A Local Depth Measure for General Data.
Lucas Fernandez-Piana* and Marcela Svarc**

*Instituto de Cálculo, FCEyN, Universidad de Buenos Aires and CONICET, Argentina.
**Departamento de Matemática y Ciencias, Universidad de San Andrés and CONICET, Argentina

Abstract

We introduce a local depth measure for data in a Banach space, based on the use of one-dimensional projections. Its theoretical properties are studied, as well as strong consistency results for it and also of the local depth regions. In addition, we propose a clustering procedure based on local depths. Applications of the clustering procedure are illustrated on some artificial and real data sets for multivariate, functional, and multifunctional data, obtaining very promising results.

Keywords: Data Depth, Cluster Analysis, Functional Data, Projection Procedures.

1Corresponding author: Marcela Svarc, Departamento de Matemática y Ciencias, Universidad de San Andrés, Vito Dumas 248, Victoria, Argentina. Email: msvarc@udesa.edu.ar
1 Introduction

Data depth measures play an important role when analyzing complex data sets, such as functional or high dimensional data. The main goal of depth measures is to provide a center-outer ordering of the data, generalizing the concept of median. Depth measures are also useful for describing different features of the underlying distribution of the data. Moreover, depth measures are powerful tools to deal with several inference problems such as, location and symmetry tests, classification, outlier detection, etc.

Nonetheless, since one of their major characteristics is that the depth values decrease along any half-line ray from the center, they are not suitable for capturing characteristics of the distribution when data is multimodal. Hence, over the last few years, there have been introduced several definitions of local depth, with the aim of revealing the local features of the underlying distribution. The basic idea is to restrict a global depth measure to a neighborhood of each point of the space. In this way, a local depth measure should behave as a global depth measure with respect to the neighborhoods of the different points. Agostinelli and Romanazzi (2011) gave the first definition of local depth for the case of multivariate data. They extended the concepts of simplicial and half-space depth so as to allow recording the local space geometry near a given point. For simplicial depth, they consider only random simplices with sizes no greater than a certain threshold, while for half-space depth, the half-spaces are replaced by infinite slabs with finite width. Both definitions strongly rely on a tuning parameter, which retains a constant size neighborhood of every point of the space, something which plays an analogous role to that of bandwidth in the problem of density estimation. Desirable statistical theoretical properties are attained for the case of univariate absolutely continuous distributions. Paindaveine and Van Bever (2013) introduce a general procedure for multivariate data that allows converting any global depth into a local depth. The main idea of their definition is to study local environments. This means regarding the local depth as a global depth restricted to some neighborhood of the point of interest. They obtain strong consistency results of the sample version with its population counterpart. All the proposals provide a continuum between definitions of local and global depth. More recently, for the case of functional data, Agostinelli (2016) gives a definition of local depth extending the ideas introduced by Lopez-Pintado and Romo (2011) of a half-region space. This definition is also suitable for finite large dimensional datasets. Asymptotic results are obtained.

Our goal is to give a general definition of local depth for random elements in a Banach space, extending the definition of global depth given by Cuevas and Fraiman.
(2009), where they introduce the Integrated Dual Depth (IDD). The main idea of IDD is based on combining one-dimensional projections and the notion of one-dimensional depth. Let $\Omega$ be a probability space and $E$ a separable Banach space. Denote by $E'$ the separable dual space. Let $X : \Omega \rightarrow E$ be a random element in $E$ with distribution $P$ and $Q$ a probability measure in $E'$ independent of $P$. The IDD is defined as,

$$IDD(x, P) = \int D(f(x), P_f) dQ(f),$$

(1)

where $D$ is an univariate depth (for instance, simplicial or Tukey depth), $f \in E'$, $x \in E$ and $P_f$ is the univariate distribution of $f(X)$.

In the present paper we define the Integrated Dual Local Depth (IDLD). The main idea is to replace the global depth measure in Equation (1) by a local one dimensional depth measure following the definition given in (2013). We study how the classical properties, introduced by Zou and Serfling (2000), should be analyzed within the framework of local depth. We prove, under mild regularity conditions, that our proposal enjoys those properties. Moreover, uniform strong consistency results are exhibited for the definition of the empirical local depth of to the population counterpart, and also for the local depth regions. The main advantages of our proposals are its flexibility in dealing with general data and also its low computational cost, which enables it to work with high-dimensional data. As a natural application, we propose a clustering procedure based on local depths, and illustrate its performance with synthetic and real data, for different kind of data.

The remainder of the paper is organized as follows. In Section 2 we define the integrated dual local depth, and study its basic properties. Section 3 is devoted to the asymptotic study of the proposed local depth measure. In Section 4 the local depth regions are defined and the consistency results are exhibited. A clustering procedure based on local depth regions is proposed in Section 5. Simulations and real data examples are given in Section 6. Some concluding remarks are given in Section 7. All the proofs appear in the Appendix.

2 General Framework and Definitions

In this section, we first review the concept of local depth for the univariate case. Then we define the Integrated Dual Local Depth, and we finally show that, under mild regularity assumptions, our proposal has good theoretical properties that correspond to those established in Paindavaine and Van Bever (2013).
Let $P^1$ be a probability measure on $\mathbb{R}$ and $x \in \mathbb{R}$. Let $LD(x, P^1)$ be the local depth measure of $x$ with respect to $P^1$, for example, the univariate simplicial depth, that is

$$LD_S^\beta(x, P^1) = \frac{2}{\beta^2} (F^1(x + \lambda_x^\beta) - F^1(x)) (F^1(x) - F^1(x - \lambda_x^\beta)), \quad (2)$$

where $F^1$ is the cumulative distribution function of $P^1$ and $\lambda_x^\beta$ is the neighborhood width defined as follows.

**Definition 1.** Let $F$ be a univariate cumulative distribution function and $x \in \mathbb{R}$. Then, for $\beta \in (0, 1]$, we define the neighborhood width $\lambda_x^\beta$ by

$$\lambda_x^\beta = \inf \{\lambda > 0 : F(x + \lambda) - F(x - \lambda) \geq \beta\}, \quad (3)$$

where $\beta$ is the locality level.

**Remark 1.** If $F$ is absolutely continuous, the infimum in Equation (3) is attained and hence,

$$\lambda_x^\beta = \min \{\lambda > 0 : F(x + \lambda) - F(x - \lambda) \geq \beta\}.$$  

Even more, it is clear that if $\beta_1 < \beta_2$, then $\lambda_x^{\beta_1} < \lambda_x^{\beta_2}$.

The locality level $\beta$ is a tuning parameter that determines the centralness of the point $x$ of the space conditional to a given window around $x$. If the value is high it approaches the regular value of the point depth whereas if it is low it will only describe the centralness in a small neighborhood of $x$. As $\beta$ tends to one, the local depth measure tends to the depth measure.

We can also define, in an analogous way, the Tukey univariate local depth,

$$LD_H^\beta(x, P^1) = \frac{1}{\beta} \min \{F^1(x + \lambda_x^\beta) - F^1(x), F^1(x) - F^1(x - \lambda_x^\beta)\}.$$  

In what follows, without loss of generality, we restrict our attention to the case of simplicial local depth, $LD_S^\beta$.

### 2.1 Integrated Dual Local Depth

Our aim in this section is to extend the IDD introduced by Cuevas and Fraiman (2009), to the local setting. The IDD is a depth measure defined for random elements in a general Banach space. The idea is to project the data according to random directions
and compute the univariate depth measure of the projected unidimensional data. To obtain a global depth measure, these univariate depths measures are integrated. Under mild regularity conditions, the IDD satisfies the basic properties of depth measures described by Zou and Serfling (2000), and it is strongly consistent. In addition, it is important to remark that its computational cost is low, even in high dimensions, since it is based on the repeated computation of one dimensional projections.

Let \( \Omega \) be a probability space and \( E \) a separable Banach space, with \( E' \) its separable dual space. Let \( X : \Omega \rightarrow E \) be a random element in \( E \) with distribution \( P \), \( Q \) a probability measure in \( E' \) independent of \( P \), \( \beta \in (0,1] \), and \( x \in E \). We define the Integrated Dual Local Depth (IDLD),

\[
IDLD^\beta(x, P) = \int LD^\beta(f(x), P_f) dQ(f),
\]

where \( LD^\beta_s \) is the univariate local depth given in Equation (2), \( f \in E' \), \( x \in E \) and \( P_f \) is the univariate distribution of \( f(X) \). As suggested by Cuevas and Fraiman, in the infinite dimensional setting \( Q \) may be chosen to be a non-degenerate Gaussian measure and in the multivariate setting as a uniform distribution in the unitary sphere. With a slight abuse of notation, we write \( F_f = F_{f(X)} \) for the cumulative distribution function of \( f(X) \). Specifically, it reduces to

\[
F_{f(X)}(t) = P_{f(X)}((-\infty, t]) = P(f(X) \leq t).
\]

It is clear that the IDLD is well-defined, since it is bounded by \( \frac{1}{2} \) and non-negative.

Zou and Serfling (2000) established the general properties that depth measures should satisfy (P. 1 - P. 6). Paindavaine and Van Bever (2013) extend those properties to the local depth framework. We describe the properties satisfied by IDLD.

The first property deals with the invariance of the local depths. For the finite dimensional case, IDLD is independent of the coordinate system. This property is inherited from the IDD. Since IDLD is a generalization of IDD, which is not in general affine invariant (i.e., let \( A \) be a non-singular linear transformation in \( \mathbb{R}^p \) and \( P_{AX} \) denote the distribution of \( AX \); then \( D(AX, P_{AX}) \) is not equal to \( D(x, P_X) \)), neither is IDLD. It is clear that IDLD is also invariant under translations and changes of scale.

**P. 1. (affine-invariance).** Let \( E \) by a finite dimensional Banach space, \( X \in E \) a random vector, \( Q \) the Haar measure on the unit sphere of \( E' \) independent of \( P_X \). Let \( A : E \rightarrow E \) be a linear transformation such that \( |\text{det}(A)| = 1 \), \( b \in E \) and \( \beta \in (0,1] \). Then \( IDLD^\beta(Ax, P_{AX}) = IDLD^\beta(x, P_X) \).
The proof appears in the Appendix A.

**Remark 2.** It is well known that the spatial median is not affine invariant, hence, transformation and retransformation methods have been designed to construct affine equivariant multivariate medians (Chakraborty, B. and Chaudhuri 1996, 1998)). IDLD can be modified following the ideas of Kotík and Hlubinka (2017) to attain this property.

Depth measures are powerful analytical tools, especially in cases where the random element enjoy symmetry properties. Local depths should locally (restricted to certain neighborhoods) inherit these properties. Hence we give an appropriate definition of local symmetry.

**Definition 2.** Let $X$ be a real random variable and $\beta \in (0, 1]$. Then $X$ is said to be $\beta$-symmetric about $\theta$ if the cumulative function distribution $F$ satisfies

$$F\left(\theta + \lambda_\beta \right) - F(\theta) = \frac{\beta'}{2}, \text{ for every } 0 < \beta' \leq \beta. \quad (5)$$

A random element $X$ in a Banach space $E$ is $\beta$-symmetric about $\theta$ if for every $f \in E'$, $f(X)$ is $\beta$-symmetric.

The notion of $\beta$-symmetry aims to locally capture the behavior of a unimodal random variable on a neighborhood of probability $\beta$, about $\theta$, the locally deepest point. Figure 1(a) and (b) exhibit a bimodal distribution, with modes at $\theta = 1$ and $\theta = 4$. On the former, both modes are local symmetry points for $\beta = 0.25$, while on the latter $\theta = 4$ is a local symmetry point for $\beta = 0.4$ but $\theta = 1$ is not a local symmetry point for $\beta = 0.4$, the shaded area around $\theta = 1$ is non-symmetrical.

An important property of depth measures is maximality at the center, meaning that if $P$ is symmetric about $\theta$, then $D(x, P)$ attains its maximum value at that point. This property should be inherited by local depths if the distribution of $P$ is unimodal and convex. Local depths are relevant for detecting local features, for instance local centers, hence our aim is to extend the property of maximality at the center to each point $\theta$, that is $\beta$-symmetry.

**P. 2. (maximality at the center).** Let $X \in E$ be a random continuous element $\beta$-symmetric about $\theta$. For $\beta \in (0, 1]$ we have that

$$IDLD^{\beta'}(\theta, P_X) = \max_{x \in E} IDLD^{\beta'}(x, P_X), \text{ for every } 0 < \beta' \leq \beta. \quad (6)$$
$\theta = 1$ and $\theta = 4$ are local symmetry points with locality level 0.25.

(b) $\theta = 4$ is a local symmetry point with locality level 0.4, while $\theta = 1$ is not a local symmetry point at local level 0.4.

Figure 1: Local symmetry points.

The proof appears in the Appendix A.

Proposition 1 bridges the definition of $\beta$-symmetry with the usual definition of $C$-symmetry (see Zhou and Serfling 2000).

**Proposition 1.** Let $X \in \mathbb{E}$ be a random continuous element $C$-symmetric about $\theta$. Then $X$ is $\beta$-symmetric about $\theta$ for each $\beta \in (0, 1]$.

The proof appears in the Appendix A.

Proposition 2 describes the $\beta$-symmetry points of $X$.

