The deepest point for distributions in infinite dimensional spaces

Anirvan Chakraborty* and Probal Chaudhuri

Theoretical Statistics and Mathematics Unit,
Indian Statistical Institute
203, B. T. Road, Kolkata - 700108, INDIA.
emails: anirvan_r@isical.ac.in, probal@isical.ac.in

Abstract

Identification of the center of a data cloud is one of the basic problems in statistics. One popular choice for such a center is the median, and several versions of median in finite dimensional spaces have been studied in the literature. In particular, medians based on different notions of data depth have been extensively studied by many researchers, who defined median as the point, where the depth function attains its maximum value. In other words, the median is the deepest point in the sample space according to that definition. In this paper, we investigate the deepest point for probability distributions in infinite dimensional spaces. We show that for some well-known depth functions like the band depth and the half-region depth in function spaces, there may not be any meaningful deepest point for many well-known and commonly used probability models. On the other hand, certain modified versions of those depth functions as well as the spatial depth function, which can be defined in any Hilbert space, lead to some useful notions of the deepest point with nice geometric and statistical properties. The empirical versions of those deepest points can be conveniently computed for functional data, and we demonstrate this using some simulated and real data sets.

Keywords: Breakdown point, coordinatewise median, functional depths, spatial depth, spatial median, strong consistency

1 Introduction

For a univariate probability distribution, median is a well-known and popular choice of its center. It has several desirable statistical properties, which include equivariance under

*Research is partially supported by CSIR SPM Fellowship

monotone transformations, asymptotic consistency under very general conditions and high breakdown point. The concept of median has been extended in several ways for probability distributions in finite dimensional Euclidean spaces (see, e.g., [8, 21] for some reviews). The median can also be defined as the point in the sample space with the highest depth value with respect to appropriate depth functions (see, e.g., [9, 14, 21, 25]). Due to the recent advances in technology and measurement devices, statisticians frequently have to analyze data for which the number of variables is much larger than the sample sizes. Such data can be conveniently viewed as random observations from probability distributions in infinite dimensional spaces, e.g., the space of real-valued functions defined on an interval.

It turns out that many medians for finite dimensional probability measures do not extend in any natural and meaningful way into infinite dimensional spaces. On the other hand, an extension of the well-known spatial median (see, e.g., [4]) into general Banach spaces was studied in [12, 23]. There has been some recent work on developing depth functions for probability measures in infinite dimensional spaces. The band depth (see, e.g., [15]) and the half-region depth (see, e.g., [16]) have been defined for data in the space of real-valued continuous functions on an interval on the real line. The authors of those papers have used these depth functions to find central and extreme curves for several real data sets, and also to construct test procedures based on depth based ranking. Recently, the authors of [22] have used the deepest point based on the band depth in the construction of boxplots for functional data. In Section 2, we shall critically investigate the deepest points associated with some of the functional depths. In Section 3, we shall demonstrate the empirical deepest points using some simulated and real functional data.

In the univariate setting, one of the main motivations for considering the median is its robustness against outlying observations. According to the traditional measures of robustness like breakdown point, the median is a more robust estimator of location than the mean. In an interesting paper [20], the author considered an alternative measure of robustness assuming that all the observations remain bounded. Using that measure, the author of [20] obtained some counter-intuitive results regarding the robustness of the median and the mean (see also [17]). In particular, it was shown that the mean may be more robust than the median under certain conditions. Unlike the univariate sample median, many multivariate depth based medians fail to achieve 50% breakdown point (see, e.g., [21] for a review). In fact, in $\mathbb{R}^d$ for $d \geq 2$, the half-space median has a breakdown point of $1/3$, while that for the simplicial median is atmost $1/(d+2)$ (see, e.g, [8] for a brief review and relevant references). In Section 4.1, we will consider the breakdown point of the empirical deepest point based on some functional depths.

The strong consistency of the empirical versions of most of the medians for finite dimensional data is well-known in the literature. However, for infinite dimensional data, the

situation is much more complex. In some cases, like the medians based on the band depth and the half-region depth, strong consistency has been proved under the assumption that the empirical medians along with the unique population median lie in a fixed equicontinuous set (see [15, 16]). For the empirical spatial median in infinite dimensions, the convergence has been shown to hold only in a weak sense, namely, for continuous real linear functions of the estimator (see, e.g., [5, 11]). Although the author of [5] proved strong consistency of the empirical spatial median in the norm topology of the Hilbert space $L_2(\mathbb{R})$, the result requires extremely strong assumptions on the underlying stochastic process, which include boundedness and differentiability of sample paths. Thus, this result is not valid for many important processes including the standard Brownian motions and fractional Brownian motions. Although the author of [12] considered the spatial median for general Banach spaces, he proved the consistency of its empirical version only in the finite dimensional setting. One of the major reasons for the difficulties in proving consistency results of such estimators in infinite dimensional spaces is the noncompactness of the closed unit ball. As a consequence, some of the standard methods of proof in the finite dimensional setup using Arzela-Ascoli Theorem (see, e.g., the proof as given in [12]) fail in infinite dimensions. Let us mention here that the strong consistency of an updation based alternative estimator of the spatial median has been proved in separable Hilbert spaces (see [6]). However, this estimator is very different from the empirical spatial median, and is more suited when data arrive sequentially. In Section 4.2, we will study the asymptotic strong consistency of the empirical deepest points for data lying in infinite dimensional function spaces.

2 The deepest point

In this section, we will study the deepest point arising from some of the depth functions that are available for probability distributions in infinite dimensional spaces. The spatial depth of $\mathbf{x} \in \mathbb{R}^d$ with respect to the probability distribution of $\mathbf{X} \in \mathbb{R}^d$ is defined as $SD(\mathbf{x}) = 1 - ||E\{(\mathbf{x} - \mathbf{X})I(\mathbf{X} \neq \mathbf{x})/||\mathbf{x} - \mathbf{X}||\}||$ (see, e.g., [19]). Recall that the spatial median (see, e.g., [4]), say $\mathbf{m}_s$, of the distribution of $\mathbf{X}$ is given by the minimizer of the function $E\{||\mathbf{x} - \mathbf{X}||\}$ (if $E||\mathbf{X}|| = \infty$, we can minimize $E\{||\mathbf{x} - \mathbf{X}|| - ||\mathbf{X}||\}$ as suggested in [12]). The definitions of both of the spatial median and the spatial depth function extend naturally to any Hilbert space, e.g., the function space $L_2[0, 1]$. It has been proved in [12] that the spatial median is unique if the Hilbert space is strictly convex, and F is not entirely supported on a line (see Theorem 2.17 in that paper). Further, if the distribution of $\mathbf{X}$ is nonatomic, the unique spatial median is the only point in that Hilbert space, which satisfies $E\{(\mathbf{x} - \mathbf{X})/||\mathbf{x} - \mathbf{X}||\} = \mathbf{0}$ (see Theorem 4.14 in [12]). Summarizing all these, we have the following fact.

Fact 2.1. *Let* $\mathbf{X}$ *be a random element in a strictly convex Hilbert space, and suppose that the probability distribution of* $\mathbf{X}$ *is nonatomic and not entirely supported on a line in* $\mathcal{X}$*. Then, the unique deepest point associated with the spatial depth function* $SD(\mathbf{x})$ *is the spatial median of* $\mathbf{X}$*, and its spatial depth is* 1*.*

Let us also mention here that an alternative definition of spatial depth in $\mathbb{R}^d$ was considered by the authors of [24]. They defined the spatial depth of $\mathbf{x} \in \mathbb{R}^d$ as $1 - \inf\{w \geq 0 :$ spatial median of $(w\delta_{\mathbf{x}} + F)/(1+w) = \mathbf{x}\}$. Here $\delta_{\mathbf{x}}$ denotes the point mass at $\mathbf{x}$. It can be shown using the characterization of spatial median given in Theorem 4.14 in [12] that one gets $1 - \max\{0, 1 - SD(\mathbf{x}) - p(\mathbf{x})\}$ as the depth of $\mathbf{x}$ according to this definition. Here $p(\mathbf{x})$ denotes the mass at $\mathbf{x}$ under the distribution of $\mathbf{X}$. So, this definition of spatial depth coincides with the previous definition if F is nonatomic.

