Allgemeines Statistisches Archiv 88, 1–20
© Physica-Verlag 2004, ISSN 0002-6018

Robustness of weighted L^p-depth and L^p-median

By Yijun Zuo*

Summary: L^p-norm weighted depth functions are introduced and the local and global robustness of these weighted L^p-depth functions and their induced multivariate medians are investigated via influence function and finite sample breakdown point. To study the global robustness of depth functions, a notion of finite sample breakdown point is introduced. The weighted L^p-depth functions turn out to have the same low breakdown point as some other popular depth functions. Their influence functions are also unbounded. On the other hand, the weighted L^p-depth induced medians are globally robust with the highest possible breakdown point for any reasonable estimator. The weighted L^p-medians are also locally robust with bounded influence functions for suitable weight functions. Unlike other existing depth functions and multivariate medians, the weighted L^p depth and medians are easy to calculate in high dimensions. The price for this advantage is the lack of affine invariance and equivariance of the weighted L^p depth and medians, respectively.

Keywords: Breakdown point, depth function, efficiency, equivariance, influence function, L^p-norm, median, robustness. JEL C10, C14.

1. Introduction

Since the introduction of the halfspace depth (Tukey, 1975; Donoho and Gasko, 1992) and the simplicial depth (Liu, 1990), data depth has become an important tool for high dimensional data ordering, analysis, and inference. The key motivation of depth functions in the location setting is to provide a center-outward ordering of observations in high dimensions where, unlike in the one-dimensional case, no natural and meaningful order principle of points exists. Some general treatments of depth functions have been provided by Liu *et al.* (1999), Zuo and Serfling (2000a, b) and Mosler (2002). Among many interesting applications of data depth, employing depth (and consequently a center-outward ordering) to define multivariate medians is a paradigm.

A legitimate concern for depth functions and especially depth induced medians is: How sensitive are they with respect to the assumed underlying distribution (data)? Are they robust, locally and globally?

In this paper, we extend the L^p-depth defined in Zuo and Serfling (2000a) and introduce a class of weighted L^p-depth functions. We then focus on the robustness of the weighted L^p-depth functions and medians induced from them. Specifically, we investigate the local and the global robustness of these depth functions and depth medians via influence function and finite sample breakdown point, respectively. The latter notion, introduced by Donoho and Huber (1983), has become the most prevailing quantitative measure of

Received: 10.06.2003 / Revised: 11.12.2003
* The author thanks the referees for their very insightful and constructive comments and suggestions which led to corrections and substantial improvements. Supported in part by NSF Grants DMS-0071976 and DMS-0134628.

global robustness of estimators, especially location and scatter estimators. We adapt the original definition and introduce in this paper a notion of breakdown point for depth functions.

It turns out that, like some popular depth functions, the weighted L^p-depth function, with a low breakdown point, is not very robust globally. Its influence function is also unbounded, although the local shift sensitivity of this depth function is bounded for suitable weight functions. On the other hand, the multivariate median induced from the weighted L^p-depth is globally robust with the best possible breakdown point and locally robust as well with a bounded influence function for suitable weight and distribution functions.

The article is organized as follows. In Section 2 we define a class of weighted L^p-depth functions and investigate the local and global robustness of the depth functions via influence function and a notion of finite sample breakdown point introduced in the same section. Section 3 is devoted to the study of the local and the global robustness of the weighted L^p-depth induced multivariate medians. The paper ends in Section 4 with some concluding remarks.

2. Weighted L^p-depth

Zuo and Serfling (2000a) defined a depth function based on the L^p-norm. Different distances (norms) relative to the underlying distribution (data) were treated with equal importance (equally weighted). In practice, the importance (weight, cost, penalty, or incentive) may not be the same for different distances (norms). This motivates us to define weighted L^p-depth as follows

$$WL^pD(x;F) = \frac{1}{1 + Ew(\|x - X\|_p)}, \tag{1}$$

where w is a suitable weight function on $[0,\infty)$, $X \sim F$ and "$\|\cdot\|_p$" stands for the L^p-norm (when $p = 2$ we have the Euclidean norm and write $\|\cdot\|$ for $\|\cdot\|_2$). We assume that w is non-decreasing and continuous on $[0,\infty)$ with $w(\infty-) = \infty$. We rule out the non-existence case of $Ew(\|x - X\|_p)$ (which gives rise to an unappealing depth 0 for all points) and assume that $Ew(\|x-X\|_p) < \infty$ for any $x \in \mathbb{R}^d$. The latter holds true if $Ew(\|X\|_p) < \infty$ and w does not increase too rapidly in the sense that $w(2\|x\|_p) \le Cw(\|x\|_p)$ for some $C > 0$ and any $x \in \mathbb{R}^d$ (such w includes $w(x) = \sum_{i=0}^n a_i x^{b_i}, n \ge 0, a_n, b_n > 0, a_i, b_i \ge 0$). An empirical version of the weighted L^p-depth function is obtained by replacing F of X in $Ew(\|x - X\|_p) = \int w(\|x - t\|_p)dF(t)$ with its empirical version F_n.

The weighted L^p-depth possesses some desirable properties of depth functions (see Zuo and Serfling, 2000a, b). For example, it is translation invariant (can be affine invariant for $p = 2$ under some modification), maximized at the center of a (centrally) symmetric distribution for convex w, decreasing when a point moves along a ray stemming from the deepest point, and vanishing at infinity; see Zuo and Serfling (2000a) for more related discussions. We now investigate the robustness of the weighted L^p-depth.

2.1. Influence function. Denote by δ_x the point mass probability distribution at a fixed point $x \in \mathbb{R}^d$. For a given distribution F in $\mathbb{R}^d$ and an $\epsilon > 0$, the distribution resulting from contaminating F with an ϵ amount of the point mass distribution δ_x is denoted by $F(\epsilon, \delta_x) = (1-\epsilon)F + \epsilon\,\delta_x$. The *influence function* of a statistical functional T at a given point $x \in \mathbb{R}^d$ for a given F is defined as (Hampel *et al.*, 1986)

$$IF(x; T, F) = \lim_{\epsilon \to 0^+} \frac{T(F(\epsilon, \delta_x)) - T(F)}{\epsilon}. \tag{2}$$

$IF(x; T, F)$ describes the relative effect (influence) on T of an infinitesimal point-mass contamination at x, and captures the local robustness of T. The supremum norm of the influence function is called the *gross error sensitivity* of T at F (Hampel *et al.*, 1986). That is,

$$GRE(T, F) = \sup_{x \in \mathbb{R}^d} \|IF(x; T, F)\|_p. \tag{3}$$

$GRE(T, F)$ is the maximum relative effect on T of an infinitesimal point-mass contamination and measures the local (and the global as well) robustness of T.

When an observation x is slightly shifted to a neighboring point y, the effect on the functional T can be measured by means of $IF(y; T, F) - IF(x; T, F)$. A measure for the worst case standardized effect of 'wiggling' is provided by the local shift sensitivity (LSS) (Hampel *et al.*, 1986)

$$LSS(T, F) = \sup_{x \neq y} \|IF(y; T, F) - IF(x; T, F)\|_p / \|y - x\|_p. \tag{4}$$

Note that in the original definitions of GRE and LSS, the Euclidean norm is employed. We adopt the L_p norm here simply for the consistency with the underlying metric.

