interés. En esta situación, parece que lo natural es estimar el conjunto mediante un estimador del soporte y calcular el área superficial de dicho estimador. La intuición que tenemos es que, si el estimador aproxima bien al conjunto, entonces su área superficial también aproximará bien al área superficial del conjunto, que es el objetivo. Con este planteamiento presentamos un nuevo estimador, consistente en medir el α-shape de la muestra. El α-shape, véase Edelsbrunner et al. (1983), es un grafo cuyas aristas son rectas uniendo puntos muestrales denominados α-vecinos. Dos puntos muestrales son α-vecinos si existe una bola de radio α de forma que los dos puntos están en su frontera y, además, ningún punto muestral está en su interior.

El segundo planteamiento para abordar el problema de la estimación del área superficial consiste en suponer que la información muestral viene dada por puntos tanto de dentro del conjunto de interés G como de fuera del conjunto. Asumimos entonces, sin pérdida de generalidad, que $G \subset (0,1)^d$ y definimos $R = [0,1]^d \setminus \text{int}(G)$, donde $\text{int}(G)$ denota el interior del conjunto. La información muestral consiste en observaciones i.i.d. $(Z_1, \xi_1), \ldots, (Z_n, \xi_n)$ de una variable aleatoria (Z, ξ), donde Z se distribuye uniformemente en $[0,1]^d$ y $\xi = \mathbb{I}_{\{Z \in G\}}$ es la función indicadora de G. Denotamos $\mathcal{X}_n = \{Z_i : \xi_i = 1\}$ e $\mathcal{Y}_n = \{Z_i : \xi_i = 0\}$. El área superficial de G se calcula a partir de su contenido de Minkowski,

$$L_0 = \lim_{\varepsilon \to 0} \frac{\mu(B(\Gamma, \varepsilon))}{2\varepsilon} = \lim_{\varepsilon \to 0} L(\varepsilon),$$

siempre que este límite exista y sea finito. En la definición anterior, Γ denota la frontera del conjunto G y $B(\Gamma, \varepsilon)$ es la dilatación de radio ε de dicha frontera, es decir,

$$B(\Gamma, \varepsilon) = \bigcup_{x \in \Gamma} B(x, \varepsilon).$$

Por lo tanto, si definimos un estimador para la dilatación de la frontera, podremos definir un estimador L_n para el contenido de Minkowski L_0 de la siguiente forma:

$$L_n = \frac{\mu(\Gamma_n)}{2\varepsilon_n},$$

siendo $\varepsilon_n > 0$ y Γ_n un estimador de $B(\Gamma, \varepsilon_n)$. La clave para definir Γ_n es que, bajo condiciones suaves,

$$B(\Gamma, \varepsilon_n) = B(G, \varepsilon_n) \cap B(R, \varepsilon_n),$$

es decir, es posible construir un estimador de L_0 a partir de estimadores de los conjuntos G y R, lo cual nos lleva de nuevo al problema de estimación del soporte. Por ejemplo, Cuevas et al. (2007) proponen estimar G y R mediante $\mathcal{X}_n$ e $\mathcal{Y}_n$, respectivamente. Nos referiremos al estimador resultante L_n como estimador empírico. De nuevo, dependiendo de las restricciones de forma de G y R, podremos definir estimadores más sofisticados. Siguiendo la línea abierta en la estimación del soporte, nos centraremos en la condición de α-convexidad.

Resultados sobre la estimación de conjuntos α-convexos

En el Capítulo 2 se aborda el problema de la estimación de conjuntos α-convexos. El estimador natural en esta situación es la envoltura α-convexa de una muestra de puntos tomada en el conjunto de interés. De manera formal, sea $S \subset \mathbb{R}^d$ un conjunto compacto, no vacío y α-convexo con $\alpha > 0$. El objetivo es estimar S a partir de una muestra $\mathcal{X}_n$ de una variable aleatoria X con distribución de probabilidad absolutamente continua P_X y soporte S. Puesto que, en general, el parámetro α es desconocido, consideraremos el estimador $C_{r_n}(\mathcal{X}_n)$, donde asumimos que r_n es menor o igual que α para todo n. ¿Es $C_{r_n}(\mathcal{X}_n)$ un estimador consistente de S? ¿Bajo qué condiciones? ¿Cuánto se aproxima $C_{r_n}(\mathcal{X}_n)$ a S? En el Capítulo 2 damos respuesta a estas preguntas. El Teorema 2.5.1 establece una condición necesaria y suficiente para la consistencia del estimador envoltura r_n-convexa. Se prueba que $\mathbb{E}(d_\mu(S, C_{r_n}(\mathcal{X}_n))) \to 0$ si y sólo si $nr_n^d \to \infty$. Merece la pena comentar que la hipótesis de α-convexidad no es esencial para probar la consistencia del estimador. De hecho, se puede probar que si $r_n \to 0$ y $nr_n^d \to \infty$, entonces se sigue cumpliendo que $\mathbb{E}(d_\mu(S, C_{r_n}(\mathcal{X}_n))) \to 0$, incluso en el caso de que S no sea α-convexo. Nótese que las hipótesis sobre r_n son idénticas a las que, impuestas sobre el parámetro de suavizado del estimador Devroye-Wise, garantizan su consistencia en probabilidad, véase Devroye y Wise (1980).