Infinite dimensional data are often obtained as realizations of functions, e.g., spectrometric data, electrocardiogram records of patients, stock price data, various meteorological data like temperature, rainfall etc. In that case, one may want to exploit the functional nature of the data in constructing depth functions. As mentioned in Section 1, the band depth and the half-region depth were introduced in the literature for such data. The band depth (BD) of $\mathbf{x} = \{x_t\}_{t\in[0,1]} \in C[0,1]$ relative to the distribution of a random element $\mathbf{X} = \{X_t\}_{t\in[0,1]} \in C[0,1]$ is given by

$$BD(\mathbf{x}) = \sum_{j=2}^{J} P\left(\min_{1\leq i\leq j} X_{i,t} \leq x_t \leq \max_{1\leq i\leq j} X_{i,t} \ \forall\ t \in [0,1]\right),$$

where for $1 \leq i \leq J$, $\mathbf{X}_i = \{X_{i,t}\}_{t\in[0,1]}$ are independent copies of $\mathbf{X}$. The half-region depth (HRD) of $\mathbf{x}$ is given by

$$HRD(\mathbf{x}) = \min\left\{P\left(X_t \leq x_t \ \forall\ t \in [0,1]\right),\ P\left(X_t \geq x_t \ \forall\ t \in [0,1]\right)\right\}.$$

However, as we shall see in the next theorem, for many well-known and commonly used stochastic models for functional data, there is no meaningful notion of deepest points associated with these two depth functions. These stochastic models include Feller processes in $C[0,1]$. Feller processes are a class of strong Markov processes, whose transition probability function satisfies certain continuity properties (see, e.g., [18] for a detailed discussion). Important examples of Feller processes include Brownian motions, Brownian bridges etc. Feller processes have been used for modelling data in physical and biological sciences (see, e.g., [3] for a review and related references).

Theorem 2.2. *Let* $\mathbf{X}$ *be a Feller process in* $C[0,1]$ *starting at* $x_0 \in \mathbb{R}$*, which is symmetric about* $\mathbf{a} = \{a_t\}_{t\in[0,1]} \in C[0,1]$*, i.e., the distribution of* $\mathbf{X} - \mathbf{a}$ *is same as that of* $\mathbf{a} - \mathbf{X}$*. Also, assume that* X_t *has a continuous distribution for all* $t \in (0,1]$ *with any finite dimensional*

marginal $(X_{t_1}, X_{t_2}, \ldots, X_{t_d})$ *of* $\mathbf{X}$ *having a positive density in a neighbourhood of its center of symmetry* $(a_{t_1}, a_{t_2}, \ldots, a_{t_d})$*, where* $0 < t_1 < t_2 < \ldots < t_d \leq 1$*. Then, the band depth vanishes identically on* $C[0,1]$*. Moreover, the half-region depth of* $\mathbf{a}$ *is zero.*

Proof of Theorem 2.2. Since $X_0 = x_0$ with probability one, if $\mathbf{y} = \{y_t\}_{t\in[0,1]} \in C[0,1]$ is such that $y_0 \neq x_0$, then $\mathbf{y}$ cannot be contained in any band formed by the $\mathbf{X}_i$'s. Consequently, $BD(\mathbf{y}) = 0$. So, it is enough to prove the result for any $\mathbf{y} = \{y_t\}_{t\in[0,1]} \in C[0,1]$ satisfying $y_0 = x_0$. Let $\mathbf{y}_m = (y_{1/2^m}, y_{2/2^m}, \ldots, y_{2^m/2^m})$. It follows from part 2 of Theorem 1 in [15] that $BD_m(\mathbf{y}_m) \leq BD_m(\mathbf{a}_m)$ for all $m \geq 1$, where the function BD_m is the m-dimensional band depth calculated using the distribution of $\mathbf{X}_m = (X_{1/2^m}, X_{2/2^m}, \ldots, X_{2^m/2^m})$. Since

$$\begin{aligned} & \{\min_{1\leq i\leq j} X_{i,k/2^{m+1}} \leq x_{k/2^{m+1}} \leq \max_{1\leq i\leq j} X_{i,k/2^{m+1}} \ \forall\ k = 1, \ldots, 2^{m+1}\} \\ \subseteq\ & \{\min_{1\leq i\leq j} X_{i,k/2^m} \leq x_{k/2^m} \leq \max_{1\leq i\leq j} X_{i,k/2^m} \ \forall\ k = 1, \ldots, 2^m\} \end{aligned}$$

for all $1 \leq j \leq J$, it follows that $BD_{m+1}(\mathbf{y}_{m+1}) \leq BD_m(\mathbf{y}_m)$ for all $m \geq 1$. So,

$$\begin{aligned} & \lim_{m\to\infty} BD_m(\mathbf{y}_m) \\ = & \sum_{j=2}^{J} P\left(\min_{1\leq i\leq j} X_{i,k/2^m} \leq x_{k/2^m} \leq \max_{1\leq i\leq j} X_{i,k/2^m} \ \forall\ k = 1, \ldots, 2^m \text{ and } m \geq 1\right) \end{aligned}$$

Then, using the almost sure uniform continuity of the sample paths, we have

$$BD(\mathbf{y}) = \lim_{m\to\infty} BD_m(\mathbf{y}_m) \leq \lim_{m\to\infty} BD_m(\mathbf{a}_m) = BD(\mathbf{a}).$$

The proof will be complete if we show that $BD(\mathbf{a}) = 0$. Let us now consider the multivariate Feller process $\{(X_{1,t}, X_{2,t}, \ldots, X_{j,t})\}_{t\in[0,1]}$ for $1 \leq j \leq J$. Define $\mathbf{Y} = \{Y_t\}_{t\in[0,1]} = \mathbf{X} - \mathbf{a}$. Since $\{X_t\}_{t\in[0,1]}$ is a Feller process starting at $x_0 = a_0$ and symmetric about $\mathbf{a}$, $\mathbf{Y}$ is a Feller process starting at 0 and symmetric about $\mathbf{0}$. Let $T_j = \inf\{t > 0 : \min_{1\leq i\leq j} X_{i,t} > a_t\} = \inf\{t > 0 : \min_{1\leq i\leq j} Y_{i,t} > 0\}$ and $S_j = \inf\{t > 0 : \max_{1\leq i\leq j} X_{i,t} < a_t\} = \inf\{t > 0 : \max_{1\leq i\leq j} Y_{i,t} < 0\}$. From the continuity of the sample paths and using Propositions 2.16 and 2.17 in [18], we get that $P(T_j = 0) = 0$ or 1 and $P(S_j = 0) = 0$ or 1 for all $1 \leq j \leq J$. Since $P(T_j = 0) = \lim_{t\downarrow 0} P(T_j \leq t) \geq 2^{-j}$ and $P(S_j = 0) = \lim_{t\downarrow 0} P(S_j \leq t) \geq 2^{-j}$, we have $P(T_j = 0) = P(S_j = 0) = 1$ for all $1 \leq j \leq J$.
The proof of the fact that the half-region depth of $\mathbf{a}$ is 0 follows from the above arguments after taking $J = 1$. □

Since any reasonable depth function should assign maximum value to the point of symmetry of a distribution, and it should preferably be the unique maximizer, it follows from the above theorem that neither of these two depth functions yields any useful notion of deepest point for such stochastic processes. It will be appropriate to note here that it was

proved in [7] that both of the band depth and the half-region depth vanish on a set of measure one for a class of Feller processes in $C[0,1]$. The assumptions of symmetry and the existence of densities of all the finite dimensional marginals of $\mathbf{X}$ yield a stronger result for the band depth here.