In this subsection we investigate the robustness of the weighted L^p-depth via influence function and gross error and local shift sensitivity. For convenience, we sometimes write $F(\epsilon, \delta_x) = F_\epsilon$ for a fixed x. It follows in a straightforward fashion that

$$\begin{aligned}
&\frac{WL^pD(y; F(\epsilon, \delta_x)) - WL^pD(y; F)}{\epsilon} \\
&= \frac{\epsilon \int w(\|y-t\|_p)dF(t) - \epsilon w(\|y-x\|_p)}{\epsilon\big(1 + \int w(\|y-t\|_p)dF(\epsilon, \delta_x)(t)\big)\big(1 + \int w(\|y-t\|_p)dF(t)\big)} \\
&\to \frac{Ew(\|y-X\|_p) - w(\|y-x\|_p)}{(1 + Ew(\|y-X\|_p))^2}, \quad \text{as } \epsilon \to 0.
\end{aligned}$$

Hence, we have

Proposition 1 *The weighted L^p-depth function has the following influence function*

$$IF(x;\, WL^pD(y;F),F) = \frac{Ew(\|y-X\|_p) - w(\|y-x\|_p)}{(1+Ew(\|y-X\|_p))^2}. \tag{5}$$

Clearly, the influence function is continuous. If w is Lipschitz continuous or differentiable, then so is the influence function. The influence of an infinitesimal amount of point mass contamination on the L^p-depth of point y, however, becomes unbounded as $x \to \infty$. That is, $GRE(WL^pD(y;F);F) = \infty$. When the point mass contamination occurs at the point y, the influence on the weighted L^p-depth of y becomes a constant $(Ew(\|y-X\|_p) - w(0))/(1+Ew(\|y-X\|_p))^2$.

The influence function of $WL^pD(y;F)$ and the asymptotic representation of $WL^pD(y;F_n)$ have the following connection

$$WL^pD(y;F_n) - WL^pD(y;F) = \frac{1}{n}\sum_{i=1}^{n} IF(X_i;\, WL^pD(y;F),F) + o_p(n^{-1/2}). \tag{6}$$

Hence, the sample weighted L^p-depth of point y is asymptotically normal with an asymptotic mean $WL^pD(y;F)$ and an asymptotic variance $E(IF(X;\, WL^pD(y),F))^2/n$.

The local shift sensitivity of the weighted L^p-depth of point y is

$$LSS(WL^pD(y;F);F) = \sup_{x_1 \neq x_2} \frac{|w(\|y-x_1\|_p) - w(\|y-x_2\|_p)|}{(1+Ew(\|y-X\|_p))^2\|x_1-x_2\|_p}. \tag{7}$$

When w is Lipschitz continuous with a constant C, then by the triangle inequality we see that $LSS(WL^pD(y;F);F) \le C/(1+Ew(\|y-X\|_p))^2 < \infty$. Thus, the local shift sensitivity of the weighted L^p-depth is bounded when w is Lipschitz continuous.

The empirical influence function of a statistical functional T at the empirical distribution function F_n of F can be defined by (see Hampel *et al.*, 1986)

$$IF(x;T(F_n),F_n) = \frac{T\Big((1-\frac{1}{n+1})F_n + \frac{1}{n+1}\delta_x\Big) - T(F_n)}{1/(n+1)} \tag{8}$$

$$= (n+1)(T(X_1,\cdots,X_n,x) - T(X_1,\cdots,X_n)). \tag{9}$$

For the weighted L^p-depth of point y, a straightforward calculation yields

$$IF(x;\, WL^pD(y;F_n),F_n) = \frac{\frac{1}{n}\sum_i w(\|y-X_i\|_p) - w(\|y-x\|_p)}{\Big(1+\frac{1}{n}\sum_i w(\|y-X_i\|_p)\Big)\Big(1+\frac{1}{n+1}\Big(\sum_i w(\|y-X_i\|_p) + w(\|y-x\|_p)\Big)\Big)}. \tag{10}$$

This empirical influence function clearly converges with probability 1 to the population counterpart $IF(x;\, WL^pD(y;F),F)$. The empirical influence function possesses many similar properties of the population counterpart. We conclude this subsection with the following remark.

REMARK (i) The influence function of the weighted L^p-depth of a point y is bounded whenever the point mass contamination occurs at a point x within a bounded set but becomes unbounded when x moves to ∞ (unlike in the halfspace depth case where the influence function is always bounded; see Romanazzi, 2001). (ii) The local shift sensitivity of the weighted L^p-depth of a point y is bounded when the weight function is Lipschitz continuous (unlike in the halfspace depth case where the local shift sensitivity is unbounded; see Romanazzi, 2001).

2.2. FINITE SAMPLE BREAKDOWN POINT. The influence function only captures the local robustness of a statistical function. To depict the entire robustness picture of a statistical function, we need a global robustness measure. The breakdown point turns out to be a prevailing one. The notion of finite sample breakdown point of an estimator was first introduced by Donoho and Huber (1983). Roughly speaking, the finite sample breakdown point of an estimator is the minimum fraction of 'bad' points in a data set that can render the estimator useless. In the location setting, if the estimator becomes unbounded under some contamination, then we say the estimator becomes useless. In the scale (or scatter matrix) setting, if the determinant of the estimator becomes arbitrary small or large under some contamination, we say the estimator becomes useless. In the statistical depth function setting, we now introduce a notion of breakdown point.

Depth functions are usually nonnegative (and bounded). In our following discussion, we assume that $D(x; F_n) \geq 0$. This indeed is true for all common depth functions and the weighted L^p-depth functions. The boundary depth value 0 corresponds to a very special location of a point and conveys very little information about the underlying data set. In the spirit of Donoho and Huber (1983, see pages 167-168), we say that the depth of a point breaks down if under some contamination its non-boundary depth value becomes the boundary value 0 and vice versa. A non-zero depth to a zero depth corresponds to an *explosion* breakdown. The minimum of all point-wise breakdown points will be called the breakdown point of the depth function. Motivated by this, we formally introduce a notion of breakdown point for depth functions. Define $\log a - \log b = 0$, if $a = b = 0$.

DEFINITION 1 The *finite sample breakdown point* of the depth $D(x; X^n)$ of a point $x \in \mathbb{R}^d$ at a sample $X^n = \{X_1, \cdots, X_n\}$, $BP(D(x; X^n))$, is defined as

$$BP(D(x; X^n)) = \min\Big\{\frac{m}{n}: \ \sup_{X^n_m} \Big|\log D(x; X^n) - \log D(x; X^n_m)\Big| = \infty\Big\}, \tag{11}$$

where X^n_m is an arbitrary contaminated data set resulting from replacing m original sample points of X^n by m arbitrary points in $\mathbb{R}^d$. The breakdown point of the depth function, $BP(D; X^n)$, is then defined to be $BP(D; X^n) = \min_{x\in\mathbb{R}^d} BP(D(x; X^n))$.

The above notion of breakdown point is based on replacement contamination. A notion based on addition of contamination can also be defined (see Donoho and Huber, 1983; and Zuo, 2001).

We remark that the above definition focuses on the special explosion robustness aspect of a depth function and that other versions of breakdown point focusing on other robustness aspects of a depth function may also be introduced. For example, one might incorporate the *implosion* breakdown concept into the above definition and assert that the depth of a point also breaks down whenever it reaches the upper boundary value (1 in many cases) under some contamination.