**Proposition 2.** Let $X$ be a $\beta$-symmetric random element in $\mathbb{E}$ and $x_0 \in \mathbb{E}$ such that $LD(x_0, P) = \frac{1}{2}$ for every $0 < \beta' \leq \beta$. Then $x_0$ is a $\beta$-symmetry point.

The proof appears in the Appendix A.

**P. 3** establishes that the local simplicial depth is monotone relative to the deepest point. Several auxiliary results that appear in the Appendix A must be stated before proving this property.

**P. 3. (monotonicity relative to the deepest point).** Let $\mathbb{E}$ be a separable Banach space and $\mathbb{E}'$ the corresponding dual separable space. Let $X$ be a random $C$-symmetric element about $\theta$ with probability measure $P$. Let $Q$ be a probability measure in $\mathbb{E}'$ independent of $P$ and assume that for every $f \in \mathbb{E}'$, $f(X)$ has unimodal density function about $f(\theta)$ and fulfills

$$f_X(t) \geq 2 \frac{f_X(t + \lambda t^\beta) f_X(t - \lambda t^\beta)}{f_X(t + \lambda t^\beta) + f_X(t - \lambda t^\beta)} \quad \forall t \in \mathbb{R}, \; Q - a.s. \quad (7)$$
Then, for every \( x \in \mathbb{E} \) and \( \beta \in (0,1] \),

\[
IDLD^\beta(x, P) \leq IDLD^\beta((1-t)\theta + xt, P) \quad \text{for every } t \in [0,1].
\]

The proof appears in the Appendix A.

**Remark 3.** It is easy to see that Inequality (15) holds for the standard normal distribution. Hence, the projections of a Gaussian process fulfill \( \mathbf{P.\ 3} \).

In what follows, we show that IDLD vanishes at infinity, under mild regularity conditions.

**P. 4. (vanishing at infinity).** Assume that

\[
\sup_{\|u\|=1} \{ f : f(u) \leq \epsilon \} = O(\epsilon),
\]

where \( O(\epsilon) \) is a function such that \( \lim_{\epsilon \to 0} O(\epsilon) = 0 \)

\[
\lim_{\|x\| \to +\infty} IDLD^\beta(x, P) = 0.
\]

The proof appears in the Appendix A.

Proposition \( \mathbf{P.\ 5} \) shows that \( IDLD^\beta(x, P) \) is continuous as a function of \( x \).

**P. 5. (continuous as a function of \( x \)).** Let \( X \in \mathbb{E} \) be a random continuous element and \( \beta \in (0,1] \). Then \( IDLD^\beta(\cdot, P) : \mathbb{E} \to \mathbb{R} \) is continuous.

The proof appears in the Appendix A.

Finally, we prove that \( IDLD^\beta(x, P) \) is continuous as a functional of \( P \).

**P. 6. (continuous as a functional of \( P \)).** For every \( \beta \in (0,1] \), \( IDLD^\beta(x, \cdot) : \mathbb{E} \to \mathbb{R} \) is continuous as a functional of \( P \).

The proof appears in the Appendix A.

### 3 Empirical Version and Asymptotic Results

In this section we introduce the empirical counterpart of the IDLD and give the main asymptotic results.
First of all, recall the definition of Paindavaine and Van Bever (2013) of the empirical local unidimensional simplicial depth: Let \( ELD_S^{β(k)}(\cdot, F_n) : \mathbb{R} \to [0, 1/2] \). Then
\[
ELD_S^{β(k)}(z, F_n) = \frac{2}{β(k)^2} \left[ F_n(z + λ_{z,n}^{β(k)}) - F_n(z) \right] \left[ F_n(z) - F_n(z - λ_{z,n}^{β(k)}) \right],
\]
where
\[
λ_{z,n}^{β(k)} = \inf_{\lambda > 0} \{ F_n(z + λ_{z,n}^{β(k)}) - F_n(z - λ_{z,n}^{β(k)}) = β(k) \}.
\]

Remark 4 entails the well-definedness of the empirical neighborhood width, \( λ_{z,n}^{β(k)} \).

**Remark 4.** Let \( β \in (0, 1] \) and \( X_1, \ldots, X_n \) be a random sample of iid variables with distribution \( F \). Given \( z \in \mathbb{R} \), put, for each \( 1 \leq j \leq n, d_j(z) = |X_j - z| \) and let \( d^j(z) \) denote the \( j \)th order statistics of \( d_1(z), \ldots, d_n(z) \). Let \( k = \lfloor nβ \rfloor \), where \( \lfloor \cdot \rfloor \) is the integer part function. It is clear that \( \#\{X_j : [z - d^k(z), z + d^k(z)]\} = k \). Hence, \( F_n(z + d^k(z)) - F_n(z - d^k(z)) = \frac{\lfloor nβ \rfloor}{n} = β(k) \), and so the empirical neighborhood width is \( λ_{z,n}^{β(k)} = d^k(z) \).

Then the empirical counterpart of IDLD is given as follows.

**Definition 3.** Let \( β \in (0, 1] \), \( X : Ω \to \mathbb{E} \) be a continuous random element and \( X_1, \ldots, X_n \) a random sample with the same distribution as \( X \). Let \( k = \lfloor nβ \rfloor \). For each \( x \in \mathbb{E} \) and \( f \in \mathbb{E}' \), define
\[
λ_{f(x),n}^{β(k)} = \inf \left\{ λ > 0 : F_{f,n}(f(x) + λ) - F_{f,n}(f(x) - λ) = \frac{k}{n} \right\}. \tag{8}
\]

Let \( β(k) = \frac{k}{n} \). The empirical version of IDLD of locality level \( β(k) \) is
\[
EIDLD^{β(k)}(x, P) = IDLD^{β(k)}(x, P_n). \tag{9}
\]

In order to establish the uniform strong convergence of the one dimensional simplicial local depth, the following lemmas must be proved in advance.

**Lemma 1.** Let \( X \) be an absolutely continuous random variable with distribution \( F \). Suppose given \( X_1, \ldots, X_n \) iid random variables, also with distribution \( F \). Let \( x_p = F^{-1}(p) \) be the quantile \( p \in (0, 1) \) from \( F \) and \( Q_{p,n} \) the quantile \( p \) from \( F_n \), which is the empirical cumulative distribution function of \( X_1, \ldots, X_n \). Then,

\[
(i) \quad Q_{p,n} = X_{\lceil np \rceil}.
(ii) \quad |F_n(Q_{p,n}) - F(x_p)| \leq \frac{1}{n} \quad ∀ \ p \in (0, 1).
\]
Lemma 2. Let $X_1, \ldots, X_n$ be a real random sample with cumulative distribution function $F$. Let $\beta \in (0, 1]$ and $z \in \mathbb{R}$. Then,

$$|ELD_S^\beta(z, F_n) - LD_S^\beta(z, F)| \leq \left| F_n - F \right|_\infty + \frac{1}{n}.$$  \hfill (10)

The proof appears in the Appendix B.

The theorems below establish the uniform strong convergence of the empirical counterpart of the univariate simplicial local depth to the population counterpart.

Theorem 1. Let $E$ be a separable Banach space with a dual separable space $E'$. Suppose given $X_1, \ldots, X_n$ a random sample of elements on $E$ with probability measure $P$ and $\beta \in (0, 1]$. Then, we have

(a) $$E \left( \sup_{x \in E} |ELD_S^\beta(f(x), P_n) - LD_S^\beta(f(x), P)| \right) \xrightarrow{n \to +\infty} 0 \quad \text{for every } f \in E'.$$  \hfill (11)

(b) $$E \left( \sup_{x \in E} |EIDLD_S^\beta(x, P_n) - IDLD^\beta(x, P)| \right) \xrightarrow{n \to +\infty} 0.$$  \hfill (12)

The proof appears in the Appendix B.

Theorem 2. Let $X$ be a random element on $E$ a separable Banach space with associated probability measure $P$ such that $E(f(X)^2) < +\infty$ for every $f \in E'$. Let $X_1, \ldots, X_n$ be a random sample following the same distribution as $X$ and $\beta \in (0, 1]$. Then,

$$P \left( \sup_{x \in E} |EIDLD_S^\beta(x, P_n) - IDLD^\beta(x, P)| \xrightarrow{n \to +\infty} 0 \right) = 1.$$  \hfill (15)

The proof appears in the Appendix B.

4 Local Depth Regions

In this section we define the $\alpha$ local depth inner region at locality level $\beta$, which will be instrumental in making applications of local depth functions. Ideally, these central
regions will be invariant of the coordinate system and nested. We also study, under mild regularity conditions, the asymptotic behavior.

Denote by $LD^\beta$ a local depth measure and $ELD^\beta$ its empirical counterpart. In particular, one can consider the integrated dual local depth defined in Section 4.

**Definition 4.** Let $\mathbb{E}$ be a separable Banach space, let $X : \Omega \to \mathbb{E}$ a random element with associated probability measure $P$. Fix $\beta \in (0, 1]$, a locality level, and $\alpha \in [0, \frac{1}{2}]$. The local inner region at locality level $\beta$ of level $\alpha$ is defined to be

$$R^\alpha_\beta = \{ x \in \mathbb{E} : LD^\beta(x, P) \leq \alpha \} .$$

(13)

Let $X_1, \ldots, X_n$ be a random sample of elements on $\mathbb{E}$. Then the empirical counterpart of $R^\alpha_\beta$ is

$$R^\alpha_n = R^\alpha_{n,\beta} = \{ x \in \mathbb{E} : ELD^\beta(x, P_n) \leq \alpha \} .$$

Throughout this section the locality level $\beta$ will remain fixed, hence we write $R^\alpha$ (respectively, $R^\alpha_n$) for $R^\alpha_\beta$ (respectively, $R^\alpha_{n,\beta}$) when no ambiguity is possible.

**Remark 5.** If $\mathbb{E}$ is a finite dimensional space, then $R^\alpha$ is invariant under orthogonal transformations.

**Remark 6.** If $\alpha_1 \leq \alpha_2$, then $R^{\alpha_2}_\beta \subset R^{\alpha_1}_\beta$.

Theorem 3 shows that the empirical $\alpha$ local depth inner region at locality level $\beta$ is strongly consistent with its corresponding population counterpart, under mild regularity conditions.

**Theorem 3.** Let $\mathbb{E}$ be a separable Banach space and let $X : \Omega \to \mathbb{E}$ be a random element with associated probability measure $P$. Assume that

a) $LD^\beta(x, P) \xrightarrow{\|x\| \to +\infty} 0$

b) $\sup_{x \in \mathbb{E}} |ELD^\beta(x, P) - LD^\beta(x, P)| \xrightarrow{n \to +\infty} 0$ a.s.

Then, for every $\epsilon > 0$, $0 < \delta < \epsilon$, $0 < \alpha$ and sequence $\alpha_n \to \alpha$:

I) There exists an $n_0 \in \mathbb{N}$ such that $R^{\alpha+\epsilon}_\beta \subset R^{\alpha_n+\delta}_n \subset R^{\alpha_n}_n \subset R^{\alpha_n-\delta}_n \subset R^{\alpha-\epsilon}_\beta$.

II) If $P (x \in \mathbb{E} : LD_\beta(x) = \alpha) = 0$, then $R^{\alpha_n}_n \xrightarrow{n \to +\infty} R^\alpha$ a.s.

The proof appears in the Appendix C.
5 A Local-Depth Based Clustering Procedure

In this section we introduce a centroid-based clustering procedure based on local depths (LDC). We propose the two-stage partition method described below. The R routines needed to compute the IDLD appear in Appendix D.

Let $X$ be a random element in a separable Banach space $\mathbb{E}$, with distribution $P$.

Step 1: Core clustering region.

a) Consider the $\alpha$ local depth inner region at locality level $\beta$, $R_\beta^\alpha$, defined in Equation (13).

b) Consider a partition of $R_\beta^\alpha$ into $k$ clusters, $\tilde{C}_1^\alpha, \ldots, \tilde{C}_k^\alpha$, such that $R_\beta^\alpha = \bigcup_{i=1}^k \tilde{C}_i^\alpha$, and $P(\tilde{C}_i^\alpha \cap \tilde{C}_j^\alpha) = 0$, for $i \neq j$.

Step 2: Final clustering allocation.

Based on the initial clustering configuration for the points in $R_\beta^\alpha$, proceed to the final clustering allocation following a minimum distance rule, i.e.

$C_i^\alpha = \{ x \in \mathbb{E} : d(x, \tilde{C}_i^\alpha) \leq d(x, \tilde{C}_j^\alpha) \text{ for every, } j \neq i \}$,

where $d(x, \tilde{C}_j^\alpha) = \inf_{y \in \tilde{C}_j^\alpha} d(x, y)$.

The main idea of the proposal is to determine the center of the cluster as a region of the space rather than a single point, even though, it is well known that there is no a “one size fits all” clustering procedure, and that the election of the clustering procedure relies heavily on the underlying distribution. Our main idea is to have centers with a flexible shape allowing a better capturing of the cluster distribution. Typically, center-based clustering proposals have very good performance under spherical distributions. More flexibility in the shape of the central region should be reflected in a better performance at detecting the true clustering structure under a wide range of distributions, including elliptical distributions. In addition, since depth measures have a close relation with robustness, the core clustering regions are expected to be resistant to the presence of outliers.

