SUPPLEMENTARY MATERIAL: DEPTH-BASED RECOGNITION OF SHAPE OUTLYING FUNCTIONS

STANISLAV NAGY[1,2], IRÈNE GIJBELS[1], AND DANIEL HLUBINKA[2]

ABSTRACT. This document complements the main paper by containing additional technical details and results. It is organized into three sections:

- in Section S.1 a review of depth functionals available in the literature is given. The emphasis of the exposition is put on the unification of these diverse concepts from the theoretical point of view. It is shown that most of the established depths fall into the general framework of projection-driven depths of either integrated, or infimal type. Within this framework, the order extended depths are introduced in full generality.
- in Section S.2 a full-scale simulation study evaluating the performance of all the depth functionals mentioned in Section S.1 can be found, along with all its technical details.
- in Section S.3 the proof of Theorem 1 from the main body of the paper, and some additional discussion on the relation of order extended depths with depths including derivatives, are given.

S.1. Unified Approach to Depth Functionals and General Order Extended Depth

For the sake of technical correctness, let us start by introducing detailed notations. Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space on which all the random quantities are defined. For a measurable space S, $\mathcal{P}(S)$ stands for the collection of all probability measures defined on S and for $P \in \mathcal{P}(S)$, $X \sim P$ means a random variable X taking values in S having distribution P. For $X_1, \ldots, X_n \in S$ a random sample of size $n \in \mathbb{N}$ from P, the empirical measure corresponding to this random sample will be denoted by $P_n(\omega) = P_n \in \mathcal{P}(S)$ for $\omega \in \Omega$.

For $\tau \colon S \to S'$ a measurable mapping to a measurable space S', the distribution of the transformed random variable $\tau(X)$ is denoted by $P_{\tau(X)} \in \mathcal{P}(S')$. To avoid confusion, the original distribution of X can then likewise be designated by $P_X = P \in \mathcal{P}(S)$.

S.1.1. **Space of Continuous Functions and its Dual.** The space of real-valued continuous functions defined on the compact interval $[0,1]$ is denoted by $\mathcal{C}([0,1])$ and is, if not stated otherwise, assumed to be equipped with the uniform norm

$$\|x\| = \sup_{t \in [0,1]} |x(t)| \qquad \text{for } x \in \mathcal{C}([0,1]). \tag{S.1}$$

[1] KU LEUVEN, DEPARTMENT OF MATHEMATICS AND LEUVEN STATISTICS RESEARCH CENTER (LSTAT), BELGIUM

[2] CHARLES UNIVERSITY, DEPARTMENT OF PROBABILITY AND MATH. STATISTICS, CZECH REPUBLIC

E-mail addresses: stanislav.nagy@kuleuven.be.

Date: February 20, 2017.

The dual of the Banach space $\mathcal{C}([0,1])$, denoted by $\mathcal{C}([0,1])^*$, plays an important role in the framework of functional depths. It can be identified with the space of all finite signed Borel measures on $[0,1]$ (cf Rudin, 1987, Theorem 6.19). This representation yields that for any bounded linear functional $\phi \in \mathcal{C}([0,1])^*$ there exists a unique finite signed Borel measure μ on $[0,1]$ such that

$$\phi(x) = \int_0^1 x(t)\,\mathrm{d}\,\mu(t) \quad \text{for all } x \in \mathcal{C}([0,1]). \tag{S.2}$$

Slightly abusing the notation, we shall identify the dual space $\mathcal{C}([0,1])^*$ with the space of finite signed Borel measures on $[0,1]$ and write merely $\mu \in \mathcal{C}([0,1])^*$, having in mind the functional ϕ defined by formula (S.2).

The simplest examples of nontrivial elements of $\mathcal{C}([0,1])^*$ are Dirac measures. For $s \in [0,1]$, the Dirac measure at s is defined as the element $\delta_s \in \mathcal{C}([0,1])^*$ such that for $B \subset [0,1]$ Borel

$$\delta_s(B) = \begin{cases} 1 & \text{if } s \in B, \\ 0 & \text{if } s \notin B. \end{cases} \tag{S.3}$$

The functional (S.2) for $\mu = \delta_s$ is the coordinate projection in $\mathcal{C}([0,1])$

$$\int_0^1 x(t)\,\mathrm{d}\,\delta_s(t) = x(s) \quad \text{for all } x \in \mathcal{C}([0,1]).$$

Other simple representatives of the dual space $\mathcal{C}([0,1])^*$ are measures that are absolutely continuous with respect to the Lebesgue measure on $[0,1]$. For a measure μ like this, the Radon-Nikodym theorem (cf Rudin, 1987, Theorem 6.10) asserts the existence of a unique function $h \in L^1([0,1])$ such that

$$\int_0^1 x(t)\,\mathrm{d}\,\mu(t) = \int_0^1 x(t)h(t)\,\mathrm{d}\,t \quad \text{for all } x \in \mathcal{C}([0,1]). \tag{S.4}$$

As usual, $L^p([0,1])$ is the space of functions for which the pth power of the absolute value is Lebesgue integrable on $[0,1]$.

As we will see later, various depth functionals tailored for data from $\mathcal{C}([0,1])$ fit into a projection-based setup. These can be well defined by means of either measures on the dual space $\mathcal{C}([0,1])^*$, or subsets of it. We will deal with these in Sections S.1.3–S.1.5. First, in Section S.1.2 we introduce three depth functionals which are frequently used, but do not fit into these general classes.

S.1.2. **h-Mode Depth, Band Depth and Halfregion Depth.** A simple approach towards depth assignment for functions is called h-mode depth, and is motivated by the likelihood estimation from the finite-dimensional case. Cuevas et al. (2006) propose to utilize the idea of kernel density estimators to measure the centrality of a curve by accounting for how closely surrounded the curve is in a random sample from a distribution on functions.

Definition. The h-mode depth of $x \in \mathcal{C}([0,1])$ with respect to $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ is defined as

$$hM(x;P) = \mathrm{E}\left[\frac{1}{h(P)} K\left(\frac{\|x - X\|_2}{h(P)}\right)\right]. \tag{S.5}$$

Here, $\|\cdot\|_2$ is a norm on $\mathcal{C}([0,1])$ (possibly different from (S.1)), K is a kernel function and $h(P) > 0$ a bandwidth parameter. In the sample case (S.5) reduces to

$$hM(x; P_n) = \frac{1}{nh(P_n)} \sum_{i=1}^{n} K\left(\frac{\|x - X_i\|_2}{h(P_n)}\right). \tag{S.6}$$

The main difference between the h-mode depth and similar density-like concepts for functional random variables (cf Ferraty and Vieu, 2006, Dabo-Niang et al., 2006) is that in this case, the bandwidth $h(P_n)$ in (S.6) is not intended to vanish as the sample size n goes to infinity. Instead, the sample bandwidth $h(P_n)$ is assumed to be a positive number approaching a fixed positive value $h(P)$ as the sampling process continues. For details see Cuevas et al. (2006, 2007). As far as we know, the only theoretical result available for the h-mode depth is its sample version consistency (Nagy, 2015).

Perhaps the most popular functional depth used in many applications is the band depth (López-Pintado and Romo, 2009).

Definition. For $K = 2, 3, \ldots$, $P \in \mathcal{P}(\mathcal{C}([0,1]))$ and $X_1, \ldots, X_K \in \mathcal{C}([0,1])$ a random sample from P, the band depth (of order K) of $x \in \mathcal{C}([0,1])$ with respect to P is defined as

$$BD(x; P) = \frac{1}{K-1} \sum_{k=2}^{K} \mathsf{P}\left(x(t) \in \left[\min_{l=1,\ldots,k} X_l(t), \max_{l=1,\ldots,k} X_l(t)\right] \text{ for all } t \in [0,1]\right). \tag{S.7}$$

The sample version of BD is defined as the U-statistic corresponding to $BD(x; P_n)$. The basic properties of this depth were established by López-Pintado and Romo (2009), see also Gijbels and Nagy (2015) for the consistency results.

The third important depth functional we consider here is the halfregion depth (López-Pintado and Romo, 2011).

Definition. The halfregion depth of $x \in \mathcal{C}([0,1])$ with respect to $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ is defined as

$$HR(x; P) = \min\left\{\mathsf{P}(X(t) \geq x(t) \text{ for all } t \in [0,1]), \mathsf{P}(X(t) \leq x(t) \text{ for all } t \in [0,1])\right\}. \tag{S.8}$$

The sample version of HR is simply $HR(x; P_n)$ and some of its properties can be found in López-Pintado and Romo (2011). See also Kuelbs and Zinn (2015) for a discussion on its consistency properties.

In the rest of this section we mainly focus on depths defined by means of projections. In Section S.1.3 we start with a particular instance of a functional data depth based on the projections of the involved functions into a few first (robust) principal components computed from the dataset. Further, we discuss other projection-based functional depths, which can be well categorized either into the group of integrated (Section S.1.4) or infimal depths (Section S.1.5). Then, in Section S.1.6 we briefly recall how the derivatives of functions can be incorporated into the functional depth computation, and finally in Section S.1.7 we define the general order extensions of depth functionals.

