Model-based statistical depth with applications to functional data

Weilong Zhao, Zishen Xu, Yue Mu, Yun Yang & Wei Wu

Published online: 21 Jun 2023.

Submit your article to this journal ☑

View related articles ☑

View Crossmark data ☑

Check for updates

Model-based statistical depth with applications to functional data

Weilong Zhao[a], Zishen Xu[a], Yue Mu[a], Yun Yang[b] and Wei Wu[a]

[a]Department of Statistics, Florida State University, Tallahassee, FL, USA; [b]Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL, USA

ABSTRACT

Statistical depth, a commonly used analytic tool in nonparametric statistics, has been extensively studied for multivariate and functional observations over the past few decades. Although various forms of depth were introduced, they are mainly procedure based whose definitions are independent of the generative model for observations. To address this problem, we introduce a generative model-based approach to define statistical depth for both multivariate and functional data. The proposed model-based depth framework permits simple computation via a bootstrap sampling and improves the depth estimation accuracy. When applied to functional data, the proposed depth can capture important features such as continuity, smoothness or phase variability, depending on the defining criteria. We propose efficient algorithms to compute the proposed depths and establish estimation consistency. Through simulations and real data, we demonstrate that the proposed functional depths reveal important statistical information such as those captured by the median and quantiles, and detect outliers.

1. Introduction

The notion of statistical depth was first introduced (Tukey 1975) as a tool to visualise bivariate data sets and has later been extended to multivariate data over the last few decades. The depth is a measure of the centrality of a point with respect to certain data cloud, which helps to set up centre-outward ordering rules of ranks. Alternatively, it can be treated as a multivariate extension of the notion of quantiles for univariate distributions. For instance, a deepest point in a given data cloud can be viewed as a 'multivariate median'. Based on different criteria on centrality, a large class of depths has been proposed, including the halfspace depth (Tukey 1975), convex hull peeling depth (Barnett 1976), simplicial depth (Liu 1990), L_1-depth (Vardi and Zhang 2000), projection depth (Zuo 2003) and Monge–Kantorovich depth (Chernozhukov, Galichon, Hallin, and Henry 2017). The concept of statistical depth has been widely applied in outlier detection (Donoho and Gasko 1992), multivariate density estimation (Fraiman, Liu, and

CONTACT Wei Wu ✉ wwu@stat.fsu.edu

Meloche 1997), nonparametric description of multivariate distributions (Liu, Parelius, and Singh 1999), and depth-based classification and clustering (Christmann 2002).

In many research areas such as medicine, biology and engineering, it is natural to assume the observations being generated from infinite dimensional models and analyse them using tools from functional data analysis (FDA). Many efforts have been attempted to extend the notion of depths from finite to infinite dimensions in recent years. To name a few, Fraiman and Muniz (2001) defined the integrated data depth for functional data based on integrals of univariate depths and used it to construct an α-trimmed functional mean to measure the centrality of given data. This method can reduce the effects of outlier bias in a sample set compared to the regular mean. In addition, Cuesta-Albertos and Nieto-Reyes (2008) extended the simple random Tukey depth (also called halfspace depth) to functional data analysis on a separable Hilbert space. A more comprehensive review on different notions of depths for functional data is provided in Section 1.1.

Despite the broad variety and wide usage of statistical depths for both finite and infinite dimensional observations in exploratory data analysis, existing depth methods suffer from two apparent drawbacks: (1) They do not make use of any structural information from the generative model when defining or estimating the depths. Utilising such information may enhance the power of the depth in tasks such as hypothesis testing, outlier detection or classification. (2) For infinite-dimensional observations such as functional data, most depths are constructed via aggregating point-wise deviations, which fails to capture deviations of some more important global features such as phase variability and degree of smoothness.

In this paper, we propose a new model-based framework for defining and estimating statistical depths for both finite and infinite-dimensional data. In particular, we propose to incorporate information from the data generative model in defining and estimating the statistical depth. When applied to functional data, our development leads to a new class of depths that captures global features such as shape and smoothness level. Our new model-based depth framework overcomes the aforementioned drawbacks and possesses several attractive features:

(1) It permits properly utilising features in the generative model to define a data-dependent depth. Both computational efficiency and estimation accuracy of the depth can be benefited from the generative model via bootstrap sampling.
(2) The depth criterion is flexible and can be chosen to better capture the underlying generative mechanism or meet specific application purposes. Depending on the defining criterion, our framework can result in various forms and generalises commonly used depth functions.
(3) The criterion may properly measure the metric distance between observations. This naturally leads to the notions of centrality and variability in the given data. In contrast, traditional depth methods are often procedure based and do not provide such measurements.

1.1. Related work on functional depth

Band depth (López-Pintado and Romo 2009) is a very commonly used depth for functional data, which has been successfully used for tasks such as classification. Another important concept is half-region depth (López-Pintado and Romo 2011), which is closely related

to the band depth. It is considered to be applied to high-dimensional data with efficient computational cost. Based on the graph representation as in band depth, a number of extensions, modifications and generalisations have emerged. For example, Agostinelli and Romanazzi (2013) proposed a so-called local band depth to deal with functional data which is considered to have multiple centres. It measures centrality conditional on a neighbourhood of each point of the space and provides a tool that is sensitive to local features of the data, while retaining most features of regular depth functions. Set band depth (Whitaker, Mirzargar, and Kirby 2013) was proposed for the nonparametric analysis of random sets and a generalisation of the method of band depth. Agostinelli (2018) proposed several modified versions of half-region depth to capture more local features. They also established similarity measures based on depth value for hierarchical clustering. Balzanella and Elvira (2015) introduced the spatial variability among the curves in the definition of band depth and proposed a method – spatially weighted band depth to incorporate the spatial information in the curves ordering. Dutta, Sarkar, and Ghosh (2016) introduced localised spatial depth to do multi-classification on high dimensional data.
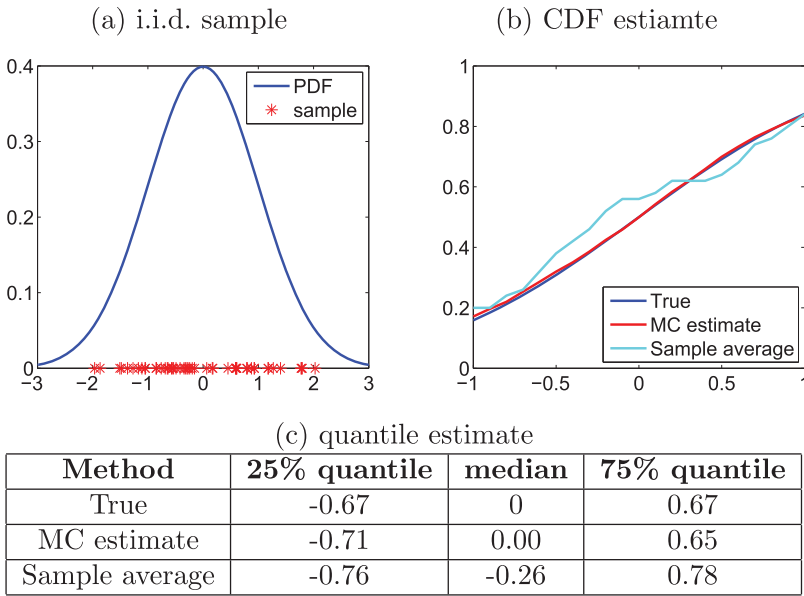
More progress has been made in recent study of functional depth. Chakraborty and Chaudhuri (2014b) used the spatial distribution to define a so-called spatial depth, since the spatial distribution possesses an invariance property under a linear affine transformation. Einmahl, Li, and Liu (2015) proposed to refine the empirical halfspace depth by setting extreme value to a so-called 'tail' to avoid the problem of vanishing value outside the convex hull of the data, which benefits for inference on extremity. Narisetty and Nair (2016) introduced a notion called extremal depth, which satisfies the desirable properties of convexity and 'null at the boundary', for which integrated data depth and band depth lack. These properties lead to a central region more resistant to outliers. Based on an elastic-metric-based measure of centrality for functional data, Cleveland, Zhao, and Wu (2018) adopted band depth and modified band depth to estimate the template for functional data with phase variability. They also showed their performance on outlier detection with new defined boxplots for time warping functions.

The rest of this article is organised as follows: in Section 2, we first introduce our model-based framework for statistical depth. We then illustrate several forms of depth and their relations to commonly used depths. In Section 3, we elaborate on the application of our framework to functional data as generated from a second-order stochastic process. In Section 4, we investigate the statistical consistency of our depth estimation procedure. Simulations and real data analysis are provided in Section 5. Section 6 includes a summary and discusses some future directions. Other computational details and proofs are deferred to appendices.

2. Model-based statistical depth

In this section, we introduce our model-based framework for statistical depth, where the model-based has two meanings: (1) the depth is defined based on a statistical model and (2) the depth estimation procedure is two stage, where we first estimate the model parameter and then use a plug-in procedure for estimating the depth. The former view allows the depth definition itself to be data dependent and automatically capture features underlying the data generating process, and the latter may lead to improved estimation accuracy of the depths due to the estimation efficiency of the model parameter.

(a) i.i.d. sample (b) CDF estiamte

(c) quantile estimate

Method	25% quantile	median	75% quantile
True	-0.67	0	0.67
MC estimate	-0.71	0.00	0.65
Sample average	-0.76	-0.26	0.78

Figure 1. Toy example to compare the Monte Carlo method and sample average: (a) 30 i.i.d. sample points from a standard normal distribution; (b) cumulative distribution functions of true model (blue), estimated using Monte Carlo method (red), and estimated using the sample average (cyan) in the range $[-1, 1]$; (c) quantile values at 25%, 50% and 75% of true and two estimate methods.

To illustrate the benefit in estimation accuracy via model-based procedures, we may compare the Monte Carlo (MC) method with the simple sample average approach. A toy example is shown in Figure 1, where we generate 30 i.i.d. sample points from a standard normal distribution (Figure 1 a). In the MC method, we estimate mean and standard deviation from the sample and then generate 2000 Monte Carlo sampling points to estimate the cumulative distribution within $[-1, 1]$. In contrast, the sample average method estimates the cumulative distribution with the empirical distribution of the 30 points. This comparison is shown in Figure 1(b). Moreover, we compare the true and estimated quantiles at 25%, 50% and 75% in Figure 1(c). It is apparent that the MC method provides more accurate and robust result.

To begin with, we provide a general definition of depth by considering it as a functional of the underlying data generating model. Then, we provide a two-stage estimation procedure for the depth via a Monte Carlo or bootstrap sampling. In the rest of this paper, we primarily focus on functional data for illustration, and the development naturally applies to finite-dimensional data.

2.1. Depths within statistical models

Let $\mathcal{P} = \{\mathbb{P}_\theta : \theta \in \Theta\}$ be a family of probability measures indexed by a parameter θ over a function (vector) space $\mathcal{F} \subset \mathcal{L}_2([0, 1]) := \{f : [0, 1] \to \mathbb{R} : \|f\|_2^2 = \int_0^1 f^2(x) \, dx < \infty\}$. For example, $\mathbb{P}_\theta$ can be the measure of a Gaussian Process $\mathrm{GP}(m, C)$ with parameter $\theta = (m, C)$ collecting the mean function $m : [0, 1] \to \mathbb{R}$ and the covariance function $C : [0, 1] \times [0, 1] \to \mathbb{R}$. Statistical depth should quantify how large a particular observed

trajectory $f_{obs} \in \mathcal{F}$ deviates from certain notion of centre $f_c \in \mathcal{F}$ under $\mathbb{P}_\theta$. For example, in the case of the Gaussian Process (GP), a natural choice of the centre would be its mean function.

2.1.1. Definitions of depths

We will now provide the formal definition of a model-based functional depth, as well as the associated depth contour and central region. All these statistical terms can be considered as infinite-dimensional generalisation of the uni-variate survival function/p-value, quantiles and highest-probability region.

Our proposed definition can be either norm based or inner-product based. We refer to the norm or inner product as the criterion in the definition. The norm-based depth is a generalisation over various distance-based forms (see the discussion after the following definition). In contrast, the inner-product depth is motivated with the classical halfspace depth by Tukey (1975). We at first define the norm-based depth in the following general form:

Definition 2.1 (Norm-based statistical depth (general form)): The statistical depth D_{ng} of $f_{obs} \in \mathcal{F}$ in the model $\mathbb{P}_\theta \in \mathcal{P}$ relative to the norm $\| \cdot \|$ and centre $f_c \in \mathcal{F}$ is defined as

$$D_{ng}(f_{obs}, \mathbb{P}_\theta, \| \cdot \|, f_c) \in [0, 1],$$

where D_{ng} is strictly decreasing with respect to $\|f_{obs} - f_c\|$ and $D_{ng} \to 0$ when $\|f_{obs} - f_c\| \to \infty$.

Norm-based depths are commonly used in statistics literature. For example, the h-depth (Nieto-Reyes 2011) and spatial depth (Sguera, Galeano, and Lillo 2014) are based on the $\mathbb{L}^2$ norm, the $\mathbb{L}^p$-depth is based on the $\mathbb{L}^p$ norm (Zuo and Serfling 2000; Long and Huang 2015) and the Mahalanobis depth is based on the Mahalanobis norm (Liu et al. 1999). The depth in Definition 2.1 generalises these concepts and provides a broader framework for norm-based methods. In this paper, we study one specific form of the this general definition. This specific form more resembles conventional depths and satisfies more desirable mathematical properties. The norm in the definition can be considered as a criterion function to compute the distance between any observation f_{obs} and the centre f_c and we denote the criterion function as $\zeta(f, f_c)$ in the rest.

Definition 2.2 (Norm-based statistical depth (specific form)): The statistical depth D_n of $f_{obs} \in \mathcal{F}$ in model $\mathbb{P}_\theta \in \mathcal{P}$ relative to norm $\| \cdot \|$ and centre $f_c \in \mathcal{F}$ is defined as

$$D_n(f_{obs}, \mathbb{P}_\theta, \| \cdot \|, f_c) := \mathbb{P}_\theta \left[f \in \mathcal{F} : \|f - f_c\| \geq \|f_{obs} - f_c\| \right].$$

Remark 2.1: We point out that this specific form of depth is a representative of all norm-based depth in the general form as defined in Definition 2.1. In fact, $D_n(f_{obs})$ measures the degree of extremeness of the observed function $f_{obs} \in \mathcal{F}$ under any normal-based depth $D_{ng}(f_{obs})$ in the following sense,

$$\mathbb{P}_\theta \left[D_{ng}(f) \leq D_{ng}(f_{obs}) \right] = \mathbb{P}_\theta \left[D_n(f) \leq D_n(f_{obs}) \right]$$
$$= \mathbb{P}_\theta \left[\|f - f_c\| \geq \|f_{obs} - f_c\| \right] = D_n(f_{obs}).$$

Remark 2.2: One proper way to choose the centre f_c is to minimise $P(\|f - f_c\| \geq a)$ for any given $a > 0$. Note that

$$P(\|f - f_c\| \geq a) \leq \frac{E\|f - f_c\|^2}{a^2}.$$

When the norm $\| \cdot \|$ is inner product induced (e.g. the classical $\mathbb{L}^2$ norm), it is easy to know that the optimal f_c should be the expectation Ef. However, f_c in general can take different form, dependent on different selection of the norm.

Based on the definitions of the norm-based depth, we can naturally introduce the notions of depth contour and central region as follows. We adopt the specific form in Definition 2.2 to simplify notation (same notion can be directly applied to the general form).

Definition 2.3 (Depth contour and central region for norm-based depth): For any $\alpha \in [0, 1]$, the αth depth contour in the model $\mathbb{P}_\theta \in \mathcal{P}$ relative to the norm $\| \cdot \|$ and centre $f_c \in \mathcal{F}$ is defined as

$$C_n(\alpha, \mathbb{P}_\theta, \| \cdot \|, f_c) := \left\{ f \in \mathcal{F} : D_n(f, \mathbb{P}_\theta, \| \cdot \|, f_c) = \alpha \right\}.$$

Also, the αth central region in the model $\mathbb{P}_\theta \in \mathcal{P}$ relative to the norm $\| \cdot \|$ and centre $f_c \in \mathcal{F}$ is defined as

$$R_n(\alpha, \mathbb{P}_\theta, \| \cdot \|, f_c) := \left\{ f \in \mathcal{F} : D_n(f, \mathbb{P}_\theta, \| \cdot \|, f_c) \geq \alpha \right\}.$$

Based on the multivariate halfspace depth, we now define the inner-product-based depth. In contrast to the general and specific forms in the norm-based case, the inner-product-based norm is defined only in a specific form as follows.

Definition 2.4 (Inner-product-based statistical depth): The statistical depth D_{ip} of $f_{obs} \in \mathcal{F}$ in the model $\mathbb{P}_\theta \in \mathcal{P}$ relative to the inner product $< \cdot, \cdot >$ and a subset $\mathcal{G}$ of $\mathcal{F}$ is defined as

$$D_{ip}(f_{obs}, \mathbb{P}_\theta, \langle \cdot, \cdot \rangle, \mathcal{G}) := \inf_{g \in \mathcal{G}, \|g\| = 1} \mathbb{P}_\theta \left[f \in \mathcal{F} : \langle f, g \rangle \geq \langle f_{obs}, g \rangle \right]$$

Remark 2.3: There are two apparent differences between Definitions 2.2 and 2.4: (1) Definition 2.2 depends on the centre f_c, whereas Definition 2.4 is independent of it. However, we will point out in Section 2.1.2 that when the distribution function has a centre under certain form of symmetry, this centre should be the deepest point under Definition 2.4. (2) Definition 2.4 involves an infimum to match the half-space depth (Tukey 1975) for finite-dimensional Euclidean data. Different from the usual half-space depth where $\mathcal{G}$ as the range of the infimum is taken as the entire function space $\mathcal{F}$, the following lemma shows that for infinite-dimensional functional data, $\mathcal{G}$ is necessary to be a proper, finite-dimensional subset to avoid depth value degeneracy. A proof is provided in Appendix 5.

Lemma 2.1: *Let $\mathbb{P}_C$ be the probability measure of a zero-mean Gaussian process $GP(0, C)$, where the eigensystem $\{(\lambda_j, \phi_j)\}_{j=1}^{\infty}$ of the covariance operator C has infinite number of positive eigenvalues $\{\lambda_j\}_{j=1}^{\infty}$. If $\langle \cdot, \cdot \rangle$ is an inner product over $\mathcal{F}$ such that the $P \times P$ Gram matrix $[\langle \phi_j, \phi_k \rangle]_{j,k=1}^{P}$ of the first P eigenfunctions $\{\phi_j\}_{j=1}^{P}$ is positive definite for any $P \in \mathbb{N}$, then*

$$D_{ip}(f, \mathbb{P}_C, \langle \cdot, \cdot \rangle, \mathcal{F}) = 0$$

almost surely for $f \in GP(0, C)$.

Remark 2.4: This lemma indicates that special attention is needed for defining an inner-product-based depth for infinite-dimensional space $\mathcal{F}$. Dutta, Ghosh, and Chaudhuri (2011) also observed this anomalous behaviour of halfspace depth in infinite-dimensional spaces. As a consequence, the halfspace depth (where $\mathcal{G} = \mathcal{F}$) is only meaningful for finite-dimensional space. In contrast, the norm-based depth can be effective for both finite- or infinite-dimensional space. To have a proper inner-product-based depth, either $\mathcal{F}$ itself is finite-dimensional, or we use a finite-dimensional subset $\mathcal{G}$ as shown in Definition 2.4.

Under this model-based framework, we can naturally estimate the proposed statistical depth $D(f_{obs}, \mathbb{P}_\theta, \cdot, f_c)$ via the following two-stage procedure: (1) find an estimate $\widehat{\theta}$ of the parameter θ and (2) compute the estimated depth $D(f_{obs}, P_{\widehat{\theta}}, \cdot, f_c)$ by either using a closed-from expression of the depth or by a bootstrap method for an approximation. For example, when $\mathbb{P}_\theta$ is a GP measure and the depth as a functional of parameter θ may not admit a closed-form expression, we may resort to bootstrap method for estimating the depth. More details of the estimation will be provided in Appendix 1.

2.1.2. Mathematical properties

Zuo and Serfling (2000) introduced a list of favourable mathematical properties to be satisfied by good multivariate statistical depths. Based on this, Nieto-Reyes and Battey (2016) further explored the extensions of these properties for functional data. Gijbels and Nagy (2017) discussed these properties on commonly used methods such as the random Tukey depth, band depth and spatial depth. In this part, we discuss these properties on our norm-based and inner-product-based depths.

Before discussing basic properties of these two types of depths, we need to clarify the concept of 'halfspace' with the following definition:

Definition 2.5: A closed halfspace $H_{h,g}$ for $g, h \in \mathcal{F}$ is defined in the form

$$H_{h,g} = \left\{ f \in \mathcal{F} : \langle f - h, g \rangle \geq 0 \right\}.$$

To make the inner-product-based depth satisfy favourable properties, we need the following assumption on the ' centre' function f_c.

Assumption 2.1: *The distribution $\mathbb{P}_\theta$ of a random function $f \in \mathcal{F}$ is halfspace symmetric, or H-symmetric, about a unique function f_c. That is, $\mathbb{P}(f \in H) \geq 1/2$ for every closed halfspace H containing f_c. Moreover, we assume that $\mathbb{P}(f \in H) < 1/2$ for every closed halfspace H that does not contain f_c.*

Now we list four basic properties of the norm-based depth (Definition 2.2) and the inner-product-based depth (Definition 2.4), respectively, as follows:

Norm-based depth:

P-1. (*Linear invariance*) Let $\mathbb{P}_{\theta,F}$ denote the distribution P_θ of a random variable $F \in \mathcal{F}$. Then for any $a \in \mathbb{R} \setminus \{0\}$ and $h \in \mathcal{F}$,

$$D(af_{obs} + h, \mathbb{P}_{\theta,aF+h}, \| \cdot \|, af_c + h) = D(f_{obs}, \mathbb{P}_{\theta,F}, \| \cdot \|, f_c).$$

P-2. (*Maximality at centre*) $D(f_c, \mathbb{P}_\theta, \| \cdot \|, f_c) = \sup_{f_{obs} \in \mathcal{F}} D(f_{obs}, \mathbb{P}_\theta, \| \cdot \|, f_c)$.

