Depth-Based Recognition of Shape Outlying Functions[*]

Stanislav Nagy

KU Leuven, Belgium, and Charles University, Czech Rep.,

Irène Gijbels

KU Leuven, Belgium,

and

Daniel Hlubinka

Charles University, Czech Rep.

May 19, 2017

Abstract

A major drawback of many established depth functionals is their ineffectiveness in identifying functions outlying merely in shape. Herein, a simple modification of functional depth is proposed to provide a remedy for this difficulty. The modification is versatile, widely applicable, and introduced without imposing any assumptions on the data, such as differentiability. It is shown that many favourable attributes of the original depths for functions, including consistency properties, remain preserved for the modified depths. The powerfulness of the new approach is demonstrated on a number of examples for which the known depths fail to identify the outlying functions.

Keywords: consistency, data depth, functional data, infimal depth, integrated depth, outlying functions, shape outliers.

1

1 Introduction

In the multivariate data setup, depth functions were proposed as a tool providing the identification of typical and outlying observations with respect to a given probability distribution. Such recognition of sample points, based on their centrality with respect to a distribution (high depth for centrality), allows for natural generalization of ranks, order statistics, central regions, or quantiles to the multivariate setting (see Zuo and Serfling, 2000 for a review, or Liu et al., 1999 for an outline of possible applications).

In the past years, a great deal of attention has focused also on depths designed specifically for high-dimensional, functional, and Banach space-valued data. In this paper, we focus on functional data, that is random variables taking values in $\mathcal{C}\left([0,1]\right)$ — the space of continuous real-valued functions over the compact interval $[0,1]$ (Ramsay and Silverman, 2002). The vast majority of depths in high-dimensional settings are designed specially for functions.

The first depth for functions (depth functional) that appeared in the literature was the integrated depth proposed by Fraiman and Muniz (2001). The idea is for a function $x \in \mathcal{C}\left([0,1]\right)$ and P in $\mathcal{P}\left(\mathcal{C}\left([0,1]\right)\right)$ (the space of all probability measures on $\mathcal{C}\left([0,1]\right)$) to compute the univariate depths D of all the functional values $x(t)$ with respect to $P_{X(t)}$, the corresponding marginal distributions of $X \sim P$, for all $t \in [0,1]$. Eventually the "average" of these univariate depth values $D\left(x(t); P_{X(t)}\right)$ — the integral of them — is taken in

$$FD\left(x; P\right) = \int_0^1 D\left(x(t); P_{X(t)}\right) \mathrm{d}\,t \tag{1}$$

to assess the depth of the function x. Similarly, Mosler (2013) recently, in the definition of Φ-depth, proposed taking an infimum of all the depths of the projected quantities as in

$$ID\left(x; P\right) = \inf_{t \in [0,1]} D\left(x(t); P_{X(t)}\right) \tag{2}$$

instead of an integral. The latter depth will be called the infimal depth for functions.

Many depth functionals available in the literature follow one of the two basic ideas behind (1) and (2). For a comprehensive review see Section S.1 in the Supplementary Material.

The present work is motivated by the performance of the existing depth functionals when evaluated for their capability of identifying outlying curves. Unlike other authors considering functional outlier detection (cf Febrero et al., 2007) we focus on a more complex notion of outlyingness represented by shape outlying functions (cf Hubert et al., 2015). We show that the majority of the established depth functionals is incapable of identifying non-trivial outliers (1st order outliers in the terminology of Section 2). It is pointed out that this drawback is caused by inadequate reduction of the functionality inherently present in the data, when computing data depths.

As shown, a partial remedy for this problem might be the consideration of derivatives of the functional data, along with the proper usage of functional values. Namely, if x and the random function X from P are both (almost surely) differentiable, their derivatives x' and X', respectively, can be plugged into (1) and (2), obtaining for D a two dimensional depth function

$$FD^{(2)}\left(x;P\right) = \int_0^1 D\left(\left(x(t), x'(t)\right)^{\mathsf{T}}; P_{\left(X(t),X'(t)\right)^{\mathsf{T}}}\right)\,\mathrm{d}\,t, \qquad (3)$$

$$ID^{(2)}\left(x;P\right) = \inf_{t\in[0,1]} D\left(\left(x(t), x'(t)\right)^{\mathsf{T}}; P_{\left(X(t),X'(t)\right)^{\mathsf{T}}}\right). \qquad (4)$$

However, these depth-based procedures are constructed ad hoc, and depend crucially not only on the differentiability of the data curves, but also on the method chosen to estimate the derivatives. The latter shortcoming plays an enormous role especially when affiliating also higher order derivatives.

To overcome these difficulties, in Section 3 we propose a general modification technique

3

applicable to both integrated and infimal depths. It allows to define natural extensions of many known depth functionals, having the ability to capture shape properties of functions.

Overall, as will be seen, the practical advantages of having the extended versions of depth functionals at hand is that (i) they maintain the capability of recognizing shape properties of functional data of depths including derivatives; (ii) can be defined without the need to actually estimate the derivatives; (iii) are well defined also for non-differentiable and discontinuous observed functions; (iv) the simplicity and tractability of the original depths remain preserved; (v) enable the use of simple depth-based diagnostic plots; and finally, (vi) for the proposed integrated depth, a variety of theoretical results, available for FD, hold immediately.

In Section 2, the definition of a functional outlier is given. Two simulated data examples are presented, each of them presenting a random sample of curves contaminated by a different type of an outlier. On a small simulation study it is shown that the majority of depth functionals is indeed unable to identify shape outliers.

The extended versions of integrated and infimal depths are introduced in Section 3. Their performances are evaluated for the simulated examples from the previous section. A theoretical result establishing the relationship of these depths with depths including derivatives, and an informal test for shape outlyingness, are given. A real data application illustrating the use of the developed depth is presented in Section 4. Here, we analyse the historical population-by-country data, and identify countries with atypical population growth patterns, impossible to be flagged by any of the competing approaches. Section 5 provides a number of theoretical properties of the extended integrated depth, including consistency results for both continuous and discrete sampling setups. In Section 6 some concluding remarks are given. The paper is accompanied with a Supplementary Material file containing the introduction of the new depths in their most general forms, and some

4

additional technical details and results. An `R` package (R Core Team, 2014) `depth.fd` containing implementations of the proposed methodology is provided.

For the sake of clarity, in the sequel, when employing a finite-dimensional depth D, only the halfspace depth (cf Zuo and Serfling, 2000) $D\left(u;Q\right) = \inf_{H\in\mathcal{H},u\in H} Q\left(H\right)$, where $u \in \mathbb{R}^d$, $Q \in \mathcal{P}\left(\mathbb{R}^d\right)$, and $\mathcal{H}$ is the collection of all closed halfspaces in $\mathbb{R}^d$, is considered. A similar treatment for a general depth D is, however, easily obtained by obvious modifications.

Finally, let us note that despite sharing the topics of shape outlyingness and depth with Hubert et al. (2015), the two contributions are in fact of very different nature. While Hubert et al. (2015) consider a fixed, integrated type depth functional, and seek for tools providing optimal outlier detection based on this depth, we propose a new class of depths having the inherent ability to capture the shape properties of functions.