The previous theorem provides a mathematical explanation for an observation made in [15, 16] that both of the band depth and the half-region depth tend to take small values if the sample paths cross each other often. This led the authors of those papers to consider modified versions of these depth functions, called modified band depth (MBD) and modified half-region depth (MHRD), respectively. The definitions of both of these depths directly extend to random elements in the space, say $\mathcal{F}(I)$, of real-valued functions defined on a set I. Here, for an appropriate σ-field $\mathcal{I}$ and a probability measure Λ on I, we consider the probability space $(I, \mathcal{I}, \Lambda)$. For example, I can be $[0,1]$ with $\mathcal{I}$ as its Borel σ-field and Λ as the Lebesgue measure on $[0,1]$. The modified band depth of $\mathbf{x} \in \mathcal{F}(I)$ relative to the distribution of $\mathbf{X} = \{X_{\mathbf{t}}\}_{\mathbf{t}\in I} \in \mathcal{F}(I)$ is defined as

$$MBD(\mathbf{x}) = \sum_{j=2}^{J} E\left[\int_I I\left(\min_{i=1,\ldots,j} X_{i,\mathbf{t}} \le x_{\mathbf{t}} \le \max_{i=1,\ldots,j} X_{i,\mathbf{t}}\right)\Lambda(d\mathbf{t})\right],$$

where $\mathbf{X}_i = \{X_{i,\mathbf{t}}\}_{\mathbf{t}\in I}$ are J independent copies of $\mathbf{X}$. A closely related notion of depth, called integrated data depth (IDD) [10], is defined on $\mathcal{F}(I)$ as $IDD(\mathbf{x}) = \int_I D_{\mathbf{t}}(x_{\mathbf{t}})\Lambda(d\mathbf{t})$, where $D_{\mathbf{t}}$'s are depth functions on the real line. Henceforth, we shall assume that for each $\mathbf{t} \in I$, $D_{\mathbf{t}}$ is maximized at the median of $X_{\mathbf{t}}$ if $X_{\mathbf{t}}$ has a continuous distribution. The modified half-region depth of $\mathbf{x}$ relative to the distribution of $\mathbf{X}$ is given by

$$MHRD(\mathbf{x}) = \min\left\{E\left[\int_I I(X_{\mathbf{t}} \le x_{\mathbf{t}})\Lambda(d\mathbf{t})\right],\ E\left[\int_I I(X_{\mathbf{t}} \ge x_{\mathbf{t}})\Lambda(d\mathbf{t})\right]\right\}.$$

The next result gives a description of points, which maximize the above three depth functions. For this, let us write $\mathbf{m}_c = \{m_{\mathbf{t}}\}_{\mathbf{t}\in I}$, where $m_{\mathbf{t}}$ denotes the median of $X_{\mathbf{t}}$ for $\mathbf{t} \in I$.

Theorem 2.3. *Let $\mathbf{X}$ be a random element in $\mathcal{F}(I)$ such that $X_{\mathbf{t}}$ has a continuous distribution, which is strictly increasing in a neighbourhood of $m_{\mathbf{t}}$ for each $\mathbf{t} \in I$. Then, $\mathbf{m}_c$ is a maximizer of the modified band depth, the modified half-region depth and the integrated data depth. Any $\mathbf{m}_c^*$, which equals $\mathbf{m}_c$ for $\mathbf{t} \in I$ except on a set of Λ-measure zero, is also a maximizer of the modified band depth and the integrated data depth. Moreover, any $\mathbf{m}_c^{**} = \{m_{\mathbf{t}}^{**}\}_{\mathbf{t}\in I}$ satisfying $\int_I F_{\mathbf{t}}(m_{\mathbf{t}}^{**})\Lambda(d\mathbf{t}) = 1/2$ is a maximizer of the modified half-region depth.*

It is clear that $\mathbf{m}_c^{**}$, which satisfies $\int_I F_{\mathbf{t}}(m_{\mathbf{t}}^{**})\Lambda(d\mathbf{t}) = 1/2$, may differ from $\mathbf{m}_c$ on a set of positive Λ-measure unlike $\mathbf{m}_c^*$. Moreover, although $\mathbf{m}_{\mathbf{c}}^{**}$ is a maximizer of the modified half-region depth, its components $m_{\mathbf{t}}^{**}$ may be far from being univariate medians. For

instance, let us consider a standard Brownian motion $\mathbf{X} = \{X_t\}_{t\in[0,1]}$. Define the function $\mathbf{f} = \{f_t\}_{t\in[0,1]}$, where f_t is the αth percentile of X_t for $t \in [0, 1/2)$, and f_t is the $(1-\alpha)^{th}$ percentile of X_t for $t \in [1/2, 1]$. Here $\alpha \in (0, 1)$ can be as small or as large as we like. Then, any such $\mathbf{f}$ is a maximizer of the modified half-region depth.

Proof of Theorem 2.3. From the definition of modified band depth we have

$$\begin{aligned} MBD(\mathbf{x}) &= \sum_{j=2}^{J} E\left[\int_I I\left(\min_{i=1,\ldots,j} X_{i,\mathbf{t}} \le x_{\mathbf{t}} \le \max_{i=1,\ldots,j} X_{i,\mathbf{t}}\right)\Lambda(d\mathbf{t})\right] \\ &= \sum_{j=2}^{J} \left[\int_I P\left(\min_{i=1,\ldots,j} X_{i,\mathbf{t}} \le x_{\mathbf{t}} \le \max_{i=1,\ldots,j} X_{i,\mathbf{t}}\right)\Lambda(d\mathbf{t})\right] \\ &= \sum_{j=2}^{J} \int_I \left[1 - F_{\mathbf{t}}^j(x_{\mathbf{t}}) - (1 - F_{\mathbf{t}}(x_{\mathbf{t}}))^j\right]\Lambda(d\mathbf{t}), \end{aligned} \tag{1}$$

where the second equality in the above follows from Fubini's theorem. Here $F_{\mathbf{t}}$ denotes the distribution of $X_{\mathbf{t}}$ for $\mathbf{t} \in I$. For each $\mathbf{t} \in I$, the integrand in (1) is maximized iff $F_{\mathbf{t}}(x_{\mathbf{t}}) = 1/2$, which can be easily verified using standard calculus. This implies that the term in the right hand side of equation (1) is maximized iff $F_{\mathbf{t}}(x_{\mathbf{t}}) = 1/2$ for all $\mathbf{t} \in I$, except perhaps on a subset of I with Λ-measure zero. Hence, the modified band depth is maximized at $\mathbf{m}_c$, and also at any $\mathbf{m}_c^*$, which equals $\mathbf{m}_c$ outside a Λ-null set.

Since $D_{\mathbf{t}}$ is maximized at $m_{\mathbf{t}}$ for each $\mathbf{t} \in I$, it follows that the integrated data depth function $\int_I D_{\mathbf{t}}(x_{\mathbf{t}})\Lambda(d\mathbf{t})$ is maximized at $\mathbf{m}_c$, or at any $\mathbf{m}_c^*$, which equals $\mathbf{m}_c$ except on a Λ-null set.