Respecto a la proximidad entre S y $C_{r_n}(\mathcal{X}_n)$, estudiamos la distancia en medida entre ambos conjuntos. Rodríguez-Casal (2006) obtuvo la tasa de convergencia casi segura para $d_\mu(S, C_{r_n}(\mathcal{X}_n))$, bajo la hipótesis de que S satisface las condiciones del Teorema 1 de Walther (1999). En concreto, se prueba que el orden de convergencia es $r_n^{-1}(\log n/n)^{2/(d+1)}$. En el Teorema 2.5.2 obtenemos la tasa de convergencia de $\mathbb{E}(d_\mu(S, C_{r_n}(\mathcal{X}_n)))$. Al igual que Rodríguez-Casal (2006), necesitamos una condición de forma adicional sobre S la cual, en particular, implica la α-convexidad. Suponemos que una bola de radio $\alpha > 0$ rueda libremente en S y en $\overline{S^c}$. Esta condición juega un papel fundamental a lo largo de nuestro trabajo y merece algunos comentarios. En primer lugar, la condición de rodamiento libre en S y en $\overline{S^c}$ excluye la posibilidad de que el conjunto S tenga picos. Nótese que si únicamente suponemos α-convexidad, no podemos asegurar que la frontera del conjunto es suave. Por otra parte, al asumir que una bola de radio $\alpha > 0$ rueda libremente en S estamos descartando, por ejemplo, conjuntos con puntos aislados. En términos generales, la condición de rodamiento libre en S obliga a los puntos de la frontera del conjunto a estar en contacto directo con el interior de S. A la vista de la importancia de esta condición, uno puede preguntarse por qué en el título del Capítulo 2 sólo hacemos referencia a la α-convexidad. Pues bien, el motivo es que la α-convexidad es la restricción de forma que motiva originalmente la definición del estimador. Además, la envoltura α-convexa de una muestra de puntos tiene sentido como estimador, independientemente de condiciones de forma más restrictivas sobre S. Esta es la razón por la que hemos decidido enfatizar la importancia de esta propiedad.

Respecto a la distribución de probabilidad, es útil suponer que P_X está acotada uniformemente en S. Formalmente, P_X está acotada uniformemente en S si existe $\delta > 0$ tal que $P_X(C) \geq \delta\mu(C \cap S)$ para todo conjunto de Borel $C \subset \mathbb{R}^d$. Es inmediato comprobar que, por ejemplo, la distribución uniforme en S está acotada uniformemente.

Una vez discutidas las hipótesis, podemos pasar a establecer los resultados más importantes del Capítulo 2. Así, sea S un subconjunto compacto no vacío de $\mathbb{R}^d$ tal que una bola de radio

$\alpha > 0$ rueda libremente en S y en $\overline{S^c}$ y supongamos que P_X está acotada uniformemente en S. En estas condiciones, el Teorema 2.5.2 establece que, si la sucesión $\{r_n\}$ satisface

$$\lim_{n\to\infty} \frac{nr_n^d}{\log n} = \infty,$$

entonces

$$\mathbb{E}(d_\mu(S, C_{r_n}(\mathcal{X}_n))) = O\left(r_n^{-\frac{d-1}{d+1}} n^{-\frac{2}{d+1}}\right).$$