Note that in the light of the breakdown point of the depth of a point we can study the breakdown point of the αth depth region $D^\alpha(X^n) := \{x : D(x; X^n) \geq \alpha\}$ $(0 < \alpha < 1)$, which can be defined as $\min_{x \in D^\alpha(X^n)}$ $\mathrm{BP}(D(x; X^n))$. Recently, Cramer (2003) and Mosler and Cramer (2004) also studied the robustness of depth functions with a focus on depth induced contours and deepest point.

A desirable property of a statistical depth function $D(\cdot; F)$, as discussed in Zuo and Serfling (2000a, b), is 'vanishing at infinity', that is, $\sup_{\|x\|_p \geq M}$ $D(x; X^n) \to 0$ as $M \to \infty$. This property not only insures the boundedness of the αth depth region but also facilitates many technical treatments of depth function related problems. Intuitively, it is also sensible: When a point moves away from the deepest point to infinity, the point becomes the least deep one. If this property fails to hold for some contaminated data X^n_m for some m, then we see that $|\log D(x, X^n) - \log D(x, X^n_m)| \to \infty$ as $\|x\| \to \infty$. Hence, the depth function breaks down in the sense of the above definition.

Another desirable feature of depth functions is the invariance property. The depth function $D(\cdot;\, \cdot)$ is *affine invariant* if $D(x; X^n) = D(Ax+b; AX^n+b)$ for any non-singular $d \times d$ matrix A and vector $b \in \mathbb{R}^d$, where $AX^n = \{AX_1, \cdots, AX_n\}$. When $A = I_{d\times d}$, an identity matrix, $D(\cdot;\cdot)$ is said to be *translation invariant*. It is seen that the breakdown point of a depth function reserves the invariance property of the underlying depth function, namely $BP(D, X^n) = BP(D; AX^n + b)$.

Now, we calculate the breakdown point of some popular depth functions.

EXAMPLE 1 The Tukey halfspace depth (Tukey, 1975) of a point $x \in \mathbb{R}^d$ with respect to a given data set X^n is the minimum fraction of sample points contained in a closed halfspace with x on its boundary. That is

$$HD(x; X^n) = \min_{H_x}\{F_n(H_x) : \; H_x \text{ closed halfspace with } x \text{ on its boundary}\}. \tag{12}$$

Clearly the points on the boundary of the convex hull formed by the sample points in X^n possess halfspace depth $1/n$ and the points outside the convex hull have halfspace depth 0. When we move one sample point on the boundary to infinity, the halfspace depth of the point keeps the same. Hence, there are points with original 0 (or positive) depth which have positive (or 0) depth under the contamination. It follows immediately that $BP(HD;\, X^n) = 1/n$. The halfspace depth of a point does not contain all the information about the relative 'distance' of the point with respect to the center of the data cloud. The value of the halfspace depth of a point can not be employed directly to identify outliers among the sample points. Indeed, outliers and the points on the boundary of the convex hull may all have the same depth $1/n$. This disadvantage of the halfspace depth is exactly captured by its global robustness: a low breakdown point. Note that *the depth of a single point* deep inside halfspace depth contours could be more resistant to a small amount of contamination and hence have a higher breakdown point. For example, if $HD(x; X^n) = k/n$, $k \le n$, then $BP(HD(x; X^n)) = k/n$. The depth of the deepest point has the highest breakdown point.

Example 2 The simplicial depth (Liu, 1990) of a point $x \in \mathbb{R}^d$ with respect to a given data set X^n is the fraction of simplices formed by sample points that contain the point x. That is,

$$SD(x; X^n) = \sum_{i_1, \cdots, i_{d+1}} I(x \in S[X_{i_1}, \cdots, X_{i_{d+1}}]) \Big/ \binom{n}{d+1}, \tag{13}$$

where $S[x_1, \cdots, x_{d+1}]$ stands for a simplex with x_j as its vertices $j = 1, \cdots, d+1$, $I(\cdot)$ is the indicator function, and i_j, $j = 1, \cdots, d+1$ are arbitrary $d+1$ numbers from $\{1, \cdots, n\}$. Clearly for this depth function, an argument similar to the one used for HD gives $BP(SD;\, X^n) = 1/n$.

Now, we consider the finite sample breakdown point of the weighted L^p-depth. Clearly, the weighted L^p-depth of any $x \in \mathbb{R}^d$ with respect to X^n is positive and approaches 0 as $\|x\| \to \infty$. Now moving a single sample point of X^n to infinity, we see that $WL^pD(x; X_1^n) \to 0$ for any $x \in \mathbb{R}^d$. Hence for continuous w with $w(\infty-) = \infty$, we have $BP\,(WL^pD; X^n) = 1/n$.

Proposition 2 *For continuous w with $w(\infty-) = \infty$, the sample weighted L^p-depth has a breakdown point: $BP(WL^pD; X^n) = 1/n$.*

The weighted L^p-depth function has, unfortunately, a low breakdown point $1/n$, just like the halfspace and the simplicial depth functions do. A natural question is: Is there some depth function that can have a (much) higher breakdown point? The answer is given in the following example.

Example 3 The projection depth (Liu, 1992; Zuo and Serfling, 2000a, b; and Zuo, 2003) of a point $x \in R^d$ with respect to a given data set X^n is defined based on the Stahel (1981) and Donoho (1982) outlyingness function $O(x; X^n)$ as

$$PD(x; X^n) = 1/(1 + O(x; X^n)), \tag{14}$$

where $O(x; X^n) = \sup_{\|u\|=1} |u'x - \mu(u'X^n)|/\sigma(u'X^n)$ with μ and σ being univariate location and scale estimators and $u'X = \{u'X_1, \cdots, u'X_n\}$. Consider robust μ and σ such as the median (Med) and a modified median absolute deviation (MAD_d): $\text{MAD}_d(x^n) = \text{Med}_d\{|x_i - \text{Med}(x^n)|\}$ with $\text{Med}_d = (x_{(\lfloor (n+d)\rfloor/2)} + x_{(\lfloor (n+1+d)\rfloor/2)})/2$ where $x^n = \{x_1, \cdots, x_n\}$ is a univariate data set and $x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$ are ordered values. Assume that X^n is in *general position*, that is, no more than d points from X^n lie in any $d-1$ dimensional hyperplane. Then $PD(x; X^n)$ is positive for any $x \in \mathbb{R}^d$ and approaches 0 as $\|x\| \to \infty$. Moving $m = \lfloor (n-d+1)/2 \rfloor$ points of X^n to a hyperplane determined by some other d points of X^n, we see that $O(x; X^n_m) = \infty$ for all x not on the hyperplane. On the other hand, this will never happen for any $x \in \mathbb{R}^d$ if $m < \lfloor (n-d+1)/2 \rfloor$ and furthermore $O(x, X^n_m) \to \infty$ as $\|x\| \to \infty$ since with such an m we have $\sup_{X^n_m} \sup_{\|u\|=1} \mu(u'X^n_m) < \infty$ and $0 < \inf_{X^n_m} \inf_{\|u\|=1} \sigma(u'X^n_m) \le \sup_{X^n_m} \sup_{\|u\|=1} \sigma(u'X^n_m) < \infty$ (see the proof of Theorem 3.4 of Zuo, 2003). Hence, we conclude that $BP(PD; X^n) = \lfloor (n-d+1)/2 \rfloor / n$ (a detailed proof is somewhat involved and shall be reported elsewhere).