For the modified half-region depth, we have using Fubini's theorem

$$\begin{aligned} MHRD(\mathbf{x}) &= \min\left\{E\left[\int_I I(X_{\mathbf{t}} \le x_{\mathbf{t}})\Lambda(d\mathbf{t})\right],\ E\left[\int_I I(X_{\mathbf{t}} \ge x_{\mathbf{t}})\Lambda(d\mathbf{t})\right]\right\} \\ &= \min\left\{\left[\int_I P(X_{\mathbf{t}} \le x_{\mathbf{t}})\Lambda(d\mathbf{t})\right],\ \left[\int_I P(X_{\mathbf{t}} \ge x_{\mathbf{t}})\Lambda(d\mathbf{t})\right]\right\} \\ &= \min\left\{\int_I F_{\mathbf{t}}(x_{\mathbf{t}})\Lambda(d\mathbf{t}),\ 1 - \int_I F_{\mathbf{t}}(x_{\mathbf{t}})\Lambda(d\mathbf{t})\right\}. \end{aligned} \tag{2}$$

The maximum value of the right hand side of (2) is 1/2. Since $F_{\mathbf{t}}(m_{\mathbf{t}}) = 1/2$ for all $\mathbf{t} \in I$, $\mathbf{m}_c$ is a maximizer of the modified half-region depth. Further, any $\mathbf{m}_c^{**} = \{m_{\mathbf{t}}^{**}\}_{\mathbf{t}\in I}$ satisfying $\int_I F_{\mathbf{t}}(m_{\mathbf{t}}^{**})\Lambda(d\mathbf{t}) = 1/2$ will also maximize the modified half-region depth. □

3 Computation and data analytic demonstration of the empirical deepest point

We shall now consider empirical versions of the two deepest points discussed in the previous section. Suppose that $\mathbf{X}_i$, $1 \le i \le n$, are i.i.d observations from a distribution in a Hilbert

space, e.g., the function space $L_2[0,1]$. The empirical version of $\mathbf{m}_s$, say $\widehat{\mathbf{m}}_s$, is given by the minimizer of the function $n^{-1}\sum_{i=1}^{n}||\mathbf{x}-\mathbf{X}_i||$, where $||.||$ is the norm in that Hilbert space. For the empirical version of $\mathbf{m}_c$, suppose that $\mathbf{X}_i = \{X_{i,\mathbf{t}}\}_{\mathbf{t}\in I}$, $1 \le i \le n$, are i.i.d. observations from a probability distribution in the space $\mathcal{F}(I)$. Then, the natural estimator of $\mathbf{m}_c$ is $\widehat{\mathbf{m}}_c = \{\widehat{m}_{c,\mathbf{t}}\}_{\mathbf{t}\in I}$, where $\widehat{m}_{c,\mathbf{t}}$ is the median corresponding to the empirical distribution of $X_{\mathbf{t}}$. Throughout this paper, we shall use the conventional definition of median, i.e., if n is even, $\widehat{m}_{\mathbf{t}} = (X_{(n/2),\mathbf{t}} + X_{(n/2+1),\mathbf{t}})/2$, and if n is odd, $\widehat{m}_{\mathbf{t}} = X_{((n+1)/2),\mathbf{t}}$. The empirical coordinatewise median $\widehat{\mathbf{m}}_c$ is easy to compute. Although there are some algorithms for computing the empirical spatial median $\widehat{\mathbf{m}}_s$ for data in finite dimensional spaces (see, e.g., [2, 24]), some of them, which involve the inverse of the Hessian matrix of the function being optimized, cannot be used in the infinite dimensional setting. This is due to the noninvertibility of a Hessian matrix, which occurs when the data dimension exceeds the sample size. It is known that the empirical spatial median always lies in the closed convex hull generated by the sample points (see, e.g., Remark 4.20 in [12]). This property of the empirical spatial median has been used in [11] to reduce the infinite dimensional minimization problem associated with the empirical spatial median to a convex minimization over the compact set $\{\sum_{k=1}^{n}\alpha_k = 1 : \alpha_k \in [0,1]\}$, when there are n data points.

We now demonstrate the empirical deepest points using some simulated and real data sets. For both the simulated and the real data sets, the empirical spatial median is computed using the iterative algorithm proposed in [11]. Each of the simulated data sets that we consider has 10 observations. One set is generated from the standard Brownian motion in $[0,1]$, and the two other sets are generated from fractional Brownian motions in $[0,1]$ with Hurst indices $H = 0.3$ and $H = 0.7$. Recall that a fractional Brownian motion with Hurst index $H \in (0,1)$ is a zero mean Gaussian process having covariance kernel $K(t,s) = 0.5(t^{2H} + s^{2H} - |t-s|^{2H})$ for $t,s \in [0,1]$. The samples are observed at 101 equispaced points in $[0,1]$. Figures 1, 2 and 3 show the plots of the sample curves along with the empirical coordinatewise median and the empirical spatial median. It is observed from the plots that both the empirical medians are close to the zero function, which is the center of symmetry for all three distributions.

The real data considered here is the growth acceleration data set, which is derived from the Berkeley growth data. The latter is available in the R package "fda" (see http://rss.acs.unt.edu/Rdoc/library/fda/html/growth.html) and contains two subclasses, namely, the boys and the girls. Heights of 39 boys and 54 girls were measured at 31 time points between ages 1 and 18 years. The growth acceleration curves are obtained through monotone spline smoothing available in the R package "fda", and those are recorded at 101 equispaced ages in the interval $[1,18]$ years. Figure 4 shows the plots of the acceleration data of the boys and the girls along with the empirical coordinatewise medians and the empirical spatial

medians. It is seen from Figure 4 that both of the empirical medians are close to the central curves in each data set.

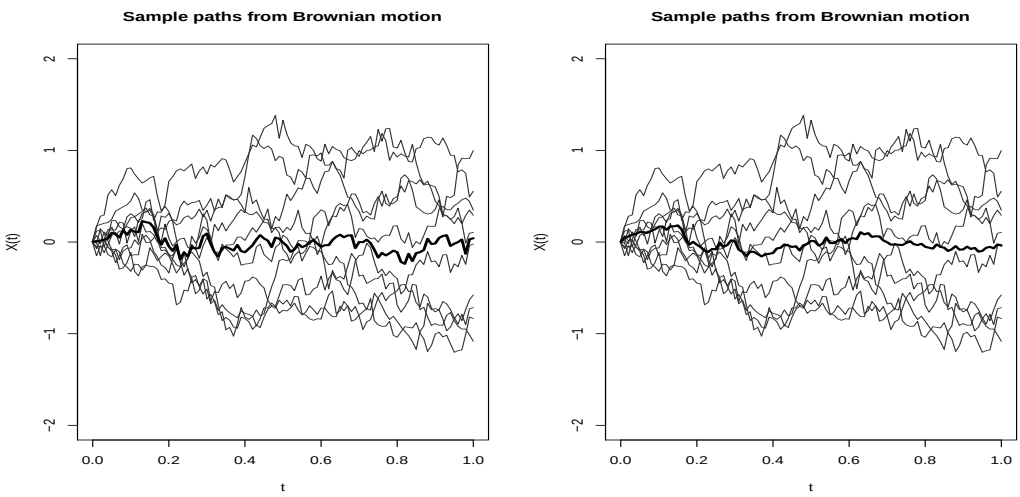

Figure 1: *Plots of sample paths from standard Brownian motion along with the empirical coordinatewise median (bold black curve in the left panel) and the empirical spatial median (bold black curve in the right panel).*

4 Geometric and statistical properties of the empirical deepest point

We shall now discuss the properties of the empirical deepest point mentioned in the previous section. The empirical spatial median is unique if the data points are not concentrated on a line, which follows from Theorem 2.17 in [12]. Further, if the empirical spatial median is not one of the data points, then using Theorem 4.14 in [12] it can be verified that the empirical spatial depth of $\widehat{\mathbf{m}}_s$ is 1, where the empirical spatial depth of $\mathbf{x}$ is given by $1 - n^{-1} \left\| \sum_{i=1}^{n} (\mathbf{x} - \mathbf{X}_i) \, I(\mathbf{X}_i \neq \mathbf{x}) / \|\mathbf{x} - \mathbf{X}\| \right\|$. In other words, in such situations, the empirical spatial median $\widehat{\mathbf{m}}_s$ is the unique maximizer of the empirical spatial depth.