P-3. (*Monotonicity with respect to the deepest point*) Let the deepest function be $f_c \in \mathcal{F}$. Then for any $f_{obs} \in \mathcal{F}$ and $\alpha \in (0,1)$, $D(f_{obs}, \mathbb{P}_\theta, \| \cdot \|, f_c) \leq D(f_c + \alpha(f - f_c), \mathbb{P}_\theta, \| \cdot \|, f_c)$.

P-4. (*Vanishing at infinity*) $D(f_{obs}, \mathbb{P}_\theta, \| \cdot \|, f_c) \to 0$ as $\|f_{obs}\| \to \infty$.

Inner-product-based depth:

P-1'. (*Linear invariance*) Let $\mathbb{P}_{\theta,F}$ denote the distribution P_θ of a random variable $F \in \mathcal{F}$. Then for any $a \in \mathbb{R} \setminus \{0\}$ and $h \in \mathcal{F}$,

$$D_{ip}(af_{obs} + h, \mathbb{P}_{\theta,aF+h}, \langle \cdot, \cdot \rangle, \mathcal{G}) = D_{ip}(f_{obs}, \mathbb{P}_{\theta,F}, \langle \cdot, \cdot \rangle, \mathcal{G}).$$

P-2'. (*Maximality at centre*) $D_{ip}(f_c, \mathbb{P}_\theta, <\cdot, \cdot>, \mathcal{G}) = \sup_{f_{obs} \in \mathcal{F}} D_{ip}(f_{obs}, \mathbb{P}_\theta, <\cdot, \cdot>, \mathcal{G})$.

P-3'. (*Monotonicity with respect to the deepest point*) Let the deepest function be $f_c \in \mathcal{F}$. Then for any $f_{obs} \in \mathcal{F}$ and $\alpha \in (0,1)$, $D_{ip}(f_{obs}, \mathbb{P}_\theta, <\cdot, \cdot>, \mathcal{G}) \leq D_{ip}(f_c + \alpha(f - f_c), \mathbb{P}_\theta, <\cdot, \cdot>, \mathcal{G})$.

P-4'. (*Vanishing at infinity*) $D_{ip}(f_{obs}, \mathbb{P}_\theta, <\cdot, \cdot>, \mathcal{G}) \to 0$ as $<f_{obs}, f_{obs}> \to \infty$.

We examine these mathematical properties of the three defined depths in Section 2.1.1, as summarised in Lemma 2.2. The detailed proof is given in Appendix 7.

Lemma 2.2: *The three depths in Definitions 2.1, 2.2, and 2.4 satisfy the mathematical properties given below:*

(1) *Norm-based depth in general form (Definition 2.1): P-2, P-3, P-4.*
(2) *Norm-based depth in specific form (Definition 2.2): P-1, P-2, P-3, P-4.*
(3) *Inner-product-based depth (Definition 2.4, given Assumption 2.1): P-1', P-2', P-3', P-4'.*

As comparison, the band depth (López-Pintado and Romo 2009) and half-region depth (López-Pintado and Romo 2011) for functional data both satisfy the linear invariance property and vanish when norm tends to infinity, but do not hold a monotonicity property. For random elements from well-known stochastic processes such as Brownian bridges, both of them will be degenerated at zero (Chakraborty and Chaudhuri 2014a; Kuelbs and Zinn 2015). For another widely discussed concept of the spatial depth (Sguera et al. 2014), it is observed that it does not decrease monotonically with respect to the deepest point (Nagy 2017). In contrast, a norm-based L^∞ depth (Long and Huang 2015) satisfies all of the desirable properties of P-1, P-2, P-3 and P-4.

2.2. Illustration of the depth definitions

We have defined two forms of model-based functional depth – norm-based (as in Definitions 2.1 and 2.2) and inner-product-based (as in Definition 2.4). In this section, we provide some examples, both finite-dimensional and infinite-dimensional, to illustrate these definitions. We will at first adopt various norms in D_n and then demonstrate the inner-product-based definition. Using these depths one can rank functional data based on their amplitude, continuity, smoothness or phase variability. Moreover, we will show that some of the functional depths can also be directly applied to multivariate data.

2.2.1. Norm-based depth

There are various norms on functional variables. One commonly used is the classical $\mathbb{L}^p$-norm, with $p \geq 1$. That is, for f in a proper space, its $\mathbb{L}^p$-norm is

$$\|f\|_p = \left(\int_0^1 |f(t)|^p \, \mathrm{d}t \right)^{1/p}.$$

In particular, $\mathbb{L}^2$-norm, the Euclidean distance from 0, is most often used in functional data analysis. Due to the nature of $\mathbb{L}^2$ norm, it is a great tool for data visualisation and ranking based on their own amplitude information.

Considering functions in a Sobolev Space (Hsing and Eubank 2015), we can also use $\mathbb{L}^p$ norm on the derivative functions to quantify continuity or smoothness feature. We may consider the norm-based depth in the following two forms:

(1) $D_n(f_{obs}, \mathbb{P}_\theta, \|\cdot\|, f_c) := \mathbb{P}_\theta[f \in \mathcal{F} : \|f - f_c\|_p \geq \|f_{obs} - f_c\|_p]$

(2) $D_n(f_{obs}, \mathbb{P}_\theta, \|\cdot\|, f_c) := \mathbb{P}_\theta[f \in \mathcal{F} : \|D^r f - D^r f_c\|_p \geq \|D^r f_{obs} - D^r f_c\|_p]$, where D^r indicates rth-order differentiation.

When we adopt the $\mathbb{L}^p$ norm, the resulting depth can approximate band depth (López-Pintado and Romo 2009) for functional observations from a distribution with mean 0.

In addition to having variability in amplitude (characterised by $\mathbb{L}^p$ norms), functional observations often exhibit variability in phase. Such variability has been extensively studied over the past two decades and various methods were proposed to separate phase and amplitude, and quantify each variability in the given data (Ramsay and Li 1998; Liu and Müller 2004; Tang and Müller 2008; Cleveland et al. 2018). In particular, phase is represented with time warping functions – Let Γ be the set of orientation-preserving diffeomorphisms of the unit interval $[0,1]$: $\Gamma = \{\gamma : [0,1] \to [0,1] | \gamma(0) = 0, \gamma(1) = 1, \dot{\gamma} > 0\}$ (the dot indicates derivative operation), and γ is called a *warping function*. Given two functions u, v, we denote γ_{uv} as the optimal warping from u to v. There are various forms to define the 'optimal' warping, and here we adopt the well-known Fisher–Rao framework (Srivastava, Wu, Kurtek, Klassen, and Marron 2011) and

$$\gamma_{uv} = \mathrm{arginf}_{\gamma \in \Gamma} \|(\dot{u} \circ \gamma)\sqrt{\dot{\gamma}} - \dot{v}\|_2$$

where ° denotes the function composition. The degree of warpingness from the identity $\gamma_{id}(t) = t$ can be properly measured by two distances, namely, the $\mathbb{L}^2$ distance and the Fisher–Rao distance. We may consider the norm criterion based on each of these distances:

(1) $D_n(f_{obs}, \mathbb{P}_\theta, \| \cdot \|, f_c) := \mathbb{P}_\theta[f \in \mathcal{F} : \|\gamma_{ff_c} - \gamma_{id}\|_2 \geq \|\gamma_{f_{obs}f_c} - \gamma_{id}\|_2]$

(2) $D_n(f_{obs}, \mathbb{P}_\theta, \| \cdot \|, f_c) := \mathbb{P}_\theta[f \in \mathcal{F} : d_{FR}(\gamma_{ff_c}, \gamma_{id}) \geq d_{FR}(\gamma_{f_{obs}f_c}, \gamma_{id})]$, where $d_{FR}(\gamma_{uv}, \gamma_{id}) = \cos^{-1}(\int_0^1 \sqrt{\dot{\gamma}_{uv}(t)} \sqrt{\dot{\gamma}_{id}(t)}\, dt) = \cos^{-1}(\int_0^1 \sqrt{\dot{\gamma}_{uv}(t)}\, dt)$.

Due to the nature of the Fisher–Rao distance, depth based on this criteria in our framework is sensitive to smoothness in the warping function.

2.2.2. Inner-product-based depth

For multivariate data, Tukey's halfspace depth (Tukey 1975) is one of the most popular depth functions available in the literature. Dutta et al. (2011) investigated an extension on any Banach space and proposed a specialisation on a Hilbert space $\mathcal{H}$. Suppose $\mathbf{X}$ is a random element in $\mathcal{H}$ having the distribution F, then the *halfspace depth* of an observation $x \in \mathcal{H}$ is defined as

$$HD(x, F) = \inf_{h \in \mathcal{H}} P\{\langle h, \mathbf{X} - x \rangle \geq 0\},$$

where $< \cdot, \cdot >$ stands for the inner product defined on $\mathcal{H}$. Note that the inner-product-based depth in Definition 2.4 can be rewritten as

$$D_{ip}(f_{obs}, \mathbb{P}_\theta, \langle \cdot, \cdot \rangle, \mathcal{G}) := \inf_{g \in \mathcal{G}, \|g\|=1} \mathbb{P}_\theta\left[f \in \mathcal{F} : \langle f, g \rangle \geq \langle f_{obs}, g \rangle\right]$$

$$= \inf_{g \in \mathcal{G}, \|g\|=1} \mathbb{P}_\theta\left[\langle f - f_{obs}, g \rangle \geq 0\right].$$

Therefore, the halfspace depth can be treated as one special case in the proposed framework. However, Lemma 2.1 illustrates that the halfspace depth may collapse to zero for infinite-dimensional functional data unless the underlying data generating model is intrinsically finite-dimensional. As a consequence, the choice of the range $\mathcal{G}$ of the infimum in the preceding display becomes important.

In general, there is no simple solution to the above minimisation process (Tukey 1975; Rousseeuw and Ruts 1996; Dutta et al. 2011). However, if the functions are samples from a finite-dimensional stochastic process, an optimal g can be found in closed forms. For illustration purpose, let us assume that the data generating process is a finite-dimensional Gaussian process, where the optimal g takes the following closed form (see detailed derivation in Appendix 6),

$$D_{ip}(f_{obs}) = 1 - \Phi(\|f_{obs}\|_{\mathbb{H}_K}),$$

where Φ is the c.d.f. of a standard normal random variable and the norm $\| \cdot \|_{\mathbb{H}_K}$ is the induced RKHS norm (formal definitions are provided in Section 3). Note that as Φ is the a c.d.f. function, the depth value of f_{obs} is in the range $[0, 1/2]$, which is consistent to the notion of halfspace depth in function space (Dutta et al. 2011). It is well known that, for the halfspace depth, if we have a symmetric distribution in a Hilbert space, then the maximum depth is 0.5, and the point of symmetry will achieve at the halfspace median. In this case, it is easy to see that $D_{ip}(f_{obs}) = 1/2 \Leftrightarrow f_{obs} = 0$. Therefore, the median (i.e. function with largest depth value) is our centre function $f_c = 0$.

Remark 2.5: The above result is based on the assumption that the stochastic process is a Gaussian process. However, the Gaussianity is only used in the step that the c.d.f. Φ is independent of g after the standardisation (i.e. $X \to \frac{X - \mu_g}{\sigma_g}$), and the results can be generalised to any second-order stochastic process.

Simplifications in Multivariate Data: The above inner-product-based depth can also be applied to multivariate data where the Gaussian process reduces to a multivariate Gaussian distribution, denoted as $\mathcal{N}(\mu, \Sigma)$. In particular, the corresponding inner-product criterion function reduces to a variant of the well-known Tukèy's *halfspace* depth, or location depth (Tukey 1975),

$$D_{ip}(x) = \inf_{u \in \mathbb{R}^d, \|u\|=1} P\{X : \langle u, X - x \rangle \geq 0\},$$

where the new halfspace depth incorporates the second moment information Σ through the (zero-mean) inner product $< x, y >= x^T \Sigma^{-1} y^T$ and the norm $\|x\|^2 = x^T \Sigma^{-1} x$ induced from the covariance matrix of the multivariate data generating distribution, and $\mathcal{G}$ becomes the unit ball of $\mathbb{R}^d$ relative to this inner product.

In the special case, when X is a random realisation from a zero-mean multivariate normal distribution, or $X \sim \mathcal{N}(0, \Sigma)$, then the halfspace depth D_{ip} admits a closed form. More concretely, using the singular value decomposition on the covariance matrix $\Sigma = U \Lambda U^T$, where Λ is a diagonal matrix with elements of eigenvalues $\{\lambda_p\}_{p=1}^d$, we can express X through a finite-dimensional version of the Karhunen Loève expansion $X = \sum_{p=1}^d \xi_p U_p$, where U_p is the pth column of U (i.e. the eigenvector corresponding to λ_p), and $\xi_p \sim \mathcal{N}(0, \lambda_p)$, $p = 1, 2, \ldots, d$ are independent random variables. Correspondingly, the depth of any $x \in \mathbb{R}^d$ is given as

$$D_{ip}(x) = 1 - \Phi\left(\sqrt{\sum_{p=1}^d \frac{\xi_p^2}{\lambda_p}}\right).$$

Note that the maximum depth value computed by this way is the same as maximum via Tukey's half space depth, which is 1/2.

3. A new model-based depth for functional data

In this section, we apply our proposed depth framework to functional data and propose a new data-dependent functional depth. As we will illustrate, our new model-based functional depths can capture and adapt to global features such as smoothness and shapes in the underlying data generating processes. Our proposed methods incorporate information from the reproducing kernel Hilbert space (RKHS) associated with the covariance operator of the underlying stochastic process.

3.1. Depths induced by reproducing kernels

We will provide a construction of norm-based depth for zero-mean second-order stochastic processes $\mathcal{F}$, where the norm itself is model dependent and learned from the data. Recall

that a stochastic process $\{f(t) : t \in [0,1]\}$ is a second-order process if $\mathbb{E}[f^2(t)] < \infty$ for all $t \in [0,1]$, so that its covariance function $\mathbb{E}[f(s)f(t)]$ is well defined. If the process has a nonzero mean function m, then we can always subtract the mean by choosing the centre f_c as m.

3.1.1. Background on covariance kernels

Since $\{f(t) : t \in [0,1]\}$ is a second-order process, its covariance kernel $K \in [0,1] \times [0,1] \mapsto \mathbb{R}, K(s,t) := \mathbb{E}[f(s)f(t)]$ is a well-defined function for all $(s,t) \in [0,1]^2$. In addition, $K(\cdot,\cdot)$ is a symmetric, positive semi-definite real-value function, that is,

$$\text{(i)} \quad K(s,t) = K(t,s),$$

$$\text{(ii)} \quad \int_0^1 \int_0^1 K(s,t)\,h(s)\,h(t)\,\mathrm{d}s\,\mathrm{d}t \geq 0 \quad \text{for any } h \in \mathcal{F}.$$

According to Mercer's Theorem (J Mercer 1909), there exists a sequence of orthonormal eigenfunctions $\{\phi_1(t), \phi_2(t), \ldots\}$ over $[0,1]$ and a sequence of corresponding non-negative eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ (Riesz and Nagy 1990) satisfying

$$\int_0^1 K(s,t)\phi_p(s)\,\mathrm{d}s = \lambda_p\phi_p(t), \quad \text{for any } p \geq 1, \quad \text{and} \quad K(s,t) = \sum_{p=1}^{\infty} \lambda_p\phi_p(s)\phi_p(t), \quad (1)$$

which implies $\int_0^1 \int_0^1 K^2(s,t)\,\mathrm{d}s\,\mathrm{d}t = \sum_{p=1}^{\infty} \lambda_p^2$. The convergence in Equation (1) is absolute and uniform on $[0,1] \times [0,1]$ (Cucker and Zhou 2007).

By the Karhunen Loève theorem (Ash 1990), a random observation f has the following representation:

$$f(t) = \sum_{p=1}^{\infty} f_p\phi_p(t) \quad (2)$$

where $f_1, f_2, \ldots$ are uncorrelated random variables with mean $\mathbb{E}f_p = 0$, and variance $\mathbb{E}f_p^2 = \lambda_p$. Each coefficient f_p is unique and can be obtained by $f_p = \int_0^1 f(s)\phi_p(s)\,\mathrm{d}s$. In particular, if the stochastic process is a GP, $f_1, f_2, \ldots$ will be independent Gaussian random variables.

3.1.2. Reproducing kernel Hilbert space and its induced norm

Any symmetric, positive semi-definite function K on $[0,1] \times [0,1]$ corresponds to a unique RKHS with K as its reproducing kernel (Wahba 1990). We denote this RKHS by $\mathbb{H}_K$ with inner product

$$\langle K(s,\cdot), K(t,\cdot)\rangle_{\mathbb{H}_K} = \langle K(t,\cdot), K(s,\cdot)\rangle_{\mathbb{H}_K} = K(s,t).$$

Moreover, the reproducing property ensures that for any $f \in \mathbb{H}_K$, $\langle f, K(t,\cdot)\rangle_{\mathbb{H}_K} = f(t)$. The inner product induces the RKHS norm $\|f\|_{\mathbb{H}_K} = \sqrt{\langle f, f\rangle_{\mathbb{H}_K}}$. This leads to an equivalent definition of the RKHS as $\mathbb{H}_K = \{f : [0,1] \to \mathbb{R}, \|f\|_{\mathbb{H}_K} < \infty\}$. Therefore, under the representations in Equations (1) and (2), we have $f \in \mathbb{H}_K$ if and only if $\|f\|_{\mathbb{H}_K}^2 = \sum_{p:\lambda_p>0} \frac{f_p^2}{\lambda_p} < \infty$.

For a random trajectory $f \in \mathcal{F}$ from a second-order stochastic process with covariance kernel K, it is important to examine if the norm $\|f\|^2_{\mathbb{H}_K}$ is finite. If K has only finite number of positive eigenvalues, this conclusion certainly holds. However, if K has infinite number of positive eigenvalues, in general $\sum_{p:\lambda_p>0} \frac{f_p^2}{\lambda_p} = \infty (a.s.)$ since by the SLLN,

$$\frac{1}{n} \sum_{p=1}^{n} \frac{f_p^2}{\lambda_p} \xrightarrow{a.s.} E\left(\frac{f_p^2}{\lambda_p}\right) = 1$$

(the case for GP is discussed in (Wahba 1990)).

Consequently, although the RKHS norm $\|\cdot\|_{\mathbb{H}_K}$ contains important global features of the underlying data generating process, we cannot use the RKHS norm to define the depth since the RKHS norm of the observations are infinite almost surely. For example, for one-dimensional integrated Brownian motions (Vaart and Zanten 2011), it is known that smoothness level of the sample trajectories is 0.5 smaller than that of its associated RKHS (for Brownian motion, see Karatzas and Shreve 2012), where the corresponding RKHS norm coincides with the Sobolev norm. In this paper, we aim to combine these global features reflected in the RKHS norm into the construction of model-based functional depth. In particular, to solve this divergent issue of the RKHS norm, we propose a modified RKHS norm in the construction of our norm induced depth for functional data by weakening the impact of high-frequency signals, which are usually hard to estimate, on the modified norm.

3.2. Depth induced by modified RKHS norm

In this section, we propose a modified inner product structure for functions in $\mathcal{F}$. This new inner product will induce a modified RKHS norm that is almost surely finite for the sample trajectories from the second-order stochastic process.

3.2.1. Modified inner product and norm

Suppose f, g are two random realisations over $\mathcal{F}$ from a second-order stochastic process with covariance kernel K. Recall the eigen-decomposition $K(s, t) = \sum_{p=1}^{\infty} \lambda_p \phi_p(s) \phi_p(t)$ for any $s, t \in [0, 1]$. Without loss of generality, we assume all eigenvalues $\{\lambda_p\}$ are positive to avoid zero appearing in the denominator.

Recall the Karhunen Loève expansion, $f(t) = \sum_{p=1}^{\infty} f_p \phi_p(t)$ and $g(t) = \sum_{p=1}^{\infty} g_p \phi_p(t)$, with $f_p = \int_0^1 f(s)\phi_p(s)\,\mathrm{d}s$ and $g_p = \int_0^1 g(s)\phi_p(s)\,\mathrm{d}s$. In addition, the RKHS-induced inner product and norm are given in the following forms, respectively.

$$\langle f, g \rangle_{\mathbb{H}_K} = \sum_{p=1}^{\infty} \frac{f_p g_p}{\lambda_p} \quad \text{and} \quad \|f\| = \langle f, f \rangle_{\mathbb{H}_K}^{1/2}$$

As we discussed earlier in Section 3.1, the RKHS norm diverges almost surely. This divergence motivates us to a modified inner product as follows:

$$\langle f, g \rangle_{mod} := \sum_{p=1}^{\infty} \frac{f_p g_p}{\lambda_p} a_p^2,$$

where $\{a_p\}_{p=1}^{\infty}$ is any real sequence satisfying $\sum_{p=1}^{\infty} a_p^2 < \infty$. In practice, we may adopt commonly used convergent sequence $\{a_p = \frac{1}{p^s}\}_{p=1}^{\infty}$ or $\{a_p = \frac{1}{\sqrt{p}(\log p)^s}\}_{p=1}^{\infty}$ with $s > 1/2$. Our idea is to assign a decaying weight to each positive eigenvalue, so that the overall sum converges after the adjustment. This modified inner product yields a squared *modified RKHS norm* as

$$\|f\|_{mod}^2 = \langle f, f \rangle_{mod} = \sum_{p=1}^{\infty} \frac{f_p^2}{\lambda_p} a_p^2.$$

Straightforward calculations yield $\mathbb{E}(\|f\|_{mod}^2) = \sum_{p=1}^{\infty} \frac{E(f_p^2)}{\lambda_p} a_p^2 = \sum_{p=1}^{\infty} a_p^2 < \infty$. As a consequence, $\|f\|_{mod} < \infty$ almost surely, and the above modified inner product and norm are well-defined for the observed trajectories. We can use this modified RKHS norm as a selected norm to define a norm-based depth in our framework as described in Section 5.1.