2 Recognition of Shape Outlying Functions

In this section, we motivate our study by providing two examples in which the established depth functionals fail to recognize an obvious outlier. As such this gives us an insight into the strengths and weaknesses of these depth functionals. For an extended, comprehensive simulation study including four other examples, and many other known depth functionals see Section S.2 in the Supplementary Material. There, also some additional technical details of the study can be found.

First of all, we formalize the notion of outlyingness for curves. Assume that there is a particular definition of a multivariate (finite-dimensional) outlier that is agreed upon. For instance, think of an outlier defined as a multivariate observation having a particularly low depth value with respect to the rest of the random sample, as will be considered here. Then the following, recursively given classification of functional outliers can be made.

5

Definition. Let $P \in \mathcal{P}(\mathcal{C}([0,1]))$, $X \sim P$ and $x \in \mathcal{C}([0,1])$. If there exists $t \in [0,1]$ such that $x(t) \in \mathbb{R}$ is outlying with respect to the distribution $P_{X(t)} \in \mathcal{P}(\mathbb{R})$, then we say that x is a 1st order outlier with respect to P.

For $J = 2, 3, \ldots$, assume that the collections of jth order outliers with respect to P are given, $j = 1, \ldots, J-1$. If there exists a set of points $(t_1, \ldots, t_J)^\mathsf{T} \in [0,1]^J$ such that $(x(t_1), \ldots, x(t_J))^\mathsf{T} \in \mathbb{R}^J$ is outlying with respect to the distribution $P_{(X(t_1), \ldots, X(t_J))^\mathsf{T}} \in \mathcal{P}(\mathbb{R}^J)$, and at the same time x is not a jth order outlier with respect to P for any $j = 1, \ldots, J-1$, then we say that x is a Jth order outlier with respect to P.

According to the previous definition, 1st order outliers are the functions usually conceived as functional outliers in the statistical literature. On the contrary, higher order outliers are those to which we refer here as to shape outliers. Since the joint distribution $P_{(X(t_1), X(t_2))^\mathsf{T}}$ relates to the difference of functional values $X(t_1)$ and $X(t_2)$, 2nd order outliers are the curves violating the pattern of functions from P in terms of growth. Likewise, 3rd order outliers are the outliers in terms of convexity/concavity, etc.

We now give two motivating examples. In both of them, in each of 100 independent runs, a functional random sample from $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ of size 200 is generated, and a single 2nd order outlier Y is evaluated for its depth-based centrality rank with respect to the sample. In the comparison, depths (1)–(4) are considered.

Example 1. Take $X(t) = A + B \arctan(t) + G(t)$, $t \in [0,1]$, for A having a centred normal distribution with variance 4, and B exponentially distributed with mean 1, A and B independent of each other. The random function $G \colon [0,1] \to \mathbb{R}$ is an independent centred Gaussian process, with $\mathrm{Cov}(G(s), G(t)) = 0.2 \exp(-|s-t|/0.3)$ for $s, t \in [0,1]$.

The outlying function Y is generated randomly as $Y(t) = 1 - 2\arctan(t) + \widetilde{G}(t)$ for $t \in [0,1]$, where $\widetilde{G}$ is a process distributed as G, but independent of it. The random sample functions exhibit a tendency to increase, while Y tends to decrease over $[0,1]$, see Fig. 1a.

6

Function Y is thus outlying in shape, and one would expect it to have a low functional depth value.

In principle it is impossible to incorporate derivatives into the computation of depth, as the functions are non-differentiable because of G and $\widetilde{G}$. Nevertheless, these "first derivatives" may be "estimated" formally as the derivatives of smooth functions approximating the observations.

In the first part of Tab. 1 the results, represented by the depth rank of Y with respect to the random sample, can be found. Depths considering only functional values perform very badly, and barely any improvement can be observed if the derivatives are approximated. Moreover, the substantial amount of noise in the data, along with the unsuitable usage of derivatives, raise the variance of the results considerably.

Another interesting finding is the great amount of functions identified as attaining simultaneously the minimal depth value for $ID^{(2)}$. This malady is common to infimal depths, especially in connection with multi-dimensional D in (4) (and (8)) and noisy data.

A visual tool explaining the bad performance of depths based on one-dimensional projections $x \mapsto x(t)$ (depths FD and ID) is the coordinate-wise D-depth graph. It is the plot of $t \in [0, 1]$ against the depth D of the coordinate projections associated with t

$$[0, 1] \to [0, 1] : t \mapsto D\left(x(t); P_{X(t)}\right). \tag{5}$$

For the current example, such a plot is displayed in Fig. 1b. There, the solid red line represents the trajectory of the univariate depth D of $Y(t)$ with respect to the marginal distribution of P_n, the empirical measure of the random sample, for $t \in [0, 1]$. The dotted lines represent the curves (5) corresponding to 10 randomly chosen functions of the random sample. The integrated depth FD is then visible as the integral of the corresponding curve (the area beneath the graph), and the infimal depth ID as the minimal value of

7

the function. It is evident that even though the function Y is outlying, the amount of information extracted from the plot (5), which is "sufficient" for the computation of both FD and ID, does not contain any hint of why Y should be identified as an outlier. Clearly, depths based solely on coordinate projections cannot capture shape outliers properly.

Similar diagnostic plots are readily available for depths including derivatives:

$$[0,1] \to [0,1]\colon t \mapsto D\left((x(t), x'(t))^{\mathsf{T}}; P_{(X(t), X'(t))^{\mathsf{T}}}\right), \tag{6}$$

where D is two-dimensional. One such plot will be presented for Example 2, see Fig. 2b.
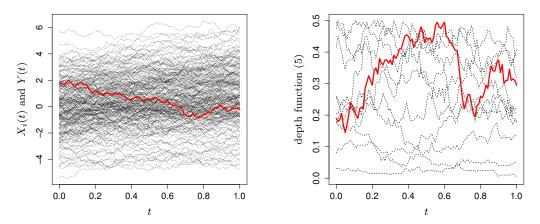

Figure 1: Example 1: (a) Random sample from P and the outlier Y (red solid); and (b) the diagnostic plot (5) of some randomly chosen sample curves and Y (red solid).

Example 2. Now, $X \sim P$ is taken to be smooth, having the form $X(t) = A + B \arctan(t)$ for $t \in [0,1]$, for A and B as in Example 1. The outlier Y is now discontinuous

$$Y(t) = \begin{cases} 0.5 + \log(2) \arctan(t) & \text{for } t \in [0, 0.5], \\ -0.5 + \log(2) \arctan(t) & \text{for } t \in (0.5, 1]. \end{cases}$$

8

Table 1: Results from Examples 1 and 2, and the computational cost of the considered depths. For each example, **Rank** indicates the average depth-based rank of Y with respect to the random sample (1–201, 201 for the lowest depth value), **Ties** stands for the average number of functions attaining the same, lowest depth value (1–201, 1 meaning a single function attains the lowest depth value). Estimated standard deviations of the quantities are in brackets.