Let $\widehat{F}_{\mathbf{t}}$ denote the empirical distribution of $X_{\mathbf{t}}$ for each $\mathbf{t} \in I$. Then, the empirical modified band depth of $\mathbf{x} = \{x_{\mathbf{t}}\}_{\mathbf{t} \in I}$ is given by $\sum_{j=2}^{J} \int_I \left[1 - \widehat{F}_{\mathbf{t}}^{j}(x_{\mathbf{t}}-) - (1 - \widehat{F}_{\mathbf{t}}(x_{\mathbf{t}}))^j \right] \Lambda(d\mathbf{t})$. Also, the empirical integrated data depth of $\mathbf{x}$ is given by $\int_I \widehat{D}_{\mathbf{t}}(x_{\mathbf{t}}) \Lambda(d\mathbf{t})$, where $\widehat{D}_{\mathbf{t}}(x_{\mathbf{t}})$'s denote the empirical versions of the depths $D_{\mathbf{t}}$'s. We can assume that $\widehat{D}_{\mathbf{t}}(x_{\mathbf{t}})$ is maximized at $\widehat{m}_{\mathbf{t}}$ for all $\mathbf{t} \in I$, which is true for almost any depth function for univariate data. The empirical half-region depth of $\mathbf{x}$ is given by $\min \left\{ \int_I \widehat{F}_{\mathbf{t}}(x_{\mathbf{t}}) \Lambda(d\mathbf{t}), \; 1 - \int_I \widehat{F}_{\mathbf{t}}(x_{\mathbf{t}}-) \Lambda(d\mathbf{t}) \right\}$. It can be verified that $\widehat{\mathbf{m}}_c$ is a maximizer of the empirical versions of all of the three depth

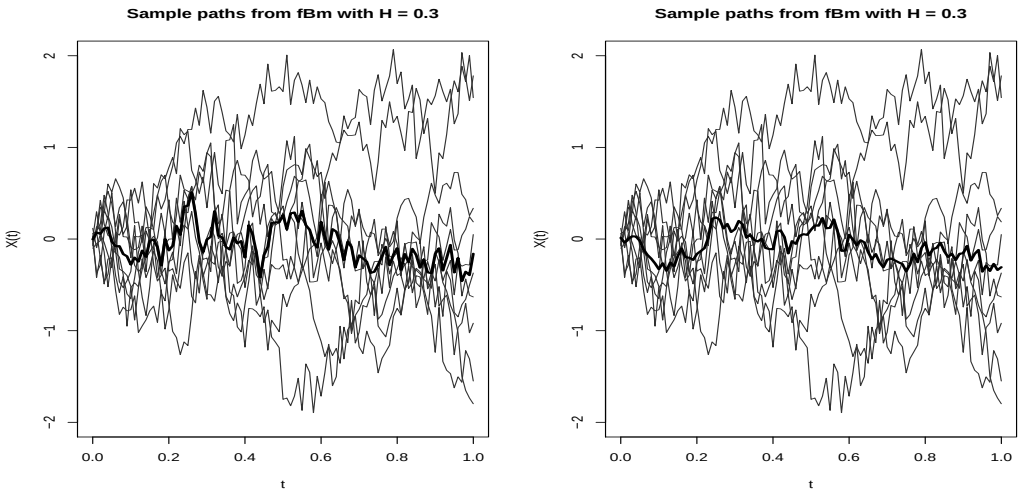

Figure 2: *Plots of sample paths from fractional Brownian motion with Hurst index* $H = 0.3$ *along with the empirical coordinatewise median (bold black curve in the left panel) and the empirical spatial median (bold black curve in the right panel).*

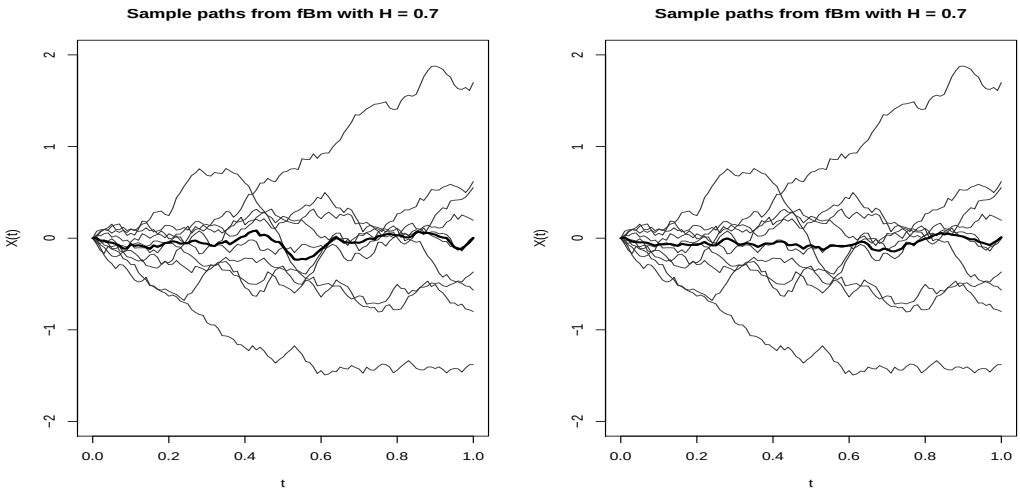

Figure 3: *Plots of sample paths from fractional Brownian motion with Hurst index* $H = 0.7$ *along with the empirical coordinatewise median (bold black curve in the left panel) and the empirical spatial median (bold black curve in the right panel).*