Recall of Definition 2.2 of depth in Section 2.1: $D(f_{obs}, \mathbb{P}_\theta, \|\cdot\|, f_c) = \mathbb{P}_\theta[\|f - f_c\| \geq \|f_{obs} - f_c\|]$. In this case, the central function $f_c = 0$ is the mean function in our model; the norm function is the modified RKHS norm $\|\cdot\| = \|\cdot\|_{mod}$; $\mathbb{P}_\theta$ is a probability measure defined by the probability density on $\|f\|_{mod}$ or $\|f\|_{mod}^2$. Apparently, with different settings of the decaying sequence $\{a_p\}_{p=1}^{\infty}$, we will have different probability density for $\|f\|_{mod}$ or $\|f\|_{mod}^2$. It is often intractable to derive a closed-from expression on the density. Fortunately, our model-based depth framework provides a natural way of estimating the depth through bootstrap sampling, where the coefficients ($\{f_p\}$ in the Karhunen–Loève expansion) can be simulated with re-sampling techniques such as the Bootstrap.

3.2.2. Depth estimation procedure and algorithm

Suppose we have n zero-mean independent sample functions $f_1, \ldots, f_n \in \mathcal{F}$ on $t \in [0, 1]$, and our goal is to compute the model-based depth of any observed sample $f_{obs} \in \mathcal{F}$. We propose an estimation algorithm as follows.

Algorithm I: (Input: functional data $\{f_1, \ldots, f_n\}$, any observation f_{obs}, a small threshold $\delta_n > 0$, and a sequence $a_1, \ldots, a_n$.

(1) Compute the sample mean function $\hat{f}(t) = \frac{1}{n} \sum_{i=1}^{n} f_i(t)$, and empirical covariance kernel $\hat{K}(s, t) = \frac{1}{n} \sum_{i=1}^{n} [f_i(s) - \hat{f}(s)][f_i(t) - \hat{f}(t)]$;
(2) Eigen-decompose $\hat{K} = \sum_{p=1}^{\infty} \hat{\lambda}_{p,n} \hat{\phi}_{p,n}(s) \hat{\phi}_{p,n}(t)$;
(3) Set $\hat{\lambda}_{p,n} = 0$ if $\hat{\lambda}_{p,n} < \delta_n$;
(4) Set $M_n = \arg\max_m\{\lambda_m \geq \delta_n\}$, and $C_n = M_n \wedge n$ (minimum of M_n and n);
(5) Compute $\hat{f}_{i,p} = \int_0^1 f_i(t) \hat{\phi}_{p,n}(t) \, dt$ for all $i = 1, \ldots, n$ and $p = 1, \ldots, C_n$, and compute $\hat{f}_p = \int_0^1 f_{obs}(t) \hat{\phi}_{p,n}(t) \, dt$;
(6) For $p = 1, \ldots, C_n$, re-sample (with replacement) a large number N of coefficients $\{\hat{g}_{j,p}\}_{j=1}^{N}$ based on $\{\hat{f}_{1,p}, \ldots, \hat{f}_{n,p}\}$;
(7) Construct $g_j(t) = \sum_{p=1}^{C_n} \hat{g}_{j,p} \hat{\phi}_{p,n}(t)$;

(8) Compute $||f_{obs}||^2_{\hat{mod}} = \sum_{p=1}^{C_n} \frac{\hat{f}_p^2}{\hat{\lambda}_{p,n}} a_p^2$, and $||g_j||^2_{\hat{mod}} = \sum_{p=1}^{C_n} \frac{\hat{g}_{j,p}^2}{\hat{\lambda}_{p,n}} a_p^2$;

(9) Estimate the depth of f_{obs} using $\{g_j\}$:

$$D_n(f_{obs}; \{g_j\}_{j=1}^N) = \frac{1}{N} \sum_{j=1}^N \mathbb{1}_{||f_{obs}||^2_{\hat{mod}} \le ||g_j||^2_{\hat{mod}}}.$$

The first four steps aim to estimate the eigen system of the covariance kernel via given observations. In particular, the Karhunen Loève expansion (Ash 1990) is used in Step 2 to decompose the covariance kernel and offer a method to reconstruct samples. Using a functional principal component analysis (Ramsay 2005), we retain the eigen functions which explain meaningful variance in our system by truncating the empirical eigenvalues in Step 3 (Nicol 2013).

Steps 5–8 are the second part of the algorithm. They estimate the depth value with the modified RKHS norm, where we need re-sampling techniques and bootstrap approximations. This algorithm can be easily adapted to the multivariate data. In such case, the dimension of the data is already given and the principal component analysis and the multivariate metric can be directly applied. Step 9 estimates the probability in the depth definition by resampling from the empirical distribution of the sample basis expansion coefficients $\{\hat{f}_{i,p}\}_{i=1}^p$ for each coordinate $p = 1, \ldots, C_n$.

In Appendix 2, we specialise these developments to finite-dimensional processes (or multivariate data).

4. Asymptotic consistency

In this section, we will prove the consistency for the new model-based depths in Section 3. We assume the functional data are fully observed over its domain $[0, 1]$. This assumption is commonly used in asymptotic theory for various depths in functional data such as the integrated data depth (Fraiman and Muniz 2001), the band depth (López-Pintado and Romo 2009), the half-region depth (López-Pintado and Romo 2011) and the extremal depth (Narisetty and Nair 2016).

As our framework is model-based, there will be a main difference in the proofs between our framework and the traditional functional depth methods. In particular, since previous depths are independent of the generative model, usually an LLN suffices to show the consistency. In contrast, our method is considerably more involved since the depth itself is data dependent – it depends on the estimated model or parameters from the observations. Despite this extra difficulty in the theory, our new model-based depth can better utilise the generative patterns in the data, and therefore yields better (discriminative) power and efficiency in a variety of applications.

We start by introducing the notation used throughout in our proofs. Recall that $\mathcal{F} \subseteq \mathbb{L}^2([0,1])$ is the function space supporting the observations, which are generated from a second-order stochastic process with covariance function $K(s,t) = \mathbb{E}[(f(s) - \mathbb{E}(f(s)))(f(t) - \mathbb{E}(f(t)))]$. Suppose we have n functional replicates $f_1, \ldots, f_n \in \mathcal{F}$. Recall that the empirical approximation of $K(s,t)$ is $\hat{K}(s,t) = \frac{1}{n} \sum_{i=1}^n [(f_i(s) -$

$\frac{1}{n}\sum_{p=1}^{n}f_p(s))(f_i(t) - \frac{1}{n}\sum_{p=1}^{n}f_p(t))]$. It is clear that $\hat{K}$ is also a symmetric positive semi-definite kernel. Recall that, by Mercer's theorem, we have

$$K(s,t) = \sum_{p=1}^{\infty}\lambda_p\phi_p(s)\phi_p(t) \quad \text{and} \quad \hat{K}(s,t) = \sum_{p=1}^{n}\hat{\lambda}_{p,n}\hat{\phi}_{p,n}(s)\hat{\phi}_{p,n}(t),$$

where eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$ and $\hat{\lambda}_{1,n} \geq \hat{\lambda}_{2,n} \geq \cdots \geq \hat{\lambda}_{n,n}$ are non-negative, and their corresponding eigenfunctions $\{\phi_p\}_{p=1}^{\infty}$ and $\{\hat{\phi}_{p,n}\}_{p=1}^{n}$ are continuous on $[0,1]$. In this section, we primarily study the consistency of the proposed depth in the infinite-dimensional case where $\lambda_p > 0$ for any $p \in \mathbb{N}$. Due to space constraint, a counterpart result in the finite-dimensional case where $\lambda_p = 0$ for all $p > P$, where $P \in \mathbb{N}$, is deferred to Appendix 4.

4.1. Depth estimation consistency

At first, we study the general case when all eigenvalues $\{\lambda_p\}_{p=1}^{\infty}$ are positive. For any $f_{obs} \in \mathcal{F}$, we have shown in Section 3.2 that the squared modified norm

$$\|f_{obs}\|_{mod}^2 = \sum_{p=1}^{\infty}\frac{\langle f_{obs},\phi_p\rangle^2}{\lambda_p}a_p^2 \tag{3}$$

where $\langle\cdot,\cdot\rangle$ is the classical $\mathbb{L}^2$ inner product and $\{a_p\}_{p=1}^{\infty}$ is a real-valued sequence satisfying $\sum_{p=1}^{\infty}a_p^2 < \infty$. Based on the modified norm, the depth of f_{obs} is given as follows:

$$d_{mod}(f_{obs}) = D_n(f_{obs},\mathbb{P},\|\cdot\|_{mod},0) = \mathbb{P}\left[f : \|f\|_{mod} \geq \|f_{obs}\|_{mod}\right]$$
$$= 1 - \mathbb{P}\left[f : \|f\|_{mod}^2 \leq \|f_{obs}\|_{mod}^2\right] = 1 - F(\|f_{obs}\|_{mod}^2), \tag{4}$$

where $F(x)$ denotes the cumulative distribution function of $\|f\|_{mod}^2$ for the random function f.

As given in Algorithm I, the sample version of the squared modified norm is given as

$$\|f_{obs}\|_{\hat{mod}}^2 = \sum_{p=1}^{M_n}\frac{\langle f_{obs},\hat{\phi}_{p,n}\rangle^2}{\hat{\lambda}_{p,n}}a_p^2 \tag{5}$$

where $M_n = \arg\max_m\{\lambda_m \geq \delta_n\}$ for a given small threshold $\delta_n > 0$. In our framework, we adopt the sample version of the depth of f_{obs} given as

$$d_{mod,n}(f_{obs}) = \mathbb{P}\left[f : \|f\|_{mod} \geq \|f_{obs}\|_{\hat{mod}}\right] = 1 - F(\|f_{obs}\|_{\hat{mod}}^2). \tag{6}$$

In this section, we focus on proving $d_{mod,n}(f_{obs})$ converges to $d_{mod}(f_{obs})$ when n is large. Before we proceed to find consistency of the modified norm, we make the following two assumptions:

Assumption 4.1: $\exists \beta > 1, C, C_1, C_2 > 0$, s.t.

$$C_1 p^{-\beta} \geq \lambda_p \geq C_2 p^{-\beta} \quad \text{and} \quad \lambda_p - \lambda_{p+1} \geq C p^{-(\beta+1)} \quad \forall p \in \mathbb{N}.$$

Assumption 4.2: *There exists a real sequence $\{b_p\}_{p=1}^{\infty}$ and some constant $\alpha > 0$, such that $\sum_p b_p^2 < \infty$, and $a_p \le b_p \, p^{-\alpha}$ as p goes to ∞.*

For convenience, we abuse the notation 'C' to denote any constant coefficient. Followed by Assumption 4.1, it is apparent that the multiplicity of each λ_p is strictly 1. We point out that Assumption 4.2 can be easily satisfied in commonly used sequences of $\{a_p\}$. For example, if we choose $a_p = p^{-(0.5+\gamma)}$ for $\gamma > 0$, then we can choose $b_p = p^{-(0.5+\gamma/2)}$ (with $\alpha = \gamma/2$). Using the sequence $\{b_p\}$, we can define another type of modified form for any $f_{obs} \in \mathcal{F}$. As compared to modified norm in Equation (3), we only change the sequence $\{a_p\}$ to $\{b_p\}$. That is,

$$\|f_{obs}\|_b^2 = \sum_{p=1}^{\infty} \frac{\langle f_{obs}, \phi_p \rangle^2}{\lambda_p} b_p^2. \tag{7}$$

Our main convergence result is given in Theorem 4.1 as follows, where the proof is given in Appendix 8.

Theorem 4.1: *Under Assumptions 4.1 and 4.2, if the covariance kernel K has infinite number of positive eigenvalues $\{\lambda_p\}$, then the following holds with probability tending to one as $n \to \infty$,*

$$\sup_{f_{obs} \in \mathcal{F}, \|f_{obs}\| \le 1, \|f_{obs}\|_b \le 1} |\|f_{obs}\|_{\hat{mod}}^2 - \|f_{obs}\|_{mod}^2| \le C \, n^{-\kappa} \to 0, \tag{8}$$

where (C, κ) are some positive constants, $\|\cdot\|$ is the classical $\mathbb{L}^2$ norm and $\|\cdot\|_{mod}$, $\|\cdot\|_{\hat{mod}}$, $\|\cdot\|_b$ are the norms defined in Equations (3), (5), and (7), respectively. Moreover, for any $f_{obs} \in \mathcal{F}$

$$\lim_{n \to \infty} d_{mod,n}(f_{obs}) = d_{mod}(f_{obs}), \tag{9}$$

where the two depths $d_{mod,n}(f_{obs})$ and $d_{mod}(f_{obs})$ are given in Equations (4) and (6), respectively.

4.2. Monte-Carlo method and sample average

We have proven the convergence of the sample depth to the population depth. In practical computation such as Algorithm I, the sample depth is obtained using samples. In the proposed model-based framework, the depth is computed using Monte-Carlo samples. Alternatively, we can simply use the given sample and the estimate will be the sample average. In this section, we will prove that either of the methods can lead to accurate estimate asymptotically.

The main result on the Monte-Carlo approximation and sample average can be summarised in the following two theorems, where the detailed proofs are given in Appendix 9. The main result will be based on the following assumption.

Assumption 4.3: *Let f denote an observed sample from the true model. Then $\|f\|_b$ is sub-Gaussian, that is, there exists some constant $\sigma > 0$, such that $\mathbb{E}[\exp\{t\,\|f_p\|_b\}] \le \exp\{\sigma^2 t^2/2\}$ for all $t \in \mathbb{R}$.*

This assumption essentially controls the tail probability bound for $\|f\|_b$ as a random variable. In particular, it controls the maximal norm $\max_{i=1,\dots,n} \|f_i\|_b$ as of order $O_p(\sqrt{\log n})$ for an i.i.d. sample $\{f_i\}_{i=1}^n$ of size n from the true model, so that we can apply Theorem 4.1 to control the approximation errors $|\|f_i\|_{\hat{mod}}^2 - \|f_i\|_{mod}^2|$ uniformly over all $i = 1, 2, \dots, n$.

Theorem 4.2: *Let the sample depth $d_{mod,n}(f_{obs}) = \mathbb{P}[f : \|f\|_{mod} \geq \|f_{obs}\|_{\hat{mod}}]$ be estimated as $\frac{1}{n}\sum_{p=1}^n 1_{\|f_p\|_{\hat{mod}} \geq \|f_{obs}\|_{\hat{mod}}}$, where $\{f_p\}$ are observed i.i.d. sample from the true model and the model parameters are estimated from this sample. Then under Assumptions 4.1, 4.2, and 4.3, we have*

$$\frac{1}{n}\sum_{p=1}^n 1_{\|f_p\|_{\hat{mod}} \geq \|f_{obs}\|_{\hat{mod}}} \to 1 - F(\|f_{obs}\|_{mod}^2),$$

in probability as $n \to \infty$.

For the Monte Carlo approximation, we consider the simpler case where the true model is a zero mean Gaussian process with covariance function given by K for technical simplicity, and the Monte Carlo samples are also from a zero mean Gaussian process, but with the estimated covariance function $\hat{K}$.

Theorem 4.3: *Assume the true model is a zero-mean Gaussian process and let the sample depth $d_{mod,n}(f_{obs}) = \mathbb{P}[f : \|f\|_{mod} \geq \|f_{obs}\|_{\hat{mod}}]$ be estimated as $\frac{1}{N}\sum_{p=1}^N 1_{\|g_p\|_{\hat{mod}} \geq \|f_{obs}\|_{\hat{mod}}}$, where $\{g_p\}$ are an i.i.d. sample from the estimated distribution. Then under Assumptions 4.1 and 4.2 we have*

$$\frac{1}{N}\sum_{p=1}^N 1_{\|g_p\|_{\hat{mod}} \geq \|f_{obs}\|_{\hat{mod}}} \to 1 - F(\|f_{obs}\|_{mod}^2)$$

almost surely as $N, n \to \infty$.

5. Simulation and real data analysis

In this section, we illustrate applications of our proposed model-based depths to synthetic data and real data.

5.1. Simulation examples

We will at first use several simulations to illustrate the uses of the norm-based and inner-product-based forms in Sections 2.2.1 and 2.2.2 for exploratory data analysis of both multivariate and functional data. In particular, Simulations 1 and 2 focus on several commonly used norms (inner-products) for model-based depth developed in Section 2, and Simulations 3 and 4 consider the new model-based functional depth introduced in Section 3. More simulation examples, including multivariate depth, are provided in Appendix 3.

Simulation 1. In this example, we illustrate the $\mathbb{L}^p$ induced norms as criteria functions which are discussed in the first part of Section 2.2.1. We demonstrate our framework by

observations from zero-mean Gaussian Process with Matérn class kernel on $[0, 1]$. The generative formula for Matérn kernel is

$$K_M(x_i, x_j) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}|x_i - x_j|}{l} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu}|x_i - x_j|}{l} \right), \quad x_i, x_j \in [0, 1]$$

where K_ν is the modified Bessel function of order ν, and the parameter l is the characteristic length scale of the process. For instance, if $\nu = \frac{1}{2}$ and $l = 1$, then the Matérn kernel $K_1(s, t) = \exp(-|s - t|)$, and if $\nu = \frac{3}{2}$ and $l = 1$, $K_2(s, t) = (1 + \sqrt{3}|s - t|) \exp(-\sqrt{3}|s - t|)$, for $s, t \in [0, 1]$.

For better visualisation, we sample only one function from $GP(0, K_1)$ on $[0, 1]$, and then mix it with another $n = 29$ simulated samples from $GP(0, K_2)$ on $[0, 1]$. All these 30 functions are shown in Figure 2(a). It is apparent that the one function from K_1 is near the zero-line, but somewhat 'noisy'. In contrast, the 29 functions from K_2 have high variability in the magnitudes, but are very smooth. We then colour-labelled them differently in Panels (b)–(d) using their depth values with respect to different criterion functions, namely, (1) $\mathbb{L}^2$ norm on each function, (2) $\mathbb{L}^2$ norm on the first-order derivative of the original function and (3) $\mathbb{L}^2$ norm on the second-order derivative of the original function. In Case 1, the observations are ranked with respect to their absolute magnitudes (well captured by the $\mathbb{L}^2$ norm). In Cases 2 and 3, the observations are ranked with respect to their degrees of smoothness (well captured by the $\mathbb{L}^2$ norms in the first- and second-order derivatives, respectively). The results clearly illustrate that criteria properly characterise the desirable features in the data. In Panel (b), we rank the function with respect to their $\mathbb{L}^2$ norm. The one function from K_1 is near the zero-line and has the highest depth value. In contrast, since this function is not smooth, it has the least depth values with derivative-based norms in Panels (c) and (d).

Simulation 2. In this example, we illustrate the time warping distance in the depth computation. We study a set of simulated functions $\{f_1, \ldots, f_{21}\}$ on $[-3, 3]$. For $i = 1, \ldots, 21$, we first simulate a set of functions by $h_i(t) = \phi_{i,1} e^{-(t-1.5)^2/2} + \phi_{i,2} e^{-(t+1.5)^2/2}$, where $\phi_{i,1}$ and $\phi_{i,2}$ are i.i.d. normal with mean one and variance $1/16$. Let the warping function $\gamma_i(t) = 6(\frac{e^{a_i(t+3)/6}-1}{e^{a_i}-1}) - 3$ if $a_i \neq 0$ otherwise $\gamma_i = \gamma_{id}$, where a_i are equally spaced between -1 and 1. The observations are $f_i(t) = h_i(\gamma_i(t))$ on $[-3, 3]$, $i = 1, \ldots, 21$. At the final step, we add some noise to the original f_{11} by $\tilde{f}_{11}(t) = f_{11}(t) + \epsilon(t)$, where $\epsilon(t)$ is a Gaussian process with mean 0 and covariance function $C(s, t) = 0.01\delta_{s,t}$. To simplify the notation, we abuse f_{11} to denote the noise contaminated $\tilde{f}_{11}$.

All these 21 functions are shown in Figure 3(a), where we use red line to represent f_{11} and blue lines to represent the others. Before computing depth values, we conduct the Fisher–Rao alignment procedure to align the observed functions and obtain the corresponding time warping functions $\{\hat{\gamma}_1(t), \ldots, \hat{\gamma}_{21}(t)\}$ (Srivastava et al. 2011). Let f_c denote the Karcher mean of $\{f_1, \ldots, f_{21}\}$ (in the sense of SRVF space). Then the optimal time warping function from f_i to f_c is $\hat{\gamma}_i$, $i = 1, \ldots, 21$.