In Step 1 part b), any clustering procedure can be considered; for the sake of simplicity in what follows, we use the classical $k$-means algorithm. If the number of clusters, $k$, is not given beforehand, it can be estimated using any procedure existing in the literature.
The empirical counterpart of the proposal is given in a straightforward way, employing a classical plug-in procedure.

Let $X_1, \ldots, X_n$ be iid observations in $\mathbb{E}$, a separable Banach space, with a $k$ cluster structure. Denote by $R^\alpha_n$ the $\alpha$ empirical local depth inner region at locality level $\beta$, and let $\tilde{C}^\alpha_{n,1}, \ldots, \tilde{C}^\alpha_{n,k}$ denote the initial partition obtained in Step 1 part b). The final allocation is given by,

$$C^\alpha_{n,i} = \{ x \in \mathbb{E} : d(x, \tilde{C}^\alpha_{n,i}) \leq d(x, \tilde{C}^\alpha_{n,j}) \text{ for every, } j \neq i \},$$

where $d(x, \tilde{C}^\alpha_{n,j}) = \min_{y \in \tilde{C}^\alpha_{n,j}} d(x, y)$.

**Remark 7.** The core observations of the clustering procedure can be selected considering any local depth, as long as the procedure is consistent.

### 6 Simulations and Real Data Examples

In this section we numerically analyze the performance of the clustering procedure introduced in Section 5. Simulations have been done both in the finite and infinite dimensional settings. In addition, real data examples are analyzed. The LDC procedure is implemented using not only the IDLD but also any other proposal available in the literature.

#### 6.1 Simulations: Multivariate data

The main aim of this section is to evaluate the performance of our clustering proposal under a wide range of clustering configurations. Specifically, we will analyze the case where the data presents sparseness, outliers or the sizes of the groups is not balanced. For this end, we will work under fourteen different scenarios. The original variable distribution has been proposed by Witten and Tibshirani (2010) and extended by Kondo et al. (2016). Our proposal will be challenged by several well known clustering procedures, which are briefly described.

In all the cases the data has a three group structure, each group has 300 observations. The data is generated as follows.

**Model 1:** The data are spherically generated, following $N(\mu_i, \Sigma)$, for $i = 1, 2, 3$, with centers $(-3, -3, 0), (0, 0, 0), (3, 3, 0)$, and the covariance matrix is the identity matrix.
Model 2: The data are ellipsoidally generated, following $N(\mu_i, \Sigma)$, for $i = 1, 2, 3$, with centers $(-3, -3, 0), (0, 0, 0), (3, 3, 0)$, and the covariance matrix $\Sigma = \text{diag}(3, 0.25, 1)$.

In these two models, the first two variables are informative while the last one is noise.

Model 3, (respectively, Model 4) are five dimensional datasets. The first three variables have the same distribution as Model 1 (respectively, Model 2), the remaining variables are two independent noisy variables, with distribution $N(0, 1)$.

We then consider two different contamination settings. In each of them we add five outliers, we only replace one coordinate by a variable generated with uniform distribution in the interval $[25, 25.01]$. In the first setting, for Models 5-8, the contamination is done by replacing the first coordinate (which is an informative variable) of the first five observations of the first cluster, while the rest of the distribution remains as in Models 1-4. In Models 9-12, the contamination has the same distribution but is situated in the last coordinate, which is a non-informative variable. The two remaining models, 13 and 14, have clusters with unbalanced sizes, the same distributions are followed as in Models 1 and 2, but instead of having 100 observations each cluster, the first cluster has 60% of the observations, while the two remaining clusters have 20% each.

The benchmark clustering procedures are:

- The $k$-means algorithm, we consider ten random initializations.
- The sparse $k$-means clustering procedure (SKM), introduce by Witten and Tibshirani (2010). The tuning parameter, $L_1$, bound is chosen, as suggested in the literature ($s = 3, 7$), and five random initializations are considered.
- The robust and sparse $k$-means clustering procedure (RSKM), proposed by Kondo et al. (2016). Two tuning parameters must be set. Both of them have been set as suggested in [12]: the parameter that corresponds to the $L_1$ norm is $L_1 = 4$ and the trimming proportion is 0.1.
- The model-based clustering procedure (MCLUST) proposed by Fraley and Raftery (2002, 2009), designed to cluster mixtures of $G$ normals distributions.

SKM is designed to cluster observations in a high dimensional setting, with a low proportion of clustering informative variables. RSKM is a robust extension of SKM.

The LDC introduced in Section 5 has been implemented using three definitions of local depth, every case the parameters where chosen following Hennig [8], and the results were very stable.
• The simplicial local depth procedure (LDCS) introduced by Agostinelli and Romanazzi (2011). We used the R package *localdepth*, the threshold value for the evaluation of the local depth, $\tau$, was calculated with the *quantile.localdepth* function, as suggested in the same R package, and the quantile order of the statistic was set to $probs = 0.1$.

• Local version of depth at locality level $\beta$ (LDCPV) according to proposals of Paindaveine and Van Bever (2013), using the R package *DepthProc*. We set $\beta = 0.2$.

• Integrated dual local depth at locality level $\beta$ (LDCI) introduced in Section 4. As with the LDCPV we set $\beta = 0.2$ and set the number of random projections $N = 50$, with standard normal distribution. Routines are available in the Appendix E of the Supplementary Material.

The parameter $\alpha$ represents the proportion of data which will contain the core regions of the clusters, if this value is very small the procedure will have a very similar behavior to $k$-means, not being able to capture the shape of the clusters. If it takes high values, the core regions will have observations with moderate local depth, that can lead to errors in the assignments. For these reasons we suggest taking values between 0.15 and 0.45. To set this parameter we perform an analysis of the sensitivity, following the resampling ideas proposed by Hennig [8], from them we could see that in all cases the method is stable, as in most cases $\alpha = 0.4$ showed slightly better performance we settled this value throughout the study. We performed $M = 500$ replicates for each model.

There is no commonly accepted criterion for evaluating the performance of a clustering procedure. Nonetheless, since we are dealing with synthetic datasets, we know the real label of each observation, hence in these cases we may use the Correct Classification Rate (CCR). We denote the original clusters by $k = 1, \ldots, K$. Let $y_1, \ldots, y_n$ be the group label of each observation, and $\hat{y}_1, \ldots, \hat{y}_n$ the class label assigned by the clustering algorithm. Let $\tilde{\Sigma}$ be the set of permutations over $1, \ldots, K$. Then the CCR is given by:

$$CCR = \min_{\sigma \in \tilde{\Sigma}} \frac{1}{n} \sum_{i=1}^{n} I_{\{y_i \neq \sigma(\hat{y}_i)\}}.$$  \hspace{1cm} (14)

The results of the simulation are exhibited in Table 1. As expected, all the clustering procedures have an exceptional performance for Models 1 and 3, where all the clusters are spherical without outliers. For Models 2 and 4, where the clusters have an elliptical distribution, MCLUST has an outstanding performance and it is clear that LDC (with
any local depth measure) performs better than the other three alternatives. In Models 5 to 12, since $k$-means, SKM and MCLUST are nonrobust procedures, they fail in the classification of the observation, typically the five outliers make up one group and the cluster with mean $(0, \ldots, 0)$ is usually split into two clusters. LDC and RSKM are based on more robust clustering criteria, hence both methods have a good performance; RSKM seems to perform better under spherical distributions while LDC performs better under elliptical distribution. It is clear, that LDC has a good performance for Models 1 to 12, and that the choice of the local depth is not crucial. Nonetheless, when cluster sizes are unbalanced the only criteria able to correctly detect the cluster structure are MCLUST and LDC considering the integrated dual local depth. It is clear that LDC combined with the other two proposals of local depths is not able to detect the center of the clusters. The remainder of the clustering procedures had a good performance on the spherical case but failed on the elliptical case. In summary, LDCI is the only clustering procedure versatile enough to detect clusters under adverse situations (sparse data, outliers and unbalanced cluster size).

In what follows we compare the computational times for the three local depths measures. The simulation were based on data generated according to Model 3, but instead of having three noise variables, we added $p-2$, $(p = 5, 35, 65)$ normal independent noise variables centered at the origin with unit standard deviation. Also we considered different sample sizes, $n = 300, 2100, 3900$ and $5700$. For ILDL 50, random directions were generated. Since the computational time increases exponentially as the dimension increases, we only performed $M = 50$ replicates under each scenario.

From Table 2 we can see that in every case IDLD is the fastest procedure, moreover it is not affected by the dimension of the dataset, while the computational efforts required by LDS and LDPV grow dramatically as $p$ increases. LDPV is overall the slowest procedure. Even though all the procedures demand more time as the sample size grows, IDLD is the one with the least pronounced growth rate.

### 6.2 Simulations: Multivariate functional data

In this section we present the results of a simulation study for multivariate functional data; for such a multivariate setting, there are scarcely any clustering procedures. We will replicate the simulation done by Schumtz et al. (2017). They present three different scenarios. In every case, the data es bivariate.

Model A. Three groups, each of them with 100 observations.

16
Table 1: Mean CCR for each clustering criterion and distribution configuration

| Model | \(k\)-means | SKM | RSKM | MCLUST | LDCS | LDCPV | LDCI |
|-------|-------------|-----|------|--------|------|--------|------|
| 1     | 0.98        | 0.98| 0.98 | 0.98   | 0.96 | 0.95   | 0.97 |
| 2     | 0.87        | 0.80| 0.86 | 0.99   | 0.91 | 0.87   | 0.91 |
| 3     | 0.98        | 0.98| 0.98 | 0.98   | 0.96 | 0.96   | 0.97 |
| 4     | 0.87        | 0.80| 0.85 | 0.99   | 0.89 | 0.90   | 0.90 |
| 5     | 0.66        | 0.70| 0.96 | 0.65   | 0.95 | 0.92   | 0.95 |
| 6     | 0.65        | 0.62| 0.84 | 0.66   | 0.90 | 0.85   | 0.87 |
| 7     | 0.67        | 0.70| 0.96 | 0.65   | 0.94 | 0.94   | 0.95 |
| 8     | 0.65        | 0.62| 0.84 | 0.66   | 0.88 | 0.89   | 0.87 |
| 9     | 0.65        | 0.68| 0.98 | 0.65   | 0.95 | 0.94   | 0.95 |
| 10    | 0.65        | 0.65| 0.86 | 0.66   | 0.91 | 0.84   | 0.89 |
| 11    | 0.65        | 0.67| 0.98 | 0.65   | 0.95 | 0.86   | 0.96 |
| 12    | 0.65        | 0.66| 0.85 | 0.66   | 0.88 | 0.95   | 0.90 |
| 13    | 0.97        | 0.98| 0.98 | 0.97   | 0.54 | 0.46   | 0.96 |
| 14    | 0.74        | 0.70| 0.69 | 0.98   | 0.52 | 0.43   | 0.82 |

Here \(a_1 \sim N(0, 0.2), a_2 \sim N(0, 0.3), e_1(t)\) is white noise with variance \(\frac{a_1^2}{2}\), and \(e_2(t)\) is white noise with variance \(\frac{a_2^2}{2}\). The curves are generated for 101 equidistant points in the interval \([0, 1]\).

Model B. Four groups, each of them with 250 observations.

Here \(t \in [1, 21], U \sim U(0, 0.1),\) and \(e_1(t)\) is white noise independent of \(U\) with variance 0.25. The functions are \(h_1(t) = (6 - |t - 7|)_+\) and \(h_2(t) = (6 - |t - 15|)_+,\) where \((\cdot)_+\) means the positive part. The curves are generated at 101 equidistant points in the interval \([0, 1]\).

Model C. Four groups, each of them with 250 observations.

Here, \(t \in [1, 21],\) while \(U, e(t), h_1\) and \(h_2\) are defined as before. The curves are generated at 101 equidistant points in the interval \([0, 1]\).

As in the original paper, the estimated partition will be compared with the theoretical one via the Adjusted Rand Index (ARI), from the function \(AdjustedRandIndex\) from the
Table 2: Mean computer time for LDS, LDPV and IDLD.

|     | n   | 300  | 2100 | 3900 | 5700 |
|-----|-----|------|------|------|------|
| 5   | LDS | 0.785| 38.27| 131.65| 280.66|
|     | LDPV| 4.236| 100.08| 292.67| 624.91|
|     | IDLD| 0.397| 20.74| 73.74| 160.43|
| 35  | LDS | 1.770| 86.88| 299.03| 638.38|
|     | LDPV| 7.840| 200.94| 629.07| 1363.97|
|     | IDLD| 0.402| 20.68| 74.41| 160.29|
| 65  | LDS | 3.788| 184.92| 641.01| 1368.31|
|     | LDPV| 10.934| 288.79| 982.79| 2094.89|
|     | IDLD| 0.406| 20.66| 75.07| 164.40|

Group 1:  
\[X_1(t) = \sin((10 + a_1)t) + (1 + a_1) + e_1(t)\]  
\[X_2(t) = \sin((5 + a_2)t) + (0.5 + a_2) + e_2(t)\]

Group 2:  
\[X_1(t) = \sin((5 + a_2)t) + (0.5 + a_2) + e_2(t)\]  
\[X_2(t) = \sin((15 + a_1)t) + (1 + a_1) + e_1(t)\]

Group 3:  
\[X_1(t) = \sin((15 + a_1)t) + (1 + a_1) + e_1(t)\]  
\[X_2(t) = \sin((10 + a_1)t) + (1 + a_1) + e_1(t).\]

mclust R package. For each model, 50 replications where carried out. Schmutz et al. (2017), report the ARI for settings settings of their proposal, and also for funclust (2014) as well as kmeans-d1 and kmeans-d2, which are two proposals introduced by Ieva et al. (2013). In Table 6.2 we present the maximum value of the ARI for Schmutz et al. and
Group 1: \[ X_1(t) = U + (1 - U)h_1(t) + e(t) \]
\[ X_2(t) = U + (0.5 - U)h_1(t) + e(t) \]

Group 2: \[ X_1(t) = U + (1 - U)h_2(t) + e(t) \]
\[ X_2(t) = U + (0.5 - U)h_2(t) + e(t) \]

Group 3: \[ X_1(t) = U + (1 - U)h_1(t) + e(t) \]
\[ X_2(t) = U + (1 - U)h_1(t) + e(t) \]

Group 4: \[ X_1(t) = U + (0.5 - U)h_2(t) + e(t) \]
\[ X_2(t) = U + (0.5 - U)h_1(t) + e(t) \].

the remainder of the procedures. It is clear that LDCI outperforms by far the rest of the proposals, since it does not misclassify any observation throughout the simulation study.