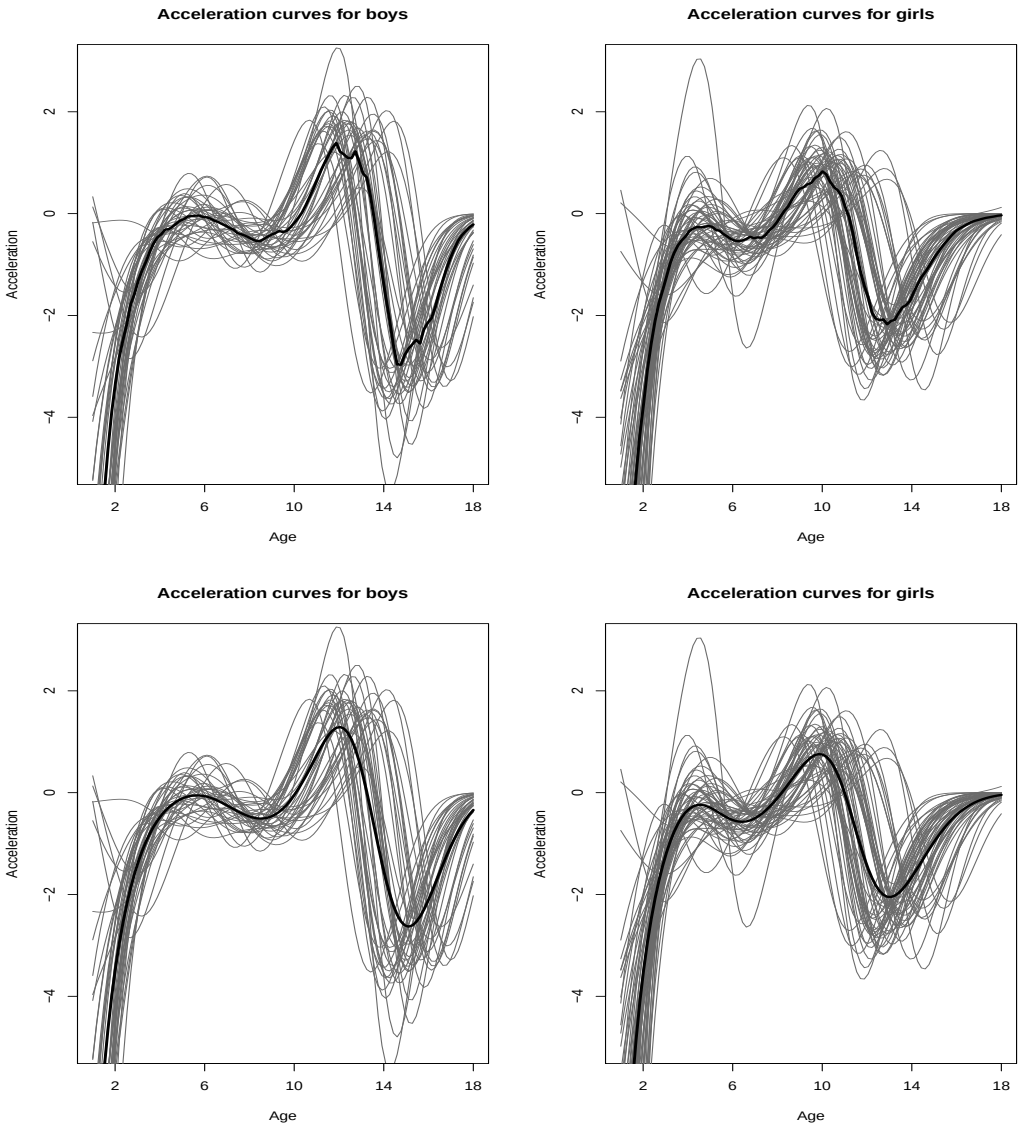

Figure 4: *Plots of acceleration curves for boys (left panel) and girls (right panel) along with the corresponding empirical coordinatewise medians (bold black curves in the top panel) and the empirical spatial medians (bold black curves in the bottom panel).*

functions mentioned above if the distribution of $X_{\mathbf{t}}$ is continuous for all $\mathbf{t} \in I$. Further, any $\widehat{\mathbf{m}}_c^*$, which differ from $\widehat{\mathbf{m}}_c$ only on a Λ-null set is also a maximizer of the empirical modified band depth and the empirical integrated data depth. Also, there is no unique maximizer of the empirical modified half-region depth, and any $\widehat{\mathbf{m}}_c^{**} = \{\widehat{m}_{\mathbf{t}}^{**}\}_{\mathbf{t}\in I}$ satisfying $\int_I \widehat{F}_{\mathbf{t}}(\widehat{m}_{\mathbf{t}}^{**}) = 1/2$ will be a maximizer of the empirical modified half-region depth like its population counterpart.

It is easy to see that the empirical coordinatewise median is equivariant under any coordinatewise monotone transformation given by $\mathbf{x} \mapsto \Psi(\mathbf{x})$, where $\mathbf{x} = \{x_{\mathbf{t}}\}_{\mathbf{t}\in I}$, $\Psi(\mathbf{x}) = \{\psi_{\mathbf{t}}(x_{\mathbf{t}})\}_{\mathbf{t}\in I}$, and $\psi_{\mathbf{t}}$ is a monotone function for each $\mathbf{t} \in I$. These include location shifts $\mathbf{x} \mapsto \mathbf{x} + \mathbf{c}$, where $\mathbf{c} \in \mathcal{F}(I)$ as well as coordinatewise scale transformations $\mathbf{x} \mapsto \mathbf{x}_1$, where $\mathbf{x}_1 = \{a_{\mathbf{t}} x_{\mathbf{t}}\}_{\mathbf{t}\in I}$ and $a_{\mathbf{t}} > 0$ for each $\mathbf{t} \in I$. On the other hand, the empirical spatial median is equivariant under location shifts and homogeneous scale transformations $\mathbf{x} \mapsto a\mathbf{x}$, where $a > 0$. Moreover, it is also equivariant under linear isometries, i.e., any linear map $\mathbf{T} : \mathcal{X} \to \mathcal{X}$ such that $||\mathbf{T}\mathbf{x}|| = ||\mathbf{x}||$ for all $\mathbf{x}$ in the Hilbert space $\mathcal{X}$.

4.1 Robustness properties of the empirical deepest point

It was mentioned in Section 1 that many finite dimensional depth based medians, e.g., the half-space median, the simplicial median, do not have 50% breakdown point unlike the univariate median. We shall now discuss the robustness properties of the two empirical deepest points discussed in Section 3 in terms of their breakdown points. Let us first consider the breakdown point of $\widehat{\mathbf{m}}_c$, and for this we assume that $\int_I |X_{\mathbf{t}}|\Lambda(d\mathbf{t}) < \infty$ with probability 1. This assumption holds if I is a compact set, and with probability one, the process $\mathbf{X}$ has continuous paths(e.g., Brownian motions on a compact interval). Then, it can be shown that $\widehat{\mathbf{m}}_c$ is a minimizer of the function $\int_I \sum_{i=1}^n |X_{i,\mathbf{t}} - x_{\mathbf{t}}|\Lambda(d\mathbf{t})$ over the set of $\mathbf{x} \in \mathcal{F}(I)$, which satisfies $\int_I |x_{\mathbf{t}}|\Lambda(d\mathbf{t}) < \infty$. Here $\mathbf{X}_i = \{X_{i,\mathbf{t}}\}_{\mathbf{t}\in I}$ are the sample observations. In other words, $\widehat{\mathbf{m}}_c$ is an empirical spatial median of the distribution of $\mathbf{X}$ in the Banach space of real-valued absolutely integrable functions defined on I. Hence, from Theorem 2.10 in [12], we get that $\widehat{\mathbf{m}}_c$ has 50% breakdown point. The same theorem also implies that the empirical spatial median has 50% breakdown point. The following result summarizes the above observations.

Fact 4.1.1. *The empirical deepest point $\widehat{\mathbf{m}}_s$ has 50% breakdown point. Further, if the random element $\mathbf{X} \in \mathcal{F}(I)$ satisfies $\int_I |X_{\mathbf{t}}|\Lambda(d\mathbf{t}) < \infty$ with probability 1, then $\widehat{\mathbf{m}}_c$ also has 50% breakdown point.*

4.2 Asymptotic consistency of empirical deepest points

We shall now investigate the strong consistency of the empirical deepest points. The following result asserts uniform strong consistency of the empirical coordinatewise median, where the uniformity is over a subset of I the size of which grows with the sample size at an appropriate rate.