The next natural question is: How high can the breakdown point of a depth function be? We answer this question for a class of translation invariant depth functions.

Proposition 3 *Let $D(\cdot;\cdot)$ be translation invariant and vanish at infinity. Assume there is a point x_0 such that $D(x_0; X^n) > 0$ for a given data set X^n. Then $BP(D; X^n) \le \lfloor (n+1)/2 \rfloor / n$.*

Proof By translation invariance, assume without loss of generality that $D(0; X^n) > 0$. Let $m = \lfloor (n+1)/2 \rfloor$. Let $X^n_{m,1} = \{X_1 + b, \cdots, X_m + b, X_{m+1}, \cdots, X_n\}$ and $X^n_{m,2} = X^n_{m\,1} - b$ for $b \in \mathbb{R}^d$. By 'vanish at infinity' property, $D(b; X^n) \to 0$ as $\|b\| \to \infty$. If this no longer holds for $D(b; X^n_{m\,1})$, then m contaminating points can break down the depth function. Otherwise, we have $D(0, X^n_{m,2}) = D(b, X^n_{m,1}) \to 0$ as $\|b\| \to \infty$, which again shows that m contaminating points can break down the depth function. Thus, we have the desired result.

3. Weighted L^p-median

A straightforward application of the weighted L^p-depth (or any other depth) is to define the deepest point based on the depth as a multivariate analogue of the univariate median. The multivariate median induced from the weighted L^p-depth is called *weighted L^p-median.* That is,

$$WL^pM(F) = \arg\sup_{x\in\mathbb{R}^d} WL^pD(x; F) = \arg\inf_{x\in\mathbb{R}^d} \phi^p(x; F), \qquad (15)$$

where $\phi^p(x; F) = \int (w(\|t - x\|_p) - w(\|t\|_p))dF(t)$, $1 \le p < \infty$. If there is a non-uniqueness problem, we take the average to take care of it.

If w is Lipschitz continuous with a constant C (such w includes $w(x) = Cx + B$), then $\phi^p(x; F)$ always exists since $|\phi^p(x; F)| \le C\|x\|_p$ for fixed x. If $Ew(\|X\|_p) < \infty$ and w does not increase too rapidly in the sense that $w(2x) \le C_1 w(x)$ (see Section 2), then $\phi^p(x; F)$ is again well defined since $w(\|t - x\|_p) \le w(2\|x\|_p) + C_1 w(\|X\|_p)$. From now on, we assume that $\phi^p(x; F)$ is well defined.

The weighted L^p median exists for 'reasonable' weight function w and is unique for 'reasonable' distribution F. It is also Fisher consistent for F symmetric about θ, that is $WL^pM(F) = \theta$. F of X is said to be symmetric about θ if $X - \theta$ and $\theta - X$ have the same distribution. If F is symmetric about 0, then the weighted L^p-median is odd, that is, $WL^pM(F_X) = -WL^pM(F_{-X})$.

PROPOSITION 4 (1) *$WL^pM(F)$ exists if w is continuous and $P\{x : w(\|x\|_p) < w(\infty-)\} > 0$;* (2) *$WL^pM(F)$ is unique if w is strictly convex or if it is convex and F is not concentrated on a line in $\mathbb{R}^d$;* (3) *$WL^pM(F)$ is Fisher consistent for symmetric F and convex w;* (4) *$WL^pM(F)$ is odd for F symmetric about the origin;* and (5) *$WL^pM(F_n)$ is odd and translation equivariant and is symmetric about the center θ of symmetric F and is unbiased for θ provided that EX exists.*

Note that the weighted L^p-median becomes regular M-estimates of multivariate location under some regularity conditions on w and F (see Huber, 1981) for discussions on M-estimates of location. When $w(x) = x^\alpha$, the sample weighted L^p-median becomes the L_q^α estimate ($q = p$) of Rao (1988).

A particularly interesting case of the general weighted L^p-median is the one with $p = 2$, although different norms are equivalent in some sense. The corresponding weighted L^2-median is

$$WL^2M(F) = \arg\inf_{x\in\mathbb{R}^d} \phi^2(x; F) = \arg\inf_{x\in\mathbb{R}^d} \int (w(\|t - x\|_2) - w(\|t\|_2))dF(t). \tag{16}$$

When $w(x) = x$, the general weighted L^2-median becomes the so-called L_1 or spatial median in the literature. An immense amount of research related to L_1-median has been carried out (see for example Hayford, 1902; Haldane, 1948; Brown, 1983; Pollard, 1984; Rao, 1988; Small, 1990; Chaudhuri, 1992; Chakraborty, Chaudhuri and Oja, 1998; Dodge and Rousson, 1999; and Hettmansperger and Randles; 2002). The uniqueness of the L_1 median is proved by Milasevic and Ducharme (1987), among others. The L_1-median is Fisher consistent even for halfspace symmetric F (see Zuo and Serfling, 2000c) for the notion of halfspace symmetry and the proof.

In the following we confine attention to the most interesting case: the weighted L^p-median with $p = 2$. Many of the following results can actually be extended for more general p, though.

3.1. Influence function. In this subsection we investigate the local robustness of the weighted L^2-median via its influence function. For simplicity, denote by $\theta(F)$ the weighted L^2-median.

Proposition 5 *Let $P(X = x) = 0$ for any $x \in \mathbb{R}^d$ and w be continuously twice differentiable. Then the influence function of the weighted L^2-median $\theta(F)$ is given by*

$$IF(x;\theta(F),F) = A^{-1}w^{(1)}(\|x-\theta(F)\|)\frac{x-\theta(F)}{\|x-\theta(F)\|}I(x\neq\theta(F)) \qquad (17)$$

as long as

$$A = \int\Big(w^{(1)}(\|t-\theta(F)\|)\Big(\frac{I_{d\times d}}{\|t-\theta(F)\|}-\frac{t-\theta(F)}{\|t-\theta(F)\|^2}\frac{(t-\theta(F))'}{\|t-\theta(F)\|}\Big) + w^{(2)}(\|t-\theta(F)\|)\frac{t-\theta(F)}{\|t-\theta(F)\|}\frac{(t-\theta(F))'}{\|t-\theta(F)\|}\Big)dF(t) \qquad (18)$$

exists and is invertible and consequently

$$GES(\theta(F);F) \le \sup_{x\in\mathbb{R}^d} w^{(1)}(\|x-\theta(F)\|)I(x\neq\theta(F))\Big/\min_{1\le i\le d}|\lambda_i(A)|, \qquad (19)$$

where $\lambda_i(A)$ are the eigenvalues of A $(1\le i\le d)$.

When $X \overset{d}{=} -X$, then $\theta(F) = 0$, where '$\overset{d}{=}$' stands for 'equal in distribution'. Furthermore, if $e_i'X \overset{d}{=} -e_i'X$, $1\le i\le d$, with $e_i' = (e_{ij})_{1\le j\le d}$ and $e_{ij} = I(i=j)$, $1\le j\le d$, then

$$A^{-1} = \mathrm{diag}\Big(1\Big/E\Big(w^{(1)}(\|X\|)/\|X\|\Big(1-X_i^2/\|X\|^2\Big) + w^{(2)}(\|X\|)X_i^2/\|X\|^2\Big)\Big)_{1\le i\le d} \qquad (20)$$

if and only if $E[w^{(1)}(\|X\|)/\|X\|(1-X_i^2/\|X\|^2)+w^{(2)}(\|X\|)X_i^2/\|X\|^2]$ exists and $\neq 0$, $1\le i\le d$.