We take the criterion function $\zeta(f_i, f_c) = \|\hat{\gamma}_i - \gamma_{id}\|_2$ for the depth computation. The 21 colour-labelled functions using depth values are shown in Figure 3(b). In general, functions in the middle along x_axis have large depth values, whereas those at each side have low values. In particular, because f_{11} stays in the middle of the observations, it has the least $\mathbb{L}^2$ warping distance from f_c and largest depth values. As comparison we also use the

(a) Given observations

(b) $\zeta(f, 0) = \|f\|_2$

(c) $\zeta(f, 0) = \|f'\|_2$

(d) $\zeta(f, 0) = \|f''\|_2$
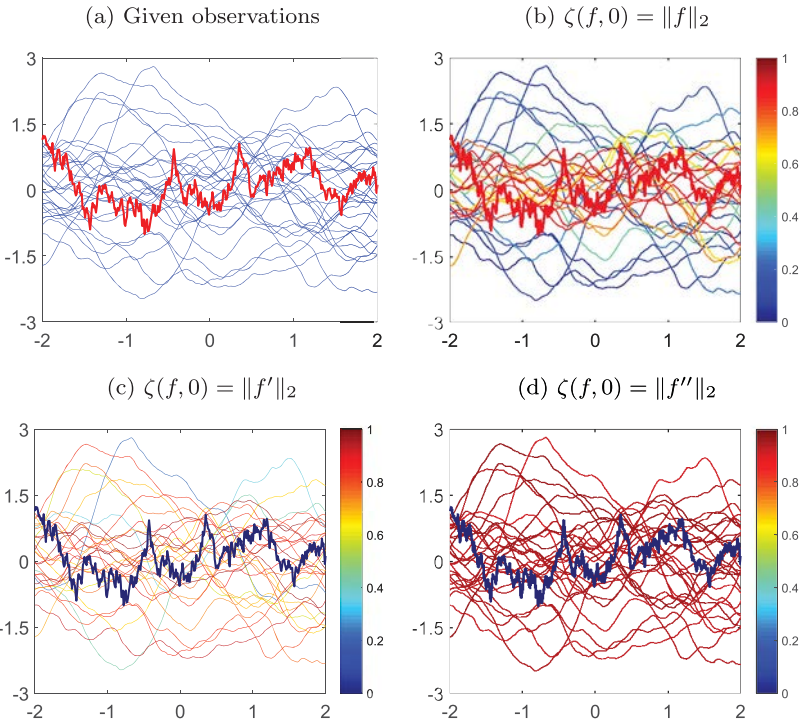
Figure 2. Simulation 1: (a) 30 observed functions, where the red one is generated from $GP(0, K_1)$ and 29 blue ones are generated from $GP(0, K_2)$. (b) The 30 functions with colour-labelled depth using $\mathbb{L}^2$ norm. Observations assigned with colour closer to red are considered to be deeper than those assigned with colour closer to blue. (c) and (d) Same as (b) except for $\mathbb{L}^2$ norm on the first and second-order derivative functions, respectively.

(a) Observed functions

(b) $\zeta(f_i, f_c) = \|\hat{\gamma}_i - \gamma_{id}\|_2$

(c) $\zeta(f_i, f_c) = d_{FR}(\hat{\gamma}_i, \gamma_{id})$
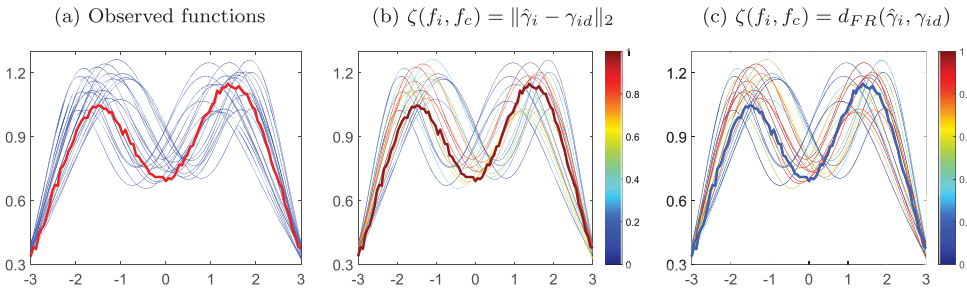
Figure 3. Simulation 2: (a) 21 observed functions, where f_{11} is emphasised in red colour; (b) 21 functions with colour-labelled depth via the $\mathbb{L}^2$ warping distance $\|\hat{\gamma}_i - \gamma_{id}\|_2$ and (c) same as (b) except for the Fisher–Rao distance $d_{FR}(\hat{\gamma}_i, \gamma_{id})$.

well-known Fisher–Rao distance function $\zeta(f_i, f_c) = d_{FR}(\hat{\gamma}_i, \gamma_{id})$ for the depth computation. The 21 colour-labelled functions using depth values are shown in Figure 3(c). As the Fisher–Rao distance is derivative based, small perturbation on time warping results in large difference. The small noise on f_{11} makes it have smallest depth value in the 21 functions. For other 20 smooth functions, their depth values are consistent to those in Panel (b).

Simulation 3. In this illustration, we demonstrate Algorithm I in Section 3 for modified norm-based depth estimation on a variant of the continuous-time Brownian Bridge on $[0,1]$, with different choices of decaying weight sequences $\{a_p\}$, where the covariance kernel function is $K(s,t) = \min(s,t) - st$ for any $s, t \in [0,1]$. According to the notation in Equations (1) and (2), we have $\lambda_p = \frac{1}{p^2\pi^2}, \phi_p(t) = \sqrt{2}\sin(\pi p t)$, and can simulate f_p from *independent Laplace distribution* with mean 0 and variance λ_p for $p = 1, 2, \ldots$ (note that this is different from the normal distribution $\mathcal{N}(0, \lambda_p)$ in a Brownian bridge).

More specifically, we sample $\{f_i(t)\}$ by the linear combination $f_i(t) = \sum_{p=1}^{500} f_{i,p}\phi_p(t), i = 1, \ldots, n(= 100)$ to approximate the infinite-dimensional stochastic process. We set $N = 500$ and $\delta_n = 4 \times 10^{-7}$ in the Monte Carlo sampling. We have three different settings to choose the weight coefficients $\{a_p\}$: (a) $a_p \equiv 1$, (b) $a_p = 1/p$ and (c) $a_p = 1/[\sqrt{p}\log(p + 1)], p = 1, \ldots, N$. In Case (a), there is actually no weight terms, and the modified norm is equal to the RKHS induced norm. In Figure 4(a), we show the 100 functions with colour-labelled using its depth value from this norm. Note that we compute this norm in a finite-dimensional setup and it will diverge to ∞ when N is large. It is straightforward to find that

$$\|f_i'\|_{\mathbb{L}^2}^2 = \sum_{p=1}^{N}\int_0^1 2\pi^2p^2 \cdot f_{i,p}^2\cos(\pi p t)^2\,\mathrm{d}t = \sum_{p=1}^{N}\pi^2p^2f_{i,p}^2 = \sum_{p=1}^{N}\frac{f_{i,p}^2}{\lambda_p} = \|f_i\|_{\mathbb{H}_K}^2. \quad (10)$$

That is, the RKHS induced norm is the same as $\mathbb{L}^2$ norm on the derivative function, a common measure of smoothness of a function.

In Cases (b) and (c), the series satisfies the convergent requirement $\lim_{N\to\infty}\sum_{p=1}^{N}a_p^2 < \infty$, and therefore the modified norms are well-defined. In particular, we find that when $a_p = 1/p$, the classical $\mathbb{L}^2$ norm

$$\|f_i\|_{\mathbb{L}^2}^2 = \sum_{p=1}^{N}\int_0^1 f_{i,p}^2\phi_p^2(t)\,\mathrm{d}t = \sum_{p=1}^{N}f_{i,p}^2 = \frac{1}{\pi^2}\sum_{p=1}^{N}f_{i,p}^2 \cdot \frac{1}{1/\pi^2p^2} \cdot \frac{1}{p^2} \propto \|f_i\|_{mod}^2. \quad (11)$$

That is, the modified norm in Case (b) is proportional to the $\mathbb{L}^2$ norm. This explains the result in Figure 4(b) where the 100 functions are colour-labelled using its depth value from this norm. We can see that high-depth functions are near the zero-line and low-depth functions are near boundary lines. This depth reflects the traditional functional depths such as band depth and half-region depth (López-Pintado and Romo 2009, 2011). In Case (c), we use another type of weight coefficient and the depth result is shown in Figure 4(c), which is very similar to the result in Case (b). The result indicates the robustness of the depth value with respect to a_p.

To evaluate the robustness with respect to δ_n in Algorithm I, we fix $a_p = 1/1/[\sqrt{p}\log(p + 1)]$ (as given in Case c) and examine depth values with three different thresholds $\{\delta_n\}$: (d) $\delta_n = 0.05$, (e) $\delta_n = 0.01$, and (f) $\delta_n = 10^{-5}$. In case (d), the estimation of depth values is relatively poor (see Figure 4 d). This is because this large threshold makes many eigen-components cut off. However, the depth values in Cases (e) and (f), shown in Figure 4(e,f), are very similar to that in Case (c), shown in Figure 4(c).
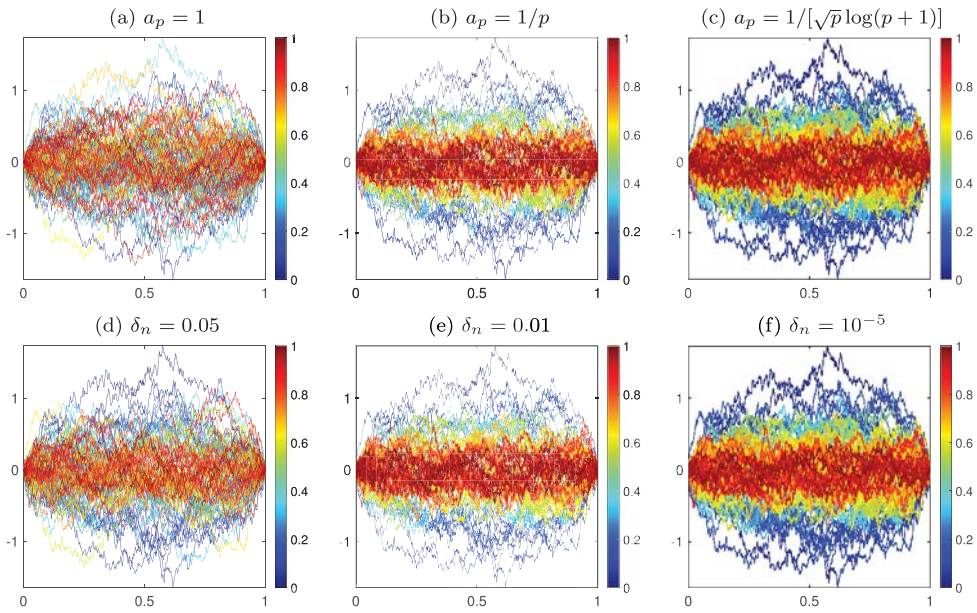
Figure 4. Simulated functions with colour-labelled depth. (a) Each function is colour-labelled using its depth value, where the sequence $\{a_p\}$ is constant 1 and the threshold $\delta_n = 4 \times 10^{-7}$. Observations assigned with colour closer to red are considered to be deeper than those assigned with colour closer to blue. (b) and (c) Same as (a) except that the coefficients $a_p = 1/p$ and $a_p = 1/[\sqrt{p}\log(p+1)], p = 1, \ldots, 500$, respectively. (d)–(f) Same as (c) except that the thresholds δ_n are 0.05, 0.01 and 10^{-5}, respectively.

In summary, we have found that (1) the modified norms can provide different forms of measurement on the centre-out rank on the given functional observation and some of the special forms are consistent to the classical norms; (2) the rank may be robust with respect to different choices of norm and (3) the depth estimation is robust with respect to weight sequence a_p and threshold δ_n.

In addition, we compare our approach to some well-known depth methods – modified band depth (López-Pintado and Romo 2009), modified half-region (HR) depth (López-Pintado and Romo 2011), extremal depth (Narisetty and Nair 2016), and integrated depth based on the half-space (HS) depth (Nagy, Gijbels, and Hlubinka 2017). All these four depths, shown in Figure 5, provide consistent centre-out ranks as in Figure 4(b,c). However, some outer observations are assigned to higher depth values in modified HR depth, as shown in Figure 5(b); the values of modified band depth and integrated HS depth concentrate on higher depth values and low depth values are assigned to only a few observations, as shown in Figure 5(a,d); and the curves with high depth values (red) and the curves with middle depth values (yellow and green) are mixed together in extremal depth, as shown in Figure 5(c). In contrast, the distributions of the proposed depth values are more evenly distributed in [0, 1], as shown in Figure 4(b,c).

Finally, we compare the computational cost for all above depth methods. It is found that all methods can finish calculations within seconds, and there is no significant difference in terms of efficiency among them.
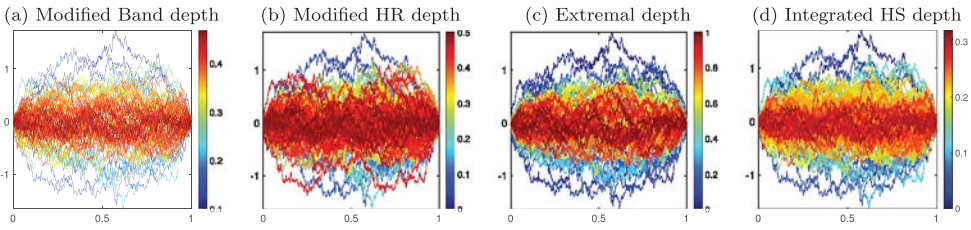
Figure 5. Comparison: (a) depth values computed using modified band depth method; (b) depth values computed using modified half-region depth method; (c) depth values computed using extremal depth method; (d) depth values computed using integrated depth method based on the half-space depth.

Simulation 4. We consider a finite-dimensional Gaussian process by selecting a sequence of orthonormal Fourier basis functions up to order $P = 10$ on $[0, 1]$ such that

$$\phi_p(t) = \begin{cases} 1 & p = 1 \\ \sqrt{2}\cos(\pi pt) & p = 2, 4, 6, 8, 10 \\ \sqrt{2}\sin(\pi(p-1)t) & p = 3, 5, 7, 9 \end{cases},$$
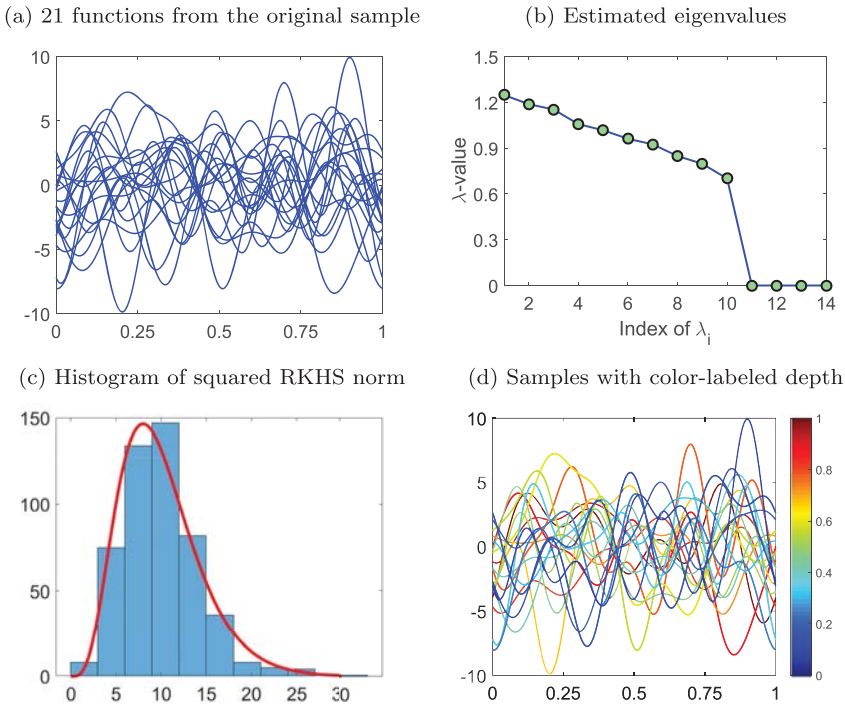
and a set of coefficients $\{a_1, \ldots, a_P\} \sim N(0, I_{10})$. Then we generate $N = 500$ functions via linear combination $f_i = \sum_{p=1}^{P} a_{i,p}\phi_p$. Panel (a) in Figure 6 shows $n = 21$ randomly selected samples from $\{f_i(t), t \in [0, 1]\}_{i=1}^{N}$.

From Panel (b) in Figure 6, it is clear that there exists a significant gap in the decreasing sequence of estimated eigenvalues, and the gap locates just after the order of the dimension $P = 10$. This indicates the correct dimension can be easily estimated. From Panel (c), we can tell that the squared RKHS norm fits $\chi^2(P)$ well. This is also consistent to the above theoretical derivation. The estimated depth values are colour-labelled in Panel (d). We note that the RKHS induced norm does not have a conventional $\mathbb{L}^2$ type of norm, so there is no direct visualisation to evaluate the depth in this example. However, we point out that if the data are generated from two random processes, these depth values can help differentiate the observations, as illustrated in the following.

In particular, we show how the model-based depth can be used for the classification purpose, and compare the performance between RKHS norm and the modified one. Specifically, we select a sequence of orthonormal Fourier basis functions up to order P on $[0, 1]$ such that for $p = 1, 2, \ldots, P$,

$$\phi_p(t) = \begin{cases} \sqrt{2}\sin(\pi(p+1)t) & p \text{ is odd} \\ \sqrt{2}\cos(\pi pt) & p \text{ is even} \end{cases},$$

Then we generate 45 functions as $f_i(t) = \sum_{p=1}^{P} a_{i,p}\phi_p(t), i = 1, \ldots, 45$ and 5 functions as $f_i(t) = \sum_{p=1}^{P} b_{i,p}\phi_p(t), i = 46, \ldots, 50$, where independent coefficients $a_{i,p} \sim N(0, 1)$ and $b_{i,p} \sim N(0, 3)$. Due to the different variance values on the coefficients, the first 45 functions are in the main cluster and the last 5 functions are outliers. One special case when only $P = 4$ low frequency basis functions are used is shown in Figure 7(a). Because of the smaller coefficient variance, the first 45 functions are in a main cluster. In contrast, some of the last 5 function have much larger amplitude and are apparently outliers. In

(a) 21 functions from the original sample

(b) Estimated eigenvalues

(c) Histogram of squared RKHS norm

(d) Samples with color-labeled depth

Figure 6. Finite Gaussian Process Illustration: (a) 21 randomly selected samples; (b) Estimated eigenvalues $\hat{\lambda}_p$ from the covariance $\hat{K}$; (c) Histogram of squared RKHS norm $\|f_i\|^2_{\mathbb{H}_{\hat{K}}} = \sum_{p=1}^{10} \frac{\hat{a}^2_{i,p}}{\hat{\lambda}_{p,n}}$, where the red line indicates a fit to chi-square distribution $\chi^2(10)$; (d) Estimated depth of the 21 samples with colour label, where blue to red indicates the depth value range of [0, 1].

another example, we use $P = 100$ basis functions shown in Figure 7(b). As compared to when $P = 4$, both main clusters functions and the five outliers have much higher frequency components.

We have shown that the RKHS induced norm can characterise the smoothness level in Equation (10) and the modified RKHS norm can characterise the amplitude level ($\mathbb{L}^2$ norm) in Equation (11). We at first use these two norms for the case when $P = 4$ and the result on depth values is shown in Figure 7(c). Note that for the simulated five outliers, only three of them show large amplitude as compared to the main cluster, and therefore only these three have relatively lower depth values by using either RKHS norm or the modified norm. In contrast, when $P = 100$, the difference on amplitude for the main cluster and the five outliers are apparent. This can be easily seen using the modified norm shown in Figure 7(d). As all high frequency basis functions can have large un-smooth level, the RKHS norm is not able to clearly differentiate five outliers from the main cluster. This is also shown in Figure 7(d).

To measure the classification performance, we set a threshold of 0.1 on the depth value for all functions. This is done for the number of basis components P being 4, 10, 20, 30, 40, 70 or 100, which varies from highly smooth to highly nonsmooth observations. The classification result on all 50 functions is shown in Figure 7(e). In particular, we also show the detection on the five outliers in Figure 7(f). When P is small, both norms produce
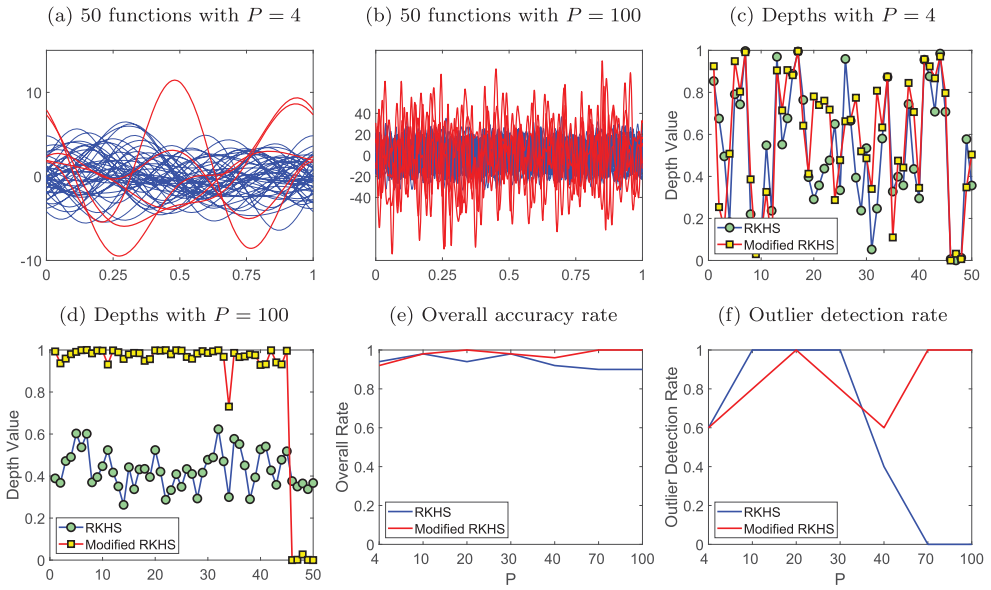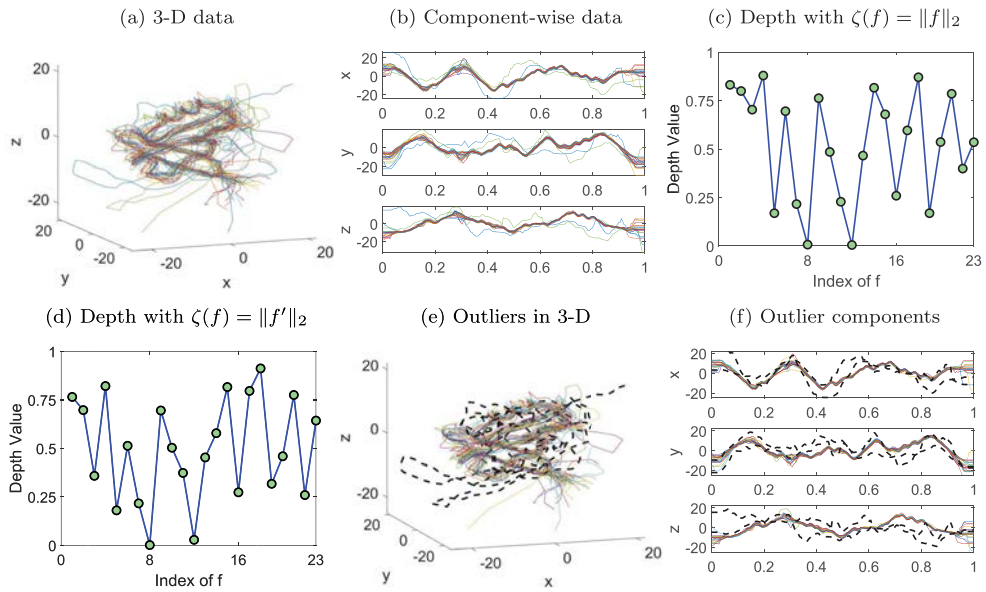
Figure 7. Classification by depth values: (a) 50 function samples for $P = 4$, where the blue ones represent 45 functions in the main cluster and the red ones represent 5 outliers. (b) Same as (a) except $P = 100$. (c) Depth values of the 50 functions with $P = 4$ using the RKHS norm (green circles over blue lines) and modified RKHS norm (yellow squares over red lines). (d) Same as (c) except $P = 100$. (e) The classification accuracy for all 50 functions by using the RKHS norm (blue line) and modified RKHS norm (red line), where P varies on seven different values 4, 10, 20, 30, 40, 70, and 100. (f) Same as (e) except the accuracy on the 5 outlier functions.

reasonable classification accuracy around 95% (a couple of errors in the outliers). When P gets larger, the modified RKHS can capture larger amplitude in the outliers and reach 100% classification accuracy. In contrast, all 50 functions have similar smoothing level which makes the RKHS norm not able to detect the outliers.