Fact 4.2.1. *Suppose that $X_{\mathbf{t}}$ has a density $f_{\mathbf{t}}$ in a neighbourhood of $m_{\mathbf{t}}$ for each $\mathbf{t} \in I$. Assume that for some $c_0, \eta_0 > 0$, we have $\inf_{|x-m_{\mathbf{t}_k}|<\eta_0} f_{\mathbf{t}_k}(x) > c_0$ for all $1 \leq k \leq d_n$ and $d_n \geq 1$. Then, if $log(d_n) = o(n)$ as $n \to \infty$, we have $\sup_{1\leq k\leq d_n} |\widehat{m}_{\mathbf{t}_k} - m_{\mathbf{t}_k}| \to 0$ as $n \to \infty$ almost surely.*

The proof of the above result can be obtained using the arguments in the proof of Corollary 6 in [13]. The authors of [13] considered the coordinatewise median for high dimensional data when the dimension increases with the sample size. Using sharp uniform bounds on the marginal empirical processes corresponding to the coordinate variables, they obtained the rate of convergence of $\widehat{\mathbf{m}}_c$ under the same set of assumptions used in Fact 4.2.1 (see Corollary 6 in [13]).

In practice, the index set I is most often a subset of $\mathbb{R}^p$ for some fixed $p \geq 1$, and the process $\mathbf{X}$ is observed at d_n grid points in I. We can then construct a functional estimator of $\mathbf{m}_c$ from $\widehat{\mathbf{m}}_c$ as follows. For any point $\mathbf{t}$ in I, which is not a grid point, we can define $\widehat{m}_{\mathbf{t}}$ by the average of the empirical medians corresponding to its k nearest grid points. Here $k \geq 1$ is a fixed integer. In that case, the uniform consistency of this functional estimator over the whole of I can be derived from Fact 4.2.1. For this derivation, let us assume that I is compact, and the grid points become dense in I as $n \to \infty$. In other words, for each $\mathbf{t} \in I$, any fixed neighbourhood of $\mathbf{t}$ will contain infinitely many grid points as $n \to \infty$. Let us also assume that the population deepest point $\mathbf{m}_c$ is a continuous function on I. Then, under the conditions assumed in Fact 4.2.1, it is straightforward to show that $\sup_{\mathbf{t}\in I} |\widehat{m}_{\mathbf{t}} - m_{\mathbf{t}}| \to 0$ as $n \to \infty$ almost surely.

We next consider the strong consistency of the empirical deepest point based on the spatial depth. As mentioned in Section 1, earlier work on the asymptotic consistency of the empirical spatial median in Hilbert spaces are restricted to showing convergence in the weak topology of the Hilbert space, i.e., convergence of $\langle \mathbf{h}, \widehat{\mathbf{m}}_s \rangle$ to $\langle \mathbf{h}, \mathbf{m}_s \rangle$ as $n \to \infty$ for each fixed $\mathbf{h}$ in the Hilbert space (see, e.g., [5, 11]). The following theorem asserts that the empirical spatial median in fact converges in the norm topology to its population version under very general conditions.

Theorem 4.2.2. *Let $\mathbf{X}$ be a random element in a strictly convex separable Hilbert space $\mathcal{X}$, and the distribution of $\mathbf{X}$ is nonatomic and not entirely supported on a line in $\mathcal{X}$. Also*

assume that for every $R > 0$, $\sup_{||\mathbf{x}||\leq R} E\{||\mathbf{x}-\mathbf{X}||^{-1}\} < \infty$. Then, $||\widehat{\mathbf{m}}_s - \mathbf{m}_s|| \to 0$ as $n \to \infty$ almost surely.

The assumption $\sup_{||\mathbf{x}||\leq R} E\{||\mathbf{x}-\mathbf{X}||^{-1}\} < \infty$ in the above theorem holds under very general circumstances. Since $\mathcal{X}$ is a separable Hilbert space, any $\mathbf{x} \in \mathcal{X}$ can be written as $\mathbf{x} = \sum_{k=1}^{\infty} x_k \phi_k$ for an orthonormal basis $\{\phi_k\}_{k\geq 1}$ of $\mathcal{X}$. Let $\mathbf{X} = \sum_{k=1}^{\infty} X_k \phi_k$. If some two dimensional marginal of $(x_1 - X_1, x_2 - X_2, \ldots)$ has a density that is bounded on bounded subsets of $\mathbb{R}^2$, then the aforementioned assumption holds.

Proof of Theorem 4.2.2. Let us first note that the assumptions in the statement of the theorem ensure that the spatial median is unique, and is the deepest point based on the spatial depth function (see Fact 2.1 in Section 2). Recall that the spatial median $\mathbf{m}_s$ is the minimizer of the function $g(\mathbf{x}) = E\{||\mathbf{x}-\mathbf{X}|| - ||\mathbf{X}||\}$. It follows from Lemma 2.1(i) in [5] that $g(\widehat{\mathbf{m}}_s) \to g(\mathbf{m}_s)$ as $n \to \infty$ almost surely. Let us denote the Hessian of g at $\mathbf{x}$ by $J_{\mathbf{x}} : \mathcal{X} \to \mathcal{X}$. It is a continuous linear operator on $\mathcal{X}$, and it is given by

$$\langle J_{\mathbf{x}}(\mathbf{z}), \mathbf{w}\rangle = E\left\{\frac{\langle \mathbf{z}, \mathbf{w}\rangle}{||\mathbf{x}-\mathbf{X}||} - \frac{\langle \mathbf{z}, \mathbf{x}-\mathbf{X}\rangle\langle \mathbf{w}, \mathbf{x}-\mathbf{X}\rangle}{||\mathbf{x}-\mathbf{X}||^3}\right\},$$

where $\mathbf{z}, \mathbf{w} \in \mathcal{X}$. A second order Taylor expansion of the function $\Phi(v) = g(\mathbf{m}_s + v\mathbf{h})$, where $\mathbf{h} = (\mathbf{x}-\mathbf{m}_s)/||\mathbf{x}-\mathbf{m}_s||$, at $v = ||\mathbf{x}-\mathbf{m}_s||$ about 0 yields

$$\begin{aligned} g(\mathbf{x}) - g(\mathbf{m}_s) &= E\left\{\frac{\langle \mathbf{m}_s - \mathbf{X}, \mathbf{x}-\mathbf{m}_s\rangle}{||\mathbf{m}_s-\mathbf{X}||}\right\} + \frac{1}{2}\langle J_{\mathbf{y}}(\mathbf{x}-\mathbf{m}_s), \mathbf{x}-\mathbf{m}_s\rangle, \\ &\quad \text{where } ||\mathbf{y}-\mathbf{m}_s|| < ||\mathbf{x}-\mathbf{m}_s|| \qquad (3) \\ &= \frac{1}{2}\langle J_{\mathbf{y}}(\mathbf{x}-\mathbf{m}_s), \mathbf{x}-\mathbf{m}_s\rangle. \qquad (4) \end{aligned}$$

The last equality holds because $E\{(\mathbf{m}_s - \mathbf{X})/||\mathbf{m}_s - \mathbf{X}||\} = \mathbf{0}$, which follows from the nonatomicity of $\mathbf{X}$ and Theorem 4.14 in [12]. It has been shown (see Proposition 2.1 in [6]) that the function g is strongly convex for all $\mathbf{x}$ in any closed and bounded ball around the origin. In other words, for each $A > 0$, there exists $c_A > 0$ such that

$$\langle J_{\mathbf{x}}(\mathbf{h}), \mathbf{h}\rangle \geq c_A, \qquad (5)$$

for every $||\mathbf{x}|| \leq A$ and $||\mathbf{h}|| = 1$. This inequality is the key argument in the proof. In the finite dimensional setup, the above inequality is obtained using the compactness of the the sets $\{\mathbf{x} : ||\mathbf{x}|| \leq A\}$ and $\{\mathbf{h} : ||\mathbf{h}|| = 1\}$ along with the positive definiteness of the Hessian J_x. It is remarkable that even when we do not have the compactness of closed and bounded balls, inequality (5) holds. Note that (5) is equivalent to $\langle J_{\mathbf{x}}(\mathbf{h}), \mathbf{h}\rangle \geq c_A||\mathbf{h}||^2$ for $\mathbf{h} \in \mathcal{X}$. Since for any $\epsilon > 0$, there exists $M > 0$ such that $P(||\mathbf{X}-\mathbf{m}_s|| > M) < \epsilon$, by Hoeffding's inequality we get that

$$\frac{1}{n}\sum_{i=1}^{n} I(||\mathbf{X}_i - \mathbf{m}_s|| > M) \leq 2\epsilon \text{ as } n \to \infty \text{ almost surely.}$$