When X is spherical symmetric with a density f, then $\theta(F) = 0$ and $A^{-1} = c^{-1}I_{d\times d}$ with

$$c = \frac{E(\frac{(d-1)w^{(1)}(\|X\|)}{\|X\|} + w^{(2)}(\|X\|))}{d} = \frac{2\pi^{d/2}}{d\,\Gamma(d/2)}\int_0^\infty [\frac{(d-1)w^{(1)}(r)}{r} + w^{(2)}(r)]\, r^{d-1}f(r^2)\,dr \qquad (21)$$

provided that c exists and $\neq 0$.

It is not difficult to see that a sufficient condition for the existence of A is the existence of $Ew^{(2)}(\|X-\theta(F)\|)$ and $Ew^{(1)}(\|X-\theta(F)\|)/\|X-\theta(F)\|$. The latter is guaranteed if $w^{(2)}(y)$ and $w^{(1)}(y)/y$ are bounded (a.s. Lebesgue measure) on $(0,\infty)$. Examples of such w include $w(y)=ay+b$ or cy^2+d for $a>0, b\geq 0, c>0$ and $d\geq 0$. For the invertibility of A, by the Cauchy-Schwarz inequality, $w^{(i)}(y)\geq 0$ on $(0,\infty)$, $i=1,2$, $\sum_{i=1}^{2} P(w^{(i)}(\|X-\theta(F)\|)>0)>0$, and $P(H_{\theta(F)})=0$ for any hyperplane $H_{\theta(F)}$ containing $\theta(F)$ are sufficient conditions. Indeed, it is readily seen that these conditions insure a positive definite matrix A.

Clearly, when A^{-1} exists and the contamination occurs over a bounded set in $\mathbb{R}^d$, the influence of a point mass contamination on the weighted L^2-median is bounded. Furthermore, the influence function of the weighted L^2-median is bounded as long as $w^{(1)}(\cdot)$ is bounded on $[0,\infty)$ (such w includes $w(x)=ax+b$ with $a>0$ and $b\geq 0$) and A^{-1} exists. Due to the jump at the point $x=\theta(F)$, the local shift sensitivity of the weighted L^2-median can be unbounded, though.

Note that for symmetric F, by translation equivariance, we can assume, w. l. o. g., that F is symmetric about the origin. Distributions F satisfying $e_i'X \stackrel{d}{=} -e_i'X$, $1\leq i\leq d$ include the so-called d-version symmetric distributions (see Zuo, 2003; and Zuo, Cui and Young, 2004): $u'X \stackrel{d}{=} a(u)Z$ with $a(u)=a(-u)>0$ for any $\|u\|=1$ and univariate random variable $Z \stackrel{d}{=} -Z$. Special cases of d-version symmetry include spherical symmetry with $a(u)=$ constant, elliptical symmetry with $a(u)=\sqrt{u'\Sigma u}$ for some positive definite matrix Σ and α-symmetry with $a(u)=(\sum_i c_i|u_i|^{\alpha})^{1/\alpha}$ with $c_i\geq 0$, $0<\alpha\leq 2$ and $u'=(u_1,\cdots,u_d)$.

In the special case $w(x)=ax+b$ with $a>0$ and $b\geq 0$ (this covers the L_1-median case), we can simplify the results in the above proposition and have

Corollary 1 *Let $P(X=x)=0$ for any $x\in\mathbb{R}^d$ and $w(x)=ax+b$, $a>0$ and $b\geq 0$. Then*

$$IF(x;\theta(F),F)=A^{-1}\frac{x-\theta(F)}{\|x-\theta(F)\|}I(x\neq\theta(F)) \tag{22}$$

provided that

$$A=\int\Big(\frac{I_{d\times d}}{\|t-\theta(F)\|}-\frac{t-\theta(F)}{\|t-\theta(F)\|^2}\frac{(t-\theta(F))'}{\|t-\theta(F)\|}\Big)dF(t) \tag{23}$$

exists and is invertible and consequently

$$GES(\theta(F);F)=1/\min_{1\leq i\leq d}|\lambda_i(A)|<\infty \tag{24}$$

where $\lambda_i(A)$ are the eigenvalues of A $(1\leq i\leq d)$.

When $X \stackrel{d}{=} -X$, then $\theta(F) = 0$. Furthermore, if $e_i' X \stackrel{d}{=} -e_i' X$, $1 \le i \le d$, with $e_i' = (e_{ij})_{1\le j\le d}$ and $e_{ij} = I(i = j)$, $1 \le j \le d$, then

$$A^{-1} = \mathrm{diag}\Big(1\Big/E(1/\|X\| - X_i^2/\|X\|^3)\Big)_{1\le i\le d} \tag{25}$$

provided that $E(1/\|X\| - X_i^2/\|X\|^3)$ exists and $\neq 0$, $1 \le i \le d$.

When X is spherical symmetric with a density f, then $\theta(F) = 0$ and

$$A^{-1} = c^{-1} I_{d\times d}, \quad c = \frac{(d-1)E(1/\|X\|)}{d} = \frac{2(d-1)\pi^{d/2}}{d\,\Gamma(d/2)} \int_0^\infty r^{d-2} f(r^2)\, dr \tag{26}$$

provided that c exists. When X is d-variate standard normal, then $\theta(F) = 0$ and

$$A^{-1} = c^{-1} I_{d\times d} = \frac{\sqrt{2}\, d\, \Gamma(d/2)}{(d-1)\,\Gamma((d-1)/2)} I_{d\times d}. \tag{27}$$

Thus, A exists if $E\|X - \theta(F)\|^{-1}$ exists. The latter is guaranteed if F has a density that is bounded in any bounded region in $\mathbb{R}^d$. The matrix A is invertible if further $P(H_{\theta(F)}) = 0$.

To the best of our knowledge, the above general results are new. A bivariate influence function of the weighted L^2-median in the special case: $d = 2$, $w(x) = x$, and $X \stackrel{d}{=} -X$ was given in Niinimaa and Oja (1995). The authors presumed the existence of the matrix

$$A = \begin{pmatrix} E(X_2^2/\|X\|^3) & -E(X_1X_2/\|X\|^3) \\ -E(X_1X_2/\|X\|^3) & E(X_1^2/\|X\|^3) \end{pmatrix} \tag{28}$$

(and its inverse) and their influence function is not defined at the special point $x = \theta(F) = 0$.