5.2. Real data illustration

In this section, we apply our proposed method to detect outliers on a real dataset. The dataset is taken from the SCOP database (Murzin, Brenner, Hubbard, and Chothia 1995). We take the subset of proteins with sample size 23 from PDZ domain using PISCES server (Wang and Dunbrack Jr 2003). The data have been pre-processed as described in Wu, Srivastava, Laborde, and Zhang (2013), and we get normalised data where the three components are properly rotated and aligned. This given data are shown in Figure 8(a) as three-dimensional curves and the three components are shown in Figure 8(b).

This given data has been applied with two different norms: one is the classical $\mathbb{L}^2$ norm on three-dimensional functions and the other is the $\mathbb{L}^2$ norm on the first derivative functions. The depth values computed by these two different norms are shown in Figure 8(c,d), respectively. We note that the depth results are very close to each for the two norms – both methods indicate that the 8th and 12th protein sequences are outliers in our dataset by using a detection threshold $\alpha = 0.05$. The two outliers are shown in Figure 8(e,f) as

(a) 3-D data   (b) Component-wise data   (c) Depth with $\zeta(f) = \|f\|_2$

(d) Depth with $\zeta(f) = \|f'\|_2$ (e) Outliers in 3-D   (f) Outlier components

Figure 8. Real data example: (a) three-dimensional data of 23 PDZ domain observations; (b) Three-coordinate components of given observations; (c) depth values computed using the classical $\mathbb{L}^2$ norm; (d) depth values computed using the $\mathbb{L}^2$ norm on the first derivative functions; (e) detected outliers (black dash lines) in three dimensions by using depth values; (f) three-coordinate components of the detected outliers (black dashed lines).

three-dimensional curves and for three coordinate components, respectively. It is apparent that the depth values successfully detect the outliers in the given data.

For comparison, the depth values obtained by the RKHS norm in our framework are shown in Figure 9(a), where a lot of functions have low depth values and the two outliers cannot be clearly identified. Figure 9(b) shows the depth values computed by modified norm with $a_p = 1/p$ in our model-based depth framework. It is clearly to see that the 8th and 12th functions have lowest depth values, though not as close to 0 as the two $\mathbb{L}^2$ norms in Figure 8.

We compare our approach to the well-known depth methods – band depth and its modified version for three-dimensional functional data (López-Pintado and Romo 2009; Ieva and Paganoni 2013), modified half-region (HR) depth (López-Pintado and Romo 2011), h-mode depth (Cuevas, Febrero, and Fraiman 2007), and integrated depth based on simplicial depth and half-space (HS) depth (Nagy et al. 2017). With no prior information, the three components contribute equally in the overall depth estimation. The performance of band depth shown in Figure 9(c) is very poor, and it implies that there are great difficulties in applying the band depth due to large variation in the three coordinates. On the other hand, the result in Figure 9(d) shows that depth values obtained by modified band depth have a clear large gap between the two outliers and main portion of the data, consistent to the result in Figure 8, whereas the depth values are distributed in a very narrow range. The modified HR depth also gives poor performance as illustrated in Figure 9(e). However, the result for h-mode depth in Figure 9(f) shows three outliers are clearly identified. In particular, those two with the lowest depth values agree with our approach and the modified
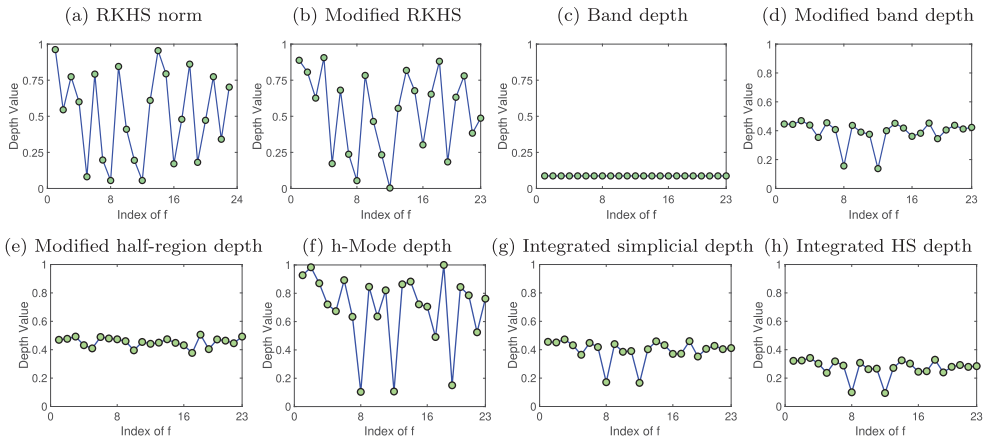
Figure 9. Comparison: (a) depth values using RKHS induced norm; (b) depth values using modified RKHS norm with $a_p = 1/p$; (c) depth values computed using band depth method; (d) depth values computed using modified band depth method; (e) depth values computed using modified half-region depth method; (f) depth values computed using h-mode depth method; (g) depth values computed using integrated depth method based on the simplicial depth and (h) depth values computed using integrated depth method based on the half-space depth.

band depth, and the observation with the third lowest depth also has a low depth value in our method. Finally, the integrated depths (Figure 9 g,h) give similar results as the modified band depth.

6. Summary and future work

In this article, we have proposed a new framework to define model-based statistical depth for functional as well as multivariate observations. Our definitions have two forms: norm-based and inner-product-based. Depending on the selection (of norms), the norm-based depth can have various centre-outward ranks. For the inner-product depth, it is mainly the generalisation of the multivariate halfspace depth. We then focus on using norms which are naturally defined with the generative model. That is, we use induced RKHS norm from the finite-dimensional covariance kernel in a second-ordered stochastic process. For an infinite-dimensional kernel, we have introduced a modified version to avoid the infinity value on the induced norm. For practical use, we propose efficient algorithms to compute the proposed depths. Through simulations and real data, we demonstrate the proposed depths reveal important statistical properties of given observations, such as median and quantiles. Furthermore, we establish the consistency theory on the estimators.

We would like to emphasise that although our proposed methods are derived under the Gaussian assumption, they apply to any second-order (stochastic) process. This is because the methods (depth, inner-product, norm) are solely based on the covariance function of the process. For a Gaussian process, it is completely determined by its first- and second-order moments. For a general process, it is not. The same expansion (the well-known Karhunen–Loève expansion given in Equation 2), however, still holds. That is indeed why we use the resampling step (e.g. Step 6 in Algorithm I) to generate the coefficients instead of

sampling from normals. For Gaussian process, the coefficients must follow a normal distribution and a conventional Monte Carlo method can be used, while in our bootstrap method, we do resampling to preserve the non-Gaussianity of their distributions. This means our method actually can be applied to non-Gaussian case, robustly.

In addition, our method uses an induced RKHS norm that is adaptively defined through the covariance function of the process, which should better capture structures than a pre-specified norm. Actually, it is easy to come up with some example where the conventional $\mathbb{L}^2$ distance is worse. For example, Simulation 1 shows that blindly using a fixed norm such as the $\mathbb{L}^2$ norm may fail to capture some important features of functional data such as smoothness. We also point out that even if an RKHS method is used to define the norm, selecting the right space will have a significant impact on the performance of the procedure. If the space is the RKHS induced by the covariance function, then the procedure uses the second-order moment information, which coincides with our data adaptive method. Otherwise, for any pre-specified RKHS norm such as the Matérn RKHS, it is easy to have some examples where smoothness is not important, and blindly using the Matérn RKHS may lead to misleading results. Even if the RKHS induced by the covariance function of the data generating process is the Matérn RKHS, selecting a right smoothness level is not straightforward without learning from the data. Our method estimates the kernel matrix from the data, which is data-adaptive, and can also be viewed as providing a data-adaptive way of selecting the smoothness parameter.

Statistical depth is an extensively-studied area. However, all previous methods are either procedure-based or properties-based. To the best of our knowledge, this is the first model-based investigation. This paper introduced the basic framework, but the model is limited to covariance-based method. Due to the nature of covariance kernel, our framework has a tendency to deal with second-order stochastic process like Gaussian family well. We plan to work in a space where higher order statistics can also be important in the future. In addition, we have discussed four important properties for the proposed depths. As Gijbels and Nagy (2017) provided an elaboration on more desirable properties (such as receptivity and continuity) of statistical depths for functional data, our future work is to investigate whether our proposed framework would meet those properties. Moreover, since we obtain median and quantiles by the proposed depths, we can also extend our method to construct boxplot visualisation. Last but not least, we are seeking broader applications of the new framework in real world problems such as clustering, classification and outlier detection.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Agostinelli, C. (2018), 'Local Half-region Depth for Functional Data', *Journal of Multivariate Analysis*, 163, 67–79.

Agostinelli, C., and Romanazzi, M. (2013), 'Ordering Curves by Data Depth', in *Statistical Models for Data Analysis*, Springer, pp. 1–8.

Ash, R.B. (1990), *Information Theory*, Dover Publications, corrected reprint of the 1965 original.

Balzanella, A., and Elvira, R. (2015), 'A Depth Function for Geostatistical Functional Data', in *Advances in Statistical Models for Data Analysis*, Springer, pp. 9–16.

Barnett, V. (1976), 'The Ordering of Multivariate Data', *Journal of the Royal Statistical Society. Series A (General)*, 139(3), 318–355.

Bosq, D. (2012), *Linear Processes in Function Spaces: Theory and Applications* (Vol. 149), Springer Science & Business Media.

Chakraborty, A., and Chaudhuri, P. (2014a), 'On Data Depth in Infinite Dimensional Spaces', *Annals of the Institute of Statistical Mathematics*, 66(2), 303–324.

Chakraborty, A., and Chaudhuri, P. (2014b), 'The Spatial Distribution in Infinite Dimensional Spaces and Related Quantiles and Depths', *The Annals of Statistics*, 42(3), 1203–1231.

Chernozhukov, V., Galichon, A., Hallin, M., and Henry, M. (2017), 'Monge–Kantorovich Depth, Quantiles, Ranks and Signs', *The Annals of Statistics*, 45(1), 223–256.

Christmann, A. (2002), "Classification Based on the Support Vector Machine and on Regression Depth", in *Statistical Data Analysis Based on the L1-Norm and Related Methods*, Springer, pp. 341–352.

Cleveland, J., Zhao, W., and Wu, W. (2018), 'Robust Template Estimation for Functional Data with Phase Variability Using Band Depth', *Computational Statistics & Data Analysis*, 125, 10–26.

Cucker, F., and Zhou, D.X. (2007), *Learning Theory: An Approximation Theory Viewpoint* (Vol. 24), Cambridge University Press.

Cuesta-Albertos, J.A., and Nieto-Reyes, A. (2008), 'The Random Tukey Depth', *Computational Statistics & Data Analysis*, 52(11), 4979–4988.

Cuevas, A., Febrero, M., and Fraiman, R. (2007), 'Robust Estimation and Classification for Functional Data Via Projection-based Depth Notions', *Computational Statistics*, 22(3), 481–496.

Dauxois, J., Pousse, A., and Romain, Y. (1982), 'Asymptotic Theory for the Principal Component Analysis of a Vector Random Function: Some Applications to Statistical Inference', *Journal of Multivariate Analysis*, 12(1), 136–154.

Donoho, D.L., and Gasko, M. (1992), 'Breakdown Properties of Location Estimates Based on Halfspace Depth and Projected Outlyingness', *The Annals of Statistics*, 20(4), 1803–1827.

Dutta, S., Ghosh, A.K., and Chaudhuri, P. (2011), 'Some Intriguing Properties of Tukey's Half-space Depth', *Bernoulli*, 17(4), 1420–1434.

Dutta, S., Sarkar, S., and Ghosh, A.K. (2016), 'Multi-scale Classification Using Localized Spatial Depth', *The Journal of Machine Learning Research*, 17(1), 7657–7686.

Einmahl, J.H., Li, J., and Liu, R.Y. (2015), 'Bridging Centrality and Extremity: Refining Empirical Data Depth Using Extreme Value Statistics', *The Annals of Statistics*, 43(6), 2738–2765.

Fraiman, R., Liu, R.Y., and Meloche, J. (1997), 'Multivariate Density Estimation by Probing Depth', *Lecture Notes-Monograph Series*, 415–430.

Fraiman, R., and Muniz, G. (2001), 'Trimmed Means for Functional Data', *Test*, 10(2), 419–440.

Gijbels, I., and Nagy, S. (2017), 'On a General Definition of Depth for Functional Data', *Statistical Science*, 32(4), 630–639.

Hsing, T., and Eubank, R. (2015), *Theoretical Foundations of Functional Data Analysis, with An Introduction to Linear Operators*, John Wiley & Sons.

Ieva, F., and Paganoni, A.M. (2013), 'Depth Measures for Multivariate Functional Data', *Communications in Statistics-Theory and Methods*, 42(7), 1265–1276.

J Mercer, B. (1909), 'Xvi. Functions of Positive and Negative Type, and Their Connection the Theory of Integral Equations', *Philosophical Transactions of the Royal Society A*, 209(441–458), 415–446.

Karatzas, I., and Shreve, S. (2012), *Brownian Motion and Stochastic Calculus* (Vol. 113), Springer Science & Business Media.

Kuelbs, J., and Zinn, J. (2015), 'Half-region Depth for Stochastic Processes', *Journal of Multivariate Analysis*, 142, 86–105.

Liu, R.Y. (1990), 'On a Notion of Data Depth Based on Random Simplices', *The Annals of Statistics*, 18(1), 405–414.

Liu, X., and Müller, H.-G. (2004), 'Functional Convex Averaging and Synchronization for Time-warped Random Curves', *Journal of the American Statistical Association*, 99(467), 687–699.

Liu, R.Y., Parelius, J.M., and Singh, K. (1999), 'Multivariate Analysis by Data Depth: Descriptive Statistics, Graphics and Inference,(with Discussion and a Rejoinder by Liu and Singh)', *The Annals of Statistics*, 27(3), 783–858.

Long, J.P., and Huang, J.Z. (2015), 'A Study of Functional Depths', *arXiv preprint arXiv:1506.01332*.

López-Pintado, S., and Romo, J. (2009), 'On the Concept of Depth for Functional Data', *Journal of the American Statistical Association*, 104(486), 718–734.

López-Pintado, S., and Romo, J. (2011), 'A Half-region Depth for Functional Data', *Computational Statistics & Data Analysis*, 55(4), 1679–1695.

Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. (1995), 'Scop: a Structural Classification of Proteins Database for the Investigation of Sequences and Structures', *Journal of Molecular Biology*, 247(4), 536–540.

Nagy, S. (2017), 'Monotonicity Properties of Spatial Depth', *Statistics & Probability Letters*, 129, 373–378.

Nagy, S., Gijbels, I., and Hlubinka, D. (2017), 'Depth-based Recognition of Shape Outlying Functions', *Journal of Computational and Graphical Statistics*, 26(4), 883–893.

Narisetty, N.N., and Nair, V.N. (2016), 'Extremal Depth for Functional Data and Applications', *Journal of the American Statistical Association*, 111(516), 1705–1714.

Nicol, F. (2013), 'Functional Principal Component Analysis of Aircraft Trajectories', Ph.D. thesis, ENAC.

Nieto-Reyes, A. (2011), 'On the Properties of Functional Depth', in *Recent Advances in Functional Data Analysis and Related Topics*, Springer, pp. 239–244.

Nieto-Reyes, A., and Battey, H. (2016), 'A Topologically Valid Definition of Depth for Functional Data', *Statistical Science*, 31(1), 61–79.

Ramsay, J. (2005), 'Functional Data Analysis', *Encyclopedia of Statistics in Behavioral Science*

Ramsay, J.O., and Li, X. (1998), 'Curve Registration', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2), 351–363.

Riesz, F., and Nagy, B.S. (1990), *Functional Analysis*, New York: Frederick Ungar. 1955'. English Translation.

Rousseeuw, P.J., and Ruts, I. (1996), 'Algorithm As 307: Bivariate Location Depth', *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 45(4), 516–526.

Rousseeuw, P.J., and Struyf, A. (1998), 'Computing Location Depth and Regression Depth in Higher Dimensions', *Statistics and Computing*, 8(3), 193–203.

Sguera, C., Galeano, P., and Lillo, R. (2014), 'Spatial Depth-based Classification for Functional Data', *Test*, 23(4), 725–750.

Srivastava, A., Wu, W., Kurtek, S., Klassen, E., and Marron, J. (2011), 'Registration of Functional Data using Fisher-Rao Metric', *arXiv preprint arXiv:1103.3817*.

Tang, R., and Müller, H.-G. (2008), 'Pairwise Curve Synchronization for Functional Data', *Biometrika*, 95(4), 875–889.

Tran, N.M. (2008), 'An Introduction to Theoretical Properties of Functional Principal Component Analysis', *Department of Mathematics and Statistics, The University of Melbourne, Victoria, Australia*.

Tukey, J.W. (1975), 'Mathematics and the Picturing of Data', in *Proceedings of the International Congress of Mathematicians* (Vol. 2), Vancouver, 1975', pp. 523–531.

Vaart, A.v.d., and Zanten, H.v. (2011), 'Information Rates of Nonparametric Gaussian Process Methods', *Journal of Machine Learning Research*, 12(Jun 6), 2095–2119.

Vardi, Y., and Zhang, C.-H. (2000), 'The Multivariate L1-median and Associated Data Depth', *Proceedings of the National Academy of Sciences*, 97(4), 1423–1426.

Wahba, G. (1990), *Spline Models for Observational Data*, SIAM.

Wang, G., and Dunbrack Jr, R.L. (2003), 'Pisces: a Protein Sequence Culling Server', *Bioinformatics*, 19(12), 1589–1591.

Whitaker, R.T., Mirzargar, M., and Kirby, R.M. (2013), 'Contour Boxplots: A Method for Characterizing Uncertainty in Feature Sets From Simulation Ensembles', *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2713–2722.

Wu, W., Srivastava, A., Laborde, J., and Zhang, J. (2013), 'An Efficient Multiple Protein Structure Comparison Method and Its Application to Structure Clustering and Outlier Detection', in *2013 IEEE International Conference on Bioinformatics and Biomedicine*, IEEE, pp. 69–73.

Zuo, Y. (2003), 'Projection-based Depth Functions and Associated Medians', *The Annals of Statistics*, 31(5), 1460–1490.

Zuo, Y. (2018), 'A New Approach for the Computation of Halfspace Depth in High Dimensions', *Communications in Statistics-Simulation and Computation*, 48, 1–22.

Zuo, Y., and Serfling, R. (2000), 'General Notions of Statistical Depth Function', *Annals of Statistics*, 28(2), 461–482.

Appendices

Appendix 1. Algorithms for general model-based depth

Suppose we have n zero-mean independent sample functions $f_1, \ldots, f_n \in \mathcal{F}$ on $t \in [0, 1]$, and our goal is to compute the model-based depth of any observed sample $f_{obs} \in \mathcal{F}$. We first describe the *norm-based depth* estimation algorithm as follows:

Algorithm II: (Input: observations $\{f_1, \ldots, f_n\}$, any observation f_{obs}, a threshold $\epsilon > 0$, the centre function f_c, and the selected norm $\| \cdot \|$, which means $\zeta(f, f_c) = \|f - f_c\|$ for any observation f.