So,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}||\mathbf{x}-\mathbf{X}_i|| &\geq \frac{1}{n}\sum_{i=1}^{n}(||\mathbf{x}-\mathbf{m}_s||-||\mathbf{X}_i-\mathbf{m}_s||)I(||\mathbf{X}_i-\mathbf{m}_s||\leq M) \\ &\quad -2\epsilon||\mathbf{x}-\mathbf{m}_s|| \\ &\geq ||\mathbf{x}-\mathbf{m}_s||(1-2\epsilon)-M-2\epsilon||\mathbf{x}-\mathbf{m}_s|| \;>\; 0\end{aligned}$$

if $\epsilon < 1/5$ and $||\mathbf{x}-\mathbf{m}_s|| > 5M$. Since $\widehat{\mathbf{m}}_s$ is the minimizer of $n^{-1}\sum_{i=1}^{n}||\mathbf{x}-\mathbf{X}_i||$, there exists $R > 0$ such that $P(||\widehat{\mathbf{m}}_s-\mathbf{m}_s|| \leq R \text{ as } n\to\infty) = 1$. This inequality along with (3) and (5) imply that

$$\langle J_{\mathbf{y}}(\widehat{\mathbf{m}}_s-\mathbf{m}_s), \widehat{\mathbf{m}}_s-\mathbf{m}_s\rangle \geq c_R||\widehat{\mathbf{m}}_s-\mathbf{m}_s||^2 \tag{6}$$

for some $c_R > 0$ as $n\to\infty$ almost surely. Since $g(\widehat{\mathbf{m}}_s)\to g(\mathbf{m}_s)$ as $n\to\infty$ almost surely, (4) and (6) together now imply that $||\widehat{\mathbf{m}}_s-\mathbf{m}_s||\to 0$ as $n\to\infty$ almost surely. This completes the proof of the theorem. □

The strong convergence of the empirical spatial median in the norm topology is not restricted to separable Hilbert spaces only. For a class of Banach spaces, which include L_p spaces for $1 < p < \infty$, it can be shown that the empirical spatial median converges in the norm topology to its population counterpart. As noted in the proof of the previous theorem, the key requirements there are the Frechet differentiability and the strong convexity of g at $\mathbf{m}_s$. In other words, we need

$$g(\mathbf{x})-g(\mathbf{m}_s) \geq g_1(\mathbf{m}_s)(\mathbf{x}-\mathbf{m}_s)+\phi(||\mathbf{x}-\mathbf{m}_s||),$$

where $g_1(\mathbf{m}_s) \in \mathcal{X}^*$ is the Frechet derivative of g at $\mathbf{m}_s$, and $\phi : [0,\infty)\to[0,\infty)$ is a convex lower semicontinuous function such that $\phi(0) = 0$ and $\phi(t) > 0$ if $t > 0$. Here $\mathcal{X}^*$ is the dual space of the Banach space $\mathcal{X}$. When $\mathcal{X}$ is a separable Hilbert space, $g_1(\mathbf{x}) = E\{(\mathbf{m}_s-\mathbf{X})/||\mathbf{m}_s-\mathbf{X}||\}$ and $\phi(t) = ct^2$ for some $c > 0$. The strong convexity of g in a class of separable Banach spaces, which includes L_p spaces for $1 < p < \infty$, follows from Proposition 1 and Theorem 3 in [1]. Then, using similar arguments as in the proof of Theorem 4.2.2, we get that $||\widehat{\mathbf{m}}_s-\mathbf{m}_s||\to 0$ as $n\to\infty$ almost surely in such spaces.

References

[1] E. Asplund. Fréchet differentiability of convex functions. *Acta Math.*, 121:31–47, 1968.

[2] F. K. Bedall and H. Zimmermann. Algorithm AS 143: The mediancentre. *J. Roy. Statist. Soc. Ser. C. Appl. Stat.*, 28(3):325–328, 1979.

[3] B. Böttcher. Feller processes: The next generation in modeling. Brownian motion, Lévy processes and beyond. *PLoS ONE*, 5(12):e15102, 2010.

[4] B. M. Brown. Statistical uses of the spatial median. *J. Roy. Statist. Soc. Ser. B*, 45(1):25–30, 1983.

[5] B. Cadre. Convergent estimators for the L_1-median of a Banach valued random variable. *Statistics*, 35(4):509–521, 2001.

[6] H. Cardot, P. Cénac, and P.-A. Zitt. Efficient and fast estimation of the geometric median in hilbert spaces with an averaged stochastic gradient algorithm. To appear in Bernoulli, 2012.

[7] A. Chakraborty and P. Chaudhuri. On data depth in infinite dimensional spaces. Technical Report No. R4/2012, Theoretical Statistics and Mathematics Unit. Indian Statistical Institute, Kolkata, India, 2012.

[8] S. S. Dhar and P. Chaudhuri. On the statistical efficiency of robust estimators of multivariate location. *Stat. Methodol.*, 8(2):113–128, 2011.

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

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

[11] D. Gervini. Robust functional estimation using the median and spherical principal components. *Biometrika*, 95(3):587–600, 2008.

[12] J. H. B. Kemperman. The median of a finite measure on a Banach space. In *Statistical data analysis based on the L_1-norm and related methods (Neuchâtel, 1987)*, pages 217–230. North-Holland, Amsterdam, 1987.

[13] M. R. Kosorok and S. Ma. Marginal asymptotics for the "large p, small n" paradigm: with applications to microarray data. *Ann. Statist.*, 35(4):1456–1486, 2007.

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

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

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

[17] G. Maguluri and K. Singh. On the fundamentals of data robustness. In *Robust inference*, volume 15 of *Handbook of Statist.*, pages 537–549. North-Holland, Amsterdam, 1997.

[18] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1991.

[19] R. Serfling. A depth function and a scale curve based on spatial quantiles. In *Statistical data analysis based on the L_1-norm and related methods (Neuchâtel, 2002)*, Stat. Ind. Technol., pages 25–38. Birkhäuser, Basel, 2002.

[20] K. Singh. Paradoxes in robustness. In *Statistics and Probability: A Raghu Raj Bahadur Festschrift, ed. J. K. Ghosh, S. K. Mitra, K. R. Parthasarathy and B. L. S. Prakasa Rao*, pages 531–537. Wiley Eastern Limited Publishers, 1993.

[21] C. G. Small. A survey of multidimensional medians. *Int. Statist. Rev.*, 58(3):263–277, 1990.

[22] Y. Sun and M. G. Genton. Functional boxplots. *J. Comput. Graph. Statist.*, 20(2):316–334, 2011.

[23] M. Valadier. La multi-application médianes conditionnelles. *Z. Wahrsch. Verw. Gebiete*, 67(3):279–282, 1984.

[24] Y. Vardi and C.-H. Zhang. The multivariate L_1-median and associated data depth. *Proc. Natl. Acad. Sci. USA*, 97(4):1423–1426 (electronic), 2000.

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