Debemos también mencionar que el concepto de familia inevitables de conjuntos, discutido en detalle en las Secciones 2.3 y 2.4 es fundamental en el desarrollo del Capítulo 2 y juega un papel esencial en la prueba del Teorema 2.5.2. Finalmente, en el Teorema 2.5.3 probamos que la tasa de convergencia obtenida para $\mathbb{E}(d_\mu(S, C_{r_n}(\mathcal{X}_n)))$ no puede ser mejorada ya que existen conjuntos bajo las condiciones establecidas para los cuales

$$\liminf_{n\to\infty} r_n^{\frac{d-1}{d+1}} n^{\frac{2}{d+1}} \mathbb{E}(d_\mu(S, C_{r_n}(\mathcal{X}_n))) > 0.$$

Estos resultados nos llevan a comparar la tasa de convergencia de $\mathbb{E}(d_\mu(S, C_{r_n}(\mathcal{X}_n)))$, establecida en el Teorema 2.5.2, con la de $d_\mu(S, C_{r_n}(\mathcal{X}_n))$ (tasa de convergencia casi segura) obtenida por Rodríguez-Casal (2006). Observamos que la convergencia de $\mathbb{E}(d_\mu(S, C_{r_n}(\mathcal{X}_n)))$ es más rápida puesto que el logaritmo del numerador desaparece y el factor de penalización $r_n^{-(d-1)/(d+1)}$ es asintóticamente menor que r_n^{-1}.

Resultados sobre la estimación del área superficial

El Capítulo 3 aborda el problema de la estimación del área superficial de un conjunto. Al presentar este problema distinguimos entre el caso en que la información muestral viene dado por puntos dentro del conjunto de interés y el caso en el que la información muestral viene dada por puntos tanto dentro del conjunto de interés $G \subset (0,1)^d$ como dentro de $R = [0,1]^d \setminus \text{int}(G)$. Como comentamos, la primera situación se puede entender como un paso más dentro de la estimación del soporte. A pesar de que dicho planteamiento es más elemental e intuitivo, resulta más difícil de abordar desde el punto de vista teórico ya que no es inmediato determinar cuándo un punto está próximo a la frontera del conjunto. Por tanto, en el Capítulo 3 nos restringimos al caso en el que la información muestral viene dada por observaciones i.i.d. $(Z_1, \xi_1), \ldots, (Z_n, \xi_n)$ de una variable aleatoria (Z, ξ), donde, como ya comentamos, Z sigue una distribución uniforme en $[0,1]^d$ y $\xi = \mathbb{I}_{\{Z\in G\}}$. Siguiendo la notación introducida anteriormente consideramos

$$L_n = \frac{\mu(\Gamma_n)}{2\varepsilon_n},$$

donde Γ_n es un estimador de $B(\Gamma, \varepsilon_n)$ y $\varepsilon_n > 0$. Recordemos que la expresión de L_n viene motivada por la definición del contenido de Minkowski L_0. Así, para valores pequeños de ε_n, el estimador L_n se aproxima a L_0. Siguiendo con la restricción de forma estudiada, suponemos que G y R son α-convexos. Entonces, proponemos estimar $B(\Gamma, \varepsilon_n)$ mediante

$$\Gamma_n = B(C_\alpha(\mathcal{X}_n), \varepsilon_n) \cap B(C_\alpha(\mathcal{Y}_n), \varepsilon_n)$$

donde $\mathcal{X}_n = \{Z_i : \xi_i = 1\}$ e $\mathcal{Y}_n = \{Z_i : \xi_i = 0\}$. Una cuestión de importancia teórica es la existencia del contenido de Minkoski L_0. Este hecho está relacionado con el comportamiento de la función $\mu(B(\Gamma, \varepsilon))$ y, por lo tanto, con las hipótesis sobre el conjunto G. En cuanto al estimador, la cuestión más relevante es si L_n aproxima con exactitud a L_0. De forma análoga al problema de estimación del soporte, los resultados del Capítulo 3 se obtienen bajo una condición adicional de rodamiento libre. De nuevo, asumimos que una bola de radio $\alpha > 0$ rueda libremente en G y en $\overline{G^c}$. Esta condición garantiza que el contenido de Minkowski está bien definido. De todas formas, no debemos olvidar que L_n tiene sentido bajo condiciones más suaves. Por ejemplo, la α-convexidad de G y R es suficiente para garantizar que, con probabilidad uno, $\Gamma_n \subset B(\Gamma, \varepsilon_n)$. Esta última propiedad indica que L_n es sesgado, tendiendo a infraestimar el valor de L_0. Las propiedades asintóticas de L_n se estudian y comparan con las del estimador del área superficial propuesto por Cuevas et al. (2007). Los Teoremas 3.3.1 y 3.3.2 nos dan, respectivamente, la tasa de convergencia casi segura y la tasa de convergencia L_1 del estimador L_n a L_0. Bajo las hipótesis establecidas se prueba que, con probabilidad uno,