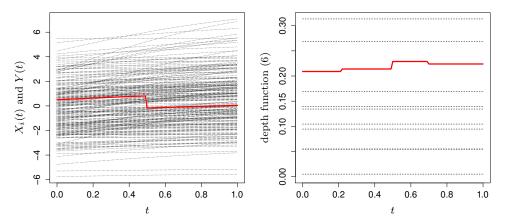|		FD	ID	$FD^{(2)}$	$ID^{(2)}$	FD_2	ID_2
Example 1	**Rank**	55.5 (16.1)	73.1 (25.2)	67.0 (28.6)	155.0 (70.0)	122.7 (28.0)	201.0 (0.0)
	Ties	1.1 (0.4)	6.3 (4.4)	1.2 (0.5)	126.4 (7.7)	1.2 (0.6)	199.1 (1.4)
Example 2	**Rank**	41.7 (6.8)	39.7 (8.7)	42.8 (10.4)	50.0 (10.9)	62.0 (8.5)	201.0 (0.0)
	Ties	1.6 (1.1)	3.6 (2.7)	10.8 (4.4)	10.8 (4.4)	10.8 (4.4)	11.8 (4.4)
Comput. Time (in s)		0.03	0.03	1.52	1.52	18.72	18.72

Figure 2: Example 2: (a) Random sample from P and the outlier Y (red solid); and (b) the diagnostic plot (6) of some randomly chosen sample curves and Y (red solid).

The reason for considering a discontinuous outlier (see Fig. 2a) is to demonstrate that even if all the functions are (almost everywhere) differentiable and their derivatives are incorporated into the depth computation, the established depth functionals may still fail to identify the outlier. This is indeed seen in Tab. 1. All the established functional depths (FD, ID, $FD^{(2)}$ and $ID^{(2)}$) assign to Y rather high values of depth, claiming it to be central. Moreover, the incorporation of derivatives does not make the results better, as the function Y is indistinguishable from sample curves, if looked upon locally. This is confirmed by Fig. 2b, where the two-dimensional marginal depth plot (6) is shown.

3   Shape Recognizing Depth Functionals

Now we propose a simple extension of depths FD and ID capable of recognizing shape outliers. The new depths are defined without the requirement of differentiability of the data, though recover the adequate results of depths considering derivatives. For a more general exposition, see Section S.1 in the Supplementary Material.

As we saw in Figs. 1b and 2b in Section 2, the integrated depths FD and $FD^{(2)}$, and infimal depths ID and $ID^{(2)}$, may face problems when dealing with shape outlying curves. The common feature causing this is their over-reliance on the coordinate projections $x \mapsto x(t)$, or $x \mapsto (x(t), x'(t))^{\mathsf{T}}$. A remedy may be found if the covariance structure of the multi-dimensional marginals of $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ gets involved into the computation of the depth. If Y violates the shape of X, then there must exist a finite set of time points $\{t_1, \ldots, t_J\} \subset [0,1]$ such that the vector $(Y(t_1), \ldots, Y(t_J))^{\mathsf{T}}$ does not fit into the pattern of the random vector $(X(t_1), \ldots, X(t_J))^{\mathsf{T}} \in \mathbb{R}^J$. More concretely, consider the setup of Example 1. Here, no one-dimensional projection $x \mapsto x(t)$ recognizes Y as an outlier, but virtually any two-dimensional projection $x \mapsto (x(t_1), x(t_2))^{\mathsf{T}}$ for $t_1 \neq t_2$ does so. Likewise,

three-dimensional projections may be used to emulate second derivatives, and so on.

Motivated by these observations, the following depths are defined.

Definition. For $J = 1, 2, \ldots$, the Jth order integrated depth of $x \in \mathcal{C}([0,1])$ with respect to $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ is defined as

$$FD_J(x; P) = \int_0^1 \cdots \int_0^1 D\left((x(t_1), \ldots, x(t_J))^\mathsf{T}; P_{(X(t_1), \ldots, X(t_J))^\mathsf{T}}\right) \, dt_J \ldots dt_1. \qquad (7)$$

Moreover, the Jth order infimal depth of x with respect to $X \sim P$ is defined as

$$ID_J(x; P) = \inf_{t_1, \ldots, t_J \in [0,1]} D\left((x(t_1), \ldots, x(t_J))^\mathsf{T}; P_{(X(t_1), \ldots, X(t_J))^\mathsf{T}}\right). \qquad (8)$$

In fact, in the definition above, the depth D needs to be evaluated only for points $(t_1, \ldots, t_J)^\mathsf{T} \in [0,1]^J$ such that $0 \leq t_J \leq \cdots \leq t_1 \leq 1$ due to the invariance of D with respect to coordinate permutations.

The definitions of sample versions of the new depths FD_J and ID_J depend on the sampling setup of the functional data. Two approaches towards this can be adopted, depending whether the sampled functions are assumed to be observed continuously in time, or discretely. This will be discussed in Section 5.2 in more detail, but in both cases a measure of an empirical type constructed from the random sample curves is plugged into (7), or (8), respectively, to replace P.

Considering multiple integrals and infima in the previous definition imposes in practice a substantial computational burden to the problem (see the last row of Tab. 1). To overcome this difficulty, one could take advantage of finite approximations of the quantities in (7) and (8). For $M \geq 1$, let $(T_{1,1}, \ldots, T_{1,J})^\mathsf{T}, \ldots, (T_{M,1}, \ldots, T_{M,J})^\mathsf{T} \in [0,1]^J$ be a random sample of size M from the uniform distribution on the cube $[0,1]^J$. Then the approximated versions of Jth order integrated and infimal depths can be computed as

$$FD_J^A(x; P) = \frac{1}{M} \sum_{m=1}^M D\left((x(T_{m,1}), \ldots, x(T_{m,J}))^\mathsf{T}; P_{\left(X(T_{m,1}), \ldots, X(T_{m,J})\right)^\mathsf{T}}\right), \qquad (9)$$

11

and

$$ID_J^A(x; P) = \inf_{m=1,\ldots,M} D\left((x(T_{m,1}), \ldots, x(T_{m,J}))^\mathsf{T}; P_{\left(X(T_{m,1}),\ldots,X(T_{m,J})\right)^\mathsf{T}}\right), \qquad (10)$$

respectively. Of course, these quantities are no longer deterministic and depend on the random sample of time points. In most non-pathological cases, however, the approximations of these depths provide reliable results, if the approximation size M is taken large enough. As a rule of thumb, for a single dataset one can approximate the depths as in (9) and (10) independently $m^* > 1$ times, using a different choice of the parameter $M = M_j$ for each set of replications. Then, a suitable M may be chosen as the smallest M_j such that the average (Spearman) correlation coefficient of the m^* vectors of depths (9) and (10) corresponding to M_j is high enough (say, 0.99). This way, M is guaranteed to have little effect on the resulting ordering of the curves. In our simulations, $M = 100$ proved to be enough to obtain sufficiently close approximations, in combination with excellent computational cost (see Section S.2 in the Supplementary Material).