Table 3: ARI for different clustering procedures for multivariate functional data.

|                | Model A | Model B | Model C |
|----------------|---------|---------|---------|
| LDCI           | 1       | 1       | 1       |
| Best Schmutz   | 0.96    | 0.92    | 0.80    |
| funclust       | 0.23    | 0.36    | 0.45    |
| kmeans - $d_1$| 0.90    | 0.37    | 0.32    |
| kmeans - $d_2$| 0.90    | 0.37    | 0.32    |

Computational results functional data, considering synthetic and real examples appear in Appendix D.

### 6.3 Real data examples for mixed-type datasets

Our aim in this Section is to analyze data set AEMET, from the R library *fda.esc*. This dataset contains series of daily summaries of 73 spanish weather stations selected for the period 1980-2009. We will analyze the clustering structure of the dataset conformed by the variables: mean daily wind speed during between 1980 and 2009 (which is a functional variable) and geographic information of each station: altitud, latitud and height, which are real variables. Analyzing these variables together is relevant given that height influences in the intensity of the winds. Although the sensors are located at the same height, it is possible that phenomena related to the climate of the region
generate deformations in the curves given by the intensity of the wind. To apply the LDC clustering criterion, we must be precise in the definition of the IDLD in data sets that have these characteristics. Our proposal is to project the functional variable as we have done in Section 6.2 and the multivariate variables as in Section 6.1. Then, we join those two projections with equal weight, and compute the IDLD. We look for two clusters, the parameters of the clustering procedure are $\alpha = 0.15$ and $\beta = 0.3$. These parameters have been settled upon visual considerations of the dataset. After performing the clustering analysis we obtained two groups, one of them corresponds to the coastal stations (orange stations) while the other one corresponds to the continental ones (red stations), as it can be seen in Figure 2. This classification corresponds to the well-known fact that the wind speed is more constant over the coastal areas. An example can be found in the use made of wind farms.

![Image of geographical position of meteorological stations](image)

**Figure 2**: Geographical position of each meteorological station. The stations that belong to the coastal group are in orange, while the ones that belong to the continental stations appear in red.

Finally, to understand the conformation of the groups in an integral way, it is convenient to analyze the core regions for the mean speed of the wind and the height of the stations. It can be seen that the stations corresponding to the core continental region are at higher altitudes, suffer more variability in wind intensity, as shown in the left and right panels of Figure 3. However, the coast stations that they are in lower zones have
less daily variability and apparently the wind has greater intensity, as can be seen in the central and right panels of Figure 3.

![Figure 3](image)

Figure 3: Left: The red curves correspond to the core observations of the mean wind speed for the coast cluster. Center: The yellow curves are the core observations of the mean wind speed for the continental cluster. Right: Grouping conformation for the height, coast cluster in red and continental cluster in yellow.

7 Final remarks

In this paper, we introduced a local depth measure, IDLD, suitable for data in a general Banach space with low computational burden. It is an exploratory data analysis tool, which can be used in any statistical procedure that seeks to study local phenomena. From the theoretical perspective, local depths are expected to be generalizations of a global depth measure. Our proposal has this property. Additionally, they are expected to inherit good properties from global depths: this point has been overlooked for local depths. Strong consistency results for the local depth and local depth regions have been proved.

From the practical point of view, we explored the use of local depth measures in cluster analysis, introducing a simple clustering procedure. The first stage is to split into \( k \) groups the \( \alpha \) local inner region. The points are assigned to the closest group of the \( \alpha \) local inner region. The flexibility of shape of the groups made up by the points in the \( \alpha \) local inner region, produces a flexibility of the shapes in the groupings of the entire space. Computational experiments reflect this fact by showing an extraordinary performance under a wide range of clustering configurations.
References

[1] Agostinelli, C. (2018). “Local half-region depth for functional data.” *Journal of Multivariate Analysis* 163, 67-79.

[2] Agostinelli, C., and M. Romanazzi. (2011). “Local Depth.” *Journal of Statistical Planning and Inference*, 141, 817-830.

[3] Chakraborty, B., and P. Chaudhuri (1996). “On transformation and retransformation technique for constructing an affine equivariant multivariate median.” *Proceedings of the American Mathematical Society* 124, 2539-2547.

[4] Chakraborty, B., and P. Chaudhuri (1998). “Operating transformation retransformation on spatial median angle test.” *Statistica Sinica* 8, 767-784.

[5] Cuevas, A., and R. Fraiman (2009). “On depth measures and dual statistics. A methodology for dealing with general data.” *Journal of Multivariate Analysis* 100(4), 753-766.

[6] Fraley C., and A. E. Raftery (2002). “Model-based clustering, discriminant analysis, and density estimation.” *Journal of the American Statistica Association* 97, 611631.

[7] Fraley C., and A. E. Raftery (2009). “MCLUST Version 3 for R: Normal Mixture Modeling and Model-based Clustering,” Technical Report No. 504, Department of Statistics, University of Washington.

[8] Henning, C (2007). “Cluster-wise assesment of cluster stability.” *Computational Statistics and Data Analysis*, 52, 258?271.

[9] Ieva F., A. M. Paganoni, D. Pigoli, and V. Vitelli (2013). “Multivariate functional clustering for the morphological analysis of electrocardiograph curves.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 62,401-418. doi:10.1111/j.1467-9876.2012.01062.x

[10] Jacques, J., and C.Preda (2014). “Model-based clustering of functional data.” *Computational Statistics and Data Analysis* 71, 92-106. DOI:10.1016/j.csda.2012.12.004

[11] Kotík, L., and D. Hlubinka (2017). “A weighted localization of halfspace depth and its properties.” *Journal of Multivariate Analysis* 157, 53-69.

[12] Kondo, Y., M. Salibian-Barrera, and R. H. Zamar (2016). “A robust and sparse K-means clustering algorithm.” *Journal of Statistical Software* 72(5).
[13] Lopez-Pintado, S., and J. Romo (2011). “A half-region depth for functional data.” *Computational Statistics and Data Analysis* 55(4), 1679-1695.

[14] Paindavaine, D., and G. Van Bever (2013). “From depth to local depth: A focus in centrality.” *Journal of the American Statistical Association* 108(503), 1105-1119.

[15] Schmutz, A., J. Jacques, C. Bouveyron, L. Cheze, and P. Martin (2017). “Clustering Multivariate functional data in group-specific functional subspaces.” (Unpublished) [https://hal.inria.fr/hal-01652467/file/Clusteringmultivariatedata.pdf](https://hal.inria.fr/hal-01652467/file/Clusteringmultivariatedata.pdf)

[16] Witten, D., and R. Tibshirani (2010). “A framework for feature selection in clustering.” *Journal of the American Statistical Association* 105(490), 713-726.

[17] Zou, Y., and R. Serfling (2000). “General Notion of Statistical Depth Function.” *The Annals of Statistics* 28(2), 461-482.
8 Appendix A: Proofs of properties P.1-6.

Proof: P. 1. (affine - invariance). Since $\mathbb{E}$ has finite dimension, without loss of generality we assume that $\mathbb{E} = \mathbb{R}^d$.

$$IDLD^\beta(Ax, P_{AX}) = \int LD^\beta_S(f(Ax), P_{f\circ A})Q(f) = \int LD^\beta_S(A^*(f)(x), P_{A^*(f)})Q(f).$$

By the change of variables theorem

$$IDLD^\beta(Ax, P_{AX}) = \int LD^\beta_S(f(x), P_f)|\det(A)|^{-1}Q(A^*(f)).$$

Since Haar measure is invariant under unitary linear transformations and $|\det(A)| = 1$, we have that $LD^\beta(Ax, P_{AX}) = LD^\beta(x, P_X)$.

\[
\]

Proof (Proposition 1). Let $\beta \in (0, 1]$, $X$ is $C$-symmetric about $\theta$ if for every $f \in \mathbb{E}'$, $f(X)$ is symmetric about $f(\theta)$. Then, for every $0 < \beta' \leq \beta$,

$$\beta' = F_f(f(\theta) + \lambda_f^\beta) - F_f(f(\theta) - \lambda_f^\beta) = 2 \left( F_f(f(\theta) + \lambda_f^\beta) - F_f(f(\theta)) \right).$$

Finally,

$$\frac{\beta'}{2} = F_f(f(\theta) + \lambda_f^\beta) - F_f(f(\theta)),$$

which is what we wanted to show.

Proof (Proposition 2). First note that, given $f \in \mathbb{E}'$, $x \in \mathbb{E}$ and $\beta \in (0, 1]$,

$$\beta = F_f \left( f(x) + \lambda_f^\beta(x) \right) - F_f \left( f(x) - \lambda_f^\beta(x) \right)$$

$$F_f \left( f(x) + \lambda_f^\beta(x) \right) - F_f(f(x)) = \beta - \left( F_f(f(x)) - F_f \left( f(x) - \lambda_f^\beta(x) \right) \right).$$
From the definition of $LD_{S}^{\beta}(x, P)$ is clear that,

$$LD_{S}^{\beta}(f(x), P_f) = \frac{2}{\beta^2} \left[ \beta - \left( F_f(f(x)) - F_f\left(f(x) - \lambda^{\beta}_{f(x)}\right) \right) \right] \left[ F_f \left(f(x) + \lambda^{\beta}_{f(x)}\right) - F_f(f(x)) \right].$$

Let $h : [0, \beta] \to \mathbb{R}, h(t) = \frac{2}{\beta^2}(\beta - t)t$, attains a global maximum at $t = \frac{\beta}{2}$, hence $LD_{S}^{\beta}$ attains its maximum when $F_f\left(f(x) + \lambda^{\beta}_{f(x)}\right) - F_f(f(x)) = \frac{\beta}{2}$. If this property is satisfied for every $0 < \beta' \leq \beta$, then, $x$ is a $\beta$-symmetry point of $f(X)$.

Then, let $0 < \beta' \leq \beta$,

$$\frac{1}{2} = IDLD_{S}^{\beta'}(x_0, P) = \int LD_{S}^{\beta'}(f(x_0), P_f)dQ(f) = \int \left( \frac{1}{2} - LD_{S}^{\beta'}(f(x_0), P_f) \right) dQ(f).$$

For every $f \in \mathbb{E}$, $LD_{S}^{\beta}(f(x_0), P_f)$ is bounded by $\frac{1}{2}$. Hence, $LD_{S}^{\beta}(f(x_0), P_f) = \frac{1}{2} Q - a.s.$ From the first part of the proof we know that $f(x_0)$ is a $\beta$-symmetry point of $f(X)$, hence $x_0$ is a $\beta$-symmetry point of $X$.

We now focus on the proof of P. 3 that establishes that the local depth is monotone relative to the deepest point. We first show that this results holds if the distribution is unimodal. We begin by proving several auxiliary results that we will need to prove P. 3.

**Lemma 3 (P.3).** Let $X$ be an absolutely continuous, symmetric and unimodal about $t = 0$ random variable with cumulative distribution function $F$. Let $\beta \in (0, 1]$, define the functions,

- $U(t) = F(t + \lambda^{\beta}_{t}) - F(t)$.
- $V(t) = F(t) - F(t - \lambda^{\beta}_{t})$.

Then, for every $t \in \mathbb{R}$ we have that:

a) If $t \geq 0 \Rightarrow U(t) \leq \beta / 2 \leq V(t)$.

b) If $t \leq 0 \Rightarrow V(t) \leq \beta / 2 \leq U(t)$.