$$\inf_{\varepsilon_n} |L_n - L_0| = O\left(\frac{\log n}{n}\right)^{\frac{1}{d+1}},$$

donde la tasa óptima se obtiene para $\varepsilon_n = (\log n/n)^{1/(d+1)}$. Respecto a la tasa de convergencia L_1, probamos que se puede eliminar el logaritmo en la tasa anterior y, por tanto,

$$\inf_{\varepsilon_n} \mathbb{E}\, |L_n - L_0| = O\left(n^{-\frac{1}{d+1}}\right).$$

El orden óptimo en este caso se obtiene para $\varepsilon_n = n^{-1/(d+1)}$. La convergencia L_1 del estimador propuesto es así más rápida que la del estimador empírico propuesto por Cuevas et al. (2007), de orden $n^{-1/2d}$.

Aspectos computacionales

Una vez que hemos discutido las propiedades teóricas de diferentes estimadores del soporte y el área superficial de un conjunto, el Capítulo 4 se centra en cómo se puede llevar a cabo el análisis práctico de dichos problemas. El cálculo de la envoltura α-convexa de una muestra no es un problema de solución inmediata y, por este motivo, dedicamos parte del Capítulo 4 a describir el algoritmo de implementación propuesto por Edelsbrunner (1983).

Además de la envoltura α-convexa, hemos programado el estimador de la longitud de la frontera propuesto en el Capítulo 3 para el caso particular de $\mathbb{R}^2$. Ilustramos el problema de estimación del área superficial mediante un estudio de simulación en el que comparamos nuestro estimador con el propuesto por Cuevas et al. (2007). Puesto que los resultados del estudio no son tan satisfactorios como cabría esperar tras el análisis teórico, hemos planteado una solución alternativa al problema de la estimación del área superficial. Dada la envoltura α-convexa de una muestra, podemos calcular su perímetro sumando las longitudes de los arcos que conforman su frontera. De forma análoga, se consideran otros estimadores como, por ejemplo, el α-shape para los cuales medimos la longitud de su frontera. En el Capítulo 4 se muestran los resultados de un estudio de simulación que pretende mostrar el comportamiento en la práctica de este tipo de estimadores de la longitud de la frontera.

A la vista de los resultados obtenidos, no podemos concluir que los modelos basados en la noción de contenido de Minkowski, sean significativamente mejores que los basados en la idea más intuitiva de medir la frontera de un estimador del soporte. Los prometedores resultados obtenidos en este último estudio de simulación nos animan a afrontar en el futuro la justificación teórica que explique el buen comportamiento observado. Así, existe una línea abierta a la investigación en este contexto.

Finalmente, merece la pena comentar que, como consecuencia de la implementación en R de los estimadores estudiados, hemos desarrollado una nueva librería denominada `alphahull`. La documentación completa del paquete, incluyendo la descripción de todas sus funciones, se puede consultar en el Apéndice C. Nos gustaría resaltar algunas de las características más destacables de la librería. Además de las funciones que calculan los estimadores del soporte y de la longitud de la frontera utilizados en los estudios de simulación, el paquete `alphahull` incluye otras funciones que pueden ser de utilidad en diferentes contextos. Por ejemplo, hemos programado el diagrama de Voronoi y la triangularización de Delaunay. El diagrama de Voronoi y la triangularización de Delaunay se usan con mucha frecuencia en varios campos de investigación y, por lo que nosotros sabemos, no existía un código depurado en R que calculase estas estructuras geométricas. Por lo tanto, pretendemos que el paquete `alphahull` se entienda, no sólo como una colección de funciones programadas para la realización de un estudio de simulación aislado, sino como una herramienta útil para la investigación más allá del contexto de esta tesis.

Agradecimientos

Esta investigación fue financiada por los proyectos MTM2005-00820 del Ministerio de Educación y Ciencia y PGIDIT06PXIB207009PR de la Xunta de Galicia.

Bibliography

AURENHAMMER, F. (1991). Voronoi diagrams - a survey of a fundamental geometric data structure. *ACM Comput. Surv.*, vol. 23(3), pp. 345–405.