By the corollary, if $X \sim N_d(\mathbf{0}, I_{d\times d})$ and $w(x) = ax + b$, $a > 0$, $b \ge 0$, then the gross error sensitivity of the weighted L^2-median is

$$GES(\theta(F); F) = \frac{d\sqrt{2}\,\Gamma(d/2)}{(d-1)\,\Gamma((d-1)/2)}$$

which is finite for any $d \ge 2$ and increases in a rate of $\sqrt{d}$ as $d \to \infty$ (indeed $GES(\theta(F); F)/\sqrt{d} \to 1$ as $d \to \infty$). When $d = 2$, the GES of these bivariate weighted median (including the L_1-median) is in-between those of the halfspace median (HM) (see Chen and Tyler, 2002) and the projection median (PM) (see Zuo, Cui and Young, 2004):

$$GRE(HM; F) = \frac{\sqrt{2\pi}}{2} < GRE(\theta(F); F) = \frac{2\sqrt{2}}{\sqrt{\pi}} < GRE(PM; F) = \sqrt{2\pi}.$$

Under some regularity conditions, we have the asymptotic representation of the weighted L^2-median (the proof is slightly involved and not the focus of this paper)

$$\theta(F_n) - \theta(F) = \frac{1}{n}\sum_{i=1}^{n} IF(X_i; \theta(F), F) + o_p(\frac{1}{\sqrt{n}}). \tag{29}$$

Hence, $\sqrt{n}\,(\theta(F_n)$ has the asymptotic mean $\theta(F)$ and the asymptotic covariance matrix

$$E(IF(X; \theta(F), F)IF(X; \theta(F), F)').$$

When $X \sim N_d(\mathbf{0}, I_{d\times d})$ and $w(x) = ax + b$, $a > 0$, $b \geq 0$, the covariance matrix becomes

$$\mathrm{cov}\,(\sqrt{n}\,\theta(F_n)) = \frac{1}{d}\Big(\frac{d\,\sqrt{2}\,\Gamma(d/2)}{(d-1)\,\Gamma((d-1)/2)}\Big)^2 I_{d\times d} = k_d I_{d\times d}. \tag{30}$$

Thus, the asymptotic relative efficiency (ARE) of these weighted L^2-median with respect to the mean is $1/k_d$ which increases to 1 as $d \to \infty$. Indeed, we have the following ARE results:

$d = 2$	$d = 3$	$d = 4$	$d = 5$
$\frac{\pi}{4} \approx 0.79$	$\frac{8}{3\pi} \approx 0.85$	$\frac{9\pi}{32} \approx 0.88$	$\frac{128}{45\pi} \approx 0.91$
$d = 6$	$d = 7$	$d = 8$	
$\frac{75\pi}{256} \approx 0.92$	$\frac{4608}{1575\pi} \approx 0.93$	$\frac{11025\pi}{36864} \approx 0.94$	

Table 1. The asymptotic relative efficiency of the weighted L^2-median relative to the mean.

We note that the asymptotic distribution and the efficiency of the L_1- (or spatial) median, as a special case of the weighted L^2-median here, have been discussed by many authors (see for example Huber, 1981; Brown, 1983; Pollard, 1984; and Chaudhuri, 1992). Indeed, we realized that among others Brown (1983) also obtained the k_d and provided an efficiency table of the L_1-median relative to the mean for d from 2 to 7.

To end this subsection we remark that the weighted L^2-median possesses a bounded influence function for suitable weight and distribution functions. At standard normal model, it also has a high asymptotic efficiency (relative to the mean) which approaches 100% rapidly as d increases.

3.2. Finite sample breakdown point. The weighted L^2-median is locally robust in the sense it possesses a bounded influence function for properly chosen weight functions. In this section we investigate the global robustness of the weighted L^2-median via its finite sample breakdown point.

The finite sample breakdown point of the L_1-median (the special case of the weighted L^2-median with $w(x) = x$) has been studied in Lopuhaä and Rousseeuw (1991). They proved that the L_1-median has a breakdown point $\lfloor (n+1)/2 \rfloor / n$, the highest possible value for any translation equivariant location estimator. We end this section with the following slightly more general result.

Proposition 6 *Let w be non-decreasing, $w(x) < \infty$ for any $x \in [0, \infty)$ and $w(\infty-) = \infty$, and $w(\|a+b\|) \leq w(\|a\|) + w(\|b\|)$ for any $a, b \in \mathbb{R}^d$. Then the sample weighted L^2-median has the best possible breakdown point $\lfloor (n+1)/2 \rfloor / n$ of any translation equivariant location estimator.*

4. Concluding remarks

For studying the global robustness of weighted L^p-depth, a notion of finite sample breakdown point for general depth functions is introduced. The weighted L^p-depth turns out to have a low breakdown point, just as some popular depth functions. The influence function of the weighted L^p-depth is also unbounded. The weighted L^p-depth thus is not very robust.

On the other hand, the weighted L^p-depth induced medians possess the best possible global robustness for suitably chosen weight functions. These weighted medians can be locally robust as well in the sense that they have bounded influence functions for appropriate weight functions. Robustness and efficiency of location estimators are uncooperative in general. But the weighted L^2-medians somehow can keep a very good balance between the two. Indeed, the asymptotic relative efficiency of the medians tends to 100% rapidly as d increases.

A remarkable advantage of the weighted L^P-depth and L^p-medians is the ease in computation in high dimensions. The price for gaining this big advantage in computation is the lack of affine invariance and equivariance, respectively, although the L^p-depth is translation (and even orthogonal) invariant and the weighted L^p-medians are translation and scale and can be orthogonal equivariant.

Affine invariance (or equivariance) is certainly a desirable and ideal property for depth functions (or location estimators)(see for related discussions Liu, 1990; and Zuo and Serfling, 2000a, c). This is especially true when the underlying variables are measurements of the same quantity and are on the same scale and when linear combinations are actually employed in practice. However, there is a trade-off between affine invariant depth (or affine equivariant location estimators) and computability in high dimensions. High dimensional affine invariant depth or affine equivariant location estimators are typically computationally challenging. Fortunately, in many (not all) practical applications, the coordinates have specific means and represent measurements of very different types of variables (such as blood pressure, education level, and marriage status) and the linear combination of the underlying variables may not be very meaningful. The equivariance property of the L^p- (especially L^2-) medians seems adequate in those practical applications, if one is willing to sacrifice the ideal property for the accuracy and the ease in computation.

That said, it is still possible to have the ideal and the desirable affine invariance (or equivariance) property for the weighted L^p depth (or medians) if one is willing to pay a higher price in the computing. For example, one can modify the weighted L^2-depth and medians by replacing $\|x\|$ in their definitions with $\|x\|_\Sigma = \sqrt{x'\Sigma^{-1}x}$ for $x \in \mathbb{R}^d$, where Σ is a covariance matrix of the underlying distribution F (see for example Rao, 1988; and Zuo and Serfling, 2000a). Also one can employ the transformation-retransformation technique of Chakraborty, Chaudhuri, and Oja (1998) to achieve the ideal property. The latter technique is utilized in Hettmansperger and Randles (2002).

Appendix

Proof (of Proposition 4) The function $\phi(x; F)$ does achieve its minimum at the weighted L^p-median if w is continuous and $P\{x : w(\|x\|_p) < w(\infty-)\} > 0$. Assume that there is a sequence $\{x_n\}$ in $\mathbb{R}^d$ such that $\liminf_{n\to\infty} \phi^p(x_n; F) = \inf_{x\in\mathbb{R}^d} \phi^p(x; F)$ and assume that there is a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that $x_{n_k} \to x_0$ (in the L^p norm sense), then $\|x_0\|_p < \infty$. Since otherwise, by Fatou's lemma,

$$\begin{aligned}
&\inf_{x\in\mathbb{R}^d} \phi^p(x; F)\\
&= \liminf_{k\to\infty} \int w(\|t - x_{n_k}\|_p)dF(t) - \int w(\|t\|_p)dF(t)\\
&\geq \int w(\infty-)dF(t) - \int w(\|t\|_p)dF(t)\\
&> \int_{w(\|t\|_p)<w(\infty-)} w(\|t\|_p)dF(t) + \int_{w(\|t\|_p)=w(\infty-)} w(\|t\|_p)dF(t)\\
&\quad - \int w(\|t\|_p)dF(t)\\
&= \phi^p(0; F),
\end{aligned}$$

which is a contradiction. This completes the proof of part (1).