(1) Compute the sample mean function $\bar{f}(t) = \frac{1}{n}\sum_{i=1}^{n} f_i(t)$, and empirical covariance kernel $\hat{K}(s, t) = \frac{1}{n}\sum_{i=1}^{n}[f_i(s) - \bar{f}(s)][f_i(t) - \bar{f}(t)]$

(2) Eigen-decompose $\hat{K} = \sum_{p=1}^{n} \hat{\lambda}_{p,n}\hat{\phi}_{p,n}(s)\hat{\phi}_{p,n}(t)$

(3) Choose a number P if $\hat{\lambda}_{P+1}$ is the first eigenvalue such that $\hat{\lambda}_{P+1} < \epsilon$ for sufficiently small ϵ; then $\hat{K}(s, t) = \sum_{p=1}^{P} \hat{\lambda}_{p,n}\hat{\phi}_{p,n}(s)\hat{\phi}_{p,n}(t)$

(4) Compute $\hat{f}_{i,p} = \int_0^1 f_i(t)\hat{\phi}_{p,n}(t) \, \mathrm{d}t$ for all $i = 1, \ldots, n$ and $p = 1, \ldots, P$, and compute $\hat{f}_p = \int_0^1 f_{obs}(t)\hat{\phi}_{p,n}(t) \, \mathrm{d}t$

(5) Re-sample (with replacement) a large number N of coefficients $\{\hat{g}_{j,p}\}_{j=1}^{N}$ based on the coefficients of $\{\hat{f}_{i,p}\}$, and construct $g_j(t) = \sum_{p=1}^{P} \hat{g}_{j,p}\hat{\phi}_p(t)$

(6) Estimate the sample depth of f_{obs} w.r.t. $\{g_j\}$:

$$D_n(f_{obs}; \{g_j\}_{j=1}^{N}) = \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}_{(\zeta(f_{obs}, f_c) \leq \zeta(g_j, f_c))} = \frac{1}{N}\sum_{j=1}^{N}\mathbf{1}_{(\|f_{obs} - f_c\| \leq \|g_j - f_c\|)},$$

where $\mathbf{1}_{(\cdot)}$ is the indicator function.

Steps 1–4 are the first part in the algorithm. They aim to estimate the eigen system of the covariance kernel via given observations. In particular, the Karhunen Loève expansion (Ash 1990) is used in Step 2 to decompose the covariance kernel and offer a method to reconstruct samples (the background on the Karhunen Loève expansion will be provided in Section 3). Using a functional principal component analysis (Ramsay 2005), we retain the eigen functions which explain meaningful variance in our system. Steps 5 and 6 are the second part of the algorithm. They estimate the depth value with the given norm, where we need re-sampling techniques and bootstrap approximations. This algorithm can be easily adapted to the multivariate data. In such case, the dimension of the data is already given and the principal component analysis and the multivariate metric can be directly applied.

In general, computing a halfspace depth in $\mathbb{R}^d$ is a very challenging task. So far, exact computations can be given only when $d = 2$ (Rousseeuw and Ruts 1996) and $d = 3$ (Rousseeuw and Struyf 1998). There are approximation algorithms when $d \geq 4$ (Zuo 2018). However, if the data distribution is a multivariate normal, our framework will result in an optimal solution similar to that obtained for the Gaussian process. For infinite dimensional GP, Lemma 2.1 shows that the inner-product-based depth can only be feasible for finite-dimensional space. Fortunately, when the random samples are from a finite-dimensional zero-mean Gaussian process, the depth has simple

closed-form (see detail in Appendix 1). We adopt this special case and modify the above algorithm for halfspace depth as follows, where Steps 4–6 are simplified as follows:

(4)   Compute $\hat{f}_p = \int_0^1 f_{obs}(t)\hat{\phi}_{p,n}(t)\,dt$ for $p = 1,\ldots,P$

(5)   Compute the induced RKHS norm $\|f_{obs}\|_{\mathbb{H}_{\hat{K}}}^2 = \sum_{p=1}^P \frac{\hat{f}_p^2}{\hat{\lambda}_{p,n}}$

(6)   Compute the depth as $D_{ip}(f_{obs}) = 1 - \Phi(\|f_{obs}\|_{\mathbb{H}_{\hat{K}}})$, where $\Phi(x)$ denotes the cumulative distribution function of a standard normal random variable.

Appendix 2. Applications of norm-based depth in finite-dimensional process

Finite-dimensional process is a commonly used stochastic process in practical applications. In particular, this includes any finite-dimensional Gaussian Process (GP) and multivariate Gaussian distribution as special cases. In this appendix, we simplify our model in Section 3.2 into a zero-mean finite-dimensional process, which means that K has a finite number P of positive eigenvalues, and P will be referred as the dimension of this process. That is, $K(s,t) = \sum_{j=1}^P \lambda_j \phi_j(s)\phi_j(t)$. For convenience we denote this kernel as $K = K_P$. One important benefit in this process is that the associated RKHS norm is always finite and can be directly used in our construction of norm-induced depth as described in Section 5.1.

A.1   RKHS norm induced depth for finite-dimensional process

Suppose we have functional observation $f \in \mathbb{L}^2([0,1])$ from a zero-mean stochastic process with covariance kernel $K_P(s,t) = \sum_{p=1}^P \lambda_p \phi_p(s)\phi_p(t)$ on $[0,1] \times [0,1]$. Then $f(t) = \sum_{p=1}^P f_p \phi_p(t)$, where $f_1,\ldots,f_P$ are uncorrelated and $E(f_p) = 0, E(f_p^2) = \lambda_p, j = 1,\ldots,P$. In particular, when the process is a Gaussian process, $f_1,\ldots,f_P$ are independent. In this case, $\{X_p = f_p/\sqrt{\lambda_p}\}_{p=1}^P$ are i.i.d. samples from a standard normal distribution, and the squared induced norm $\|f\|_{\mathbb{H}_K}^2 = \sum_{p=1}^P f_p^2/\lambda_p = \sum_{p=1}^P X_p^2$ follows a χ^2 distribution with P degrees of freedom, denoted as $\chi^2(P)$.

The computation of depth still depends on Definition 2.2: $D(f_{obs}, \mathbb{P}_\theta, \|\cdot\|, f_c) = \mathbb{P}_\theta[\zeta(f,f_c) \geq \zeta(f_{obs},f_c)]$. The central function $f_c = 0$ is the mean function in our model; the criterion function $\zeta(f,g) = \|f-g\|_{H_K}$; $\mathbb{P}_\theta$ is a probability measure.

We can now rewrite the definition of depth in the following form:

$$D(f_{obs}, \mathbb{P}_P, \|\cdot\|_{\mathbb{H}_K}, 0) = \mathbb{P}_P \left[f : \|f\|_{\mathbb{H}_K} \geq \|f_{obs}\|_{\mathbb{H}_K} \right]$$
$$= 1 - \mathbb{P}_P \left[f : \|f\|_{\mathbb{H}_K}^2 \leq \|f_{obs}\|_{\mathbb{H}_K}^2 \right] = 1 - F(\|f_{obs}\|_{\mathbb{H}_K}^2), \qquad (A1)$$

where $F(\cdot)$ denotes the cumulative distribution function (c.d.f.) of $\|f\|_{\mathbb{H}_K}^2$. In the case of Gaussian process, this is a c.d.f. of $\chi^2(P)$. Moreover, for any $\alpha \in [0,1]$, the αth depth contour is rewritten as

$$C(\alpha, \mathbb{P}_P, \|\cdot\|_{\mathbb{H}_K}, 0) = \left\{ f \in \mathcal{F} : F(\|f_{obs}\|_{\mathbb{H}_K}^2) = 1 - \alpha \right\},$$

and central region for this model is

$$R(\alpha, \mathbb{P}_P, \|\cdot\|_{\mathbb{H}_K}, 0) = \left\{ f \in \mathcal{F} : F(\|f_{obs}\|_{\mathbb{H}_K}^2) \leq 1 - \alpha \right\}.$$

Based on the above derivation, it is easy to see that the depth contours defined via induced RKHS norm on a Gaussian process are P-dimensional ellipsoids, and the centre of all ellipsoids is the origin in $\mathbb{R}^P$. For illustrative purpose, we let $P = 2$ and $(f_1, f_2) \sim \mathcal{N}(0, \Sigma)$, with $\Sigma = diag(\lambda_1, \lambda_2)$. For any random samples $f(t) = \sum_{j=1}^2 f_p \phi_p(t)$, we could use a point $(f_1, f_2) \in \mathbb{R}^2$ to represent random function $f(t)$, because the coefficients set for each $f(t)$ is unique with respect to the eigen-functions basis. In Figure A1, if we have any (f_1, f_2) locating on the same ellipsoid, their corresponding random observations will have the same depth defined by the induced RKHS norm. In particular, when $\Sigma = I_2$, the depth contours are concentric circles. Moreover, any random observations $f(t)$, whose coefficients (f_1, f_2) locates inside of αth contour, will have a larger depth than α.

(a) $\Sigma = I_2$ (b) $\Sigma = diag(1, 0.25)$
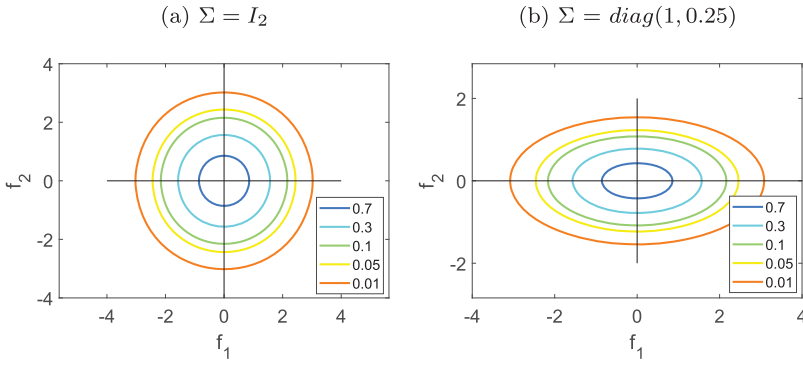
Figure A1. Illustration depth contours in the 2-D case where *x*-axis and *y*-axis represent the values of f_1 and f_2, respectively: (a) depth contours at different levels for zero mean Gaussian Process with $(f_1, f_2) \sim N(0, I_2)$ and (b) same as (a) except that $(f_1, f_2) \sim diag(1, 0.25)$.

A.2 Depth estimation procedure and algorithm

Similar to the infinite-dimensional case, we can derive algorithm to compute depth on a finite-dimensional stochastic process. Suppose we have n independent random sample functions $\{f_1, \ldots, f_n\} \subseteq \mathcal{F}$ on $t \in [0, 1]$, and $\mathcal{F}$ is a zero-mean P-dimensional stochastic process. The following algorithm is to compute the depth based on $\mathcal{F}$ of any observed sample $f_{obs} \in \mathcal{F}$. In practice when P is unknown, we can set a small threshold ϵ to identify it.

Algorithm III: (Input: functional data $\{f_1, \ldots, f_n\}$, any observation f_{obs}, and a threshold $\epsilon > 0$.

(1) Compute the sample mean function $\bar{f}(t) = \frac{1}{n} \sum_{i=1}^{n} f_i(t)$, and empirical covariance kernel $\hat{K}(s, t) = \frac{1}{n} \sum_{i=1}^{n} [f_i(s) - \bar{f}(s)][f_i(t) - \bar{f}(t)]$;

(2) Eigen-decompose $\hat{K} = \sum_{p=1}^{n} \hat{\lambda}_{p,n} \hat{\phi}_{p,n}(s) \hat{\phi}_{p,n}(t)$;

(3) Choose a number P if $\hat{\lambda}_{P+1}$ is the first eigenvalue such that $\hat{\lambda}_{P+1} < \epsilon$ for sufficiently small ϵ; then $\hat{K}(s, t) = \sum_{p=1}^{P} \hat{\lambda}_{p,n} \hat{\phi}_{p,n}(s) \hat{\phi}_{p,n}(t)$;

(4) Compute $\hat{f}_{i,p} = \int_0^1 f_i(t) \hat{\phi}_{p,n}(t) \, dt$ for all $i = 1, \ldots, n$ and $p = 1, \ldots, P$, and compute $\hat{f}_p = \int_0^1 f_{obs}(t) \hat{\phi}_{p,n}(t) \, dt$;

(5) For each $p \in 1, \ldots, P$, re-sample (with replacement) a large number N of coefficients $\{\hat{g}_{j,p}\}_{j=1}^{N}$ based on $\{\hat{f}_{1,p}, \ldots, \hat{f}_{n,p}\}$;

(6) Construct $g_j(t) = \sum_{p=1}^{P} \hat{g}_{j,p} \hat{\phi}_{p,n}(t)$;

(7) Compute $||f_{obs}||^2_{\mathbb{H}_{\hat{K}}} = \sum_{p=1}^{P} \frac{\hat{f}_p^2}{\hat{\lambda}_{p,n}}$, and $||g_j||^2_{\mathbb{H}_{\hat{K}}} = \sum_{p=1}^{P} \frac{\hat{g}_{j,p}^2}{\hat{\lambda}_{p,n}}$;

(8) Estimate the depth of f_{obs} using $\{g_j\}$:

$$D(f_{obs}; \{g_j\}_{j=1}^{N}) = \frac{1}{N} \sum_{j=1}^{N} 1_{||f_{obs}||^2_{\mathbb{H}_{\hat{K}}} \leq ||g_j||^2_{\mathbb{H}_{\hat{K}}}}.$$

This algorithm is very similar to Algorithm II. The first three steps are to estimate the eigensystem of the covariance kernel via our observations. As there are only finite number P of positive eigenvalues, we can set small threshold to estimate P. Steps 4–8 are to estimate the modified RKHS norm by resampling based on the eigen-decomposition on the covariance.
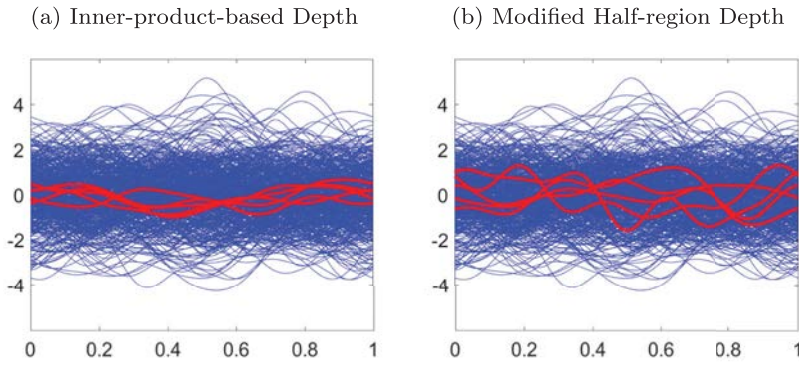
Figure A2. Simulation 3: (a) all 500 simulated functions, where the red curves have the five deepest values obtained by the proposed method and (b) same as (a) except that the depth values are obtained by the modified half-region depth.

An important special case is when the process is a Gaussian process. In this case, we have pointed out the squared norm $\|f\|_{\mathbb{H}_{\hat{K}}}^2$ has a Chi-square distribution. Therefore resampling will not be needed and the estimation of depth will be more robust and efficient. Steps 4–8 can be simplified and modified to the following three steps:

(4) Compute $\hat{f}_p = \int_0^1 f_{obs}(t)\hat{\phi}_{p,n}(t)\,\mathrm{d}t$ for all $i = 1, \ldots, n$ and $p = 1, \ldots, P$

(5) Compute the induced RKHS norm $\|f_{obs}\|_{\mathbb{H}_{\hat{K}}}^2 = \sum_{p=1}^{P} \frac{\hat{f}_p^2}{\hat{\lambda}_{p,n}}$

(6) Compute the depth as $D = 1 - F(\|f_{obs}\|_{\mathbb{H}_{\hat{K}}}^2)$, where $F(x)$ denotes the cumulative distribution function of $\chi^2(P)$.

Appendix 3. More simulation examples

Simulation 5. In this example, we illustrate the inner-product-based depth computation. We first select a sequence of orthonormal Fourier basis functions up to order $P = 10$ on $[0, 1]$ such that

$$\phi_p(t) = \begin{cases} 1 & p = 1 \\ \sqrt{2}\cos(\pi p t) & p = 2, 4, 6, 8, 10 \\ \sqrt{2}\sin(\pi(p-1)t) & p = 3, 5, 7, 9 \end{cases}$$

Next we random generate $N = 500$ coefficient vectors $\{(a_{i,1}, \ldots, a_{i,10})\}_{i=1}^N$ following a multivariate normal distribution $\mathcal{N}(0, diag(1, ((P-1)/P)^2, \ldots, (1/P)^2))$. Then we generate N functions via linear combination $f_i = \sum_{p=1}^{P} a_{i,p}\phi_p$. We apply Algorithm I for inner-product depth discussed in the above section on this simulated data. We display these 500 functions in Figure A2(a), where the five deepest curves are represented in bold red. We see that these five red ones stay in the middle of the sample, which illustrate the effectiveness of the depth measurement. As a comparison, we also show the result obtained by modified half-region depth (López-Pintado and Romo 2011) and the result is shown in Figure A2(b). Visually, the five deepest functions displayed in Panel (a) seem to be more centralised near x-axis, and our method provide better centre-outward rank than the modified half-region depth.

Simulation 6. In this example, we illustrate the norm-based depth on a multivariate data set and compare the Monte Carlo estimate with sample average (as indicated in the beginning of this section). We at first generate $n = 50$ random samples in $\mathbb{R}^2$ from multivariate normal distribution

(a) Given observations   **(b) Depth value comparison 1**   **(c) Depth value comparison 2**
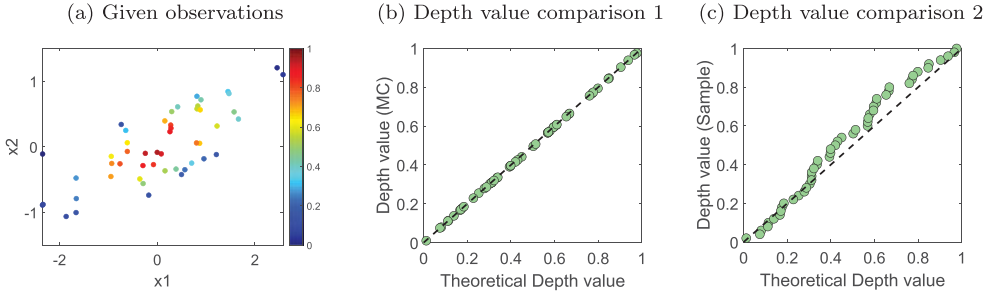
Figure A3. Simulation 4: (a) 50 points from multivariate normal density with colour-labelled depth by Algorithm I; (b) depth value comparison: closed-form depth function (*x*-axis) vs. Monte-Carlo-based Algorithm I (*y*-axis) and (c) same as (b) except for sample average method in *y*-axis.

$\mathcal{N}(\mu, \Sigma)$, where

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 1 & 1/3 \\ 1/3 & 1/4 \end{pmatrix}.$$

We choose a Mahalanobis distance as the criterion function, that is, for any $x \in \mathbb{R}^2$, $\zeta(x, \mu) = \sqrt{(x - \mu)^T \Sigma^{-1}(x - \mu)}$. Therefore, it is straightforward to derive the closed form of the depth function $D(x) = 1 - F(\zeta(x, \mu)^2)$, where $F(\cdot)$ denotes the cumulative distribution function of chi-square distribution with 2 degrees of freedom.

We compute the depth value for each of these 50 points by Monte-Carlo-based Algorithm I, and then compare the result to the algorithm integrated with sample average of these points. We display these 50 points with colour label of their depth values in Figure A3(a). Note that the depth value using the Mahalanobis distance criterion ranges from 0 to 1 and the distribution of these depth values approximately follow elliptic contours for a two dimensional normal distribution. Since we obtain the closed-form depth values, we can use them to compare the performance of Monte Carlo and sample average method. In Algorithm I, we generate 5000 re-sampling points in step 5. The results in Figure A3(b,c) show that the depth values computed by Algorithm I are very close to the theoretical ones, whereas the sample average method does not have the same level of accuracy.

Appendix 4. Depth estimation consistency in finite-dimensional data

In this case, $\exists P \in \mathbb{N}$ such that $\lambda_P > 0$ and $\lambda_p = 0$, $\forall p > P$ under the notation setup in Section 4 of the main paper. Then $K(s, t) = \sum_{p=1}^{P} \lambda_p \phi_p(s) \phi_p(t)$ and $\hat{K}(s, t) = \sum_{p=1}^{n} \hat{\lambda}_{p,n} \hat{\phi}_{p,n}(s) \hat{\phi}_{p,n}(t)$. Therefore, for any $f_{obs} \in \mathcal{F}$, we have the squared RKHS induced norm

$$\|f_{obs}\|_{\mathbb{H}_K}^2 = \sum_{p=1}^{P} \frac{\langle f_{obs}, \phi_p \rangle^2}{\lambda_p} < \infty, \tag{A2}$$

where $\langle \cdot, \cdot \rangle$ indicates the inner product operation in RKHS with K as reproducing kernel.

Based on the RKHS norm, the depth of f_{obs} is given as follows:

$$d(f_{obs}) = D_n(f_{obs}, \mathbb{P}, \| \cdot \|_{\mathbb{H}_K}, 0) = \mathbb{P}\left[f : \|f\|_{\mathbb{H}_K} \geq \|f_{obs}\|_{\mathbb{H}_K} \right] = 1 - F(\|f_{obs}\|_{\mathbb{H}_K}^2), \tag{A3}$$

where $F(x)$ denotes the cumulative distribution function of $\|f\|_{\mathbb{H}_K}^2$ for all $f \in \mathcal{F}$.

As given in Algorithm II in Appendix 7, the sample version of the squared modified norm is given as

$$\|f_{obs}\|_{\mathbb{H}_{\hat{K}}}^2 = \sum_{p=1}^{n} \frac{\langle f_{obs}, \hat{\phi}_{p,n} \rangle^2}{\hat{\lambda}_{p,n}}. \tag{A4}$$

Similar to Case I, we adopt the sample version of the depth of f_{obs}

$$d_n(f_{obs}) = \mathbb{P}\left[f : \|f\|_{\mathbb{H}_K} \geq \|f_{obs}\|_{\mathbb{H}_{\hat{K}}}\right] = 1 - F(\|f_{obs}\|^2_{\mathbb{H}_{\hat{K}}}). \tag{A5}$$

We focus on proving $d_n(f_{obs})$ converges to $d(f_{obs})$ when n is large. This is shown in Theorem A.1 as follows. In this case, neither a convergent weight series $\{a_p\}$ nor Assumption 4.1 is needed in the proof of consistency.