**Proof.** a) It is clear that if $t = 0$ then by symmetry the equality is attained.

Let $t > 0$ and $f_X$ be the density function of $X$. There are two possible cases to analyze:
i) If \(-t < t - \lambda_t^\beta\) :

Since \(f_X(s)\) decreases on \((0, +\infty)\), we have that

\[
\min_{0 \leq s \leq t} f_X(s) \geq \max_{t \leq s \leq t + \lambda_t^\beta} f_X(s).
\]

By symmetry \(f_X(-s) = f_X(s)\) for every \(s \in [0, t]\), then

\[
\min_{-t \leq s \leq t} f_X(s) = \min_{0 \leq s \leq t} f_X(s).
\]

On the other hand, \([t - \lambda_t^\beta, t] \subset [-t, t]\), since \(-t < t - \lambda_t^\beta\), which implies that

\[
\min_{-t \leq s \leq t} f_X(s) \leq \min_{t - \lambda_t^\beta \leq s \leq t} f_X(s).
\]

Thus,

\[
U(t) - V(t) = \int_t^{t + \lambda_t^\beta} f_X(s) ds - \int_{t - \lambda_t^\beta}^t f_X(s) ds \leq \int_t^{t + \lambda_t^\beta} \max_{t \leq s \leq t + \lambda_t^\beta} f_X(s) - \int_{t - \lambda_t^\beta}^t \min_{t - \lambda_t^\beta \leq s \leq t} f_X(s) =
\]

\[
= \lambda_t^\beta \max_{t \leq s \leq t + \lambda_t^\beta} f_X(s) - \lambda_t^\beta \min_{t - \lambda_t^\beta \leq s \leq t} f_X(s) =
\]

\[
= \lambda_t^\beta \left( \max_{t \leq s \leq t + \lambda_t^\beta} f_X(s) - \min_{t - \lambda_t^\beta \leq s \leq t} f_X(s) \right) \leq 0
\]

ii) If, \(t - \lambda_t^\beta < -t\) :

Observe that if \(s \in [t, -t + \lambda_t^\beta]\), we have that \(-s \in [t - \lambda_t^\beta, -t]\) and since the density function is symmetric about \(t = 0\), we know that \(f_X(s) = f_X(-s)\). Hence,

\[
\int_t^{-t + \lambda_t^\beta} f_X(s) ds = \int_{t - \lambda_t^\beta}^{-t} f_X(s) ds.
\]

Since \(f_X\) decreasing, we obtain the following inequalities,

\[
\max_{\lambda_t^\beta \leq s \leq t + \lambda_t^\beta} f_X(s) \leq \max_{-t + \lambda_t^\beta \leq s \leq \lambda_t^\beta} f_X(s) \leq \min_{0 \leq s \leq t} f_X(s).
\]
Then,

\[ U(t) - V(t) = \int_{t}^{t+\lambda_t^\beta} f_X(s)ds - \int_{t-\lambda_t^\beta}^{t} f_X(s)ds = \]

\[ = \int_{t}^{t+\lambda_t^\beta} f_X(s)ds + \int_{t-\lambda_t^\beta}^{t+\lambda_t^\beta} f_X(s)ds - \int_{t-\lambda_t^\beta}^{t} f_X(s)ds - \int_{-t}^{t} f_X(s)ds = \]

\[ = \int_{t-\lambda_t^\beta}^{t+\lambda_t^\beta} f_X(s)ds - \int_{-t}^{t} f_X(s)ds = \]

\[ = \int_{-t}^{t} f_X(s)ds + \int_{t-\lambda_t^\beta}^{t+\lambda_t^\beta} f_X(s)ds - 2 \int_{0}^{t} f_X(s)ds \leq \]

\[ \leq t \max_{-t+\lambda_t^\beta \leq s \leq t} f_X(s) + t \max_{\lambda_t^\beta \leq s \leq t+\lambda_t^\beta} f_X(s) - 2t \min_{0 \leq s \leq t} f_X(s) \leq \]

\[ \leq 2t \left( \max_{-t+\lambda_t^\beta \leq s \leq t} f_X(s) - \min_{0 \leq s \leq t} f_X(s) \right) \leq 0. \]

Finally, since \( U(t) + V(t) = \beta \) and \( U(t) \leq V(t) \Rightarrow U(t) \leq \beta/2 \leq V(t) \).

b) Consider the random variable \(-X\) which is absolutely continuous, symmetric and unimodal about \( t = 0 \). Denote \( F_X \) the cumulative distribution function of \( X \) and \( F_{-X} \) the cumulative distribution function of \(-X\). In addition, observe that given \( t \in \mathbb{R} \)

\[ F_{-X}(t) = P(-X \leq t) = P(X \geq -t) = 1 - F_X(-t). \]

\[ U_{-X}(t) = F_{-X}(t + \lambda_t^\beta) - F_{-X}(t) = 1 - F_X(-t - \lambda_t^\beta) - (1 - F_X(-t)) = \]

\[ = F_X(-t) - F_X(-t - \lambda_t^\beta) = V_X(-t). \]

Analogously,

\[ V_{-X}(t) = F_{-X}(t) - F_{-X}(t - \lambda_t^\beta) = 1 - F_X(-t) - \left( 1 - F_X(-(t - \lambda_t^\beta)) \right) = \]

\[ = F_X(-t + \lambda_t^\beta) - F_X(-t) = U_X(-t). \]

Then, if \( t < 0 \) we have that \(-t > 0\) and since part (a) of the proof holds we have that,

\[ U_{-X}(-t) \leq \beta/2 \leq V_{-X}(-t) \Rightarrow V_X(t) \leq \beta/2 \leq U_X(t). \]

\[ \square \]
Lemma 4 (P.3). Let $X$ be an absolutely continuous random variable with $C^1$ cumulative distribution function, $F_X$. Let $\beta \in (0, 1]$. Let $t_0 \in \mathbb{R}$ such that the density function $f_X$ satisfies \( f(t_0 - \lambda_{t_0}) \in Sop(f_X) \) and \( f(t_0 + \lambda_{t_0}) \in Sop(f_X) \). Then, there exists an interval $I$ centred at $t_0$ and function $\lambda^\beta : I \to \mathbb{R}_{\geq 0}$ such that $\lambda$ is $C^1$ on $I$, $\lambda^\beta(t_0) = \lambda_{t_0}^\beta$. Even more, for each $s \in I$,

\[
\frac{\partial \lambda^\beta}{\partial t}(s) = -\frac{f_X(t + \lambda^\beta(s)) - f_X(t - \lambda^\beta(s))}{f_X(t + \lambda^\beta(s)) + f_X(t - \lambda^\beta(s))}
\]

Proof. The proof follows straightforward applying the implicit function theorem to the function $g : \mathbb{R} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$,

\[
g(x, \lambda) = F_X(x + \lambda) - F_X(x - \lambda) - \beta.
\]

Then we have,

\[
\frac{\partial g}{\partial t}(t, \lambda) = \frac{\partial}{\partial t} (F_X(t + \lambda) - F_X(t - \lambda) - \beta) = f_X(t + \lambda) - f_X(t - \lambda).
\]

\[
\frac{\partial g}{\partial \lambda}(t, \lambda) = \frac{\partial}{\partial \lambda} (F_X(t + \lambda) - F_X(t - \lambda) - \beta) = f_X(t + \lambda) + f_X(t - \lambda).
\]

Lemma 5 (P.3). Let $X$ be an absolutely continuous, symmetric and unimodal about $t = 0$ random variable, such that the cumulative distribution function $F_X$ is $C^1$. Let $\beta \in (0, 1]$, and $f_X$ the density function such that $f_X(t + \lambda^\beta_t)f_X(t - \lambda^\beta_t) > 0$, which in addition satisfies that

\[
f_X(t) \geq 2 \frac{f_X(t + \lambda^\beta_t)f_X(t - \lambda^\beta_t)}{f_X(t + \lambda^\beta_t) + f_X(t - \lambda^\beta_t)} \quad \forall t \in \mathbb{R}. \tag{15}
\]

Then,

a) $LD_{S}^{\beta}(t, F_X)$ is non increasing if $t > 0$.

b) $LD_{S}^{\beta}(t, F_X)$ is non decreasing if $t < 0$.

Proof. Following Lemma 4 and for the sake of simplicity denote $\lambda^\beta_t = \lambda^\beta(t)$.

It is clear that,

\[
LD_{S}^{\beta}(t, F_X) = \frac{2}{\beta^2} \frac{[F_X(t + \lambda(t)) - F_X(t)] [\beta - (F_X(t + \lambda(t)) - F_X(t))]}{[\beta - (F_X(t + \lambda(t)) - F_X(t))]} = \frac{2}{\beta^2} U(t)(\beta - U(t)).
\]
The derivative of $LD_S^\beta$ with respect to $t$ is:

$$\frac{\partial LD_S^\beta}{\partial t}(t, F_X) = \frac{2}{\beta^2} \left[ \beta \frac{\partial U}{\partial t}(t) - 2U(t) \frac{\partial U}{\partial t}(t) \right] = \frac{2}{\beta^2} \frac{\partial U}{\partial t}(t)[\beta - 2U(t)].$$

By Lemma 4 and considering the derivative of $U(t)$ respect to $t$, we have that:

$$\frac{\partial U}{\partial t}(t) = \frac{\partial}{\partial t} (F_X(t + \lambda(t)) - F_X(t)) = fx(t + \lambda(t)) \left(1 + \frac{\partial \lambda}{\partial t}(t)\right) - fx(t) =$$

$$= fx(t + \lambda(t)) + fx(t + \lambda(t)) \frac{\partial \lambda}{\partial t}(t) - fx(t) =$$

$$= fx(t + \lambda(t)) - fx(t) - fx(t + \lambda(t)) \frac{fx(t + \lambda(t)) - fx(t - \lambda(t))}{fx(t + \lambda(t)) + fx(t - \lambda(t))} =$$

$$= 2 \frac{fx(t + \lambda(t)) - fx(t - \lambda(t))}{fx(t + \lambda(t)) + fx(t - \lambda(t))} - fx(t) \leq 0.$$

From Lemma 3 we have that:

a) If $t < 0$, $U(t) > \beta_2 \Rightarrow \frac{\partial LD_S^\beta}{\partial t}(t, F_X) = \frac{2}{\beta^2} \frac{\partial U}{\partial t}(t)[\beta - 2U(t)] \geq 0.$

b) if $t > 0$, $U(t) < \beta_2 \Rightarrow \frac{\partial LD_S^\beta}{\partial t}(t, F_X) = \frac{2}{\beta^2} \frac{\partial U}{\partial t}(t)[\beta - 2U(t)] \leq 0.$

\[\square\]

**Lemma 6 (P.3).** Let $X$ be a random variable and $\mu \in \mathbb{R}$. Let $\beta \in (0,1]$ and $Y = X - \mu$.
Denote $F_X$ and $F_Y$ to the corresponding cumulative distribution functions, then, $LD_S^\beta(t, F_X) = LD_S^\beta(t - \mu, F_Y)$.

**Proof.** Let $t \in \mathbb{R}$, we have that $F_X(t) = F_Y(t - \mu)$. Then,

$$U_X(t) = F_X(t + \lambda^\beta_t) - F_X(t) = F_Y(t - \mu - \lambda^\beta_t) - F_Y(t - \mu) = U_Y(t - \mu).$$

entails the desired equality.

\[\square\]

Finally we prove P. 3.

**Proof: P. 3.** (monotonicity relative to the deepest point). Let $t \in \mathbb{R}$ and $Y = X - \theta$.
Suppose that $t > \theta$ then $t - \theta > 0$. On the other hand, $(1 - s)\theta + st = \theta + s(t - \theta)$ and $s(t - \theta) < t - \theta$. Then, Lemmas 5 and 6 entail that,

$$LD^\beta_S(t, F_X) = LD^\beta_S(t - \theta, F_Y) \leq LD^\beta_S(s(t - \theta), F_Y) = LD^\beta_S(s(t - \theta) + \theta, F_X) = LD^\beta_S((1-s)\theta + st, F_X).$$

\[\square\]
Proof: P. 4. (vanishing at infinity). Let $\beta \in (0, 1]$, it is clear that, $0 \leq F_f(f(x)) - F_f(f(x) - \lambda_{f(x)}) \leq \frac{\beta}{2}$. Given that $F_f$ is a cumulative distribution function then, for every $f \in \mathcal{E}'$, $F_f(f(x) - \lambda_{f(x)}) \leq 1$.

Hence,

$$IDLD^{\beta}(x, P) = \int \frac{2}{\beta^2} \left[ F_f(f(x)) - F_f(f(x) - \lambda_{f(x)}) \right] \left[ F_f(f(x) + \lambda_{f(x)}) - F_f(f(x)) \right] dQ(f)$$

$$\leq \int \frac{2}{\beta^2} \left[ F_f(f(x) + \lambda_{f(x)}) - F_f(f(x)) \right] dQ(f)$$

$$\leq \frac{1}{\beta} \int [1 - F_f(f(x))] dQ(f) = \frac{1}{\beta} \int P(f(X) > f(x)) dQ(f).$$

Let $\epsilon > 0$, $M > 0$ and $x \in \mathcal{E}$ such that $\|x\| \leq M$, thus

$$\int P(f(X) > f(x)) dQ(f) \leq \int \mathcal{I}_{\{f : \frac{f(x)}{\|x\|} \leq \epsilon\}} dQ(f) + \int P(f(X) > f(x)) \mathcal{I}_{\{f : \frac{f(x)}{\|x\|} \geq \epsilon\}} dQ(f)$$

$$\leq \epsilon + \int P(f(X) \geq M\epsilon) dQ(f).$$

Then $\lim_{M \rightarrow +\infty} P(f(X) \geq M\epsilon) = 0$ for every $f \in \mathcal{E}'$, by the Dominated Convergence Theorem we have that $\lim_{M \rightarrow +\infty} \int P(f(X) \geq M\epsilon) = 0$. \hfill \Box

Before proving P. 5. the following result must be stated.