AURENHAMMER, F. AND KLEIN, R. (2000). Voronoi diagrams. In *Handbook of computational geometry*, pp. 201–290. North-Holland, Amsterdam.

BADDELEY, A. AND JENSEN, E. B. V. (2005). *Stereology for statisticians*, vol. 103 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, FL.

BAÍLLO, A. AND CUEVAS, A. (2001). On the estimation of a star-shaped set. *Adv. in Appl. Probab.*, vol. 33(4), pp. 717–726.

BENSON, R. V. (1966). *Euclidean geometry and convexity*. McGraw-Hill Book Co., New York.

BILLINGSLEY, P. (1995). *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third ed. A Wiley-Interscience Publication.

BRÄKER, H. AND HSING, T. (1998). On the area and perimeter of a random convex hull in a bounded convex set. *Probab. Theory Related Fields*, vol. 111(4), pp. 517–550.

CHEVALIER, J. (1976). Estimation du support et du contour du support d'une loi de probabilité. *Ann. Inst. H. Poincaré Sect. B (N.S.)*, vol. 12(4), pp. 339–364.

CROFT, H. T., FALCONER, K. J., AND GUY, R. K. (1991). *Unsolved problems in geometry*. Problem Books in Mathematics. Springer-Verlag, New York. Unsolved Problems in Intuitive Mathematics, II.

CRUZ-ORIVE, L. M. (2001/02). Stereology: meeting point of integral geometry, probability, and statistics. *Math. Notae*, vol. 41, pp. 49–98 (2003). Homage to Luis Santaló. Vol. 1 (Spanish).

CUEVAS, A., FRAIMAN, R., AND RODRÍGUEZ-CASAL, A. (2007). A nonparametric approach to the estimation of lengths and surface areas. *Ann. Statist.*, vol. 35(3), pp. 1031–1051.

CUEVAS, A. AND RODRÍGUEZ-CASAL, A. (2004). On boundary estimation. *Adv. in Appl. Probab.*, vol. 36(2), pp. 340–354.

DEVROYE, L. (1983). The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates. *Ann. Statist.*, vol. 11(3), pp. 896–904.

DEVROYE, L. AND WISE, G. L. (1980). Detection of abnormal behavior via nonparametric estimation of the support. *SIAM J. Appl. Math.*, vol. 38(3), pp. 480–488.

DÜMBGEN, L. AND WALTHER, G. (1996). Rates of convergence for random approximations of convex sets. *Adv. in Appl. Probab.*, vol. 28(2), pp. 384–393.

EDELSBRUNNER, H., KIRKPATRICK, D. G., AND SEIDEL, R. (1983). On the shape of a set of points in the plane. *IEEE Trans. Inform. Theory*, vol. 29(4), pp. 551–559.

EDGAR, G. A. (1990). *Measure, topology, and fractal geometry*. Undergraduate Texts in Mathematics. Springer-Verlag, New York.

EGGLESTON, H. G. (1958). *Convexity*. Cambridge Tracts in Mathematics and Mathematical Physics, No. 47. Cambridge University Press, New York.

FEDERER, H. (1959). Curvature measures. *Trans. Amer. Math. Soc.*, vol. 93, pp. 418–491.

GEFFROY, J. (1964). Sur un problème d'estimation géométrique. *Publ. Inst. Statist. Univ. Paris*, vol. 13, pp. 191–210.

KOROSTELËV, A. P. AND TSYBAKOV, A. B. (1993). *Minimax theory of image reconstruction*, vol. 82 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

MATHERON, G. (1975). *Random sets and integral geometry*. John Wiley & Sons, New York-London-Sydney. With a foreword by Geoffrey S. Watson, Wiley Series in Probability and Mathematical Statistics.

MATTILA, P. (1995). *Geometry of sets and measures in Euclidean spaces. Fractals and rectifiability*, vol. 44 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

MØLLER, J. (1994). *Lectures on random Voronoi tessellations*, vol. 87 of *Lecture Notes in Statistics*. Springer-Verlag, New York.

PATEIRO-LÓPEZ, B. AND RODRÍGUEZ-CASAL, A. (2008). Length and surface area estimation under smoothness restrictions. *Adv. Appl. Prob.* To appear.

PREPARATA, F. P. AND SHAMOS, M. I. (1985). *Computational geometry: An Introduction*. Texts and Monographs in Computer Science. Springer-Verlag, New York.