S.1.3. **Dimension Reduction and Depth for Principal Components.** In what follows, assume that D is a generic finite-dimensional depth function applicable to d-dimensional datasets, for arbitrary $d \in \mathbb{N}$

$$D\colon \mathbb{R}^d \times \mathcal{P}\left(\mathbb{R}^d\right) \to [0,1]\colon\ (u, Q) \mapsto D(u; Q). \tag{S.9}$$

The most important finite-dimensional depth is the halfspace depth (Tukey, 1975, Donoho and Gasko, 1992) defined in Section 1 of the main paper. To provide a unifying exposition, similarly as in the main paper, when a concrete choice of a depth D is necessary we consider the halfspace depth, though any other finite-dimensional depth could be utilized as well.

A general technique applicable to functional observations rests in a two-step procedure, starting with a dimension reduction technique as the initial step. Assume that the data generated by a random function $X \sim P$ come from the Hilbert space $L^2([0,1])$. Denote the scalar product of $f, g \in L^2([0,1])$ by $\langle f, g \rangle = \int_0^1 f(t)g(t)\,\mathrm{d}\,t$. If certain moment assumptions on X may be justified, all the involved observations may initially be projected into the first few (robust) principal component curves $\psi_1, \ldots, \psi_d$ of P estimated from the dataset. The resulting vectors

$$(\langle x, \psi_1 \rangle, \ldots, \langle x, \psi_d \rangle)^{\mathsf{T}} \in \mathbb{R}^d \quad \text{for } x \in L^2([0,1]) \tag{S.10}$$

of the projections are often called the (robust) scores of the functions. In the second step of the analysis, the scores (S.10) of all the functions are used as substitutes (or approximations) of the original curves, and ordinary multivariate statistical analysis of these vectors is performed. This way, a simple functional depth may be obtained as

$$PCD_d(x; P) = D\left((\langle x, \psi_1 \rangle, \ldots, \langle x, \psi_d \rangle)^{\mathsf{T}}; P_{(\langle X, \psi_1 \rangle, \ldots, \langle X, \psi_d \rangle)^{\mathsf{T}}}\right). \tag{S.11}$$

As we will see in Section S.1.7, this depth is also very close to the general family of order extended depths. Nonetheless, the single projection on the pre-specified vector of the principal components of P makes it more a finite-dimensional depth function applied to a particular d-dimensional approximation of the functional data, rather than a typical functional data depth. As far we know, no systematic study of the theoretical properties of the depths (S.11) was performed.

S.1.4. **Integrated Depth and Random Projection Depth.** Cuevas and Fraiman (2009) proposed a general depth functional of integrated type.

Definition. Consider a probability measure ν defined on the space $\mathcal{C}([0,1])^*$. Then the integrated dual depth of $x \in \mathcal{C}([0,1])$ with respect to $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ is defined as

$$IDD(x; P, \nu) = \int_{\mathcal{C}([0,1])^*} D\left(\phi(x); P_{\phi(X)}\right)\,\mathrm{d}\,\nu(\phi). \tag{S.12}$$

In (S.12), $P_{\phi(X)} \in \mathcal{P}(\mathbb{R})$ stands for the $\mathbb{R}$-valued measure of the transformed random variable $\phi(X)$. Notice also that in the definition we assume ν defined as a measure on the space $\mathcal{C}([0,1])^*$, not as an element of the space $\mathcal{C}([0,1])^*$.

The general notion (S.12) naturally encompasses many depth functionals. For ν restricted to the set of Dirac measures (S.3) $\delta_t \in \mathcal{C}([0,1])^*$ for $t \in [0,1]$, the canonical bijection $\delta_t \mapsto t$ provides a major simplification. The measure ν can then be considered as defined on the space $[0,1]$, and for D univariate, (S.12) reduces to

$$IDD(x; P, \nu) = \int_0^1 D\left(x(t); P_{X(t)}\right)\,\mathrm{d}\,\nu(t). \tag{S.13}$$

Here, again, $P_{X(t)} \in \mathcal{P}(\mathbb{R})$ designates the distribution of the random variable $X(t) = \delta_t(X)$ for $t \in [0,1]$ fixed. The depth defined by (S.13) for ν the Lebesgue measure on $[0,1]$ is the integrated depth FD defined in (1). For different choices of D, it appears commonly in the literature:

- the original integrated depth of Fraiman and Muniz (2001),
- the modified band depth of López-Pintado and Romo (2009), and
- the modified halfregion depth of López-Pintado and Romo (2011),

are all special cases of (S.13) (Nagy et al., 2016) and thus, naturally, all provide very similar results for concrete examples.

As in Section S.1.3, the series of inclusions $\mathcal{C}([0,1]) \subset L^2([0,1]) \subset L^1([0,1])$ and the Riesz representation (S.4) enable for the exploitation of the Hilbert space structure of the space $L^2([0,1])$. Given a measure ν on the space $L^2([0,1])$, another special case of the definition (S.12) is the depth

$$IDD(x; P, \nu) = \int_{L^2([0,1])} D\left(\langle x, f\rangle; P_{\langle X,f\rangle}\right) \mathrm{d}\,\nu(f). \tag{S.14}$$

Here again, D is a general depth (S.9) for $d = 1$ and $P_{\langle X,f\rangle}$ the distribution of $\langle X, f\rangle$ for $f \in L^2([0,1])$ fixed. In practice, the integral in (S.14) is not possible to be evaluated precisely, as we integrate over a space that is too large. However, it is possible to approximate it. For $M \in \mathbb{N}$ and $f_1, \ldots, f_M \in L^2([0,1])$ a random sample from the measure ν we can define a depth

$$RP(x; P, \nu) = \frac{1}{M} \sum_{m=1}^{M} D\left(\langle x, f_m\rangle; P_{\langle X,f_m\rangle}\right). \tag{S.15}$$

This integrated dual depth with the choice of a Gaussian distribution in $L^2([0,1])$ for ν was proposed in Cuevas et al. (2007), who called it the random projection depth.

While many integrated type depths proved to be useful in applications, a major remaining concern is their theoretical background. An essential issue that needs to be resolved when defining a general depth in (S.12) is the choice of the reference measure ν on $\mathcal{C}([0,1])^*$. The Lebesgue measure suits well for the integrated depth (S.13), but the domain in (S.14) is already an infinite-dimensional space and approximations as in (S.15) must be used (and justified). The measurability of the integrand function in (S.12) turns out to be another nontrivial problem. For depths in form (S.13) this was recently resolved (Nagy et al., 2016). However, in general, even for depths defined by (S.14), these issues definitely deserve more attention. Apart from some scattered results for particular instances of IDD (taking the form (S.13)) in Fraiman and Muniz (2001), Cuevas and Fraiman (2009) and Nagy et al. (2016), not much is available from theoretical point of view for the general IDD in (S.12).

S.1.5. **Infimal Depth and Random Functional Depth.** A unifying approach similar to that of integrated dual depths was recently proposed by Mosler (2013) by introducing the class of Φ-depths for general data. The method stems from (S.12) and differs only by replacing the integral in the formula by an infimum. The Φ-depth, or as we will also refer to it, the infimal type depth for functional data, can then be defined as follows. Again, assume that D is a depth as in (S.9).

Definition. For $\Phi \subset \mathcal{C}([0,1])^*$ given, the Φ-depth of $x \in \mathcal{C}([0,1])$ with respect to $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ is

$$ID(x; P, \Phi) = \inf_{\phi\in\Phi} D\left(\phi(x); P_{\phi(X)}\right).$$

Contrary to the integrated dual depth (S.12), for the Φ-depth one needs to predefine only a set $\Phi \subset \mathcal{C}([0,1])^*$, instead of a measure ν on $\mathcal{C}([0,1])^*$.

Infimal type depths resembling the integrated depths reviewed in Section S.1.4 are immediate. For $\Phi = \{\delta_t \colon t \in [0,1]\}$ one finds the infimal depth for functions ID defined in (2) (with Φ omitted from the arguments).

A depth of infimal type corresponding to the random projection depth (S.15) is the random functional depth of Cuesta-Albertos and Nieto-Reyes (2008a,b). Given a random sample $f_1, \ldots, f_M \in L^2([0,1])$ of size $M \in \mathbb{N}$ from a measure ν on $L^2([0,1])$ one can define it as

$$RF(x; P) = \inf_{m=1,\ldots,M} D\left(\langle x, f_m\rangle; P_{\langle X, f_m\rangle}\right). \tag{S.16}$$

Theoretical properties of infimal type depths are even more sparse in the literature than those for integrated type depths. Apart from observations made by Mosler and Polyakova (2012) and the consistency result of (Cuesta-Albertos and Nieto-Reyes, 2008b, Theorem 2.10) for random functional depths (conditionally on M, a random sample $f_1, \ldots, f_M$ and using univariate halfspace depth), no general theory applicable to these functionals is available. Consistency results for general infimal type depths have not been established, and there are only few consistency results for concrete representatives of these functionals (cf Gijbels and Nagy, 2015, Kuelbs and Zinn, 2013).