Let us now return to Examples 1 and 2 from Section 2 and explore how the two depths FD_J and ID_J perform when employed in the outlier recognition task.

Considering $J = 2$ for integrated/infimal depths should result in an increased ability of recognizing functions violating the 2nd order (first derivative) shape properties, i.e. monotonicity outliers. It could be expected that, even though the derivatives need not be estimated, the extended depths FD_2 and ID_2 would mimic the behaviour of $FD^{(2)}$ and $ID^{(2)}$, respectively, and do a good job in spotting such outliers. This can indeed be seen from the last columns of Tab. 1, and will be confirmed theoretically in Section 3.1.

In Example 1, the wiggly nature of the functions slightly prevents the spotting of the overall monotonicity tendencies. FD_2 performs only moderately better than most of the depths not considering derivatives and still tends to rank Y too deeply into the random sample curves. However, the improvement in performance when compared to $FD^{(2)}$ is
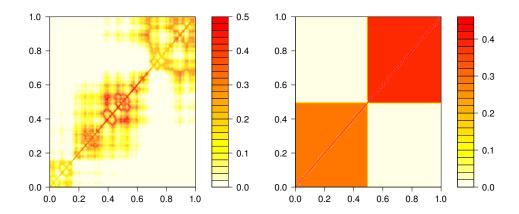
12

Figure 3: Diagnostic plots (11) for the outlier Y for (a) Example 1, and (b) Example 2.

substantial.

To gain better insight into how depths based on two-dimensional projections $x \mapsto (x(t_1), x(t_2))^{\mathsf{T}}$ actually work, a diagnostic tool similar to the coordinate-wise D-plots (5) and (6) is possible to be devised, if the contours of the function

$$[0,1]^2 \to [0,1]\colon (t_1, t_2)^{\mathsf{T}} \mapsto D\left((x(t_1), x(t_2))^{\mathsf{T}}; P_{(X(t_1), X(t_2))^{\mathsf{T}}}\right) \tag{11}$$

are plotted. Then, similarly to usual functional depths, the integral of this function over its domain determines FD_2, and its minimal value determines ID_2. This two-dimensional D-plot of the realization of random functions from Example 1, depicted in Fig. 1, can be found in Fig. 3a. Here it is evident why FD_2 provides rather faltering results for Y: while couples of points $(t_1, t_2)^{\mathsf{T}} \in [0,1]^2$ far from each other (near the NW and SE corners of the plot) contribute to the overall depth by very low values, for points close to each other (near the axis of the quadrant) the noise element dominates the contribution of slight monotonicity outlyingness, and at these points rather high values of depth D are attained.

13

ID_2, on the other hand, assigns to practically every function zero depth. This is, of course, caused by the local noise element of the functions, which is obviously too big for the more sensitive infimal type depths to handle the properties of functions accurately. Still, in such cases, one could also measure for each x the size of the set like

$$\left\{ (t_1, t_2)^{\mathsf{T}} : D\left((x(t_1), x(t_2))^{\mathsf{T}} ; P_{(X(t_1), X(t_2))^{\mathsf{T}}} \right) = \min_{i=1,\dots,n} ID_2\left(X_i; P \right) \right\} \subset [0,1]^2 \qquad (12)$$

to obtain ranking also for the functions attaining the lowest ID_2 depth. For details of this adjusted infimal-depth-based ranking of functional data we refer to the Supplementary Material document, Section S.2.

For Example 2, better results can be obtained. High D depth values are given to couples of points which both come from the same subinterval of continuity of Y ($[0, 0.5]$ or $(0.5, 1]$), and low values to couples for which each point is from a different interval. Thus, the integral in FD_2 smooths the effect of outlyingness in some coordinates, while the infimum in ID_2 provides perfect shape outlier identification. In this case, however, there is no local noise in the data, and only a small portion of functions is assigned to the smallest depth value. This makes ID_2 decidedly the best recognizing depth functional in the present case.

3.1 Relation to depths including derivatives

An interesting observation concerning the relation of FD_J and ID_J to depths considering derivatives ((3) and (4)) can be made. To show this, take $K = J - 1 = 1, 2, \dots$ and denote the derivative of $x \in \mathcal{C}\left([0,1]\right)$ at $t \in [0,1]$ of order $K > 1$ by $x^{(K)}(t)$, if it exists (taking it as one-sided at $t \in \{0, 1\}$). Then the following can be shown (the proof, along with some additional discussion on this result, is provided in the Supplementary Material).

Theorem 1. *Let $P \in \mathcal{P}\left(\mathcal{C}\left([0,1]\right)\right)$ be such that $X \sim P$ is almost surely K-times differentiable at $t \in [0,1]$. Then for any $x \in \mathcal{C}\left([0,1]\right)$ K-times differentiable at t, using the nota-*

14

tion $x^{(0...K)}(t) = \left(x(t), x'(t), \ldots, x^{(K)}(t) \right)^{\mathsf{T}}$ *and* $x_{h,K}(t) = (x(t), x(t+h), \ldots, x(t+Kh))^{\mathsf{T}}$,

it holds that

$$D\left(x^{(0...K)}(t); P^{(0...K)}(t) \right) \geq \limsup_{h \to 0} D\left(x_{h,K}(t); P_{h,K}(t) \right) \tag{13}$$

for $P^{(0...K)}(t) = P_{\left(X(t), X'(t), \ldots, X^{(K)}(t) \right)^{\mathsf{T}}}$ *and* $P_{h,K}(t) = P_{(X(t), X(t+h), \ldots, X(t+Kh))^{\mathsf{T}}}$. *If, moreover,*

$$\sup_{u \in \mathbb{R}^{K+1}, u \neq 0, a \in \mathbb{R}} \mathsf{P}\left(u^{\mathsf{T}} X^{(0...K)}(t) = a \right) = 0, \tag{14}$$

then it holds true that

$$D\left(x^{(0...K)}(t); P^{(0...K)}(t) \right) = \lim_{h \to 0} D\left(x_{h,K}(t); P_{h,K}(t) \right). \tag{15}$$

Condition (14) is satisfied for X if, for instance, the probability that X comes from any $(K+1)$-dimensional subspace of $\mathcal{C}\left([0,1]\right)$ is zero.

Theorem 1 reveals the actual motivation for taking J-dimensional vectors of functional values in the definition of the new depths. Indeed, this enables us to infer into the shape properties of functions, and in particular approximate well the behaviour of derivatives, if they exist.

Take first the two-dimensional plots of the coordinate-wise depth (11) displayed in Fig. 3. From the plot, it is possible to obtain information about the values of both (1st order) functional depths $FD\left(Y; P\right)$ and $ID\left(Y; P\right)$. To see this, consider the axis $(t,t)^{\mathsf{T}} \in [0,1]^2$ for $t \in [0,1]$. Here by the affine invariance of D (comparing the values of D defined in $\mathbb{R}^2$ on the left hand side and D defined in $\mathbb{R}$ on the right hand side)

$$D\left((Y(t), Y(t))^{\mathsf{T}}; P_{(X(t), X(t))^{\mathsf{T}}} \right) = D\left(Y(t); P_{X(t)} \right), \tag{16}$$

and the resulting depths (1) and (2) can be recovered as the path integral, and the infimum value, of depth (16) over the line $(t,t)^{\mathsf{T}}$, respectively.