RENKA, R. J. (1996). Algorithm 751: Tripack: a constrained two-dimensional delaunay triangulation package. *ACM Trans. Math. Softw.*, vol. 22(1), pp. 1–8.

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 (1963).

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 (1964).

RODRÍGUEZ-CASAL, A. (2007). Set estimation under convexity type assumptions. *Annales de l'I.H.P.- Probabilités & Statistiques*, vol. 43, pp. 763–774.

SCHNEIDER, R. (1988). Random approximation of convex sets. *J. Microscopy*, vol. 151, pp. 211–227.

SERRA, J. (1984). *Image analysis and mathematical morphology*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London. English version revised by Noel Cressie.

WALTHER, G. (1999). On a generalization of Blaschke's rolling theorem and the smoothing of surfaces. *Math. Methods Appl. Sci.*, vol. 22(4), pp. 301–316.

Notation

$\mathcal{X}_n$	Sample $(X_1, \ldots, X_n)$, 3
P_X	Distribution probability function of X, 3
$\mathbb{R}^d$	d-dimensional Euclidean space, 5
$\langle \cdot, \cdot \rangle$	Inner product in $\mathbb{R}^d$, 5
$\lVert \cdot \rVert$	Euclidean norm in $\mathbb{R}^d$, 5
$d_H(A, C)$	Hausdorff distance between A and C, 5, 6, 8
$d(a, C)$	Distance from the point a to the set C, 5
$\mathring{B}(A, \varepsilon)$	open ε-neighbourhood, 6
$B(A, \varepsilon)$	closed ε-neighbourhood, 6
$\oplus$	Minkowski addition, 7
$\ominus$	Minkowski subtraction, 7
$B(x, r)$	Closed ball with centre x and radius r, 7
$\mathring{B}(x, r)$	Open ball with centre x and radius r, 7
B	Closed ball with centre 0 and radius 1, 7
$\mathring{B}$	Open ball with centre 0 and radius 1, 7
A^c	Complement of A, 7
$\mathrm{int}(A)$	Interior of A, 7
$\overline{A}$	Closure of A, 7
∂A	Boundary of A, 7
$\mathcal{B}$	Borel σ-algebra, 9
μ	Lebesgue measure, 9
$d_\mu(A, C)$	Distance in measure between A and C, 9
$A \Delta C$	Symmetric difference between A and C, 9
$\mathbb{I}_A$	Indicator function of A, 10
H_n	Convex hull of the sample $\mathcal{X}_n$, 11
$C_\alpha(A)$	α-convex hull of A, 12
$\mathrm{reach}(S)$	Reach of S, 16
$L_0(A)$	Minkowski content of the body $A \subset \mathbb{R}^d$, 20
Γ	Boundary of a set, 21
$\mathcal{E}_{x,r}$	$\{B(y, r) : y \in B(x, r)\}$, 29
$\mathcal{U}_{x,r}$	Unavoidable family of sets for $\mathcal{E}_{x,r}$, 29
$\mathbb{S}_d$	Unit sphere $\{u \in \mathbb{R}^d : \lVert u \rVert = 1\}$, 31, 48
$\varphi_{u,v}$	Angle between the vectors u and v. $\varphi_{u,v} \in [0, \pi]$, 31, 48

e_d	Unit vector $(0,\ldots,0,1)\in\mathbb{R}^d$, 31, 48
C_u^θ, C_u	$C_u^\theta=\{x\in\mathbb{R}^d:\langle x,u\rangle\geq\|x\|\cos\theta\}$, $C_u^{\pi/6}$, 31, 48
$C_{u,r}^\theta$, $C_{u,r}$	Circular sector $C_{u,r}^\theta=C_u^\theta\cap B(0,r)$, $C_{u,r}^{\pi/6}$, 31, 48
$\mathcal{R}_\theta$, $\mathcal{R}$	Counter-clockwise rotation of angle θ, $\mathcal{R}_{\pi/6}$, 31
$P_\varGamma x$	Metric projection of x onto $\varGamma$, 37
$\mathcal{O}$	Orthogonal transformation, 38
ω_d	Measure of the unit ball in $\mathbb{R}^d$, 48
$\varGamma_n$	$\varGamma_n=B(G_n,\varepsilon_n)\cap B(R_n,\varepsilon_n)$, 86

Index