Theorem A.1: *If the covariance kernel K has only $P(\in \mathbb{N})$ positive eigenvalues $\{\lambda_p\}$, then we have*

$$\sup_{f_{obs} \in \mathcal{F}, \|f_{obs}\| \leq 1} |\|f_{obs}\|^2_{\mathbb{H}_{\hat{K}}} - \|f_{obs}\|^2_{\mathbb{H}_K}| \xrightarrow{a.s.} 0.$$

Moreover, for any $f_{obs} \in \mathcal{F}$

$$\lim_{n \to \infty} d_n(f_{obs}) = d(f_{obs}).$$

Proof: As convergence almost surely implies convergence in distribution, it is apparent that we only need to prove the first convergence $\|f_{obs}\|^2_{\mathbb{H}_{\hat{K}}} \xrightarrow{a.s.} \|f_{obs}\|^2_{\mathbb{H}_K}$.

Based on the work done by Dauxois, Pousse, and Romain (1982) and Bosq (2012), when n is large, we have $\hat{\lambda}_{p,n} \xrightarrow{a.s.} \lambda_p > 0$ for $p \in \{1, \ldots, P\}$, while $\hat{\lambda}_{p,n} \xrightarrow{a.s.} 0$ for $p \in \{P+1, \ldots, n\}$(Tran 2008).

We denote $\check{K}(s,t) = \sum_{p=1}^{P} \hat{\lambda}_{p,n}\hat{\phi}_{p,n}(s)\hat{\phi}_{p,n}(t)$, and we will get $\|\check{K} - \hat{K}\| \xrightarrow{a.s.} 0$ as $n \to \infty$. Besides, we have $\|\hat{K} - K\| \xrightarrow{a.s.} 0$(Dauxois et al. 1982), hence $\|\check{K} - K\| \xrightarrow{a.s.} 0$.

If we denote $K^{-1}(s,t) = \sum_{p=1}^{P} \frac{\phi_p(s)\phi_p(t)}{\lambda_p}$ and $\check{K}^{-1}(s,t) = \sum_{p=1}^{P} \frac{\hat{\phi}_{p,n}(s)\hat{\phi}_{p,n}(s)}{\hat{\lambda}_{p,n}}$,

$$\|\check{K}^{-1} - K^{-1}\| = \|\check{K}^{-1}(\check{K} - K)K^{-1}\|$$

$$\leq \|\check{K}^{-1}\|\|\check{K} - K\|\|K^{-1}\|$$

$$= \frac{1}{\hat{\lambda}_{p,n}}\frac{1}{\lambda_p}\|\check{K} - K\| \xrightarrow{a.s.} 0.$$

In Algorithm II, the estimated depth is written as

$$\|f_{obs}\|^2_{\mathbb{H}_{\check{K}}} = \sum_{p=1}^{P} \frac{\langle f_{obs}, \hat{\phi}_{p,n}\rangle^2}{\hat{\lambda}_{p,n}}$$

$$= \sum_{p=1}^{P} \int_0^1 \int_0^1 f_{obs}(s)f_{obs}(t)\frac{\hat{\phi}_{p,n}(s)\hat{\phi}_{p,n}(s)}{\hat{\lambda}_{p,n}} \, ds \, dt$$

$$= \int_0^1 \int_0^1 f_{obs}(s)f_{obs}(t)\check{K}^{-1}(s,t) \, ds \, dt$$

Therefore,

$$| \|f_{obs}\|^2_{\mathbb{H}_{\check{K}}} - \|f_{obs}\|^2_{\mathbb{H}_K} |$$

$$= | \int_0^1 \int_0^1 f_{obs}(s)f_{obs}(t)\check{K}^{-1}(s,t) \, ds \, dt - \int_0^1 \int_0^1 f_{obs}(s)f_{obs}(t)K^{-1}(s,t) \, ds \, dt |$$

$$= | \langle f_{obs}, (\check{K}^{-1} - K^{-1})f_{obs}\rangle |$$

$$\leq \|f_{obs}\|^2\|\check{K}^{-1} - K^{-1}\| \xrightarrow{a.s.} 0.$$

∎

Appendix 5. Proof of Lemma 2.1

To better streamline the proof, we first prove the claimed result when the inner-product is induced from the reproducing kernel Hilbert space (RKHS) associated with the covariance function of the GP (which satisfies the Gram matrix condition of the lemma), and then extend the proof to a general inner product. Specifically, we show:

(1) (basic form) If we take the induced RKHS (reproducing-kernel Hilbert space) inner-product $< \cdot, \cdot >$ using the covariance function C, then

$$D_{ip}(f_{obs}, \mathbb{P}_C, \langle \cdot, \cdot \rangle, \mathcal{F}) = 0$$

almost surely for $f_{obs} \in \mathrm{GP}(0, C)$

(2) (general form) The above result will in fact hold for any inner-product on $\mathcal{F}$ that satisfies the condition in the lemma.

Proof: (Part 1) Based on the result in Sec 3.1.2, assume the covariance function $C(\cdot, \cdot)$ in a Gaussian process $\mathrm{GP}(0, C)$ has infinite number of positive eigenvalues. Then the covariance can be represented as $C(s, t) = \sum_{p=1}^{\infty} \lambda_p \phi_p(s) \phi_p(t)$. For any $f_{obs} \in GP(0, C)$, let $f_{obs,p} = \int_0^1 f_{obs}(s) \phi_p(s) \, ds$. We have $f_{obs}(t) = \sum_{p=1}^{\infty} f_{obs,p} \phi_p(t)$. Hence, the induced RKHS norm

$$\|f_obs\|_{\mathbb{H}_C} = \sum_{p=1}^{\infty} \frac{f_{obs,p}^2}{\lambda_P} = \infty \quad (a.s.)$$

For any integer $P > 0$ and function $f \in \mathcal{F}$, we let f^P represent the finite cutoff of f at the Pth order. That is, $f^P(t) = \sum_{p=1}^{P} f_p \phi_p(t)$. Let $\mathcal{G}_P$ denote the finite-dimensional space expanded by $\{\phi_p(t)\}_{p=1}^{P}$. Using the result in Appendix 1, the inner-product depth

$$D_{ip}(f_{obs}^P, \mathbb{P}_C, \langle \cdot, \cdot \rangle, \mathcal{G}_P) = 1 - \Phi(\|f_{obs}^P\|_{\mathbb{H}_C}). \tag{A6}$$

Note that $\|f_{obs}^P\|_{\mathbb{H}_C} = \sum_{p=1}^{P} \frac{f_{obs,p}^2}{\lambda_p} \to \infty (a.s.)$ as $P \to \infty$. Then $1 - \Phi(\|f_{obs}^P\|_{\mathbb{H}_C}) \to 1 - \Phi(\infty) = 1 - 1 = 0$ $(a.s.)$ Finally, we have

$$D_{ip}(f_{obs}, \mathbb{P}_C, \langle \cdot, \cdot \rangle, \mathcal{F}) \leq \inf_P D_{ip}(f_{obs}^P, \mathbb{P}_C, \langle \cdot, \cdot \rangle, \mathcal{G}_P) \to 0. \quad (a.s.)$$

(Part 2) We see that the proof of Part 1 mainly relies on the result in Appendix 1 (Equation A6), where we use the induced RKHS inner product. Let f be a realisation from the Gaussian process $\mathrm{GP}(0, C)$. Here we just need to show that using the new inner product, such equation will still hold. Again, we consider the finite cut-off of f_{obs}^P at the Pth order and will show that Equation (A6) remains valid with the new inner product $\langle \cdot, \cdot \rangle$. Therefore, we suppress the superscript P in proving this equation in the rest of this part. Under this notation, we can write

$$f(t) = \sum_{p=1}^{P} f_p \phi_p(t) \in \mathbb{H}_K,$$

$$g(t) = \sum_{p=1}^{P} g_p \phi_p(t) \in \mathbb{H}_K,$$

$$f_{obs}(t) = \sum_{p=1}^{P} f_{obs,p} \phi_p(t) \neq \mathbf{0} \in \mathbb{H}_K,$$

where f_p are independent normal random variables with $\mathbb{E} f_p = 0$ and $Var f_p = \lambda_p, p = 1, \ldots, P$.

For this new inner product $< \cdot, \cdot >$ on $\mathcal{F}$, let $r_{ij} =< \phi_i, \phi_j >$ for $1 \leq i, j \leq P$. We also denote

$$X = \langle f - f_{obs}, g \rangle = \sum_{i=1}^{P} \sum_{j=1}^{P} (f_i - f_{obs,i}) g_j r_{ij} = - \sum_{i=1}^{P} \sum_{j=1}^{P} f_{obs,i} g_j r_{ij} + \sum_{i=1}^{P} \sum_{j=1}^{P} f_i g_j r_{ij}.$$

It is straightforward to know that X is normally distributed with $\mathbb{E}X = -\sum_{i=1}^{P} \sum_{j=1}^{P} f_{obs,i} g_j r_{ij} := \mu_g$ and $VarX = \sum_{i=1}^{P} (\sum_{j=1}^{P} g_j r_{ij})^2 \lambda_i := \sigma_g^2$. Now we can compute the probability

$$\mathbb{P}_{\theta} \left[\langle f - f_{obs}, g \rangle \geq 0 \right] = \mathbb{P}_{\theta} \left[X \geq 0 \right] = \mathbb{P}_{\theta} \left[\frac{X - \mu_g}{\sigma_g} \geq -\frac{\mu_g}{\sigma_g} \right] = 1 - \mathbb{P}_{\theta} \left[\frac{X - \mu_g}{\sigma_g} \leq -\frac{\mu_g}{\sigma_g} \right].$$

With the normal assumption, $\mathbb{P}_{\theta} [< f - f_{obs}, g >_{mod} \geq 0] = 1 - \Phi(-\frac{\mu_g}{\sigma_g})$ where Φ is the c.d.f. of a standard normal random variable $\frac{X - \mu_g}{\sigma_g}$ (it does not depend on g). To minimise the probability with respect to g, we need to maximise $-\mu_g/\sigma_g$, or μ_g^2/σ_g^2.

By the Cauchy inequality, we have

$$\frac{\mu_g^2}{\sigma_g^2} = \frac{\left(\sum_{i=1}^{P} \sum_{j=1}^{P} f_{obs,i} g_j r_{ij} \right)^2}{\sum_{i=1}^{P} \left(\sum_{j=1}^{P} g_j r_{ij} \right)^2 \lambda_i} = \frac{\left(\sum_{i=1}^{P} \frac{f_{obs,i}}{\sqrt{\lambda_i}} \sqrt{\lambda_i} \sum_{j=1}^{P} g_j r_{ij} \right)^2}{\sum_{i=1}^{P} \left(\sum_{j=1}^{P} g_j r_{ij} \right)^2 \lambda_i}$$

$$\leq \frac{\sum_{i=1}^{P} \left(\frac{f_{obs,i}}{\sqrt{\lambda_i}} \right)^2 \sum_{i=1}^{P} (\sqrt{\lambda_i} \sum_{j=1}^{P} g_j r_{ij})^2}{\sum_{i=1}^{P} \left(\sum_{j=1}^{P} g_j r_{ij} \right)^2 \lambda_i} = \sum_i \frac{f_{obs,i}^2}{\lambda_i}.$$

The equality holds if and only if there exists $c > 0$ such that $c \frac{f_{obs,i}}{\sqrt{\lambda_i}} = \sqrt{\lambda_i} \sum_{j=1}^{P} g_j r_{ij}, i = 1, \ldots, P$. That is,

$$\sum_{j=1}^{P} g_j r_{ij} = c \frac{f_{obs,i}}{\lambda_i}, \quad i = 1, 2, \ldots, P.$$

Under the condition on the inner-product in the lemma, this set of linear equations always admits a unique solution. By plugging-in this solution, the maximum of $-\frac{\mu_g}{\sigma_g}$ is obtained at

$$-\frac{\sum_{i=1}^{P} f_{obs,i} \sum_{j=1}^{P} g_j r_{ij}}{\sqrt{\sum_{i=1}^{P} \left(\sum_{j=1}^{P} g_j r_{ij} \right)^2 \lambda_i}} = \frac{\sum_{i=1}^{P} f_{obs,i} c \frac{f_{obs,i}}{\lambda_i}}{\sqrt{\sum_{i=1}^{P} (c \frac{f_{obs,i}}{\lambda_i})^2 \lambda_i}} = \sqrt{\sum_{i=1}^{P} \frac{f_{obs,i}^2}{\lambda_i}} = \|f_{obs}\|_{\mathbb{H}_C}.$$

Finally, the depth of f_{obs} is still given in the following form:

$$D_{ip}(f_{obs}, \mathbb{P}_{\theta}, \langle \cdot, \cdot \rangle, \mathcal{F}) = 1 - \Phi(\|f_{obs}\|_{\mathbb{H}_C}).$$

∎

Appendix 6. Optimal solution of the inner-product-based depth

We assume that $f \in \mathcal{F} (\subset \mathbb{L}^2([0, 1]))$ is random realisations from one zero-mean Gaussian process with covariance kernel K in a finite Karhunen Loève expansion $K(s, t) = \sum_{p=1}^{P} \lambda_p \phi_p(s) \phi_p(t), s, t \in [0, 1]$. As discussed in Section 3.1.2, the realisations from the Gaussian process form an RKHS $\mathbb{H}_K$. Let

$$f(t) = \sum_{p=1}^{P} f_p \phi_p(t) \in \mathbb{H}_K,$$

$$g(t) = \sum_{p=1}^{P} g_p \phi_p(t) \in \mathbb{H}_K,$$

$$f_{obs}(t) = \sum_{p=1}^{P} f_{obs,p} \phi_p(t) \neq \mathbf{0} \in \mathbb{H}_K,$$

where f_p are independent normal random variables with $\mathbb{E} f_p = 0$ and $Var f_p = \lambda_p, p = 1, \ldots, P$.

Using the inner product, we denote

$$X = \langle f - f_{obs}, g \rangle_{\mathbb{H}_K} = \sum_{p=1}^{P} \frac{(f_p - f_{obs,p}) g_p}{\lambda_p} = -\sum_{p=1}^{P} \frac{f_{obs,p} g_p}{\lambda_p} + \sum_{p=1}^{P} \frac{f_p g_p}{\lambda_p}.$$

It is straightforward to know that X is normally distributed with $\mathbb{E}X = -\sum_{p=1}^{P} \frac{f_{obs,p} g_p}{\lambda_p} := \mu_g$ and $Var X = \sum_{p=1}^{P} \frac{g_p^2}{\lambda_p} := \sigma_g^2$. Now we can compute the probability

$$\mathbb{P}_\theta \left[\langle f - f_{obs}, g \rangle_{\mathbb{H}_K} \geq 0 \right] = \mathbb{P}_\theta \left[X \geq 0 \right] = \mathbb{P}_\theta \left[\frac{X - \mu_g}{\sigma_g} \geq -\frac{\mu_g}{\sigma_g} \right] = 1 - \mathbb{P}_\theta \left[\frac{X - \mu_g}{\sigma_g} \leq -\frac{\mu_g}{\sigma_g} \right].$$

With the normal assumption, $\mathbb{P}_\theta[< f - f_{obs}, g >_{\mathbb{H}_K} \geq 0] = 1 - \Phi(-\frac{\mu_g}{\sigma_g})$ where Φ is the c.d.f. of a standard normal random variable $\frac{X - \mu_g}{\sigma_g}$ (it does not depend on g). To minimise the probability with respect to g, we need to maximise $-\mu_g/\sigma_g$ or μ_g^2/σ_g^2.

Let

$$a_p = \frac{f_{obs,p}}{\lambda_p}, \quad b_p = \frac{1}{\sqrt{\lambda_p}}.$$

Then use the Cauchy inequality,

$$\frac{\mu_g^2}{\sigma_g^2} = \frac{\left(\sum_p a_p g_p \right)^2}{\sum_p b_p^2 g_p^2} = \frac{\left(\sum_p \frac{a_p}{b_p} b_p g_p \right)^2}{\sum_p b_p^2 g_p^2} \leq \frac{\sum_p \left(\frac{a_p}{b_p} \right)^2 \sum_p (b_p g_p)^2}{\sum_p b_p^2 g_p^2} = \sum_{p=1}^{P} \left(\frac{a_p}{b_p} \right)^2.$$

The equality holds if and only if there exists $c > 0$ such that $c \frac{a_p}{b_p} = b_p g_p, p = 1, \ldots, P$. That is,

$$g_p = c \frac{a_p}{b_p^2} = c \cdot \frac{f_{obs,p}}{\lambda_p} \cdot \lambda_p = c f_{obs,p}.$$

With the constraint $||g||_{\mathbb{H}_K} = 1$,

$$1 = \langle g, g \rangle_{\mathbb{H}_K} = \sum_{p=1}^{P} \frac{g_p^2}{\lambda_k} = \sum_{p=1}^{P} \frac{c^2 f_{obs,p}^2}{\lambda_p} = c^2 \sum_{p=1}^{P} \frac{f_{obs,p}^2}{\lambda_p} = c^2 ||f_{obs}||_{\mathbb{H}_K}^2.$$

Therefore, $c^2 = \frac{1}{||f_{obs}||_{\mathbb{H}_K}^2}$ and $g_p = \frac{f_{obs,p}}{||f_{obs}||_{\mathbb{H}_K}}$. We have found the optimal solution

$$g^*(t) = \arginf_{g \in \mathcal{F}, ||g||_{\mathbb{H}_K} = 1} \mathbb{P}_\theta \left[\langle f - f_{obs}, g \rangle_{\mathbb{H}_K} \geq 0 \right] = \sum_{p=1}^{P} \frac{f_{obs,p}}{||f_{obs}||_{\mathbb{H}_K}} \phi_p(t) = \frac{f_{obs}(t)}{||f_{obs}||_{\mathbb{H}_K}}.$$

With this optimal g^*,

$$-\frac{\mu_{g^*}}{\sigma_{g^*}} = -\frac{-\sum_{p=1}^{P} \frac{f_{obs,p} g_p}{\lambda_p}}{\sqrt{\sum_{p=1}^{P} \frac{g_p^2}{\lambda_p}}} = \frac{\sum_{p=1}^{P} \frac{f_{obs,p}}{\lambda_p} \cdot \frac{f_{obs,p}}{||f_{obs}||_{\mathbb{H}_K}}}{\sqrt{\sum_{p=1}^{P} \frac{f_{obs,p}^2}{||f_{obs}||_{\mathbb{H}_K}^2} \cdot \frac{1}{\lambda_p}}} = ||f_{obs}||_{\mathbb{H}_K}.$$

Finally, the depth of f_{obs} is given in the following form:

$$D_{ip}(f_{obs}) := D_{ip}(f_{obs}, \mathbb{P}_\theta, \langle \cdot, \cdot \rangle, \mathcal{F}) = 1 - \Phi(\|f_{obs}\|_{\mathbb{H}_K}).$$

Appendix 7. Proof of Lemma 2.2

Proof: (1) Norm-based depth in general form:
- P-2: By definition, the general depth $D_n(f_{obs}, \mathbb{P}_\theta, \|\cdot\|, f_c)$ is strictly decreasing with respect to $\|f_{obs} - f_c\|$. As $\|f_{obs} - f_c\| \geq \|f_c - f_c\| = 0$, we have $D_n(f_{obs}, \mathbb{P}_\theta, \|\cdot\|, f_c) \geq D_n(f_c, \mathbb{P}_\theta, \|\cdot\|, f_c)$.
- P-3: For any $\alpha \in (0,1)$, $\|f_c + \alpha(f_{obs} - f_c) - f_c\| = \alpha\|f_{obs} - f_c\| \leq \|f_{obs} - f_c\|$. By Definition 2.1, $D_n(f_c + \alpha(f_{obs} - f_c), \mathbb{P}_\theta, s, f_c) \geq D_n(f_{obs}, \mathbb{P}_\theta, \|\cdot\|, f_c)$.
- P-4: Obvious.

(2) Norm-based depth in specific form:
- P-1: $D_n(af_{obs} + h, \mathbb{P}_{\theta,aF+h}, \|\cdot\|, af_c + h) = \mathbb{P}_\theta[f : \|af + h - (af_c + h)\| \geq \|af_{obs} + h - (af_c + h)\|] = \mathbb{P}_\theta[f : \|f - f_c\| \geq \|f_{obs} - f_c\|] = D_n(f_{obs}, \mathbb{P}_{\theta,F}, \|\cdot\|, f_c)$
- P-2: $D_n(f_{obs}, \mathbb{P}_\theta, \|\cdot\|, f_c) = \mathbb{P}_\theta[f : \|f - f_c\| \geq \|f_{obs} - f_c\|] \leq \mathbb{P}_\theta[f : \|f - f_c\| \geq 0] = D_n(f_c, \mathbb{P}_\theta, \|\cdot\|, f_c)$
- P-3: For any $\alpha \in (0,1)$, $D_n(f_c + \alpha(f_{obs} - f_c), \mathbb{P}_\theta, s, f_c) = \mathbb{P}_\theta[f : \|f - f_c\| \geq \|f_c + \alpha(f_{obs} - f_c) - f_c\|] = \mathbb{P}_\theta[f : \|f - f_c\| \geq \alpha\|(f_{obs} - f_c)\|] \geq \mathbb{P}_\theta[f : \|f - f_c\| \geq \|f_{obs} - f_c\|] = D_n(f_{obs}, \mathbb{P}_\theta, \|\cdot\|, f_c)$.
- P-4: $D_n(f_{obs}, \mathbb{P}_\theta, \|\cdot\|, f_c) = \mathbb{P}_\theta[f : \|f - f_c\| \geq \|f_{obs} - f_c\|] \leq \mathbb{P}_\theta[f : \|f\| \geq \|f_{obs}\| - 2\|f_c\|] \to 0$ (as $\|f_{obs}\| \to \infty$).