15

By Theorem 1, even more can be said. Namely, both integrated $FD^{(2)}$ and infimal $ID^{(2)}$ depths including derivatives can be obtained by considering the function $t \mapsto \lim_{h \to 0} D\left((x(t), x(t+h))^\mathsf{T} ; P_{(X(t),X(t+h))^\mathsf{T}} \right)$, or the points in Fig. 3 of the form $(t, t+h)^\mathsf{T} \in [0,1]^2$ for h small.

Led by these observations, it is possible to define further modifications of the proposed functional depths by taking into account only J-tuples of points $(t_1, \ldots, t_J)^\mathsf{T} \in [0,1]^J$ close to the diagonal $(t, \ldots, t)^\mathsf{T} \in [0,1]^J$ for $t \in [0,1]$, instead of in the whole hypercube $[0,1]^J$ in the definitions. Likewise, proper weighting by a function $w \colon [0,1]^J \to [0,\infty)$ could be introduced into the integral (7) to form a depth

$$\int_0^1 \cdots \int_0^1 w\,(t_1, \ldots, t_J)\, D\left((x(t_1), \ldots, x(t_J))^\mathsf{T} ; P_{(X(t_1),\ldots,X(t_J))^\mathsf{T}} \right) \, \mathrm{d}\,t_J \ldots \mathrm{d}\,t_1.$$

Both these approaches lead to a depth mimicking integrated depth using derivatives, without the need to actually compute these, or even having differentiability at hand.

Returning to Theorem 1, there is no need in restricting ourselves to the study of depth of points $x^{(0 \ldots K)}(t)$ as in (13) and (15). Taking, for instance, $J = 4$, by similar arguments it is possible to show that for any $t_1, t_2, t_3 \in [0,1]$, under conditions such as those in Theorem 1, it holds true that (omitting the argument of the distribution in D)

$$D\left((x(t_1), x'(t_1), x^{(2)}(t_1), x(t_2))^\mathsf{T} \right) = \lim_{h \to 0} D\left((x(t_1), x(t_1+h), x(t_1+2h), x(t_2))^\mathsf{T} \right),$$

$$D\left((x(t_1), x'(t_1), x(t_2), x'(t_2))^\mathsf{T} \right) = \lim_{h \to 0} D\left((x(t_1), x(t_1+h), x(t_2), x(t_2+h))^\mathsf{T} \right),$$

$$D\left((x(t_1), x'(t_1), x(t_2), x(t_3))^\mathsf{T} \right) = \lim_{h \to 0} D\left((x(t_1), x(t_1+h), x(t_2), x(t_3))^\mathsf{T} \right).$$

Thus, the approach using the full domain $[0,1]^J$ encapsulates much more information about the functional random variable than the usual depths using derivatives. It is this great versatility, allowing the aggregation of information from various functional sources, that makes it beneficial to consider the full domains in FD_J and ID_J, instead of smaller ones, or weighting.

16

3.2   Choice of the order parameter J

A practical problem arising with the introduction of order extended functional depths is when it is preferable to use them instead of the original depths. The choice of the parameter J should be first of all problem-driven — the experimenter should, taking into account the nature of the collected data, determine if shape outliers are harmful to the analysis, and if the data are prone to contain measurements differing in shape. In most situations, such prior knowledge is available to the researcher and the order extended depths enable reliable identification of functions differing in monotonicity, growth acceleration (convexity), etc.

    Another good rule of thumb for the choice of J is given by the similarity of the new depths with the original depths incorporating derivatives. As observed in Section 3.1, the Jth order integrated and infimal depth usually emulate very well the behaviour of the integrated and infimal depth, respectively, when the derivatives of all orders up to J are incorporated. Thus, always when derivatives are to be used for depth computation, the order extended depths provide a more versatile, tractable, and reliable alternative.

    As from the practical point of view, just as the Kth derivatives of functions are rarely estimated reliably for $K > 2$, the usage of depths including derivatives is not encouraged for values of K larger than 1 or 2. On the other hand, order extended depths, independent of numerical estimation of derivatives, are computationally more stable than depths including derivatives, making the use of moderate values $J = 3$ or 4 (corresponding to $K = 2$ and 3, respectively) feasible. Thus, for a functional dataset possibly contaminated by hidden shape outliers, it might be beneficial to consider the extended depths to check for their presence. For instance, if the 1st order depth of a function is high, but the 2nd order depth falls drastically, there may be issues with the monotonicity of the function. Just like this, a drop in the value of the 3rd order depth indicates problems with convexity properties.

17

4 Application: Finding Population Trend Anomalies

For a real data example, consider the world population data (United Nations, 2016). A part of this dataset (`Total Population-Both Sexes`) consists of estimates of the total population in 233 countries, areas or regions as of July 1 in the years 1950–2010.

For the current analysis we restrict ourselves to the total of $n = 105$ countries (and Hong Kong SAR), with population on July 1, 1980 in the range of one and fifteen million people. For each of these countries, 61 data points corresponding to the estimates for each year constitute a single discretized realization of a random function X taking trajectories in $\mathcal{C}([0,1])$. The data curves are plotted in Fig. 4a.

Computing the Jth order integrated depth of the sample curves for $J = 1, 2, 3$ (the non-approximated version (7) for $J = 1$ and 2, and using the approximation (9) with[1] $M = 1000$ for $J = 3$), it is possible to identify the presence of potential outliers in the dataset (see also the scatter plot of FD_J in Fig. 5).

For $J = 1$, outliers can be identified as curves attaining small values of the depth $FD \equiv FD_1$ with respect to the remaining functions. This way, the 5 % of curves attaining the lowest FD_1 depth value are those visualised in Fig. 4b. These countries are not too interesting in terms of outlier detection. They tend to have the overall population either very low, or very high compared to the others, during the whole observed period, and constitute typical examples of what we called 1st order outliers in Section 2 — functions outlying in location (vertical shift), but not in shape.

A much more interesting situation occurs when examining higher order outliers. From the FD_J-depth scatter plot in Fig. 5, it can be seen that all the points in the plots lie below the axis of the first quadrant, i.e. $FD_J(x; P) \geq FD_{J+1}(x; P)$, for $J = 1$ and 2 here.

---

[1] $M = 1000$ is chosen to minimize the effect of randomness introduced by approximation. For practical purposes, much smaller values of M are sufficient.