Part (2) can be proved in a straightforward fashion in virtue of the triangle inequality of the L^p norm and the convexity of w. Part (3) follows from the fact that for any $\theta^* \in \mathbb{R}^d$

$$Ew(\|X - \theta\|_p) \leq Ew(\|X - \theta^*\|_p/2 + \|X + \theta^* - 2\theta\|_p/2) \leq Ew(\|X - \theta^*\|_p).$$

Parts (4) and (5) are straightforward.

Proof (of Proposition 5) For simplicity, write $\theta(F_\epsilon)$ for the weighted L^2-median under $F(\epsilon, \delta_x)$. When $x = \theta(F)$, then $\theta(F_\epsilon) = \theta(F)$. This is because for any $y \in \mathbb{R}^d$

$$\begin{aligned}
&\phi\ (y; F(\epsilon, \delta_{\theta(F)})) \\
&= (1-\epsilon)\int (w(\|t-y\|) - w(\|t\|))dF(t) + \epsilon(w(\|\theta(F)-y\|) - w(\|\theta(F)\|) \\
&\geq (1-\epsilon)\phi(\theta(F); F) + \epsilon(w(\|\theta(F)-\theta(F)\|) - w(\|\theta(F)\|)) \\
&= \phi(\theta(F); F(\epsilon, \delta_{\theta(F)})),
\end{aligned}$$

for any non-decreasing w on $[0, \infty)$. When $x \neq \theta(F)$, $\theta(F_\epsilon) \neq x$ for sufficiently small $\epsilon > 0$. To prove this assertion, we first note that by uniqueness and continuity, there is a $\theta^* \in \mathbb{R}^d$ close enough to $\theta(F)$ such that $\phi(x; F) > \phi(\theta^*; F)$. Consequently, there exists some $\epsilon > 0$ small enough such that $\epsilon w(\|x-\theta^*\|) < (1-\epsilon)(\phi(x; F) - \phi(\theta^*; F))$. Thus

$$\begin{aligned}
&(1-\epsilon)\phi(\theta^*; F) + \epsilon(w(\|x-\theta^*\|) - w(\|x\|)) \\
&< (1-\epsilon)\phi(x; F) + \epsilon(w(\|x-x\|) - w(\|x\|)).
\end{aligned}$$

That is, $\phi(\theta^*; F(\epsilon, \delta_x)) < \phi(x; F(\epsilon, \delta_x))$. Hence $x \neq \theta(F_\epsilon)$.

Now we are in a position to derive the influence function of $\theta(F)$ at the point x $(\neq \theta(F))$. Write θ_ϵ for $\theta(F_\epsilon)$. Then

$$(1-\epsilon)\int w^{(1)}(\|t-\theta_\epsilon\|)\frac{-(t-\theta_\epsilon)}{\|t-\theta_\epsilon\|}dF(t) + \epsilon w^{(1)}(\|x-\theta_\epsilon\|)\frac{-(x-\theta_\epsilon)}{\|x-\theta_\epsilon\|} = 0.$$

Taking derivative with respect to ϵ in both sides and letting $\epsilon \to 0$, we have

$$\begin{aligned}
0 = &\int \Big(w^{(1)}(\|t-\theta(F)\|)\frac{d}{d\epsilon}\Big(\frac{-(t-\theta_\epsilon)}{\|t-\theta_\epsilon\|}\Big)\Big|_{\epsilon=0} \\
&+ w^{(2)}(\|t-\theta(F)\|)\frac{-(t-\theta_\epsilon)}{\|t-\theta_\epsilon\|}\frac{d\|t-\theta_\epsilon\|}{d\theta_\epsilon}\Big|_{\epsilon=0}\Big) dF(t)\ \frac{d\theta_\epsilon}{d\epsilon}\Big|_{\epsilon=0} \\
&+ w^{(1)}(\|x-\theta(F)\|)\frac{-(x-\theta(F))}{\|x-\theta(F)\|}.
\end{aligned}$$

Thus,

$$A\ IF(x; \theta(F), F) = w^{(1)}(\|x-\theta(F)\|)\frac{x-\theta(F)}{\|x-\theta(F)\|}$$

with

$$\begin{aligned}
A = &\int \Big(w^{(1)}(\|t-\theta(F)\|)\Big(\frac{I_{d\times d}}{\|t-\theta(F)\|} - \frac{t-\theta(F)}{\|t-\theta(F)\|^2}\frac{(t-\theta(F))'}{\|t-\theta(F)\|}\Big) \\
&+ w^{(2)}(\|t-\theta(F)\|)\frac{t-\theta(F)}{\|t-\theta(F)\|}\frac{(t-\theta(F))'}{\|t-\theta(F)\|}\Big) dF(t). \qquad (31)
\end{aligned}$$

Hence,

$$IF(x;\theta(F),F) = \begin{cases} A^{-1}w^{(1)}(\|x-\theta(F)\|)\frac{x-\theta(F)}{\|x-\theta(F)\|}, & x \neq \theta(F) \text{ and } A^{-1} \text{ exists}, \\ 0, & x = \theta(F). \end{cases} \tag{32}$$

When A^{-1} exists, it can be seen that the gross error sensitivity of $\theta(F)$ is

$$GES(\theta(F);F) \leq \frac{\sup_{x\in\mathbb{R}^d} w^{(1)}(\|x-\theta(F)\|)I(x \neq \theta(F))}{\min_{1\leq i\leq d}|\lambda_i(A)|}, \tag{33}$$

where $\lambda_i(A)$ are the eigenvalues of A $(1 \leq i \leq d)$.

Now we focus on the matrix A and its inverse. When $X \stackrel{d}{=} -X$, then $\theta(F) = 0$ by Fisher consistency. If we further have $e_i'X \stackrel{d}{=} -e_i'X$, $1 \leq i \leq d$, then the matrix A can be simplified to

$$\begin{aligned} A &= E\Big[\frac{w^{(1)}(\|X\|)}{\|X\|}\Big(I_{d\times d} - \frac{X}{\|X\|}\frac{X'}{\|X\|}\Big) + w^{(2)}(\|X\|)\frac{X}{\|X\|}\frac{X'}{\|X\|}\Big] \\ &= \mathrm{diag}\Bigg(E\Big[\frac{w^{(1)}(\|X\|)}{\|X\|}\Big(1-\frac{X_i^2}{\|X\|^2}\Big) + w^{(2)}(\|X\|)\frac{X_i^2}{\|X\|^2}\Big]\Bigg)_{1\leq i\leq d}, \end{aligned} \tag{34}$$

with $X' = (X_1, \cdots, X_d)$. Hence, A has an inverse

$$\begin{aligned} A^{-1} = \mathrm{diag}\Big(1\Big/E\Big(&w^{(1)}(\|X\|)/\|X\|\Big(1 - X_i^2/\|X\|^2\Big) \\ &+ w^{(2)}(\|X\|)X_i^2/\|X\|^2\Big)\Big)_{1\leq i\leq d} \end{aligned} \tag{35}$$

if and only if

$$\begin{aligned} &E[w^{(1)}(\|X\|)/\|X\|(1 - X_i^2/\|X\|^2) + w^{(2)}(\|X\|)X_i^2/\|X\|^2] \\ &\text{exists and} \neq 0 \text{ for } 1 \leq i \leq d. \end{aligned} \tag{36}$$