S.1.6. **Functional Depths Incorporating Derivatives.** If all the considered functions are differentiable, it is well known that taking the derivatives into account might lead to a significant improvement in the performance of statistical procedures. Thus, assume in this section that all trajectories of a random function $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ are continuously differentiable on $[0,1]$ (with one-sided derivatives at the endpoints). Focusing, for the sake of clarity, on the case of once differentiable functions, one can now represent the realized random function $X \in \mathcal{C}([0,1])$ by a couple $(X, X')^{\mathsf{T}} \in \mathcal{C}([0,1])^2$, or alternatively, by a vector-valued function

$$(X, X')^{\mathsf{T}} : [0,1] \to \mathbb{R}^2 \colon t \mapsto (X(t), X'(t))^{\mathsf{T}}. \tag{S.17}$$

Here, $X'(t)$ stands for the derivative of X at $t \in [0,1]$.

Having the curves represented in this fashion, extensions of some depth functionals towards differentiable functions are readily available. Assume that the trajectories of $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ and $x \in \mathcal{C}([0,1])$ are differentiable and suppose that the D depth is given by (S.9) for $d = 2$.

The integrated depth (S.13) can be easily extended (Nagy, 2012, Claeskens et al., 2014):

$$IDD(x; P, \nu) = \int_0^1 D\left((x(t), x'(t))^{\mathsf{T}}; P_{(X(t), X'(t))^{\mathsf{T}}}\right) \mathrm{d}\,\nu(t).$$

For ν the Lebesgue measure on $[0,1]$, this depth is called the integrated depth including derivatives and denoted in (3) by $FD^{(2)}$.

Similarly, the infimal depth (2) takes in this case the form (4).

Two extended versions of the random projection depth (S.15) to this setup were introduced in Cuevas et al. (2007). In both of them, the pair of functions $(x, x')^{\mathsf{T}}$ is first projected by a function $f \in L^2([0,1])$ onto $(\langle x, f\rangle, \langle x', f\rangle)^{\mathsf{T}} \in \mathbb{R}^2$. Then, two possibilities were outlined by the authors:

- either a two-dimensional depth D is employed to assess the depth of the once projected quantity $(\langle x, f\rangle, \langle x', f\rangle)^{\mathsf{T}}$ directly, or

- a second random projection of this two-dimensional quantity is made to obtain the final, real-valued projected value. A univariate depth D of these twice projected quantities is then computed.

If the first mentioned method is applied in the same way as in the random projection depth RP in (S.15), we continue to speak about the random projection depth (now involving derivatives) and denote the depth by $RP^{(2)}$. If the functions are projected twice and then RP is applied, we shall denote this depth as the double random projection depth and denote it by RPD, respecting the notation of Cuevas et al. (2007).

For the depths described in Section S.1.2, a simple extension to the setup of differentiable functions can be made in the spirit of Ieva and Paganoni (2013). For the band depth BD, for instance, the band depth involving derivatives can be defined by

$$BD^{(2)}(x;P) = BD(x;P_X) + BD(x';P_{X'}), \tag{S.18}$$

and similar extensions for the halfregion depth HR and the h-mode depth hM are evident. In all these cases, the version of the depth incorporating the derivatives is distinguished by a superscript $^{(2)}$. Another option for the extension of band depth BD would be the recent proposal of simplicial band depth of López-Pintado et al. (2014) applied to bivariate functions (S.17). However, this depth and (S.18) give in practice similar results.

Although the applicability of the depths proposed in this section is limited to smooth functions, Cuevas et al. (2007) estimate the "derivatives" of non-differentiable functions. By doing so they include some information about the shape of the functions into the procedures, though, strictly speaking, the estimated derivative curves lose their interpretation. In Example S.2 in Section S.2 (and Example 1 in the main paper) we proceed in this way as well, and formally estimate derivatives for non-differentiable functional data from the discretized functional values.

S.1.7. **General Order Extended Depths.** Taking into account the general exposition on integrated and infimal type depth functionals in Sections S.1.4 and S.1.5, respectively, the order extensions of (7) and (8) can be defined in full generality.

In both following definitions, D denotes a generic finite-dimensional depth (S.9) for $d = J$, $\left(\mathcal{C}\left([0,1]\right)^*\right)^J$ is the product space of J copies of $\mathcal{C}\left([0,1]\right)^*$, and $P_{\phi(X)} \in \mathcal{P}\left(\mathbb{R}^J\right)$ stands for the measure of the transformed random variable $\phi(X)$, $\phi \in \left(\mathcal{C}\left([0,1]\right)^*\right)^J$.

Definition. For $J = 1, 2, \ldots$ consider a probability measure ν defined on $\left(\mathcal{C}\left([0,1]\right)^*\right)^J$. Then the Jth order integrated dual depth of $x \in \mathcal{C}\left([0,1]\right)$ with respect to $X \sim P \in \mathcal{P}\left(\mathcal{C}\left([0,1]\right)\right)$ is defined as

$$IDD_J(x;P,\nu) = \int_{\left(\mathcal{C}([0,1])^*\right)^J} D\left(\phi(x);P_{\phi(X)}\right) \mathrm{d}\,\nu(\phi), \tag{S.19}$$

if the integral on the right hand side is defined.

Definition. For $J = 1, 2, \ldots$, let $\Phi \subset \left(\mathcal{C}\left([0,1]\right)^*\right)^J$ be given. Then the Jth order Φ-depth of $x \in \mathcal{C}\left([0,1]\right)$ with respect to $X \sim P \in \mathcal{P}\left(\mathcal{C}\left([0,1]\right)\right)$ is

$$ID_J(x;P,\Phi) = \inf_{\phi\in\Phi} D\left(\phi(x);P_{\phi(X)}\right). \tag{S.20}$$

The sample versions of these depths are defined just as those for order extended depth functionals from Section 3 in the main paper, and depend on the sampling design of the data curves. Moreover, finite approximations of these quantities in the spirit of Cuesta-Albertos and Nieto-Reyes (2008b,a) can be defined analogously to (9) and (10).

In general, the depths defined above may take on various forms and extensions of any depth functional mentioned in Sections S.1.3, S.1.4 and S.1.5 can be derived as their special case. For simplicity, in Section 3 of the main paper, in (8) and (7), respectively, we consider the following instances

- the set Φ in (S.20) is for δ_t Dirac measures (S.3) set to be

$$\Phi = \left\{ (\delta_{t_1}, \ldots, \delta_{t_J})^\mathsf{T} : t_1, \ldots, t_J \in [0,1] \right\} \subset \left(\mathcal{C}\left([0,1]\right)^*\right)^J,$$

- and the reference measure ν in (S.19) is defined on Φ, such that its image in the canonical representation

$$\left(\mathcal{C}\left([0,1]\right)^*\right)^J \to [0,1]^J \colon\ (\delta_{t_1}, \ldots, \delta_{t_J})^\mathsf{T} \mapsto (t_1, \ldots, t_J)^\mathsf{T}$$

  is the Lebesgue measure on the unit cube $[0,1]^J$.

Note that also the principal components depth (S.11) is very similar to both order extended depths (S.19) and (S.20), if just a single projection on the set of first d (robust) principal components is taken. Nonetheless, formally they are not members of neither of the families, as the choice of the projections depends on the data, and P.

S.2. Simulation Study

S.2.1. **Technical Details.** Let us start by listing the full technical details of the simulation study to be performed in Section S.2.2 in Examples S.1–S.6 (and Examples 1 and 2 from the main paper).

- For each example, a total of 100 independent runs are generated. In each run, two independent random samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_n$ of size $n = 200$ curves from distribution $P \in \mathcal{P}\left(\mathcal{C}\left([0,1]\right)\right)$ are generated. The X_i curves constitute the random sample of functions generated from P, Y_i are additional curves generated from P whose depths with respect to the X_i curves are computed, $i = 1, \ldots, n$.
- A single, fixed (Examples S.1, S.3, S.4, S.5 and S.6) or random (Example S.2) outlying function Y contaminates the random sample of test functions Y_i, $i = 1, \ldots, n$.
- The functional depths from Tab. S.1 are used to assess the depth values of the set of curves $Y_1, \ldots, Y_n$ appended with Y, with respect to an empirical measure supported in the random sample functions $X_1, \ldots, X_n$.
- To each of the depths, in the corresponding line of the tables of results (Tabs. S.2, S.4 and S.6) the mean and standard deviation (in brackets) over 100 runs is computed from the following characteristics:
  - **Rank:** depth-based centrality rank of the outlying function Y (the rank of the depth of Y with respect to the depths of functions $Y_1, \ldots, Y_n, Y$, ranging from 1 to 201, 201 meaning the smallest depth value, that is the most outlying function);
  - **Ties:** the total number of functions attaining simultaneously the lowest depth value (ranging from 1 to 201, 201 meaning that all the functions attained the same, low value of depth).

  The actual value of the depth of Y is not displayed, since the very diverse nature of the considered depth functionals makes it difficult to compare these values directly from depth to depth.

- In the associated tables of results for infimal depths only (Tabs. S.3, S.5 and S.7) the same characteristics of **Rank** and **Ties** are obtained also for the adjusted ranking for infimal depths, as outlined in formula (12) in Section 3 of the main paper.