(3) Inner-product-based depth:
  (a) P-1:

$$D_{ip}(af_{obs} + h, \mathbb{P}_{\theta,aF+h}, \langle \cdot, \cdot \rangle, \mathcal{G})$$

$$= \inf_{g \in \mathcal{G}, \|g\|=1} \mathbb{P}_\theta\left[f \in \mathcal{F} : \langle af + h, g \rangle \geq \langle af_{obs} + h, g \rangle\right]$$

$$= \inf_{g \in \mathcal{G}, \|g\|=1} \mathbb{P}_\theta\left[f \in \mathcal{F} : a\langle f, g \rangle \geq a\langle f_{obs}, g \rangle\right]$$

$$= D_{ip}(f_{obs}, \mathbb{P}_{\theta,F}, \langle \cdot, \cdot \rangle, \mathcal{G})$$

  (b) P-2: It is straightforward to prove this property followed by Assumption 2.1. For any $g \in \mathcal{G}$, it is easy to verify that the set $\{f \in \mathcal{F} : <f - f_c, g >\geq 0\}$ is a closed halfspace that contains f_c. By Assumption 2.1, $D_{ip}(f_c, \mathbb{P}_{\theta,F}, < \cdot, \cdot >, \mathcal{G}) = \inf_{g \in \mathcal{G}, \|g\|=1} \mathbb{P}_\theta[f \in \mathcal{F} : <f - f_c, g >\geq 0] \geq 1/2$. Assume $h(\neq f_c) \in \mathcal{F}$ satisfies that $D_{ip}(h, \mathbb{P}_{\theta,F}, < \cdot, \cdot >, \mathcal{G}) > 1/2$. Then for any $g \in \mathcal{G}$, $\mathbb{P}_\theta[f \in \mathcal{F} : <f - h, g >\geq 0] > 1/2$. Hence, $\mathbb{P}_\theta$ is also H-symmetric about h, contradicting to Assumption 2.1 that f_c is unique. Therefore, $D_{ip}(f_c, \mathbb{P}_\theta, < \cdot, \cdot >, \mathcal{G}) = \sup_{f_{obs} \in \mathcal{F}} D_{ip}(f_{obs}, \mathbb{P}_\theta, < \cdot, \cdot >, \mathcal{G})$.

  (c) P-3: For any $f_{obs}(\neq f_c) \in \mathcal{F}$, we need to prove that for any $\alpha \in (0,1)$,

$$\inf_{g \in \mathcal{G}, \|g\|=1} \mathbb{P}_\theta\left[f \in \mathcal{F} : \langle f - f_{obs}, g \rangle \geq 0\right]$$

$$\leq \inf_{g \in \mathcal{G}, \|g\|=1} \mathbb{P}_\theta\left[f \in \mathcal{F} : \langle f - (f_c + \alpha(f_{obs} - f_c)), g \rangle \geq 0\right].$$

In fact, note that $f_c \in \{f \in \mathcal{F} : <f - f_{obs}, g >\geq 0\} \Leftrightarrow f_c \in \{f \in \mathcal{F} : <f - (f_c + \alpha(f_{obs} - f_c)), g >\geq 0\}$. By Assumption 2.1, we only need to consider g such that the halfspace does not contain f_c. Therefore,

$$\inf_{g \in \mathcal{G}, \|g\|=1} \mathbb{P}_\theta\left[f \in \mathcal{F} : \langle f - f_{obs}, g \rangle \geq 0\right]$$

$$= \inf_{g \in \mathcal{G}, \|g\|=1, \langle f_c - f_{obs}, g \rangle < 0} \mathbb{P}_\theta\left[f \in \mathcal{F} : \langle f - f_{obs}, g \rangle \geq 0\right]$$

$$\leq \inf_{g \in \mathcal{G}, ||g||=1, \langle f_c - f_{obs}, g \rangle < 0} \mathbb{P}_\theta \left[f \in \mathcal{F} : \langle f - f_{obs}, g \rangle \geq (1-\alpha)\langle f_c - f_{obs}, g \rangle \right]$$

$$= \inf_{g \in \mathcal{G}, ||g||=1, \langle f_c - f_{obs}, g \rangle < 0} \mathbb{P}_\theta \left[f \in \mathcal{F} : \langle f - (f_c + \alpha(f_{obs} - f_c)), g \rangle \geq 0 \right]$$

$$\leq \inf_{g \in \mathcal{G}, ||g||=1} \mathbb{P}_\theta \left[f \in \mathcal{F} : \langle f - (f_c + \alpha(f_{obs} - f_c)), g \rangle \geq 0 \right].$$

(d) P-4:

$$D_{ip}(f_{obs}, \mathbb{P}_{\theta,F}, \langle \cdot, \cdot \rangle, \mathcal{G}) = \inf_{g \in \mathcal{G}, ||g||=1} \mathbb{P}_\theta \left[f \in \mathcal{F} : \langle f - f_c, g \rangle \geq \langle f_{obs} - f_c, g \rangle \right]$$

$$= \inf_{g \in \mathcal{G}, ||g||=1} \mathbb{P}_\theta \left[f \in \mathcal{F} : \langle f, g \rangle \geq \langle f_{obs}, g \rangle \right] \leq \mathbb{P}_\theta \left[f \in \mathcal{F} : \langle f, f_{obs} \rangle \geq \langle f_{obs}, f_{obs} \rangle \right]$$

$$\leq \mathbb{P}_\theta \left[f \in \mathcal{F} : \sqrt{\langle f, f \rangle \langle f_{obs}, f_{obs} \rangle} \geq \langle f_{obs}, f_{obs} \rangle \right] \quad \text{(Cauchy inequality)}$$

$$= \mathbb{P}_\theta \left[f \in \mathcal{F} : ||f|| \geq ||f_{obs}|| \right] \to 0 \quad \text{(as } ||f_{obs}|| \to \infty)$$

∎

Appendix 8. Proof of Theorem 4.1

Proof: Throughout the proof, we use letter C to denote some constant whose meaning may change from line to line. According to Lemma 14 in Tran (2008), we have $\mathbb{E}\|\hat{K} - K\|_\infty^2 \leq C n^{-1}$. Therefore, by Markov's inequality

$$\mathbb{P}(\|\hat{K} - K\|_\infty \geq \frac{\log n}{\sqrt{n}}) \leq \left(\frac{\sqrt{n}}{\log n} \right)^2 \mathbb{E}(\|\hat{K} - K\|_\infty^2) \leq C(\log n)^{-2} \to 0.$$

Let $\mathcal{A}$ denote the event $\{\|\hat{K} - K\|_\infty \leq \frac{\log n}{\sqrt{n}}\}$. Then, $\mathbb{P}(\mathcal{A}) \to 1$ as $n \to \infty$.

Recall that in Algorithm I, we set $\hat{\lambda}_{p,n} = 0$ if $\hat{\lambda}_{p,n}$ is less than a threshold δ_n satisfying $\delta_n \to 0$ and $\delta_n \geq C(\frac{\sqrt{n}}{\log n})^{-\frac{\beta}{2\beta+1}}$. Then for a sufficiently large n, we have $\delta_n \geq 2\frac{\log n}{\sqrt{n}}$. Consequently, the $M_n = \arg\max_m\{\lambda_m \geq \delta_n\}$ as defined in Equation (5) satisfies $M_n \to \infty$ as $n \to \infty$. In addition, using Assumption 2.1, we have

$$C_1 M_n^{-\beta} \geq \lambda_{M_n} \geq \delta_n \geq C \left(\frac{\sqrt{n}}{\log n} \right)^{-\frac{\beta}{2\beta+1}}, \quad C_2(M_n+1)^{-\beta} \leq \lambda_{M_n+1} < \delta_n,$$

implying $C_1' \delta_n^{-1/\beta} \leq M_n \leq C_2' \delta_n^{-1/\beta} \leq C(\frac{\sqrt{n}}{\log n})^{\frac{1}{2\beta+1}}$. In addition, under event $\mathcal{A}$, we have, by Weyl's theorem and the definition of M_n, that

$$\arg\max_{1 \leq p \leq M_n} |\hat{\lambda}_{p,n} - \lambda_p| \leq \|\hat{K} - K\|_\infty \leq \frac{\log n}{\sqrt{n}} \leq \frac{\delta_n}{2} \leq \frac{\lambda_{M_n}}{2} \leq \frac{\lambda_p}{2},$$

where in the last step we have used the fact that λ_p is a nonincreasing sequence. Consequently, $\hat{\lambda}_{p,n} \geq \frac{\lambda_p}{2}$ holds for each $p = 1, \dots, M_n$.

By Proposition 16 in Tran (2008) and our Assumption 2.1 on the eigenvalues, we obtain that for each $p = 1, \dots, M_n$,

$$\|\hat{\phi}_{p,n} - \phi_p\| \leq \frac{C}{\min\{\lambda_{p-1} - \lambda_p, \lambda_p - \lambda_{p+1}\}} \|\hat{K} - K\|_\infty$$

$$\leq \frac{C}{p^{-(\beta+1)}} \|\hat{K} - K\|_\infty \leq C p^{\beta+1} \frac{\log n}{\sqrt{n}} \leq C M_n^{\beta+1} \frac{\log n}{\sqrt{n}}.$$

By combining this with the bound on M_n, we obtain

$$\|\hat{\phi}_{p,n} - \phi_p\| \leq C\delta_n^{-\frac{\beta+1}{\beta}} \frac{\log n}{\sqrt{n}} \leq C \frac{n^{\frac{\beta+1}{4\beta+2} - \frac{1}{2}}}{(\log n)^{(\beta+1)/(2\beta+1)}} \log n = Cn^{\frac{-\beta}{4\beta+2}} (\log n)^{\frac{\beta}{2\beta+1}} \to 0.$$

By the Cauchy–Schwarz inequality, for any $p \in \{1, \ldots, M_n\}$,

$$|\langle f_{obs}, \hat{\phi}_{p,n}\rangle^2 - \langle f_{obs}, \phi_p\rangle^2| = |\langle f_{obs}, \hat{\phi}_{p,n} + \phi_p\rangle\langle f_{obs}, \hat{\phi}_{p,n} - \phi_p\rangle|$$
$$\leq \|f_{obs}\|^2 (\|\hat{\phi}_{p,n}\| + \|\phi_p\|)\|\hat{\phi}_{p,n} - \phi_p\|$$
$$= 2\|f_{obs}\|^2\|\hat{\phi}_{p,n} - \phi_p\|.$$

Combing the last two displays, we obtain

$$|\langle f_{obs}, \hat{\phi}_{p,n}\rangle^2 - \langle f_{obs}, \phi_p\rangle^2| \leq C\|f_{obs}\|^2 p^{\beta+1} \frac{\log n}{\sqrt{n}} \leq C\|f_{obs}\|^2 n^{\frac{-\beta}{4\beta+2}} (\log n)^{\frac{\beta}{2\beta+1}},$$

and for each $p = 1, 2, \ldots, M_n$,

$$\left|\frac{\langle f_{obs}, \hat{\phi}_{p,n}\rangle^2 - \langle f_{obs}, \phi_p\rangle^2}{\lambda_p}\right| \leq C\|f_{obs}\|^2 p^{2\beta+1} \frac{\log n}{\sqrt{n}} \leq C\|f_{obs}\|^2 M_n^{2\beta+1} \frac{\log n}{\sqrt{n}} \leq C\|f_{obs}\|^2.$$

Now we are ready to prove Equation (8). By Assumption 4.2, it is easy to verify that the series $\sum_{p=1}^{\infty} \frac{\langle f_{obs}, \phi_p\rangle^2}{\lambda_p} a_p^2$ is uniformly convergent for any $\|f_{obs}\|_b \leq 1$ (as for N sufficiently large, $\sum_{p\geq N} \frac{\langle f_{obs}, \phi_p\rangle^2}{\lambda_p} a_p^2 \leq N^{-2\alpha} \sum_{p\geq N} \frac{\langle f_{obs}, \phi_p\rangle^2}{\lambda_p} b_p^2 \leq N^{-2\alpha} \|f_{obs}\|_b$). Therefore, according to Assumption 4.1, for each $N \geq 1$, we have $\sum_{p=N+1}^{\infty} \frac{\langle f_{obs}, \phi_p\rangle^2}{\lambda_p} a_p^2 < N^{-2\alpha}$ and $\sum_{p=N+1}^{\infty} a_p^2 < N^{-2\alpha} \sum_p b_p^2 \leq CN^{-2\alpha}$. According to the error bounds on $\hat{\lambda}_{p,n}$ and $\langle f_{obs}, \hat{\phi}_{p,n}\rangle$, we have that under event $\mathcal{A}_n$,

$$\left|\frac{\langle f_{obs}, \hat{\phi}_{p,n}\rangle^2}{\hat{\lambda}_{p,n}} - \frac{\langle f_{obs}, \phi_p\rangle^2}{\lambda_p}\right| < C\|f_{obs}\|^2 \left(\delta_n + N^{2\beta+1} \frac{\log n}{\sqrt{n}}\right), \quad p = 1, \ldots, N,$$

$$\left|\sum_{p=N+1}^{M_n \wedge n} \frac{\langle f_{obs}, \hat{\phi}_{p,n}\rangle^2 - \langle f_{obs}, \phi_p\rangle^2}{\lambda_p} a_p^2\right| \leq C\|f_{obs}\|^2 \sum_{p=N+1}^{M_n \wedge n} a_p^2 \leq C\|f_{obs}\|^2 N^{-2\alpha}.$$

Therefore, we obtain

$$\sum_{p=N+1}^{M_n \wedge n} \frac{\langle f_{obs}, \hat{\phi}_{p,n}\rangle^2}{\hat{\lambda}_{p,n}} a_p^2 \leq 2 \sum_{p=N+1}^{M_n \wedge n} \frac{\langle f_{obs}, \hat{\phi}_{p,n}\rangle^2}{\lambda_p} a_p^2$$

$$\leq 2 \left|\sum_{p=N+1}^{M_n \wedge n} \frac{\langle f_{obs}, \hat{\phi}_{p,n}\rangle^2 - \langle f_{obs}, \phi_p\rangle^2}{\lambda_p} a_p^2\right| + \sum_{p=N+1}^{M_n \wedge n} \frac{\langle f_{obs}, \phi_p\rangle^2}{\lambda_p} a_p^2 \leq CN^{-2\alpha},$$

where the first inequality is due to $\hat{\lambda}_{p,n} \geq \lambda_p/2$ for all $p \leq M_n$.

Putting pieces together, we can conclude that

$$\left|\|f_{obs}\|_{\hat{mod}}^2 - \|f_{obs}\|_{mod}^2\right| = \left|\sum_{p=1}^{M_n \wedge n} \frac{\langle f_{obs}, \hat{\phi}_{p,n}\rangle^2}{\hat{\lambda}_{p,n}} a_p^2 - \sum_{p=1}^{\infty} \frac{\langle f_{obs}, \phi_p\rangle^2}{\lambda_p} a_p^2\right|$$

$$\leq \left|\sum_{p=1}^{N} \left(\frac{\langle f_{obs}, \hat{\phi}_{p,n}\rangle^2}{\hat{\lambda}_{p,n}} - \frac{\langle f_{obs}, \phi_p\rangle^2}{\lambda_p}\right) a_p^2\right| + \sum_{p=N+1}^{\infty} \frac{\langle f_{obs}, \phi_p\rangle^2}{\lambda_p} a_p^2$$

$$+ \sum_{p=N+1}^{M_n \wedge n} \frac{\langle f_{obs}, \hat{\phi}_{p,n} \rangle^2}{\hat{\lambda}_{p,n}} a_p^2$$

$$< C \left(N^{-2\alpha} + \delta_n + N^{2\beta+1} \frac{\log n}{\sqrt{n}} \right).$$

By choosing $N = (\frac{\log n}{\sqrt{n}})^{-(2\alpha+2\beta+1)}$, we have $|\|f_{obs}\|_{\hat{mod}}^2 - \|f_{obs}\|_{mod}^2| \le n^{-\kappa}$ for $\kappa = 2\alpha/(2\alpha + 2\beta + 1) > 0$ under event $\mathcal{A}$. ∎

Appendix 9. Proof of Theorem 4.2 and Theorem 4.3

A.3 Proof of Theorem 4.2

Proof: According to the proof of theorem 4.1, there exists some event $\mathcal{A}$ whose probability tending to one as $n \to \infty$, such that under this event

$$\left| \|g\|_{\hat{mod}} - \|g\|_{mod} \right| \le C n^{-\kappa} \|g\|_b$$

for all g such that $\|g\|_b \le \infty$ (note that $\|g\|_{mod}$ is always dominated by $\|g\|_b$ according to Assumption 4.2). Under this condition, we have the following inclusion relationships:

$$\left\{ \|g_p\|_{mod} \ge \|f_{obs}\|_{mod} - Cn^{-\kappa} (\|g_p\|_b + \|f_{obs}\|_b) \right\}$$
$$\subset \left\{ \|g_p\|_{\hat{mod}} \ge \|f_{obs}\|_{\hat{mod}} \right\}$$
$$\subset \left\{ \|g_p\|_{mod} \ge \|f_{obs}\|_{mod} + Cn^{-\kappa} (\|g_p\|_b + \|f_{obs}\|_b) \right\}.$$

According to Assumption 4.3 and a standard tail probability bound for the max of sub-Gaussian random variables, we have $\mathbb{P}(\max_{p=1,\dots,n} \|g_p\|_b \le C\sigma \sqrt{\log n}) \ge 1 - n^{-1}$ for some constant $C > 0$. Let $\mathcal{B}$ to denote this event. Then, under event $\mathcal{A} \cap \mathcal{B}$, we have

$$U_n = n^{-1} \sum_{p=1}^{n} 1_{\|g_p\|_{mod} \ge \|f_{obs}\|_{mod} + \varepsilon_n} \le \frac{1}{n} \sum_{p=1}^{n} 1_{\|g_p\|_{\hat{mod}} \ge \|f_{obs}\|_{\hat{mod}}} \le V_n = \frac{1}{n} \sum_{p=1}^{n} 1_{\|g_p\|_{mod} \ge \|f_{obs}\|_{mod} - \varepsilon_n},$$

where $\varepsilon_n = Cn^{-\kappa} \sqrt{\log n}$. By Markov inequality, we have

$$\mathbb{P} \left(|U_n - (1 - F((\|f_{obs}\|_{mod} + \varepsilon_n)^2))| \le \frac{\log n}{\sqrt{n}} \right) \ge 1 - \frac{C}{\log^2 n},$$

$$\mathbb{P} \left(|V_n - (1 - F((\|f_{obs}\|_{mod} - \varepsilon_n)^2))| \le \frac{\log n}{\sqrt{n}} \right) \ge 1 - \frac{C}{\log^2 n}.$$

Let $\mathcal{C}$ denote the intersection of the two events inside above probabilities, and $\mathcal{E} = \mathcal{A} \cap \mathcal{B} \cap \mathcal{C}$. Then $\mathbb{P}(\mathcal{E}) \to 1$ as $n \to \infty$, and under this event $\mathcal{E}_n$, we have

$$1 - F((\|f_{obs}\|_{mod} + \varepsilon_n)^2) \le \frac{1}{n} \sum_{p=1}^{n} 1_{\|g_p\|_{\hat{mod}} \ge \|f_{obs}\|_{\hat{mod}}} \le 1 - F((\|f_{obs}\|_{mod} - \varepsilon_n)^2).$$

This implies the claimed result by using the fact that F is a continuous function and $\varepsilon_n \to 0$ as $n \to \infty$. ∎

A.4 Proof of Theorem 4.3

Proof: By the Markov inequality, given the data D, the conditional probability

$$\mathbb{P} \left(\left| \frac{1}{N} \sum_{p=1}^{N} 1_{\|g_p\|_{\hat{mod}} \ge \|f_{obs}\|_{\hat{mod}}} - (1 - F_n(\|f_{obs}\|_{\hat{mod}})) \right| \le \frac{\log N}{\sqrt{N}} \,\middle|\, D \right) \ge 1 - C/(\log N)^2,$$

where the randomness in $\mathbb{P}$ is due to the bootstrap sampling, and for any $t > 0$,

$$F_n(t) = \mathbb{P}\left(\sum_{p=1}^{M_n} a_p^2 Z_p^2 \leq t^2 \,\middle|\, D \right),$$

only dependent on M_n (defined in Equation 5), is the probability that a weighted sum of squares of the first M_n standard normal random variables $\{Z_p\}_{p=1}^{\infty}$ are less than or equal to t. By taking expectation with respect to D on both sides, we can further obtain

$$\mathbb{P}\left(\left| \frac{1}{N} \sum_{p=1}^{N} \mathbf{1}_{\|g_p\|_{\hat{mod}} \geq \|f_{obs}\|_{\hat{mod}}} - \left(1 - F_n(\|f_{obs}\|_{\hat{mod}})\right) \right| \leq \frac{\log N}{\sqrt{N}} \right) \geq 1 - C/(\log N)^2,$$

where now the randomness in $\mathbb{P}$ is due to both the randomness in data D and the randomness in the bootstrap sampling. In addition, function F in the desired limit is

$$F(t) = \mathbb{P}\left(\sum_{p=1}^{\infty} a_p^2 Z_p^2 \leq t^2 \right).$$

According to Theorem 4.1, we have $\left| \|f_{obs}\|_{\hat{mod}} - \|f_{obs}\|_{mod} \right| \leq C n^{-\kappa}$ with probability tending to one as $n \to \infty$. Therefore, due to the continuity of F in t, it remains to show that for each $t \in \mathbb{R}$,

$$F_n(t) \to F(t) \quad \text{in probability as } n \to \infty.$$

In fact, according to Assumption 4.2 and the fact that $M_n \to \infty$ as $n \to \infty$, we have

$$\mathbb{E}\left[\sum_{p=M_n+1}^{\infty} a_p^2 Z_p^2 \right] = \sum_{p=M_n+1}^{\infty} a_p^2 \leq M_n^{-2\alpha} \sum_{p=M_n+1}^{\infty} b_p \to 0$$

as $n \to \infty$. This implies the convergence in probability of $\sum_{p=1}^{M_n} a_p^2 Z_p^2$ to $\sum_{p=1}^{\infty} a_p^2 Z_p^2$ as $n \to \infty$. Then the desired convergence of F_n to F is a consequence of the fact that convergences in probability imply convergences in distribution. $\blacksquare$