This certainly must be satisfied for FD_J based on the halfspace depth D, as it is easy to verify that $D\left(u;Q\right) \geq D\left((u,v)^{\mathsf{T}};\boldsymbol{Q}\right)$ for all $u \in \mathbb{R}^J$, $v \in \mathbb{R}$, $Q \in \mathcal{P}\left(\mathbb{R}^J\right)$, $\boldsymbol{Q} \in \mathcal{P}\left(\mathbb{R}^{J+1}\right)$, such that Q is the appropriate marginal distribution of $\boldsymbol{Q}$. This observation leads us to the construction of the following informal test for shape outlyingness:
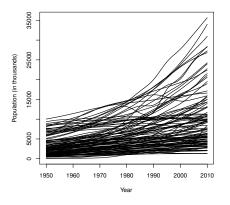
1. To identify Jth order outliers in the dataset for $J > 1$, suppose that the sets of jth order outliers for $j = 1, \ldots, J-1$ are already available.

2. Compute the transformed ratios of Jth and jth order integrated depths

$$
S_j(x_i) = \begin{cases} -\log\left(\frac{FD_J(x_i;P_{n,-i})}{FD_j(x_i;P_{n,-i})}\right) & \text{if } FD_J\left(x_i; P_{n,-i}\right) > 0, \\ \infty & \text{otherwise,} \end{cases}
$$

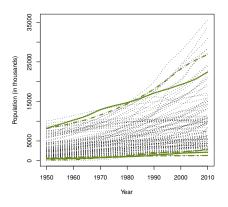   for $j = 1, \ldots, J-1$ and each sample function x_i. Here, $P_{n,-i}$ designates the empirical measure of all the sample functions, excluding x_i.

3. For $j = 1, \ldots, J-1$ denote by $\tilde{s}_j$ the value of the highest datum $S_j(x_i)$ still within 1.5 IQR of the upper quartile of $\{S_j(x_i) : i = 1, \ldots, n\}$.

4. Label the function x_i as the Jth order outlier, if it is not a jth order outlier for any $j = 1, \ldots, J-1$, and if there exists $j = 1, \ldots, J-1$ such that $S_j(x_i) > \tilde{s}_j$.
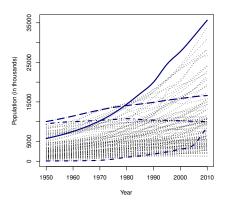
The previous test is indeed very simple; it emulates the one-sided version of the well-known Tukey's informal test for univariate outliers based on the boxplot (for a test for functional data based on a related idea, but using other depth functionals see Arribas-Gil and Romo, 2014). An observation x_i is flagged as a potential Jth order outlier, if it is not a lower-order outlier, and at the same time has a very low value of the Jth order depth compared to some of its lower-order functional depths.
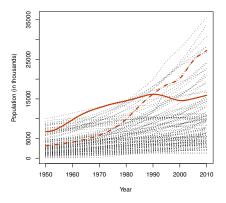
19

(a) Data curves — the total population in 105 countries between years 1950 and 2010.

(b) 1st order outliers: Kuwait, Oman, Namibia, Trinidad and Tobago, Nepal, and Australia.

(c) 2nd order outliers: United Arab Emirates, Sudan, Hungary, and the Netherlands.

(d) 3rd order outliers: Saudi Arabia, and Kazakhstan.

Figure 4: World population example. In Figs. (b)–(d) are the potentially outlying curves plotted in thick lines. The names of the countries below the figures are sorted in an increasing order, according to the total population in 1950.

20

Applying this procedure to the world population data example, four functions are identified as 2nd order outliers, see Fig. 4c. These are more interesting than the curves from Fig. 4b, as now the outliers violate the pattern of the sample in terms of growth (and vertical shift). United Arab Emirates provides, in most of the domain, an example of a rather flat curve, not unlike many other sample curves for less-populated countries. The outlyingness is accounted for by the consecutive rapid growth, especially since 2003, providing an extraordinarily conspicuous behaviour compared to the other curves. Hungary and the Netherlands are prime examples of countries with almost flat population curves (even a decreasing one for Hungary in parts of the domain). Finally, Sudan presents a curve with an abrupt, steady growth during the whole observed time period.

Two curves are flagged as 3rd order outliers: Saudi Arabia and Kazakhstan. For Saudi Arabia, this is because of the rapid growth since the 1970s, combined with the convex-concave course; for Kazakhstan it is its atypical concave-like structure from 1960 to 2000.

Analogously to the step-wise procedure described above, in the outlier recognition task we may simply restrict to the highest considered value of J (i.e. $J = 3$ here), and mention as potential outliers only the curves flagged for this particular choice of J (not excluding the outliers flagged for smaller values of J). The flagged curves then typically greatly overlap with the functions flagged using the step-wise procedure (here, all the step-wise outliers except for Oman, Namibia and Nepal are identified for $J = 3$). Nonetheless, this way we lose information about the degree of outlyingness of the curve, as it is no longer possible to say if the curve is atypical because of its location, growth, or convexity.

Similar outlier detection diagnostics can be performed also with other functional depths (see Section S.1 in the Supplementary Material). The depths not taking derivatives into account fail to identify shape outliers completely (2nd and 3rd order outliers). On the other hand, depths considering derivatives (namely $FD^{(2)}$) identify 2nd order outliers reasonably
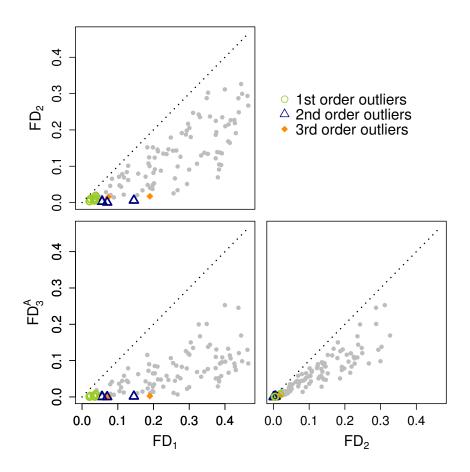
21

Figure 5: Scatter plot of Jth order integrated depths for $J = 1, 2, 3$ for the world population dataset. The observations flagged as outliers are labelled by different plot characters. The dotted line in each plot represents the axis of the first quadrant.

well, and these more or less coincide with FD_2 outliers. The observations flagged as 3rd order outliers (computed by the integrated depth incorporating second order derivatives) already differ from those obtained by FD_3 (or FD_3^A, in our case), and depend severely on the used estimators for second order derivatives. For instance, Kazakhstan was in our computations never flagged as an outlier when integrated depths depending on derivatives were used, reflecting the sensitivity of functional depths incorporating derivatives to the choice of the estimation of derivatives. This serious drawback is not a problem for order extended functional depths, as these do not depend on the estimated derivatives.

5 Theoretical Properties and Consistency

Now we provide theoretical properties of the extended depths. We focus on the integrated depth FD_J, and omit theoretical properties for the infimal depth ID_J, as there is a considerable gap in the literature concerning the theoretical results already for the base case of ID_1. These properties still need to be established.

Firstly, we deal with the measurability of the integrand function in definition (7) of FD_J. Then, the attention is focused on functional analogues of the desirable properties of depth functions. Finally, the uniform consistency of the proposed depth in both settings of continuously and discretely observed random sample functions is shown.