  The adjusted ranking for infimal depths is defined as follows:

  Definition. Consider a general depth of infimal type ID from Sections S.1.5 and S.1.7, defined as an infimum of depths D over a set of projections $\Phi \subset \left(\mathcal{C}\left([0,1]\right)^*\right)^J$. For a given measure ν on Φ, one says that the function x has higher adjusted rank based on ID (is less central for ID) than y if
  – either $ID(x;P) < ID(y;P)$,
  – or $c = ID(x;P) = ID(y;P)$ and

  $$\nu\left(\left\{\phi \in \Phi\colon D(\phi(x); P_{\phi(X)}) = c\right\}\right) \geq \nu\left(\left\{\phi \in \Phi\colon D(\phi(y); P_{\phi(X)}) = c\right\}\right).$$

  This way, x is less central than y also if their infimal depths are the same, but for x the minimal value of D is attained over a larger subset of Φ. For the infimal depths ID from (2) and $ID^{(2)}$ from (4), ν is taken to be the Lebesgue measure on $[0,1]$ (or, more precisely, its canonical copy in $\mathcal{C}\left([0,1]\right)^*$). For the order extended infimal depths ID_J and their approximations ID_J^A, the Lebesgue measure on $[0,1]^J$ (that is, in $\left(\mathcal{C}\left([0,1]\right)^*\right)^J$) is used.
- A depth well recognizing shape outliers should be assessing low (adjusted) depth values to Y. The depth-based centrality rank of Y should be indicating outlyingness (i.e. as high as possible). At the same time there should not be too many functions of the random sample attaining very low depth values (the number of ties in the lowest depth value should be low), as this would result in a large false alarm rate in the outlier flagging process. Thus, an ideal functional depth should be having mean characteristics as close as possible to $(201, 1)$ with as low standard deviations as possible.
- The computations are made in `R 3.0.2` (R Core Team, 2013).
- All the functions involved are discretized on an equidistant grid of 101 points in the interval $[0,1]$. Thus, each curve is represented by a 101-dimensional vector of its functional values.
- The numerical derivatives of functions are computed from the associated discrete vectors as the derivatives of fitted cubic splines (function `D1ss` from the `R` package `sfsmisc` of Maechler (2013)). In Examples S.3 and S.6, the derivative of the discontinuous outlying function Y is computed in the same manner, but separately on the subintervals, where the function Y is continuous and differentiable.

Details for the choice of parameters of the functional depths described in Section S.1 are as follows.

- For random and double random projection depths (RP, $RP^{(2)}$ and RPD), 50 random functions $f \in L^2\left([0,1]\right)$ are generated from a Gaussian distribution. This was chosen in agreement with Cuevas et al. (2007).
- For the random functional depth RF, $M = 15$ projections were taken from a Gaussian distribution. As argued by Cuesta-Albertos and Nieto-Reyes (2008b), the number of projections should be kept rather low compared to the dimensionality of the random sample (which is 101 in the case of our simulations).
- For the h-mode depth, the parameters were chosen similarly to those in Cuevas et al. (2007): $\|\cdot\|_2$ is the $L^2\left([0,1]\right)$ metric, K the Gaussian kernel and $h(P)$ the

TABLE S.1. *The full list of functional depths considered in the comparison.*

Type	No	Depth	Abbreviation	Referenced in
Without Derivatives	1	Integrated	FD	(1) in Section 1
	2	Infimal	ID	(2) in Section 1
	3	h-Mode	hM	(S.6) in Section S.1.2
	4	Band	BD	(S.7) in Section S.1.2
	5	Halfregion	HR	(S.8) in Section S.1.2
	6	Random Projection	RP	(S.15) in Section S.1.4
	7	Random Functional	RF	(S.16) in Section S.1.5
	8	Principal Components	PCD_1	(S.11) in Section S.1.3
	9	Principal Components	PCD_2	(S.11) in Section S.1.3
With Derivatives	10	Integrated	$FD^{(2)}$	(3) in Section 1
	11	Infimal	$ID^{(2)}$	(4) in Section 1
	12	h-Mode	$hM^{(2)}$	(S.6) in Section S.1.2, Section S.1.6
	13	Band	$BD^{(2)}$	(S.7) in Section S.1.2, Section S.1.6
	14	Halfregion	$HR^{(2)}$	(S.8) in Section S.1.2, Section S.1.6
	15	Random Projection	$RP^{(2)}$	Sections S.1.4, S.1.6
	16	Double Random Projection	RPD	Sections S.1.4, S.1.6
Newly Proposed	17	2nd Order Integrated	FD_2	(7) in Section 3
	18	2nd Order Integrated*	FD_2^A	(9) in Section 3
	19	3rd Order Integrated*	FD_3^A	(9) in Section 3
	20	2nd Order Infimal	ID_2	(8) in Section 3
	21	2nd Order Infimal*	ID_2^A	(10) in Section 3
	22	3rd Order Infimal*	ID_3^A	(10) in Section 3

* Approximated version with $M = 100$ chosen in (9) and (10).

20-percentile of the distribution of $\|X_1 - X_2\|_2$ for X_1, X_2 random, independent and distributed as $P \in \mathcal{P}(\mathcal{C}([0,1]))$, estimated by the sample quantile from the X_i functions, $i = 1, \ldots, n$.

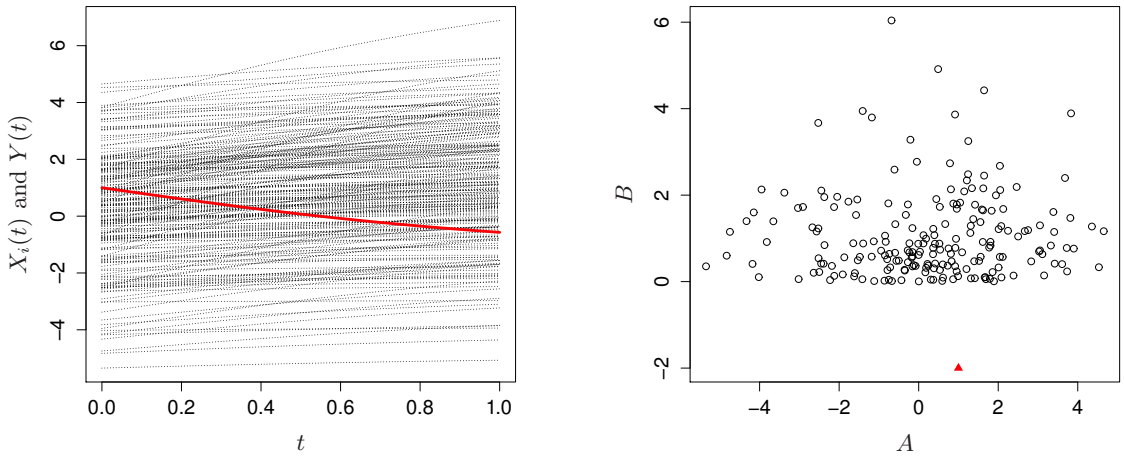
- For the band depth BD we take $K = 2$.

S.2.2. **Simulated Examples.** The first two examples exhibit difficulties when aiming to capture a function with different monotonicity properties than all the observations, i.e. a 2nd order outlier. In Example S.3 we deal with the identification of a discontinuous function in a random sample of continuous ones. Example S.4 presents an instance of a 3rd order outlier. In Examples S.5 and S.6 we illustrate some difficulties of the depths based on projecting into the principal components, in the outlier recognition task.

Example S.1. Consider $X \sim P$ generated by the process $X(t) = A + B\arctan(t)$ for $t \in [0,1]$, for A and B as in Example 1. Each random function generated by such process is almost surely increasing in $[0,1]$ with intercept A.

The single fixed function Y is given by $Y(t) = 1 - 2\arctan(t)$ for $t \in [0,1]$. One such random sample, together with the outlying function can be found in Fig. S.1A.

The whole set of sample functions and Y comes from a two-dimensional subspace of $\mathcal{C}([0,1])$ spanned by a constant function and the function $\arctan(t)$, $t \in [0,1]$. This makes the whole dataset two-dimensional, and the coefficients of each function with respect to this basis can be visualized in a two-dimensional plot (Fig. S.1B). While the graph of the function Y (red solid line in Fig. S.1A) lies within the cluster of graphs of functions from



(A) (B)

FIGURE S.1. *Example S.1: (A) Random sample from P and the outlier Y (red solid); and (B) the basis coefficients of these functions (red triangle representing Y).*

P, the point representing the coefficients of Y (red triangle in Fig. S.1B) is obviously the single outlier in the sample.

From columns 3 and 4 of Tab. S.2 we see that, with the exception of PCD_2, none of the depths without derivatives can capture the shape difference of Y reliably. For these, the central position of the graph of Y is enough to conclude that it must be central in the random sample.

If the first derivative is considered, the recognition of Y is much easier, as the large discrepancy between the derivatives of X and Y shows up. Along with PCD_2, three depths — $FD^{(2)}$, $ID^{(2)}$ and $RP^{(2)}$ — achieve very good results, with no variance in rank, and with a relatively small amount of functions identified as having minimal depth value. For $FD^{(2)}$ and $ID^{(2)}$ this is because of their non-random nature and the proper use of derivatives, for $RP^{(2)}$ and PCD_2 this is related to the choice of two-dimensional projections and the matching two-dimensional nature of the dataset.