Lemma 7. Let $Z$ be an absolutely continuous random variable with cumulative distribution function $F$. Let $(z_n)_{n \geq 1}$ be a real sequence such that $z_n \xrightarrow{n \rightarrow +\infty} z$ and $\beta \in (0, 1]$. Then,

$$LD^\beta_S(z_n, F) \xrightarrow{n \rightarrow +\infty} LD^\beta_S(z, F).$$

Proof. Since $F$ is continuous is enough to show that $\lambda^\beta_{z_n} \xrightarrow{n \rightarrow +\infty} \lambda^\beta_z$.

Let $t \in \mathbb{R}$, denote $F_z(t) = \frac{1}{2} F(t) + \frac{1}{2} (1 - F(2z - t))$ to the symmetrize version of $F$ about $z$. Recall that,

$$\lambda^\beta_z : F(z + \lambda^\beta_z) - F(z - \lambda^\beta_z) = \beta$$

$$\lambda^\beta_{z_n} : F(z_n + \lambda^\beta_{z_n}) - F(z_n - \lambda^\beta_{z_n}) = \beta \quad \text{for each } n \in \mathbb{N}$$

Then,

$$F_z(z + \lambda^\beta_z) = \frac{1}{2} F(z + \lambda^\beta_z) + \frac{1}{2} (1 - F(2z - (z + \lambda^\beta_z))) = \frac{1}{2} F(z + \lambda^\beta_z) + \frac{1}{2} (1 - F(z - \lambda^\beta_z))$$

$$= \frac{1}{2} [F(z + \lambda^\beta_z) - F(z - \lambda^\beta_z)] + \frac{1}{2} = \frac{\beta}{2} + \frac{1}{2}.$$
Meaning that \( z + \lambda_z^\beta = F_z^{-1}\left(\frac{\beta}{2} + \frac{1}{2}\right) \), analogously for \( z_n + \lambda_{z_n}^\beta \). Given that \( F_{z_n}(t) \xrightarrow{n \to +\infty} F_z(t) \) is clear that
\[
z_n + \lambda_{z_n}^\beta = F_{z_n}^{-1}\left(\frac{\beta}{2} + \frac{1}{2}\right) \xrightarrow{n \to +\infty} F_z^{-1}\left(\frac{\beta}{2} + \frac{1}{2}\right) = z + \lambda_z^\beta.
\]

**Proof: P. 5.** (continuous as a function of \( x \)). Let \( (x_n)_{n \geq 1} \) be a sequence on \( \mathbb{E} \) such that \( x_n \xrightarrow{n \to +\infty} x \), specifically \( f(x_n) \xrightarrow{n \to +\infty} f(x) \) for every \( f \in \mathbb{E}' \). For each \( f \) fixed, Lemma 7 states that
\[
LD_S^\beta(f(x_n), P_f) \xrightarrow{n \to +\infty} LD_S^\beta(f(x), P_f).
\]

As a consequence of the Dominated Convergence Theorem is clear that,
\[
LD^\beta(x_n, P) \xrightarrow{n \to +\infty} LD^\beta(x, P).
\]

**Proof: P. 6.** (continuous as a functional of \( P \)). Let \( x \in \mathbb{E} \) and \( \beta \in (0, 1] \). Our aim is to prove that if \( P_n \) is a sequence of probability measures that converges to \( P \), then
\[
IDLD^\beta(x, P_n) \xrightarrow{n \to +\infty} IDLD^\beta(x, P).
\]

Let \( (X_n)_{n \geq 1} \subset \mathbb{E} \) be a sequence of continuous random elements on \( \mathbb{E} \) with associated probability measure \( P_n \), such that, \( X_n \xrightarrow{n \to +\infty} X \) in distribution. Van der Vaart and Wellner \[8\] show that the convergence holds in the dual space, thus, \( f(X_n) \xrightarrow{n \to +\infty} f(X) \) in distribution for every \( f \in \mathbb{E}' \).

Since \( F_{n,f} \) is the cumulative distribution function of \( f(X_n) \) which converges pointwise to \( F_f \), then is clear that
\[
LD_S^\beta(f(x), P_{n,f}) \xrightarrow{n \to +\infty} LD_S^\beta(f(x), P).
\]
The result that we want to show is an straightforward consequence of the Dominated Convergence Theorem.
9 Appendix B: Uniform Strong Consistency of the IDLD.

In order to establish the uniform strong convergence of the one dimensional simplicial local depth. The following Lemma must be proved in advanced.

First of all, it is important to note the following facts. Assuming that the conditions stated in Remark 4 hold. For the sake of simplicity denote, \( \lambda = \lambda^\beta \), \( p_+ = F(z + \lambda) \), \( p_- = F(z - \lambda) \) and \( p = F(z) \). Let \( p \in (0, 1) \), then,

(i) \( Q_{p,n} = X_{([n\beta]+1)} \), \( Q_{p+,n} = X_{([n\beta]+1)} \) and \( Q_{p-,n} = X_{([n\beta]+1)} \).

(ii) \( F_n(Q_{p+,n}) - F_n(Q_{p-,n}) = \frac{[np_+] + 1}{n} - \frac{[np_-] + 1}{n} = \frac{[np_+] - [np_-]}{n} \).

Moreover, \( \frac{[np_+] - [np_-]}{n} \leq \frac{[np_+] - [np_-]}{n} \leq \frac{[np_+] - [np_-]}{n} + 1 \).

(vi) \( F_n(Q_{p+,n}) - F_n(z + d^{(k)}(z)) \leq \frac{1}{n} \) and \( F_n(Q_{p-,n}) - F_n(z - d^{(k)}(z)) \leq \frac{1}{n} \).

(vii) \( \beta(k) \leq \beta \leq \beta(k) + 1 \).

\textbf{Proof: Lemma 1.} (i) It follows straight forward by definition.

(ii) Let \( p \in (0, 1) \),

\[ |F_n(Q_{p,n}) - F(x_p)| = |F_n(Q_{p,n}) - p| = \frac{[np] + 1}{n} - p = \frac{[np] - np + 1}{n} \leq \frac{1}{n} \]

(iii) Let \( p \in (0, 1) \),

\[ |F(Q_{p,n}) - F(x_p)| \leq |F(Q_{p,n}) - F_n(Q_{p,n})| + |F_n(Q_{p,n}) - F(x_p)| \]

\[ \leq \sup_{t \in \mathbb{R}} |F_n(t) - F(t)| + \frac{1}{n} = ||F_n - F||_\infty + \frac{1}{n} \]
**Proof: Lemma 2.** For the sake of simplicity denote $\lambda = \lambda_2^2$ and $d^k = d^{(k)}(z)$.

\[
\left| (F(z + \lambda) - F(z)) (F(z) - F(z - \lambda)) - (F_n(z + d^k) - F_n(z)) (F_n(z) - F_n(z - d^k)) \right| = \\
= \left| F(z + \lambda)F(z) - F(z + \lambda)F(z - \lambda) - F(z)^2 + F(z)F(z - \lambda) \right| - \\
- \left[ F_n(z + d^k)F_n(z) - F_n(z + d^k)F_n(z - d^k) - F_n(z)^2 + F_n(z)F_n(z - d^k) \right] = \\
= \left| F(z + \lambda)F(z) - F(z + \lambda)F(z - \lambda) - F(z)^2 + F(z)F(z - \lambda) - \\
- F_n(z + d^k)F_n(z) + F_n(z + d^k)F_n(z - d^k) + F_n(z)^2 - F_n(z)F_n(z - d^k) \right| = \\
\leq \left| F(z + \lambda)F(z) - F_n(z + d^k)F_n(z) \right| + \left| F_n(z + d^k)F_n(z - d^k) - F(z + \lambda)F(z - \lambda) \right| + \\
+ \left| F_n(z)^2 - F(z)^2 \right| + \left| F(z - \lambda)F(z) - F_n(z - d^k)F_n(z) \right| \\
\tag{16}
\]

We analyze each term of Equation (16),

(a)

\[
\left| F(z + \lambda)F(z) - F_n(z + d^k)F_n(z) \right| = \\
= \left| F(z + \lambda)F(z) - F(z)F_n(Q_{p+,n}) + F(z)F_n(Q_{p+,n}) - F_n(z + d^k)F_n(z) \right| \leq \\
\leq F(z)\left| F(z + \lambda) - F_n(Q_{p+,n}) \right| + \\
+ \left| F(z)F_n(Q_{p+,n}) - F(z)F_n(z + d^k) + F(z)F_n(z + d^k) - F_n(z + d^k)F_n(z) \right| \leq \\
\leq \frac{1}{n} + F(z)\left| F_n(Q_{p+,n}) - F_n(z + d^k) \right| + \left| F(z) - F_n(z) \right|F_n(z + d^k) \leq \\
\leq \frac{1}{n} + \frac{1}{n} + \|F - F_n\|_\infty = \frac{2}{n} + \|F - F_n\|_\infty.
\]
(b) 
\[
\left| F_n(z + d^k)F_n(z - d^k) - F(z + \lambda)F(z - \lambda) \right| = \\
= \left| F_n(z + d^k)F_n(z - d^k) - F(z + \lambda)F_n(z - d^k) + \\
+ F(z + \lambda)F_n(z - d^k) - F(z + \lambda)F(z - \lambda) \right| \leq \\
\leq F_n(z - d^k)\left| F_n(z + d^k) - F(z + \lambda) \right| + F(z + \lambda)\left| F_n(z - d^k) - F(z - \lambda) \right| \leq \\
\leq \left| F_n(z + d^k) - F(z + \lambda) \right| + \left| F_n(z - d^k) - F(z - \lambda) \right| \leq \\
\leq \left| F_n(z + d^k) - F_n(Q_{p+,n}) \right| + \left| F_n(Q_{p+,n}) - F(z + \lambda) \right| + \\
+ \left| F_n(Q_{p-,n}) - F_n(z - d^k) \right| + \left| F_n(Q_{p-,n}) - F(z - \lambda) \right| \leq \frac{1}{n} + \frac{1}{n} + \frac{1}{n} + \frac{1}{n} = \frac{4}{n}.
\]

(c) 
\[
\left| F_n(z)^2 - F(z)^2 \right| = \left| F_n(z) - F(z) \right| \left| F_n(z) + F(z) \right| \leq \\
\leq 2 \left| F_n(z) - F(z) \right| \leq 2\|F_n - F\|_{\infty}.
\]

(d) Analogue to item (a).

Finally, denote 
\[
H = (F(z + \lambda)F(z)) (F(z)F(z - \lambda))
\]

and 
\[
G = (F_n(z + d^k) - F_n(z)) (F_n(z) - F_n(z - \lambda)).
\]

Then, 
\[
\left| ELD_{S}^{\beta}(z, F_n) - LD_{S}^{\beta}(z, F) \right| = \left| \frac{2}{\beta(k)}G - \frac{2}{\beta^2}H \right| \leq \left| \frac{2}{\beta(k)}G - \frac{2}{\beta^2}G \right| + \left| \frac{2}{\beta^2}G - \frac{2}{\beta^2}H \right| \leq \\
\leq \left( \frac{2}{\beta(k)}^2 - \frac{2}{\beta^2} \right) |G| + \frac{2}{\beta^2} |G - H|.
\]

On one hand, since Proposition 2 holds it's clear that each term of \( G \) is smaller than or equal to \( \frac{\beta(k)^2}{2} \). Hence, 
\[
\left( \frac{2}{\beta(k)}^2 - \frac{2}{\beta^2} \right) |G| \leq \left( \frac{2}{\beta(k)^2} - \frac{2}{\beta^2} \right) \frac{\beta(k)^2}{4} = \frac{1}{2} \left( 1 - \left( \frac{\beta(k)}{\beta} \right)^2 \right). \quad (17)
\]

34
On the other hand, we already know that,
\[ |G - H| \leq \frac{8}{n} + 4\|F - F_n\|_\infty. \]  
(18)

From Inequalities (17) and (18) we prove the inequality stated in the statement. \(\square\)

**Proof: Theorem 1.** (a) Let \( f \in \mathbb{E}' \) and \( x \in \mathbb{E} \). Denote \( P_f \) to the probability measure associated to \( f(X) \) where \( X \) is a random element on \( \mathbb{E} \) with probability measure \( P \). Analogously, denote \( P_{n,f} \) to the empirical probability measure of \( P_f \) based on \( f(X_1), \ldots, f(X_n) \).