The results presented in this section can be proved by modification of the comprehensive theory for general (1st order) integrated depth functionals taken from Nagy et al. (2016b). The consistency theory for discretely observed functions can be obtained by direct application of the observations provided in Nagy et al. (2016a).

The first, essential result establishes the correctness of the definition of the new integrated depth (7). It is sufficient to guarantee the existence of the proposed depth value for

any measurable function and any probability distribution.

Theorem 2. *For any* $J = 1, 2, \ldots$, $x \colon [0,1] \to \mathbb{R}$ *Borel measurable and* $P \in \mathcal{P}\left(\mathcal{C}\left([0,1]\right)\right)$,

$$D \colon [0,1]^J \to [0,1] \colon (t_1, \ldots, t_J)^\mathsf{T} \mapsto D\left((x(t_1), \ldots, x(t_J))^\mathsf{T}; P_{(X(t_1), \ldots, X(t_J))^\mathsf{T}}\right)$$

is a universally measurable function. In particular, the quantity $FD_J\left(x; P\right)$ *is well defined and takes values in the interval* $[0,1]$.

5.1 Statistical properties

The following theorem summarizes the main statistical properties of the proposed depth.

Theorem 3. *Let* $X \sim P \equiv P_X \in \mathcal{P}\left(\mathcal{C}\left([0,1]\right)\right)$. *The following properties hold for* FD_J:

(i) *Translation invariance: For any* $b, x \in \mathcal{C}\left([0,1]\right)$ *and* X *we have* $FD_J\left(x; P_X\right) = FD_J\left(x + b; P_{X+b}\right)$.

(ii) *Function-scale invariance: For any* $a, x \colon [0,1] \to \mathbb{R}$ *Borel measurable such that* $a(t) \neq 0$ *for all* $t \in [0,1]$ *we have* $FD_J\left(x; P_X\right) = FD_J\left(a.x; P_{a.X}\right)$, *where* $(a.x)\left(t\right) = a(t)x(t)$.

(iii) *Measure-preserving rearrangement invariance: For any* $\rho \colon [0,1] \to [0,1]$ *bijective and measure-preserving and any* $x \in \mathcal{C}\left([0,1]\right)$, $FD_J\left(x; P_X\right) = FD_J\left(x \circ \rho; P_{X \circ \rho}\right)$.

(iv) *Maximality at the centre of symmetry: If for all* $(t_1, \ldots, t_J)^\mathsf{T} \in [0,1]^J$ *the marginal distribution* $P_{(X(t_1), \ldots, X(t_J))^\mathsf{T}}$ *is halfspace symmetric about* $(\theta\left(t_1\right), \ldots, \theta\left(t_J\right))^\mathsf{T} \in \mathbb{R}^J$ *for* $\theta \colon [0,1] \to \mathbb{R}$ *Borel measurable, then* $FD_J\left(\cdot, P\right)$ *is maximized over the space of Lebesgue measurable functions on* $[0,1]$ *at the function* θ.

(v) *Decreasing along rays: If a maximum depth* FD_J *is attained at* θ *from part (iv), then for every* $x \in \mathcal{C}\left([0,1]\right)$ *and* $\gamma \in [0,1]$ *we have* $FD_J\left(\gamma\theta + (1-\gamma)x; P\right) \geq FD_J\left(x; P\right)$.

(vi) Vanishing at infinity: For a sequence of functions $\{x_\nu\}_{\nu=1}^{\infty} \subset \mathcal{C}([0,1])$ such that $|x_\nu(t)| \xrightarrow[\nu\to\infty]{} \infty$ almost everywhere in $t \in [0,1]$, it holds that $FD_J(x_\nu; P) \xrightarrow[\nu\to\infty]{} 0$.

(vii) Upper semicontinuity of FD_J as a function of x: for all $x_\nu \xrightarrow[\nu\to\infty]{} x$ we have that $\limsup_{\nu\to\infty} FD_J(x_\nu; P) \le FD_J(x; P)$. Moreover, if it is true that the marginal distribution $P_{(X(t_1),\ldots,X(t_J))^\mathsf{T}}$ is absolutely continuous for all $(t_1,\ldots,t_J)^\mathsf{T} \in [0,1]^J$, then $\lim_{\nu\to\infty} FD_J(x_\nu; P) = FD_J(x; P)$.

(viii) Weak continuity of FD_J as a functional of P: If for all $(t_1,\ldots,t_J)^\mathsf{T} \in [0,1]^J$

$$P_{(X(t_1),\ldots,X(t_J))^\mathsf{T}}(H) = 0 \text{ for all hyperplanes } H \subset \mathbb{R}^J, \tag{17}$$

then $\sup_{x\in\mathcal{C}([0,1])} |FD_J(x; P) - FD_J(x; P_\nu)| \xrightarrow[\nu\to\infty]{} 0$ for any $P_\nu \xrightarrow[\nu\to\infty]{w} P$.

Notice that in Theorem 3 the results concerning the existence of the points in $\mathcal{C}([0,1])$ maximizing FD_J (FD_J-depth medians) are rather weak. This is caused by the fact that while for $J = 1$ the function in the sample space attaining the highest depth value is always the coordinate-wise median function, this is no longer true for $J > 1$.

On one hand this is somewhat displeasing, because little can be said about the behaviour of the FD_J-depth median at this point. On the other hand, at least from the point of view of general data depth, this is very promising. Indeed, roughly speaking, depth was designed as a tool capturing the global properties of the probability distribution rather than local features (Paindaveine and Van Bever, 2013). For finite-dimensional distributions the depth median depends heavily on the inner covariance structure of the data set, and in most nontrivial cases it differs from the coordinate-wise median (Rousseeuw and Ruts, 1999). Most of the proposed depths for functional data are, however, too simple to capture this covariance-driven distinction between depth median and coordinate-wise median (Nagy et al., 2016b, Section 4.2). By the introduction of repeated integration into

25

integrated depth we are able to overcome this oversimplistic, undesirable phenomenon common to most functional depths by finding a depth median which does not inherently take a form of a trivial coordinate-wise median function.

5.2 Consistency properties

As indicated in Section 3, two basic approaches towards establishing consistency results for depth functionals are available — for continuous and discrete sampling designs. The consistency of FD_J can be shown without any additional assumptions for both cases. For a detailed discussion on the differences between these two approaches see Nagy et al. (2016a).

5.2.1 Continuous sampling design

In the simplest case of continuous sampling design one assumes that the random sample functions are observed continuously in time $t \in [0, 1]$, and the empirical measure P_n is directly assumed to be known. In this case the sample version of FD_J is taken simply as the population version (7) with respect to P_n.

The proof of the following theorem can be obtained along the same lines as the proofs of Theorems 10 and 11 of Nagy et al. (2016b). For the proof, it is essential to assume that the finite-dimensional depth D used in the definition of order integrated depths is the halfspace depth, having the required consistency properties.