As expected, all order extended depths do a good job in identifying the function Y as an outlier. In Tab. S.3 we can also observe that in the (relatively small) group of functions attaining the smallest infimal depths, Y is still reliably distinguished as the most atypical observation.

The random functional depth RF is also doing reasonably well, though with a huge amount of variability in the results. To account for it, notice that a fair comparison of the performance of RF to other functional depths is a tricky issue for simulated data. While other functional depths are constructed purely for infinite-dimensional datasets, the random functional depth is able to take advantage of the finite-dimensional discretization of functional values. Since the data in the simulations are in fact only 101-dimensional, the RF depth (S.16) provides a finite approximation to the 101-dimensional halfspace depth of the discretized functional values. Consequently, for very large values of M (in orders of tens of thousands) in (S.16), the depth deceptively recognizes any functional

outlier perfectly; however, this is true only because of the discretization of the data. On the other hand, for low values of M (15 in our case), the depth recognizes the outliers usually reasonably well compared to other established depth functionals, even though the derivatives are not considered, but with a great variance in the results. This phenomenon will be observed also in the rest of the examples in the paper, and is not discussed in detail there.

Another interesting trait that can be observed in Tab. S.2 is the remarkable similarity in the results for some functional depths (depths 10, 11, 15 and the order extended depths in Example S.1). This can be noticed also in other examples below, and accounted for by the low-dimensional nature of the considered functional data.

TABLE S.2. *The table of results of the simulation study in Examples S.1 and S.2.*

		Example S.1		Example S.2	
no	Abbr	Rank	Ties	Rank	Ties
1	FD	50.2 (06.8)	1.4 (0.9)	55.5 (16.1)	1.1 (0.4)
2	ID	82.0 (10.1)	3.5 (2.5)	73.1 (25.2)	6.3 (4.4)
3	hM	117.3 (13.9)	1.0 (0.0)	105.3 (29.3)	1.0 (0.0)
4	BD	144.7 (09.4)	3.2 (2.5)	119.8 (30.1)	7.6 (5.2)
5	HR	75.4 (10.1)	3.5 (2.6)	63.4 (24.9)	7.9 (5.1)
6	RP	75.9 (12.7)	1.0 (0.0)	73.7 (22.1)	1.0 (0.0)
7	RF	187.4 (24.3)	5.6 (3.1)	153.3 (51.1)	10.2 (4.8)
8	PCD_1	11.8 (07.1)	2.6 (1.8)	22.6 (14.7)	3.4 (2.2)
9	PCD_2	201.0 (00.0)	10.9 (6.4)	143.6 (58.0)	11.0 (5.3)
10	$FD^{(2)}$	201.0 (00.0)	11.1 (4.7)	67.0 (28.6)	1.2 (0.5)
11	$ID^{(2)}$	201.0 (00.0)	11.1 (4.7)	155.0 (70.0)	126.4 (7.7)
12	$hM^{(2)}$	188.7 (03.4)	1.0 (0.0)	119.3 (51.5)	1.0 (0.0)
13	$BD^{(2)}$	192.5 (03.8)	1.1 (0.3)	119.8 (30.1)	7.6 (5.2)
14	$HR^{(2)}$	163.0 (08.3)	1.1 (0.4)	63.4 (24.9)	7.9 (5.1)
15	$RP^{(2)}$	201.0 (00.0)	11.1 (4.7)	87.8 (31.4)	1.0 (0.0)
16	RPD	162.0 (13.7)	1.0 (0.0)	98.7 (47.2)	1.0 (0.0)
17	FD_2	201.0 (00.0)	11.1 (4.7)	122.7 (28.0)	1.2 (0.6)
18	FD_2^A	201.0 (00.0)	11.1 (4.7)	123.0 (31.0)	1.3 (0.7)
19	FD_3^A	201.0 (00.0)	11.1 (4.7)	147.2 (30.8)	1.8 (1.2)
20	ID_2	201.0 (00.0)	11.1 (4.7)	201.0 (00.0)	199.1 (1.4)
21	ID_2^A	201.0 (00.0)	11.1 (4.7)	192.2 (32.8)	113.3 (10.2)
22	ID_3^A	201.0 (00.0)	11.1 (4.7)	201.0 (00.0)	178.0 (5.8)

Example S.2. The simulation setting is the same as that in Example 1 from the main paper. As can be seen from Tab. S.2, the results as described in Example 1 are valid also more generally: established depths without derivatives are incapable of the identification of the outlier, and hardly any improvement can be seen when examining depths including derivatives (with greater variability in results). Order extended depths identify the atypical shape of the outlier better, but the amount of noise in the data still prevents them from achieving more plausible results.

TABLE S.3. *The table of results of the simulation study in Examples S.1 and S.2: Infimal depth ranks.*

		Example S.1		Example S.2	
no	Abbr	Rank	Ties	Rank	Ties
2	ID	80.2 (10.1)	1.0 (0.1)	72.5 (25.1)	1.1 (0.3)
11	$ID^{(2)}$	201.0 (00.0)	1.2 (0.5)	111.9 (51.5)	1.1 (0.4)
20	ID_2	201.0 (00.0)	1.2 (0.5)	176.2 (29.7)	1.5 (1.3)
21	ID_2^A	201.0 (00.0)	1.2 (0.5)	175.3 (34.1)	1.1 (0.4)
22	ID_3^A	201.0 (00.0)	1.2 (0.5)	175.5 (29.2)	1.3 (0.8)

Example S.3. The simulation setting is the same as that in Example 2 from the main paper. As in Example S.2, the analysis from Section 2 for this example is still valid: all non-extended depths, except for PCD_2, perform poorly, ranking the outlier deeply into "typical" curves. As a result of integration in FD_2 and FD_3^A, the order extended integrated depths are also not performing too good for reasons explained in Example 2. On the other hand, all the order extended infimal depths do a good job, and rank the outlier into the set of (small amount of) curves with minimal depth value. The good performance of PCD_2 is rather incidental. It can be explained by the (almost) one-dimensional nature of the data, making it easy to recognize the outlier as the only observation not fitting into the one-dimensional subspace of the linear space spanned by the first two principal components of X.

Example S.4. This example provides a dataset and an outlier Y which is exceptionally hard to be detected. Define the random function $X \sim P$ again as in Example S.1, and take the outlying function $Y(t) = \tan(t)/4$ for $t \in [0, 1]$. Both X and Y are strictly increasing, and at the same time the graph of Y is located inside the bundle of random sample curves from P (see Fig. S.2A). All the functions are differentiable, and it takes the inspection of the first derivatives (Fig. S.2B) to observe that Y is in fact the only convex function in a random sample of concave ones. Thus, Y should be identified as an outlier with respect to X, but this might be very hard to notice as neither the location of the graph, nor the location of the graph of the derivative indicate any abnormal behaviour.

Looking at columns 5 and 6 of Tab. S.4, all the depths without derivatives, as well as the vast majority of depths considering them, fail critically by claiming Y to be very central in the random sample. The single exception is the band depth with derivatives $BD^{(2)}$ assigning Y into the less central half of the data, though this is not caused by proper shape recognition, but only by the fact that the graph of the first derivative of Y crosses a lot of graphs of derivatives of random sample curves near $t = 1$.

As for the order extended depths, the choice $J = 2$ is not enough to spot the difference in the second derivative of Y, and neither of these extended depths outperforms the usual depth functionals. Nevertheless, if the choice $J = 3$ is employed, both FD_3^A and ID_3^A provide almost perfect outlier recognition, similarly as FD_2 and ID_2 did in Example S.1, see Tab. S.4. This is confirmed also in Tab. S.5.

Example S.5. Let $X \sim P$ take the form

$$X(t) = 0.3A\sin(2\pi(t - 1/2)) + 0.3B\sin(4\pi(t - 1/2)) + C \quad \text{for } t \in [0, 1],$$

for A, B and C independent, standard normal random variables.