When $X \sim F$ is spherically symmetric about the origin, the matrix A becomes

$$A = \frac{E((d-1)w^{(1)}(\|X\|)/\|X\| + w^{(2)}(\|X\|))}{d} I_{d\times d}, \tag{37}$$

because of the independence of $X/\|X\|$ and $\|X\|$ and the fact that $E(X_i^2/\|X\|^2) = 1/d$. If this X has a density f, then

$$A = c\, I_{d\times d} = \Big(\frac{2\pi^{d/2}}{d\,\Gamma(d/2)}\int_0^\infty [(d-1)w^{(1)}(r)/r + w^{(2)}(r)]\, r^{d-1} f(r^2)\, dr\Big) I_{d\times d}. \tag{38}$$

Proof (of Proposition 6) Let $X^n = \{X_1, \cdots, X_n\}$ be the original data set and $\theta(X^n)$ be the weighted sample median at X^n. Let $X^n_m = \{X^*_1, \cdots, X^*_m,\ X_{m+1}, \cdots, X_n\} = \{Y_1, \cdots, Y_n\}$ be a contaminated data set with at most $m(< \lfloor (n+1)/2 \rfloor)$ original points in X^n being contaminated (it is understood that $X_{m+1}, \cdots, X_n$ may not be the same $n-m$ points of X^n for different X^n_m). Let $\theta(X^n_m)$ be the weighted sample median at X^n_m. By the conditions on w we have

$$w(\|X^*_i - \theta(X^n_m)\|) + w(\|\theta(X^n_m)\|) \geq w(\|X^*_i - \theta(X^n_m)\| + \|\theta(X^n_m)\|) \geq w(\|X^*_i\|)$$

for $1 \leq i \leq m$ and

$$w(\|X_i - \theta(X^n_m)\|) + w(\|X_i\|) \geq w(\|X_i - \theta(X^n_m)\| + \|X_i\|) \geq w(\|\theta(X^n_m)\|)$$

for $m+1 \leq i \leq n$. Combining the last two displays yields

$$\begin{aligned}
&\sum_{i=1}^{n} w(\|Y_i - \theta(X^n_m)\|) \\
&\geq \sum_{i=1}^{n} w(\|Y_i\|) + (n-2m) w(\|\theta(X^n_m)\|) - 2(n-m) \sup_i w(\|X_i\|).
\end{aligned}$$

Hence, $\sup_{X^n_m} w(\|\theta(X^n_m)\|)$ must be bounded $[\leq 2(n-m)\sup_i w(\|X_i\|)/(n-2m)]$ in order for $\theta(X^n_m)$ to be the weighted L^2 sample median. The desired result follows immediately from the breakdown point upper bound given in Lopuhaä and Rousseeuw (1991).

Bibliography

Brown, B. M. (1983). Statistical uses of the spatial median. *Journal of the Royal Statistical Society Series B* **45** 25–30.

Chakraborty, B., Chaudhuri, P., Oja, H. (1998). Operating transformation retransformation on spatial median and angle test. *Statistical Sinica* **8** 767–784.

Chaudhuri, P. (1992). Multivariate location estimation using extension of R-estimates through U-statistics type approach. *Annals of Statistics* **20** 897–916.

Chen, Z., Tyler, D. E. (2002). The influence function and maximum bias of Tukey's median. *Annals of Statistics* **30** 1737–1759.

Cramer, K. (2003). *Multivariate Outliers and Data Depth.* Shaker, Aachen.

Dodge, Y., Rousson, V. (1999). Multivariate L_1 mean. *Metrika* **49** 127–134.

Donoho, D. L. (1982). Breakdown properties of multivariate location estimators. Ph.D. qualifying paper, Department Statistics, Harvard University.

Donoho, D. L., Huber, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann* (J. Bickel, K. A. Doksum, J. L. Hodges, Jr., eds.), 157–184. Wadsworth, Belmont.

Donoho, D. L., Gasko, M. (1992). Breakdown properties of multivariate location parameters and dispersion matrices. *Annals of Statistics* **20** 1803–1827.

HALDANE, J. B. S. (1948). Note on the median of a multivariate distribution. *Biometrika* **35** 414–415.

HAMPEL, F. R., RONCHETTI, E. Z., ROUSSEEUW, P. J. STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

HAYFORD, J. F. (1902). What is the center of an area or the center of a population? *Journal of the American Statistical Association* **8** 47–58.

HETTMANSPERGER, T. P., RANDLES, R. H. (2002). A practical affine equivariant multivariate median. *Biometrika* **89** 851–860.

HUBER, P. J. (1981). *Robust Statistics.* Wiley, New York.

LIU, R. Y. (1990). On a notion of data depth based on random simplices. *Annals of Statistics* **18** 405–414.

LIU, R. Y (1992). Data depth and multivariate rank tests. In *L_1-Statistical Analysis and Related Methods* (Y. Dodge, ed.), 279–294. North-Holland, Amsterdam.

LIU, R. Y., PARELIUS, J. M., SINGH, K. (1999). Multivariate Analysis by data depth: descriptive statistics, graphics and inference (with discussion). *Annals of Statistics* **27** 783–858.

LOPUHAÄ, H. P., ROUSSEEUW, J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *Annals of Statistics* **19** 229–248.

MILASEVIC, P., DUCHARME, G. R. (1987). Uniqueness of the spatial median. *Annals of Statistics* **15** 1332–1333.

MOSLER, K. (2002). *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid Approach.* Springer, New York.

MOSLER, K., CRAMER, K. (2004). Robustness of metric and combinatorial depth contours. *In preparation* .

NIINIMAA, A., OJA, H. (1995). On the influence functions of certain bivariate medians. *Journal of the Royal Statistical Society Series B* **57** 565–574.

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

RAO, C. R. (1988). Methodology based on the L_1-norm in statistical inference. *Sankhyā Series A* **50** 289–313.

ROMANAZZI, M. (2001). Influence function of halfspace depth. *Journal of Multivariate Analysis* **77** 138–161.

SMALL, C. G. (1990). A survey of multidimensional medians. *International Statistical Review* **58** 263–277.

STAHEL, W. A. (1981). Breakdown of covariance estimators. Research Report 31, ETH, Zürich.

TUKEY, J. W. (1975). Mathematics and picturing data. In *Proceedings of the 1974 International Congress Mathematicians, Vol. 2* (R. James, ed.), 523–531. Vancouver.

ZUO, Y. (2001). Some quantitative relationships between two types of finite sample breakdown point. *Statistics and Probability Letters* **51** 369–375.

ZUO, Y. (2003). Projection based depth functions and associated medians. *Annals of Statistics* **31** 1460–1490.

Zuo, Y., Cui, H., Young, D. (2004). Influence function and maximum bias of projection depth based estimators. *Annals of Statistics* **32** forthcoming.

Zuo, Y., Serfling, R. (2000a). General notions of statistical depth function. *Annals of Statistics* **28** 461–482.

Zuo, Y., Serfling, R. (2000b). Structural properties and convergence results for contours of sample statistical depth functions. *Annals of Statistics* **28** 483–499.

Zuo, Y., Serfling, R. (2000c). On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry. *Journal of Statistical Planning and Inference* **84** 55–79.

Yijun Zuo
Department of Statistics
and Probability
Michigan State University
East Lansing, MI 48824
USA
zuo@msu.edu