Theorem 4. *Let the random sample functions be observed completely. Then the following consistency properties hold true for FD_J:*

(i) Weak uniform consistency: For each $\varepsilon > 0$

$$\sup_{P \in \mathcal{P}(\mathcal{C}([0,1]))} \mathsf{P}\left(\sup_{x \in \mathcal{C}([0,1])} \left| FD_J\big(x; P\big) - FD_J\big(x; P_n\big) \right| > \varepsilon \right) \xrightarrow[n \to \infty]{} 0.$$

26

(ii) Strong universal consistency: For each $P \in \mathcal{P}\left(\mathcal{C}\left([0,1]\right)\right)$

$$\sup_{x \in \mathcal{C}([0,1])} \left|FD_J\left(x; P\right) - FD_J\left(x; P_n\right)\right| \xrightarrow[n \to \infty]{\text{a.s.}} 0.$$

5.2.2 Discrete sampling design

In the discrete design setting, it is assumed that each sample function X_i, $i = 1, \ldots, n$ is observed only at a finite set of time points $T_{1,i}, \ldots, T_{m_i,i} \in [0,1]$, $m_i = 1, 2, \ldots$ (in particular, each function is allowed to be observed at a different number of different time points). The sample curves are thus observable only partially, and must be reconstructed to form elements of $\mathcal{C}\left([0,1]\right)$ for the statistical inference. One such reconstruction method is the interpolation of the observed values by piecewise-linear functions (and setting them constant near the endpoints of the domain — for details, see Nagy et al., 2016a).

Denote now the empirical measure of these interpolating functions by $\widetilde{P}_n$, and write $T_{0,i} = 0$, $T_{m_i+1,i} = 1$. In Nagy et al. (2016a, Theorem 2) it was shown that if

$$\max_{j=0,\ldots,m_n} |T_{j+1,n} - T_{j,n}| \xrightarrow[n \to \infty]{\text{P}} 0 \tag{18}$$

is satisfied, then the sequence of measures $\widetilde{P}_n$ converges weakly in $\mathcal{P}\left(\mathcal{C}\left([0,1]\right)\right)$ to P almost surely. Under the same mild assumption it is possible to derive the following consistency result for FD_J (Nagy et al., 2016a, Theorem 5).

Theorem 5. *Let the random sample functions from $P \in \mathcal{P}\left(\mathcal{C}\left([0,1]\right)\right)$ be observed discretely so that (18) holds true. If for $P \in \mathcal{P}\left(\mathcal{C}\left([0,1]\right)\right)$ the condition (17) from part (viii) of Theorem 3 is satisfied, then*

$$\sup_{x \in \mathcal{C}([0,1])} \left|FD_J\left(x; P\right) - FD_J\left(x; \widetilde{P}_n\right)\right| \xrightarrow[n \to \infty]{\text{a.s.}} 0.$$

In Theorems 4 and 5 the consistency results for FD_J are provided only for completely, and discretely observed functional data under the condition that (18) is true. If (18) is valid, data observed within the latter setting are frequently called dense functional data. If the data are discretely observed but (18) cannot be guaranteed, one speaks about sparse functional data. Some methods in functional data analysis were already successfully extended also to sparse random functions (see, for instance, Hsing and Eubank, 2015). In most cases, however, treatment of sparse random curves requires strong distributional assumptions to be made. Thus, these extensions do not appear to be well suited for the analysis using data depth, intended to be nonparametric in nature. The behaviour of functional data depths for sparse random functions therefore seems to be an interesting, yet little studied problem.

6   Conclusions

By introducing order extensions to functional depths, we propose a natural alternative to depths including information about derivatives of functional data. The adapted depths match, if not surpass, the good practical performance of depths using derivatives when applicable, with an added value of better computational stability, and tractability. In contrast to the competing methods neither differentiability, nor continuity of the data curves must be assumed for the new depths. Most of the theoretical properties valid for the usual depths hold also for their order extensions.

For higher orders, the proposed depths may be computationally expensive. Here, we partially resolve this problem using approximations. This, however, comes at the cost of introducing variability into the calculation. Further research into the computational aspects of the new depths may be appropriate.

SUPPLEMENTARY MATERIAL

This paper is complemented by extensive Supplementary Material:

`Supplement.pdf:` A document, where we provide some additional details and technical results:

- in Section S.1 the newly proposed depths are introduced in full generality, within the unified framework of depth for functional data;

- in Section S.2 a comprehensive simulation study evaluating the performance of the new estimators can be found; and

- in Section S.3 the proof of Theorem 1, and some additional discussion on the relation of order extended depths with depths including derivatives, are given.

`depth.fd_1.0.tar.gz:` An R package[2] (`gzip` compressed `tar` archive), where a fully documented implementation of all the statistical methods discussed in the paper, and in its supplements, is provided. The package is accompanied by all its source files.

`Codes_depth.fd.R:` An R script which enables full replication of the analysis performed in the paper, and in the supplementary document.

`README.txt:` A detailed description of all the supplementary files, and a brief installation guide for the `depth.fd` package.

All of the supplemental files are contained in a single `zip` archive.

---

[2]available at `https://cran.r-project.org/`, see also `https://bitbucket.org/StanislavNagy/depth.fd_public`

References

Arribas-Gil, A. and Romo, J. (2014). Shape outlier detection and visualization for functional data: the outliergram. *Biostatistics*, 15(4):613–619.

Febrero, M., Galeano, P., and González-Manteiga, W. (2007). A functional analysis of NOx levels: location and scale estimation and outlier detection. *Comput. Statist.*, 22(3):411–427.

Fraiman, R. and Muniz, G. (2001). Trimmed means for functional data. *Test*, 10(2):419–440.

Hsing, T. and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators.* Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester.

Hubert, M., Rousseeuw, P. J., and Segaert, P. (2015). Multivariate functional outlier detection. *Stat. Methods Appl.*, 24(2):177–202.

Liu, R. Y., Parelius, J. M., and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference. *Ann. Statist.*, 27(3):783–858.

Mosler, K. (2013). Depth statistics. In *Robustness and complex data structures*, pages 17–34. Springer, Heidelberg.

Nagy, S., Gijbels, I., and Hlubinka, D. (2016a). Weak convergence of discretely observed functional data with applications. *Journal of Multivariate Analysis*, 146:46 – 62. *Special Issue on Statistical Models and Methods for High or Infinite Dimensional Spaces.*

Nagy, S., Gijbels, I., Omelka, M., and Hlubinka, D. (2016b). Integrated depth for functional data: statistical properties and consistency. *ESAIM Probab. Stat.*, 20:95–130.

Paindaveine, D. and Van Bever, G. (2013). From Depth to Local Depth: A Focus on Centrality. *J. Amer. Statist. Assoc.*, 108(503):1105–1119.

R Core Team (2014). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied functional data analysis: Methods and case studies.* Springer Series in Statistics. Springer-Verlag, New York.

Rousseeuw, P. J. and Ruts, I. (1999). The depth function of a population distribution. *Metrika*, 49(3):213–244.

United Nations, Department of Economic and Social Affairs, Population Division (2016). World population prospects: The 2015 revision. `https://esa.un.org/unpd/wpp/Download/Standard/Population/`. Accessed: 2016-07-16.

Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Ann. Statist.*, 28(2):461–482.