By Proposition 2 we have,
\[ |ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^\beta(f(x), P_f)| \leq \frac{1}{2} \left( 1 - \left( \frac{\beta(k)}{\beta} \right)^2 \right) + \frac{2}{\beta^2} \left( \frac{8}{n} + 4\|P_{n,f} - P_f\|_\infty \right). \]

Observe that,
\[ \frac{1}{2} \left( 1 - \left( \frac{\beta(k)}{\beta} \right)^2 \right) = \frac{1}{2} \frac{\beta^2 - \beta(k)^2}{\beta^2} = \frac{1}{2} \frac{(\beta - \beta(k))(\beta + \beta(k))}{\beta^2} \leq \frac{1}{2} \frac{2}{n} \frac{\beta^2}{\beta^2} = \frac{1}{n\beta^2}. \]

Thus,
\[ |ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^\beta(f(x), P_f)| \leq \frac{1}{\beta^2} \left( \frac{17}{n} + 8\|P_{n,f} - P_f\|_\infty \right). \]  
(19)

Since it does not depend on \( x \) the inequality hold for the supreme of the left hand side of Inequality (19).

\[ \sup_{x \in \mathbb{E}} |ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^\beta(f(x), P_f)| \leq \frac{1}{\beta^2} \left( \frac{17}{n} + 8\|P_{n,f} - P_f\|_\infty \right). \]  
(20)

Let \( \epsilon > 0 \), denote
\[ h(f(X), f(X_1), \ldots, f(X_n)) = \sup_{x \in \mathbb{E}} |ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^\beta(f(x), P_f)|. \]
Then,
\[
E \left( h(f(X), f(X_1), \ldots, f(X_n)) \right) = \int_{\Omega} h(f(X), f(X_1), \ldots, f(X_n)) dP_f = \]
\[
= \int_{\{\omega \in \Omega : \sup_{x \in \mathbb{E}} h(f(X), f(X_1), \ldots, f(X_n)) \leq \epsilon\}} h(f(X), f(X_1), \ldots, f(X_n)) dP_f + \]
\[
+ \int_{\{\omega \in \Omega : \sup_{x \in \mathbb{E}} h(f(X), f(X_1), \ldots, f(X_n)) > \epsilon\}} h(f(X), f(X_1), \ldots, f(X_n)) dP_f \leq \epsilon + \int_{\{\omega \in \Omega : \sup_{x \in \mathbb{E}} h(f(X), f(X_1), \ldots, f(X_n)) > \epsilon\}} h(f(X), f(X_1), \ldots, f(X_n)) dP_f.
\]

Given that,
\[
\sup_{x \in \mathbb{E}} \left| ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^{\beta}(f(x), P_f) \right| \leq 1 \quad \text{for every } n \in \mathbb{N} \text{ and } f \in \mathbb{E},
\]
and
\[
\left\{ \sup_{x \in \mathbb{E}} \left| ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^{\beta}(f(x), P_f) \right| > \epsilon \right\} \subset \left\{ \frac{1}{\beta^2} \left( \frac{8}{n} + 4\|P_{n,f} - P_f\|_{\infty} \right) > \epsilon \right\}.
\]

Then,
\[
E \left( \sup_{x \in \mathbb{E}} \left| ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^{\beta}(f(x), P_f) \right| \right) \leq \epsilon + P \left( \frac{1}{\beta^2} \left( \frac{8}{n} + 4\|P_{n,f} - P_f\|_{\infty} \right) > \epsilon \right) = \epsilon + P \left( \|P_{n,f} - P_f\|_{\infty} > \frac{1}{4} \left( \epsilon \beta^2 - \frac{8}{n} \right) \right).
\]

By Dvoretzky-Kiefer-Wolfowitz [4] inequality we have that
\[
P \left( \|P_{n,f} - P_f\|_{\infty} > \frac{1}{4} \left( \epsilon \beta^2 - \frac{8}{n} \right) \right) \leq 2 \exp \left\{ -\frac{n}{2} \left( \epsilon \beta^2 - \frac{8}{n} \right) \right\} \xrightarrow{n \to \infty} 0 \text{ for every } f \in \mathbb{E}'.
\]
\[
(21)
\]

The right hand side of Inequality (21) does not depend on \( f \in \mathbb{E}' \), hence there exists \( n_0 \) such that for every \( n > n_0 \), we have,
\[
E \left( \sup_{x \in \mathbb{E}} \left| ELD_S^{\beta(k)}(f(x), P_{f,n}) - LD_S^{\beta}(f(x), P_f) \right| \right) < 2\epsilon \quad \text{for every } f \in \mathbb{E}'.
\]
(b) It follows straightforward from part (a) of the theorem and the fact that it is the integral of a measurable, positive and bounded function.

\[
E \left[ \sup_{x \in E} \left| EIDLD^\beta(k)(x, P_n) - IDLD^\beta(x, P) \right| \right] \leq \\
\leq E \left[ \sup_{x \in E} \int \left| ELD^\beta_S(f(x), P_{f,n}) - LD^\beta_S(f(x), P_f) \right| dQ(f) \right] = \\
= E \left[ \int \sup_{x \in E} \left| ELD^\beta_S(f(x), P_{f,n}) - LD^\beta_S(f(x), P_f) \right| dQ(f) \right] = \\
= \int E \left[ \sup_{x \in E} \left| ELD^\beta_S(f(x), P_{f,n}) - LD^\beta_S(f(x), P_f) \right| \right] dQ(f) \leq \\
= \int 2\epsilon \ dQ(f) = 2\epsilon \text{ if } n > n_0.
\]

\[\square\]

**Proof: Theorem 2.** Note that,

\[
P \left( \sup_{x \in E} \left| EIDLD^\beta(x, P_n) - IDLD^\beta(x, P) \right| \xrightarrow{n \to +\infty} 0 \right) = \\
= P \left( \bigcap_{\epsilon > 0} \bigcup_{n \in \mathbb{N}} \bigcup_{l \geq n} \left\{ \sup_{x \in E} \left| EIDLD^\beta(k)(x, P_l) - IDLD^\beta(x, P) \right| < \epsilon \right\} \right) = \\
= 1 - P \left( \bigcup_{\epsilon > 0} \bigcap_{n \in \mathbb{N}} \bigcup_{l \geq n} \left\{ \sup_{x \in E} \left| EIDLD^\beta(k)(x, P_l) - IDLD^\beta(x, P) \right| > \epsilon \right\} \right).
\]

It is enough to show that

\[
P \left( \bigcup_{\epsilon > 0} \bigcap_{n \in \mathbb{N}} \bigcup_{l \geq n} \left\{ \sup_{x \in E} \left| EIDLD^\beta(k)(x, P_l) - IDLD^\beta(x, P) \right| > \epsilon \right\} \right) = 0.
\]

By Borell-Cantelli lemma it is enough to prove that if the probability of the sets

\[
A_n = \left\{ \sup_{x \in E} \left| EIDLD^\beta(k)(x, P_n) - IDLD^\beta(x, P) \right| > \epsilon \right\},
\]

are summable, then, for all \(\epsilon > 0\),

\[
P \left( \bigcap_{n \in \mathbb{N}} \bigcup_{l \geq n} \left\{ \sup_{x \in E} \left| EIDLD^\beta(k)(x, P_l) - IDLD^\beta(x, P) \right| > \epsilon \right\} \right) = 0 \text{ and the prove would be done.}
\]
Let $\epsilon > 0,$

$$\sup_{x \in \mathbb{E}} \left| EIDLD^{\beta(k)}(x, P_n) - IDLD^{\beta}(x, P) \right| = \sup_{x \in \mathbb{E}} \left| \int ELD^{\beta(k)}_S(f(x), P_{n,f}) - LD^{\beta}_S(f(x), P_f) dQ \right| \leq$$

$$\leq \sup_{x \in \mathbb{E}} \int \left| ELD^{\beta(k)}_S(f(x), P_{n,f}) - LD^{\beta}_S(f(x), P_f) \right| dQ =$$

$$= \int \sup_{x \in \mathbb{E}} \left| ELD^{\beta(k)}_S(f(x), P_{n,f}) - LD^{\beta}_S(f(x), P_f) \right| dQ \leq$$

$$\leq \int \frac{1}{\beta^2} \left( \frac{8}{n} + 4 \| P_{n,f} - P_f \|_\infty \right) dQ = \frac{1}{\beta^2} \frac{8}{n} + \frac{1}{\beta^2} 4 \int \| P_{n,f} - P_f \|_\infty dQ \leq$$

$$\leq \frac{8}{n\beta^2} + \frac{1}{2\beta^2} \sup_{f \in \mathbb{E}'} \| P_{n,f} - P_f \|_\infty.$$  

Given that $\sup_{f \in \mathbb{E}'} \| P_{n,f} - P_f \|_\infty < +\infty$ there exists $f_0 \in \mathbb{E}'$ such that

$$\sup_{f \in \mathbb{E}'} \| P_{n,f} - P_f \|_\infty \leq \| P_{n,f_0} - P_{f_0} \|_\infty + \beta^2 \epsilon,$$

then

$$\frac{8}{n\beta^2} + \frac{1}{2\beta^2} \sup_{f \in \mathbb{E}'} \| P_{n,f} - P_f \|_\infty \leq \frac{8}{n\beta^2} + \frac{1}{2\beta^2} \| P_{n,f_0} - P_{f_0} \|_\infty + \frac{\beta^2 \epsilon}{2\beta^2}.$$  

By Dvoretzky-Kiefer-Wolfowitz inequality,

$$P(A_n) \leq P\left( \frac{8}{n\beta^2} + \frac{1}{2\beta^2} \| P_{n,f_0} - P_{f_0} \|_\infty + \frac{\epsilon}{2} > \epsilon \right) = P\left( \| P_{n,f_0} - P_{f_0} \|_\infty > \epsilon \beta^2 - \frac{16}{n} \right) \leq$$

$$\leq 2 \exp\left\{ -2n \left( \epsilon \beta^2 - \frac{16}{n} \right)^2 \right\}.$$  

Which is bounded by Borell-Cantelli’s lemma,

$$\sum_{n \in \mathbb{N}} P(A_n) \leq 2 \sum_{n \in \mathbb{N}} \exp\left\{ -2n \left( \epsilon \beta^2 - \frac{16}{n} \right)^2 \right\} < +\infty. \quad (22)$$
10 Appendix C: Proof of strong consistency of the \(\alpha\) local depth inner region at locality level \(\beta\)

Proof: Theorem 3. (I) Let \(\epsilon > 0\), \(0 < \delta < \epsilon\), \(\alpha > 0\) and a sequence \(\alpha_n \to \alpha\). It is clear that since Remark 6 holds, then

\[ R^\alpha_{n-\delta} \subset R^\alpha_n \subset R^\alpha_{n+\delta}. \]

We want to prove that \(R^\alpha_{n-\delta} \subset R^{\alpha-\epsilon}\). Without loss of generality we assume that \(\alpha - \epsilon > 0\), otherwise the inclusion always holds.

The hypothesis entail that since \(\alpha_n \to \alpha\), there exists \(n_1 \in \mathbb{N}\) such that for \(n \geq n_1\), \(|\alpha_n - \alpha| < \frac{\epsilon}{2}\). Even more, from hypothesis \((b)\) it follows that there exists \(n_2 > n_1\) such that if \(n \geq n_2\) then,

\[
\sup_{x \in \mathbb{E}} |ELD^\beta(x, P) - LD^\beta(x, P)| \leq \frac{\epsilon - \delta}{2} \text{ a.s.}
\]

Let \(x \in R^\alpha_{n-\delta} \cap (R^{\alpha-\epsilon})^c\), if \(n \geq n_2\), we have that

\[
ELD^\beta(x, P) - LD^\beta(x, P) > \alpha_n - \delta - (\alpha - \epsilon) = \alpha_n - \alpha - \delta + \epsilon \geq -\frac{(\epsilon - \delta)}{2} + \epsilon - \delta = \frac{\epsilon - \delta}{2}.
\]

Which is a contradiction, hence the intersection is empty. Then,

\[
\text{si } x \in R^\alpha_{n-\delta} \Rightarrow x \in R^{\alpha-\epsilon} \Rightarrow R^\alpha_{n-\delta} \subset R^{\alpha-\epsilon}.
\]

The proof of \(R^{\alpha+\epsilon} \subset R^\alpha_{n+\delta}\), is analogue.

(II) We know that

\[
\{x \in \mathbb{E} : LD^\beta(x, P) > \alpha\} = \bigcup_{\epsilon \in \mathbb{Q}^+} R^{\alpha+\epsilon} \subset \bigcap_{\epsilon \in \mathbb{Q}^+} R^{\alpha-\epsilon} = \{x \in \mathbb{E} : LD^\beta(x, P) \geq \alpha\}
\]

We want to show that \(\bigcup_{\epsilon \in \mathbb{Q}^+} R^{\alpha+\epsilon} \subset \lim\inf R^\alpha_n = \bigcup_{n \geq 1} \bigcap_{k \geq n} R^\alpha_k\) a.s.