TABLE S.4. *The table of results of the simulation study in Examples S.3 and S.4.*

		Example S.3		Example S.4	
no	Abbr	Rank	Ties	Rank	Ties
1	FD	41.7 (06.8)	1.6 (1.1)	18.8 (11.1)	1.6 (1.0)
2	ID	39.7 (08.7)	3.6 (2.7)	15.7 (08.5)	3.6 (2.3)
3	hM	97.5 (14.3)	1.0 (0.0)	22.6 (14.9)	1.0 (0.0)
4	BD	96.3 (14.1)	3.6 (2.7)	13.9 (09.7)	3.8 (2.4)
5	HR	37.2 (08.5)	3.9 (2.8)	12.9 (08.0)	3.9 (2.4)
6	RP	62.4 (10.4)	1.0 (0.0)	16.2 (08.8)	1.0 (0.0)
7	RF	180.7 (34.0)	5.5 (3.6)	30.3 (24.3)	5.3 (2.8)
8	PCD_1	24.9 (08.6)	2.9 (2.4)	15.1 (07.5)	2.9 (1.9)
9	PCD_2	200.8 (01.8)	11.6 (5.8)	32.8 (06.0)	10.8 (5.4)
10	$FD^{(2)}$	42.8 (10.4)	10.8 (4.4)	26.9 (06.4)	10.6 (4.1)
11	$ID^{(2)}$	50.0 (10.9)	10.8 (4.4)	76.9 (13.3)	10.6 (4.1)
12	$hM^{(2)}$	76.3 (09.7)	1.0 (0.0)	22.2 (06.9)	1.0 (0.0)
13	$BD^{(2)}$	85.7 (11.4)	1.0 (0.2)	138.3 (08.6)	1.1 (0.3)
14	$HR^{(2)}$	52.2 (10.0)	1.1 (0.4)	56.0 (12.4)	1.1 (0.3)
15	$RP^{(2)}$	58.7 (11.0)	10.8 (4.4)	47.8 (12.5)	10.6 (4.1)
16	RPD	57.9 (13.3)	1.0 (0.0)	36.8 (10.2)	1.0 (0.0)
17	FD_2	62.0 (08.5)	10.8 (4.4)	20.9 (06.4)	10.6 (4.1)
18	FD_2^A	62.8 (09.9)	10.8 (4.4)	21.0 (06.9)	10.6 (4.1)
19	FD_3^A	112.3 (15.3)	10.8 (4.4)	201.0 (00.0)	11.6 (4.1)
20	ID_2	201.0 (00.0)	11.8 (4.4)	74.3 (13.2)	10.6 (4.1)
21	ID_2^A	201.0 (00.0)	11.8 (4.4)	58.5 (14.0)	10.6 (4.1)
22	ID_3^A	201.0 (00.0)	11.8 (4.4)	201.0 (00.0)	11.6 (4.1)

TABLE S.5. *The table of results of the simulation study in Examples S.3 and S.4: Infimal depth ranks.*

		Example S.3		Example S.4	
no	Abbr	Rank	Ties	Rank	Ties
2	ID	38.9 (08.5)	1.0 (0.1)	15.0 (08.6)	1.0 (0.1)
11	$ID^{(2)}$	47.9 (10.7)	1.1 (0.4)	74.1 (13.2)	1.2 (0.5)
20	ID_2	190.2 (04.4)	1.1 (0.4)	71.8 (12.7)	1.2 (0.5)
21	ID_2^A	190.2 (04.4)	1.1 (0.4)	56.0 (13.9)	1.2 (0.5)
22	ID_3^A	190.2 (04.4)	1.1 (0.4)	201.0 (00.0)	1.2 (0.5)

The single deterministic function Y is given by $Y(t) = 2\sin(10\pi(t-1/2))$ for $t \in [0,1]$. One such sample, together with the outlying function can be found in Fig. S.3A. The outlier Y is atypical because of its higher amplitude, and shorter period. Though, it is constructed to lie inside the main bundle of the sample curves from X.

As can be seen in Tab. S.6, most depth functionals are capable of identifying Y as atypical, especially due to its high peak-to-peak amplitude. The only exceptions are the depths based on the expansion into principal components, PCD_1 and PCD_2. The reason for this failure is obvious: X is, up to some noise, a function from the bivariate

TABLE S.6. *The table of results of the simulation study in Examples S.5 and S.6, and the computational cost of the considered depths, corresponding to one run in a single simulation setup.*

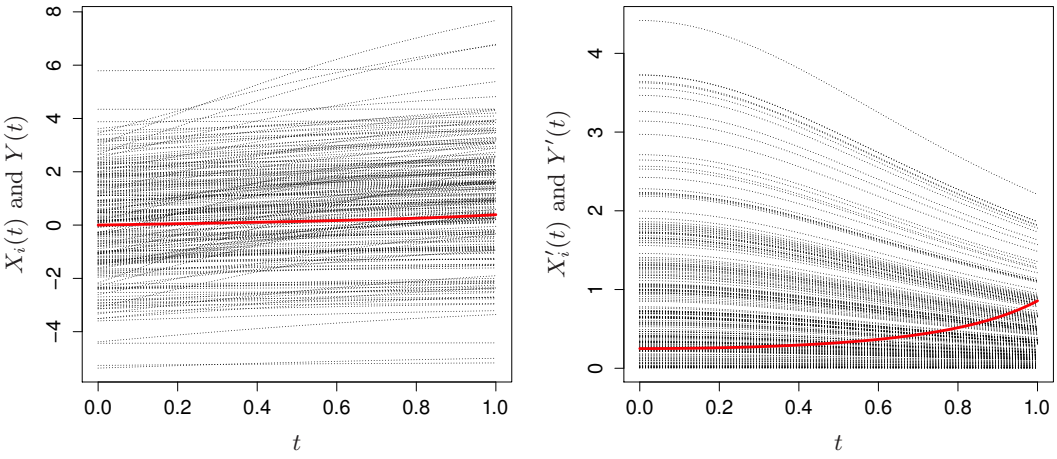
		Example S.5		Example S.6		
no	Abbr	Rank	Ties	Rank	Ties	Time (in s)
1	FD	146.1 (06.2)	1.0 (0.2)	145.9 (07.3)	1.0 (0.2)	0.03
2	ID	186.7 (05.7)	6.4 (4.0)	187.2 (06.3)	5.8 (3.3)	0.03
3	hM	199.9 (01.0)	1.0 (0.0)	200.1 (01.0)	1.0 (0.0)	0.04
4	BD	196.3 (04.2)	8.4 (5.1)	196.0 (04.0)	7.4 (4.0)	0.22
5	HR	194.3 (05.6)	8.5 (5.1)	193.5 (05.4)	7.6 (3.9)	0.02
6	RP	153.1 (13.6)	1.0 (0.0)	152.4 (13.4)	1.0 (0.0)	0.01
7	RF	200.3 (07.0)	10.7 (4.5)	200.9 (01.3)	10.3 (4.9)	0.01
8	PCD_1	9.5 (05.7)	2.7 (1.9)	9.3 (05.9)	2.4 (1.7)	0.27
9	PCD_2	1.5 (00.8)	8.2 (3.9)	1.5 (01.0)	7.4 (3.6)	0.28
10	$FD^{(2)}$	199.1 (01.5)	1.1 (0.3)	84.6 (07.9)	1.0 (0.2)	1.52
11	$ID^{(2)}$	201.0 (00.0)	26.2 (6.4)	161.0 (15.9)	24.1 (6.6)	1.52
12	$hM^{(2)}$	201.0 (00.0)	1.0 (0.0)	55.1 (06.7)	1.0 (0.0)	1.19
13	$BD^{(2)}$	200.3 (01.2)	1.5 (0.9)	184.9 (06.4)	1.4 (0.7)	1.40
14	$HR^{(2)}$	194.3 (05.6)	8.5 (5.1)	193.5 (05.4)	7.6 (3.9)	1.16
15	$RP^{(2)}$	199.4 (02.7)	1.0 (0.0)	79.7 (17.9)	1.0 (0.0)	1.67
16	RPD	201.0 (00.2)	1.0 (0.0)	43.9 (14.8)	1.0 (0.0)	1.67
17	FD_2	185.6 (03.9)	1.1 (0.2)	185.9 (04.1)	1.0 (0.2)	18.72
18	FD_2^A	185.0 (05.9)	1.2 (0.4)	185.4 (06.4)	1.1 (0.3)	0.37
19	FD_3^A	176.2 (08.2)	20.1 (9.6)	176.5 (07.1)	22.0 (8.3)	85.92*
20	ID_2	201.0 (00.0)	41.5 (9.2)	201.0 (00.0)	39.9 (9.6)	18.72
21	ID_2^A	201.0 (00.0)	26.9 (6.8)	201.0 (00.0)	25.3 (6.7)	0.37
22	ID_3^A	201.0 (00.0)	26.6 (6.6)	201.0 (00.0)	25.4 (6.5)	85.92*

* The `Fortran` implementation developed by the authors makes the computation of most of the depths very efficient. On the other hand, the evaluation of both 3rd order depths is performed using the much slower `depth` function (Genest et al., 2012) in the `R` environment, which should be taken into account when interpreting these results.