Let \(\epsilon \in \mathbb{Q}^+\) and \(x \in R^{\alpha+\epsilon}\). Part \((I)\) establishes that there exists \(n_0\) such that

\[
R^{\alpha+\epsilon} \subset R^\alpha_{k_n} \text{ for every } k \geq n_0 \text{ a.s. } \Rightarrow R^{\alpha+\epsilon} \subset \bigcap_{k \geq n_0} R^\alpha_k \text{ a.s.}
\]
Then, for every
\[ \forall \epsilon \in \mathbb{Q}^+, \quad R^{\alpha+\epsilon} \subset \bigcup_{n \leq 1} \bigcap_{k \geq n} R^{\alpha_k} \quad \text{a.s.} \Rightarrow \bigcup_{\epsilon \in \mathbb{Q}^+} R^{\alpha+\epsilon} \subset \bigcup_{n \geq 1} \bigcap_{k \geq n} R^{\alpha_k} \quad \text{a.s.} \]

It remains to prove that \( \limsup R_n^{\alpha_n} = \bigcap_{n \geq 1} \bigcup_{k \geq n} R^{\alpha_k} \subset \bigcap_{\epsilon \in \mathbb{Q}^+} R^{\alpha-\epsilon} \quad \text{a.s.} \)

From part (1) of the theorem, for \( \epsilon \in \mathbb{Q}^+ \) there exists \( n_0 \) such that if \( k \geq n_0 \), it follows that \( R^{\alpha_k} \subset R^{\alpha-\epsilon} \) for every \( k \geq n_0 \) a.s. Also, \( \bigcup_{k \geq n_0} R^{\alpha_k} \subset R^{\alpha-\epsilon} \quad \text{a.s.} \)

Thus, \( \bigcap_{n \geq 1} \bigcup_{k \geq n} R^{\alpha_k} \subset \bigcup_{k \geq n_0} R^{\alpha_k} \subset R^{\alpha-\epsilon} \quad \text{a.s.} \)

Then it holds that for every \( \epsilon \in \mathbb{Q}^+ \), \( \bigcap_{n \geq 1} \bigcup_{k \geq n} R^{\alpha_k} \subset R^{\alpha-\epsilon} \quad \text{a.s.} \)

Finally, since \( \liminf R_n^{\alpha_n} \subset \limsup R_n^{\alpha_n} \), it follows that
\[ \bigcup_{\epsilon \in \mathbb{Q}^+} R^{\alpha+\epsilon} \subset \liminf R_n^{\alpha_n} \subset \limsup R_n^{\alpha_n} \subset \bigcap_{\epsilon \in \mathbb{Q}^+} R^{\alpha-\epsilon} \quad \text{a.s.} \quad (24) \]

From (23) and (24),
\[ P \left( \limsup R_n^{\alpha_n} \neq \liminf R_n^{\alpha_n} \right) = P \left( \limsup R_n^{\alpha_n} - \liminf R_n^{\alpha_n} \right) \leq \]
\[ \leq P \left( R_\alpha - \{ x \in E : LD_\beta(x, P) > \alpha \} \right) = \]
\[ = P \left( \{ x \in E : LD(x, P) = \alpha \} = 0. \right. \]

Hence the limit exists, in an analogous way,
\[ P \left( R_\alpha \neq \lim_{n \to +\infty} R_n^{\alpha_n} \right) = P \left( R_\alpha - \lim_{n \to +\infty} R_n^{\alpha_n} \right) \leq \]
\[ \leq P \left( R_\alpha - \{ x \in E : LD_\beta(x, P) > \alpha \} \right) = 0. \]

\( \square \)
11 Appendix D: Numerical Studies for Functional Data

11.1 Simulations

This section is devoted to the empirical study of our clustering procedure when the data is functional. To the best of our knowledge, there are two local depth measures suitable for this case, the IDLD introduced in this paper (LDCI) and also the local half-region depth for functional data introduced by Agostinelli (2018), (LDCH).

When implementing the LDC clustering procedure with local depth IDLD the parameters $\alpha$ and $\beta$ take the same values as in the multivariate case, and have been chosen following the same criteria. The number of random projections following a Brownian motion distribution for each replicate remains fixed at $N = 50$.

We conduct a simulation study on four synthetic models that have been previously analyzed by Justel and Svarc (2017), when they introduced the clustering procedure for functional data DivClusFD. First of all, we analyze three different datasets that present warping, while the last one presents pointwise sampling errors.

Model A: Two clusters with $n/2$ functions generated by,

$$X_i(t) = (1 + \epsilon_{1i}) \sin (\epsilon_{3i} + \epsilon_{4i} t) + (1 + \epsilon_{2i}) \sin \left( \frac{(\epsilon_{3i} + \epsilon_{4i} t)^2}{2\pi} \right),$$

$t \in [0, 2\pi]$, for $i = 1, \ldots, n/2$,

$$X_i(t) = (1 + \epsilon_{1i}) \sin (\epsilon_{3i} + \epsilon_{4i} t) - (1 + \epsilon_{2i}) \sin \left( \frac{(\epsilon_{3i} + \epsilon_{4i} t)^2}{2\pi} \right),$$

$t \in [0, 2\pi]$, for $i = n/2 + 1, \ldots, n$.

Model B: Two clusters with $n/2$ functions generated as in the first group following (25) and in the second group as follows,

$$X_i(t) = (1 + \epsilon_{1i}) \sin \left( \epsilon_{3i} + \epsilon_{4i} \left( -\frac{1}{3} + \frac{3}{4} t \right) \right) - (1 + \epsilon_{2i}) \sin \left( \frac{(\epsilon_{3i} + \epsilon_{4i} \left( -\frac{1}{3} + \frac{3}{4} t \right))^2}{2\pi} \right),$$

$t \in [0, 2\pi]$, for $i = n/2 + 1, \ldots, n$.

Model C: Three clusters with $n/3$ functions generated in the first group following (25), in the second group following (26) and in the third group following (27).
Model D: Four groups with \( n/4 \) functions generated as follows

\[
X_{ij}(t) = f_j(t) + \epsilon_i(t),
\]

for \( t \in [0,1] \), \( i = 1, \ldots, n/4 \) and \( j = 1, \ldots, 4 \),

where

\[
f_1(t) = \min \left( \frac{2 - 5t}{2}, \left( \frac{2 - 5t^2}{2} \sin \left( \frac{5\pi t}{2} \right) \right) \right),
\]

\[
f_2(t) = -f_1(t), \quad f_3(t) = \cos(2\pi t) \quad \text{and} \quad f_4(t) = -f_4(t).
\]

The datasets are of size \( n = 90 \) for Models A, B and C, but of size \( n = 600 \) for Model D. All errors \( \epsilon_{i1}, \ldots, \epsilon_{i4} \) are independent and normally distributed with mean 0 and standard deviation 0.05. In Equation (28) the errors, \( \epsilon(t) \), are normally distributed with mean 0.4, standard deviation 0.9 and covariance structure given by,

\[
\rho(s,t) = 0.3 \exp \left( -\frac{(s-t)^2}{0.3} \right), \quad \text{for} \ s, t \in [0,1].
\]

In all these cases, except DivClusFD, the number of clusters is assumed to be known.

Recently, Yassouridis and Leisch (2017) reviewed several functional data clustering procedures, which are available in the R package \textit{funcy}. We challenged our procedure with those methods (references therein) and also with DivClusFD. Table 11.1 reports the mean CCR for each model and clustering procedure; 200 replicates have been run for each model.

For the case of LDCI, we report the mean CCR for all the values of parameters considered (\( \alpha \) and \( \beta \)) since in every case the CCR is higher than 99%, hence the variance of these results is very small. For DivClusFD we only report the mean CCR when the number of clusters is correctly estimated; in every case this happens in more than 75% of the replicates.

Models A and B are the easiest ones to classify. In fact, most of the clustering procedures achieve an almost perfect classification. Clustering the dataset of Model C is a more challenging task, as is clear from the results of waveclust. Over all, the clustering procedures that present the poorest performance are funclust and HDDC. The CCR of waveclust remarkably decreases as the difficulty of the clustering problem increases. Model D has a different pattern than the other models: for it, the only clustering procedures that present an outstanding performance are LDCI and DivClusFD.

Figure 4 shows the core observations for a dataset generated following Model A, with the same parameters as used in the simulation study. Even though the design is balanced, it can be seen that the number of observations is not the same for every class.
Table 4: Mean CCR for the different clustering procedures considered.

| Procedure       | Model A | Model B | Model C | Model D |
|-----------------|---------|---------|---------|---------|
| LDCI            | 99.98   | 99.94   | 99.07   | 99.39   |
| LDCH            | 93.11   | 100     | 66.21   | 41.51   |
| DivClusFD       | 99.67   | 99.96   | 99.45   | 99.45   |
| fitfclust       | 99.99   | 99.93   | 98.95   | 74.65   |
| distclust       | 99.79   | 100     | 99.82   | 74.61   |
| iterSubspace    | 99.80   | 100     | 98.89   | 73.92   |
| funclust        | 81.55   | 79.26   | 61.64   | 38.97   |
| funHDDC         | 87.88   | 95.99   | 72.02   | 51.15   |
| fscm            | 97.79   | 99.79   | 99.64   | 74.48   |
| waveclust       | 99.88   | 95.94   | 89.18   | 72.98   |

Figure 4: (a) A dataset simulated following Model A. (a) Core observations of the data exhibit in (a)

It is important to note that LDCI is more than 100 times faster than DivClusFD.

From Table 11.1 it is clear that LDCI and LDCH behave differently: LDCH does not always detect the center of the clusters. We generated a random sample following Model C, see Figure 5, where the clustering procedure fails. The palette in Figure 5 panels (a), respectively, (b), is descending in IDLD, respectively LDH, and it is clear that in every group there are curves with a high IDLD, but that there is one group where there are no observations with high a LDH value. Figure 5 presents the scatter plot of the sorted LDH vs. the corresponding value of the IDLD, showing practically no structure. In addition, the correlation between LDH and IDLD in this case is 0.27,
which is low. This explains the difference in the clustering performances.

11.2 Real data examples for functional data

In this section we analyze the performance of our clustering procedure on several well known, publicly available, real-world functional datasets. As a benchmark, we also use the clustering procedures of the *funcit* R package. Four real datasets are considered: Growth, Canadian Weather, ECG200, and Tecator. Since these examples arise mainly from classification problems, the clustering configuration is known, hence we can report the CCR as in the simulation studies. The tuning parameters for the different clustering procedures remain fixed at the same values as in the simulation study. The examples analyzed are considered challenging for clustering purposes.

The Berkeley Growth Study is one of the best-known long-term development investigations ever conducted. It was introduced by Tuddenham and Snyder (1954). The heights of 54 girls and 39 boys were measured between 1 and 18 years at 31 unequally spaced time points. More measurements were taken during the later years of childhood and adolescence, when growth was more rapid, and fewer during the early years, when growth was more stable. This dataset can be found in the R package *fda*.

The Canadian Weather dataset was introduced by Ramsay and Silverman (2005) and is available in the R package *fda*. The data contains daily temperatures over the course of a year, measured at 35 monitoring stations in Canada. The data is grouped by four different geographical regions.
Table 5: CCR for several real data examples.

| Method     | Growth | Canadian Weather | ECG200 | Tecator |
|------------|--------|------------------|--------|---------|
| LDCI       | 91.40  | 62.86            | 76.00  | 85.12   |
| fitfclust  | 60.22  | 60.00            | 65.00  | 66.98   |
| distclust  | 66.67  | 68.57            | 64.00  | 66.05   |
| iterSubspace | 56.99 | 64.28            | 64.00  | 73.95   |
| funclust   | 77.42  | 60.00            | 52.00  | 59.07   |
| funHDDC    | 90.32  | 71.42            | 70.00  | 50.23   |
| fscm       | 65.59  | 34.29            | 67.00  | 76.28   |
| waveclust  | 78.95  | 57.14            | 75.00  | 90.7    |

The ECG200 dataset consists of 200 electrocardiograms, which can be found on the UCR Time Series Classification and Clustering website (Chen et al. (2015)). The dataset consists of two groups: one with 133 and the other with 67 electrocardiograms, each one recorded at 96 equally spaced instants.

The Tecator dataset consists of 215 spectrometric curves of meat samples, along with their fat, water and protein contents obtained by analytic procedures. The curves are classified into two groups, one of them high in fat content (over 15) and the other one low in fat content (below 15). Our goal is to cluster the data into those two groups based on the spectrometric curves.

When analyzing the results presented in Table 11.2, it can be seen that there is no clustering procedure that outperforms the others in all cases, which is to be expected since, depending on the characteristics of the data, different methods will be more appropriate. However, we can see that LDCI has a very good performance in all the examples analyzed.

Even though it is not reported on Table 11.2 we also did the study considering DivFunFD. This procedure could not detect correctly the number of clusters for Canadian Weather, ECG200 and Tecator, conforming spureus clusters. But for the Growth data set it identified the two clusters, with CCR= 89.25%.
References

[1] Agostinelli, C. (2018). “Local half-region depth for functional data.” *Journal of Multivariate Analysis* 163, 67-79.

[2] Chen, Y., E. Keogh, B. Hu, N. Begum, A. Bagnall, A. Mueen, and G. Batista (2015). The UCR time series classification archive. [http://www.cs.ucr.edu/~eamonn/time_series_data/](http://www.cs.ucr.edu/~eamonn/time_series_data/)

[3] Justel A., and M. Svarc, “A divisive clustering method for functional data with special consideration of outliers.” *Advances in Data Analysis and Classification*, DOI: https://doi.org/10.1007/s11634-017-0290-1

[4] Massart, M. (1990). “The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality.” *Annals of Probability*, 18(3), 1269-1283.

[5] Ramsay, J., and B. W. Silverman (2005). *Functional Data Analysis*. Springer-Verlag.

[6] Serban N., and L. Wasserman (2005). “CATS: Cluster Analysis by Transformation and Smoothing.” *Journal of the American Statistical Association* 100, 990-999. DOI 10.1198/016214504000001574

[7] Tuddenham R. D., and M. M. Snyder (1954). “Physical growth of California boys and girls from birth to eighteen years.” Tech. Rep. 1, University of California Publications in Child Development.

[8] Van der Vaart, A., and A. Wellner (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer-Verlag.

[9] Yassouridis, C., and F. Leisch (2017). “Benchmarking different clustering algorithms on functional data.” *Advances in Data Analysis and Classification* 11(3), 467-492. DOI 10.1007/s11634-016-0261-y