TABLE S.7. *The table of results of the simulation study in Examples S.5 and S.6: Infimal depth ranks.*

		Example S.5		Example S.6	
no	Abbr	Rank	Ties	Rank	Ties
2	ID	182.3 (05.0)	1.0 (0.2)	182.4 (05.6)	1.0 (0.2)
11	$ID^{(2)}$	199.8 (01.3)	1.0 (0.0)	151.4 (12.8)	1.0 (0.2)
20	ID_2	197.3 (02.7)	1.0 (0.1)	197.7 (02.2)	1.1 (0.4)
21	ID_2^A	197.3 (02.7)	1.0 (0.0)	197.7 (02.3)	1.0 (0.1)
22	ID_3^A	176.3 (07.9)	1.1 (0.3)	176.7 (06.5)	1.1 (0.3)

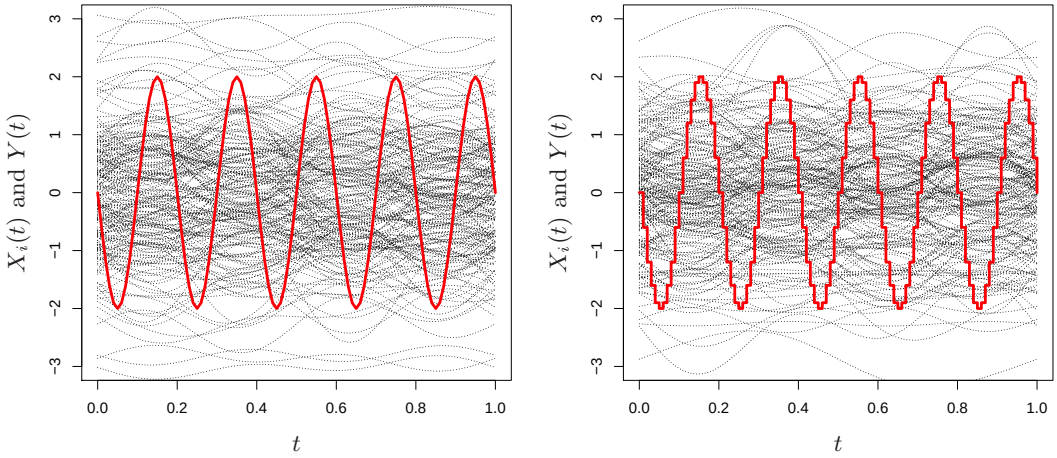
subspace of $\mathcal{C}([0,1])$ spanned by $\sin(2\pi(t-1/2))$ and $\sin(4\pi(t-1/2))$, orthogonal to Y. Thus, any projection of Y onto a function from this subspace is zero. In particular, the projection of Y onto any number of (robust) principal components of a sample from X is close to zero, and Y is mistakenly claimed to be very central w.r.t. X. This example



(A) (B)

FIGURE S.2. *Example S.4: (A) Random sample from P and the outlier Y (red solid); and (B) the first derivatives of these functions (red solid for Y).*

is intended to demonstrate that even though it may appear that the projection into principal components discriminates shape outliers quite well in Tabs. S.2 and S.4, this is true only because of the low-dimensional nature of the random samples. Generally, depths based on the projection into (robust) principal components computed from data are not capable of proper recognition of functions outlying in shape.



(A) (B)

FIGURE S.3. *Random sample from P and the outlier Y (red solid) from Example S.5 (A); and Example S.6 (B).*

Example S.6. Finally, let us conclude the simulation study with a slight modification of Example S.5, demonstrating that the good performance of depths incorporating derivatives in Example S.5 also cannot be guaranteed if the data is observed imperfectly. To see this, take the random function X from Example S.5, and modify Y to be $Y(t) = \text{round}(2\sin(10\pi(t-1/2)), 0.1)$ for $t \in [0,1]$, where $\text{round}(u, 0.1)$ stands for the closest integer multiple of 0.1 to $u \in \mathbb{R}$. A single realization of this setup is in Fig. S.3B. The function Y here is very close to Y from Example S.5, yet is observed at a grid rougher than the rest of the data. As in Example S.3, its derivatives cannot be estimated reliably, and even though the function is clearly different from the sample curves, many depths including derivatives deceivingly rank it quite centrally inside the data. In contrast, order extended depths still perform well, see Tabs. S.6 and S.7.

S.3. Theorem 1: Proof and Discussion

S.3.1. **Proof.** To begin, consider only the first derivative, i.e. $K = 1$, and $J = 2$.

Due to the affine invariance property of the halfspace depth D we can write

(S.21)
$$\begin{aligned} D\left(\begin{pmatrix} x(t) \\ \frac{x(t+h)-x(t)}{h} \end{pmatrix}; P_{\left(X(t), \frac{X(t+h)-X(t)}{h}\right)^\intercal}\right) &= D\left(M(h)\begin{pmatrix} x(t) \\ x(t+h) \end{pmatrix}; P_{M(h)(X(t),X(t+h))^\intercal}\right) \\ &= D\left(\begin{pmatrix} x(t) \\ x(t+h) \end{pmatrix}; P_{(X(t),X(t+h))^\intercal}\right) \end{aligned}$$

for any $h \in \mathbb{R} \setminus \{0\}$ and $t \in [0,1]$, where

$$M(h) = \begin{pmatrix} 1 & 0 \\ \frac{-1}{h} & \frac{1}{h} \end{pmatrix} \tag{S.22}$$

is the non-singular matrix of constants of finite differences corresponding to the zero-th (functional value) and the first derivative. Taking the limit as $h \to 0$ on both sides of (S.21) then yields the assertion of the theorem for $K = 1$. This is due to Mizera and Volauf (2002, Proposition 1), as for the halfspace depth for any $u_\nu \xrightarrow[\nu\to\infty]{} u$ in $\mathbb{R}^d$ and $Q_\nu \xrightarrow[\nu\to\infty]{w} Q$ in $\mathcal{P}\left(\mathbb{R}^d\right)$ it is true that

$$\limsup_{\nu\to\infty} D\left(u_\nu; Q_\nu\right) \leq D\left(u; Q\right),$$

and also

$$\lim_{\nu\to\infty} D\left(u_\nu; Q_\nu\right) = D\left(u; Q\right),$$

if a condition analogous to (14) holds true for the random vector distributed as Q.

The general case of $K \in \mathbb{N}$ follows analogously to that of $K = 1$ apart from the usage of a different non-singular matrix

$$M(h) = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ \frac{-1}{h} & \frac{1}{h} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \binom{K-1}{K-1}\frac{(-1)^{K-1}}{h^{K-1}} & \binom{K-1}{K-2}\frac{(-1)^{K-2}}{h^{K-1}} & \binom{K-1}{K-3}\frac{(-1)^{K-3}}{h^{K-1}} & \dots & \binom{K-1}{0}\frac{(-1)^{0}}{h^{K-1}} & 0 \\ \binom{K}{K}\frac{(-1)^{K}}{h^{K}} & \binom{K}{K-1}\frac{(-1)^{K-1}}{h^{K}} & \binom{K}{K-2}\frac{(-1)^{K-2}}{h^{K}} & \dots & \binom{K}{1}\frac{(-1)^{1}}{h^{K}} & \binom{K}{0}\frac{(-1)^{0}}{h^{K}} \end{pmatrix}$$

instead of (S.22). M is now the matrix of constants of finite differences corresponding to all derivatives of orders 0 (functional value) through K.

S.3.2. **Discussion.** Note that in Theorem 1, the relation of order extended depths with depths including derivatives is described only "locally", i.e. only in terms of the depth D at $t \in [0,1]$ fixed. For general distributions $P \in \mathcal{P}(\mathcal{C}([0,1]))$ and $x \in \mathcal{C}([0,1])$, no global inequalities between $FD_J(x;P)$ and $FD^{(J)}(x;P)$ can be shown. Let us demonstrate this on two simple examples for $J=2$. In Example S.7 we show that the depth FD_J of a function may be arbitrarily close to zero, yet its $FD^{(J)}$ depth may still be high; in Example S.8 we demonstrate the opposite situation.

Example S.7. Let $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ be given $\mathsf{P}(X \equiv 0) = 1/2$, and $\mathsf{P}(X \equiv 1) = 1/2$. For $K \in \mathbb{N}$, split the domain $[0,1]$ into K subintervals $I_k = [(k-1)/K, k/K)$, $k = 1, \ldots, K$. Consider first the function x_K defined by

$$x_K(t) = \begin{cases} k/K & \text{for } t \in I_k, \\ 1 & \text{for } t = 1. \end{cases}$$

As in Examples S.3 and S.6, x_K is discontinuous at k/K and constant on I_k for all $k = 1, \ldots, K-1$. Thus, $0 \le x_K \le 1$, and the derivative of x_K is zero almost everywhere in $[0,1]$. Thus, $FD^{(2)}(x_K;P) = 1/2$, regardless of the value of K. On the other hand, for FD_2 we have that

$$D\left((x_K(t_1), x_K(t_2))^{\mathsf{T}}; P_{(X(t_1),X(t_2))^{\mathsf{T}}}\right) = \begin{cases} 1/2 & \text{if } t_1 \text{ and } t_2 \text{ come from the same } I_k, \\ 0 & \text{elsewhere,} \end{cases}$$

from which we get $FD_2(x_K;P) = 1/(2K)$, vanishing as $K \to \infty$. The functions x_K are constructed discontinuous. Proper convolution of x_K with a mollifier (compactly supported smooth function) yields a continuous modification such that the assertion still holds true, up to small constants.

Example S.8. For $\varepsilon \in (0,1)$, consider $X \sim P \in \mathcal{P}(\mathcal{C}([0,1]))$ given by $\mathsf{P}(X \equiv 1) = \mathsf{P}(X \equiv -1) = \mathsf{P}(X(t) = 1 + \varepsilon t) = \mathsf{P}(X(t) = -1 + \varepsilon t) = 1/4$. For $K \in \mathbb{N}$ such that $K > 1/(2\pi(1-\varepsilon))$, set

$$x_K(t) = \sin(2\pi K t)/(2\pi K), \quad \text{for } t \in [0,1].$$

The joint distribution $(X(t), X'(t))^{\mathsf{T}}$ for any $t \in [0,1]$ gives probability $1/4$ to each of the four vertices of a (possibly degenerate) parallelogram $(1,0)^{\mathsf{T}}$, $(-1,0)^{\mathsf{T}}$, $(1+\varepsilon t, \varepsilon)^{\mathsf{T}}$, $(-1+\varepsilon t, \varepsilon)^{\mathsf{T}}$. Denote this parallelogram by $\Pi(t)$. For x_K, we have $(x_K(t), x_K'(t))^{\mathsf{T}} = (\sin(2\pi K t)/(2\pi K), \cos(2\pi K t))^{\mathsf{T}}$. It is easy to see that $(x_K(t), x_K'(t))^{\mathsf{T}} \in \Pi(t)$ if and only if $\cos(2\pi K t) \in [0, \varepsilon]$, where the latter is equivalent with t lying in a certain subset $C_\varepsilon \subset [0,1]$ of Lebesgue measure $1/2 - \arccos(\varepsilon)/\pi$. For the marginal depth of x_K and its derivative we therefore have

$$D\left((x_K(t), x_K'(t))^{\mathsf{T}}; P_{(X(t),X'(t))^{\mathsf{T}}}\right) = \begin{cases} 1/4 & \text{if } t \in C_\varepsilon, \\ 0 & \text{if } t \notin C_\varepsilon, \end{cases}$$

giving

$$FD^{(2)}(x_K;P) = (1/2 - \arccos(\varepsilon)/\pi)/4 = 1/8 - \arccos(\varepsilon)/(4\pi).$$

This depth value depends on ε, but not on K.

Let us now deal with the depth $FD_2(x_K;P)$. Assume, without loss of generality, that $t_1, t_2 \in [0,1]$ is such that $t_1 < t_2$. This is possible, as the diagonal $t_1 = t_2$ is of zero (two-dimensional) Lebesgue measure, and the depth D is invariant with respect to the permutation of the coordinates on its input.

The joint distribution of $(X(t_1), X(t_2))^\mathsf{T}$ is the uniform distribution on points $(1,1)^\mathsf{T}$, $(-1,-1)^\mathsf{T}$, $(1+\varepsilon\, t_1, 1+\varepsilon\, t_2)^\mathsf{T}$, $(-1+\varepsilon\, t_1, -1+\varepsilon\, t_2)^\mathsf{T}$. Denoting the parallelogram of these four points by $\Pi_2(t_1,t_2)$, we get

$$D\left((x_K(t_1), x_K(t_2))^\mathsf{T}; P_{(X(t_1),X(t_2))^\mathsf{T}}\right) = \begin{cases} 1/4 & \text{if } (x_K(t_1), x_K(t_2))^\mathsf{T} \in \Pi_2(t_1,t_2), \\ 0 & \text{otherwise.} \end{cases} \tag{S.23}$$

The set $\Pi_2(t_1,t_2)$ does not depend on K. On the other hand, $|x_K(t)| \leq 1/(2\pi K)$ for all $t \in [0,1]$, so the vector $(x_K(t_1), x_K(t_2))^\mathsf{T}$ is contained in the centred square $[-1/(2\pi K), 1/(2\pi K)]^2 \subset \mathbb{R}^2$. Take any $(t_1,t_2)^\mathsf{T} \in [0,1]^2$ such that $t_1 < t_2$ and $x_K(t_1) \neq x_K(t_2)$ for all $K \in \mathbb{N}$. We may restrict to such a pair of points because the (two-dimensional) Lebesgue measure of the set $E_K = \left\{(t_1,t_2)^\mathsf{T} \colon x_K(t_1) = x_K(t_2)\right\}$ is zero for each $K \in \mathbb{N}$, and therefore also the set $E = \bigcup_{K=1}^{\infty} E_K$ is of zero Lebesgue measure.

It is easy to verify that if $K \geq 1/(\pi\, \varepsilon\, (t_2 - t_1))$, then for any $t_1 < t_2$ it holds true that $(x_K(t_1), x_K(t_2))^\mathsf{T} \in \Pi_2(t_1,t_2)$ if and only if $x_K(t_1) < x_K(t_2)$. Denoting

$$U_K = \left\{(t_1,t_2)^\mathsf{T} \colon 0 < t_1 < t_2 < 1 \text{ and } x_K(t_1) < x_K(t_2)\right\},$$

(S.23) can be rewritten into

$$D\left((x_K(t_1), x_K(t_2))^\mathsf{T}; P_{(X(t_1),X(t_2))^\mathsf{T}}\right) = \begin{cases} 1/4 & \text{if } (t_1,t_2)^\mathsf{T} \in U_K, \\ 0 & \text{otherwise,} \end{cases}$$

if $t_1 < t_2$ and $K \geq 1/(\pi\, \varepsilon\, (t_2 - t_1))$.

By symmetry considerations, for all $K \in \mathbb{N}$ the Lebesgue measure of U_K must be $1/4$ (since the Lebesgue measure of $\left\{(t_1,t_2)^\mathsf{T} \colon 0 < t_1 < t_2 < 1\right\}$ is $1/2$). Therefore, for $K \to \infty$ we may use the Lebesgue dominated convergence theorem and write

$$\begin{aligned} FD_2(x_K; P) &= 2 \int_0^1 \int_{t_1}^1 D\left((x_K(t_1), x_K(t_2))^\mathsf{T}; P_{(X(t_1),X(t_2))^\mathsf{T}}\right) \mathrm{d}\, t_2\, \mathrm{d}\, t_1 \\ &\approx 2 \iint\limits_{U_K} 1/4\, \mathrm{d}\, t_2\, \mathrm{d}\, t_1 = 1/8 \quad \text{for } K \text{ large.} \end{aligned}$$

By taking $\varepsilon > 0$ small, we see that for K large enough $FD^{(2)}(x_K; P) \approx 0$, yet $FD_2(x_K; P) \approx 1/8$. Consequently, the depth including derivatives may also be very small compared to an order extended depth.

References

Claeskens, G., Hubert, M., Slaets, L., and Vakili, K. (2014). Multivariate functional halfspace depth. *J. Amer. Statist. Assoc.*, 109(505):411–423.

Cuesta-Albertos, J. and Nieto-Reyes, A. (2008a). A random functional depth. In *Functional and operatorial statistics*, Contrib. Statist., pages 121–126. Physica-Verlag/Springer, Heidelberg.

Cuesta-Albertos, J. A. and Nieto-Reyes, A. (2008b). The random Tukey depth. *Comput. Statist. Data Anal.*, 52(11):4979–4988.

Cuevas, A., Febrero, M., and Fraiman, R. (2006). On the use of the bootstrap for estimating functions with functional data. *Comput. Statist. Data Anal.*, 51(2):1063–1074.

Cuevas, A., Febrero, M., and Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Comput. Statist.*, 22(3):481–496.

Cuevas, A. and Fraiman, R. (2009). On depth measures and dual statistics. A methodology for dealing with general data. *J. Multivariate Anal.*, 100(4):753–766.

Dabo-Niang, S., Ferraty, F., and Vieu, P. (2006). Mode estimation for functional random variable and its application for curves classification. *Far East J. Theor. Stat.*, 18(1):93–119.

Donoho, D. L. and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.*, 20(4):1803–1827.

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis. Theory and practice*. Springer Series in Statistics. Springer, New York.

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

Genest, M., Masse, J.-C., and Plante, J.-F. (2012). *depth: Depth functions tools for multivariate analysis*. R package version 2.0-0.

Gijbels, I. and Nagy, S. (2015). Consistency of non-integrated depths for functional data. *J. Multivariate Anal.*, 140:259–282.

Ieva, F. and Paganoni, A. M. (2013). Depth measures for multivariate functional data. *Comm. Statist. Theory Methods*, 42(7):1265–1276.

Kuelbs, J. and Zinn, J. (2013). Concerns with functional depth. *ALEA. Latin American Journal of Probability and Mathematical Statistics*, 10(2):831–855.

Kuelbs, J. and Zinn, J. (2015). Half-region depth for stochastic processes. *J. Multivariate Anal.*, 142:86–105.

López-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *J. Amer. Statist. Assoc.*, 104(486):718–734.

López-Pintado, S. and Romo, J. (2011). A half-region depth for functional data. *Comput. Statist. Data Anal.*, 55(4):1679–1695.

López-Pintado, S., Sun, Y., Lin, J. K., and Genton, M. G. (2014). Simplicial band depth for multivariate functional data. *Adv. Data Anal. Classif.*, 8(3):321–338.

Maechler, M. (2013). *sfsmisc: Utilities from Seminar fuer Statistik ETH Zurich*. R package version 1.0-24.

Mizera, I. and Volauf, M. (2002). Continuity of halfspace depth contours and maximum depth estimators: diagnostics of depth-related methods. *J. Multivariate Anal.*, 83(2):365–388.

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

Mosler, K. and Polyakova, Y. (2012). General notions of depth for functional data. *arXiv preprint arXiv:1208.1981*.

Nagy, S. (2012). Nonparametric classification of noisy functions. In Komárek, A. and Nagy, S., editors, *Proceedings of the 27th International Workshop on Statistical Modelling*, pages 234–239 (Part I.).

Nagy, S. (2015). Consistency of h-mode depth. *J. Statist. Plann. Inference*, 165:91–103.

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

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rudin, W. (1987). *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition.

Tukey, J. W. (1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians (Vancouver, B. C., 1974), Vol. 2*, pages 523–531. Canad. Math. Congress, Montreal, Que.