Sankhyā: The Indian Journal of Statistics
1999, Volume 61, Series A, Pt. 3, pp. 362-380

ON MULTIVARIATE MONOTONIC MEASURES OF LOCATION WITH HIGH BREAKDOWN POINT

By SUJIT K. GHOSH
North Carolina State University, Raleigh
and
DEBAPRIYA SENGUPTA
Indian Statistical Institute, Calcutta

SUMMARY. The purpose of this article is to propose a new scheme for robust multivariate ranking by introducing a not so familiar notion called *monotonicity*. Under this scheme, as in the case of classical outward ranking, we get an increasing sequence of regions diverging away from a central region (may be a single point) as nucleus. The nuclear region may be defined as the *median region*. Monotonicity seems to be a natural property which is not easily obtainable. Several standard statistics such weighted mean, coordinatewise median and the L_1-median have been studied. We also present the geometry of constructing general monotonic measures of location in arbitrary dimensions and indicate its trade-off with other desirable properties. The article concludes with discussions on finite sample breakdown points and related issues.

1. **Introduction**

Robust handling of multivariate data typically refers to the following: (a) finding a robust measure of location, (b) finding a robust measure of dispersion matrix and (c) detection of possible outliers. The central purpose however, is to create an increasing sequence of regions (depicting increasing degree of outwardness) depending on the geometry of the data cloud. As a consequence we get a *center outward ranking* of a multivariate data (see, Liu (1990)). The method of construction through ellipsoidal regions (required by (a) and (b)) becomes therefore, one of the many similar techniques. There is a great deal of literature on finding out descriptive multivariate location measures with high finite sample breakdown point. These measures are loosely classified

Paper received. October 1997; revised December 1998.
AMS (1991) *subject classification.* Primary 62F35, secondary 62G05, 62H12
Key words and phrases. Multivariate location estimates, high Breakdown point, robustness, monotonicity.

according to their equivariance properties. Suppose $\underset{\sim 1}{x}, ..., \underset{\sim n}{x} \in I\!R^d$ denote a set of observations.

A statistic $\boldsymbol{T}(\underset{\sim 1}{x}, ..., \underset{\sim n}{x})$ is *translation equivariant* if $\boldsymbol{T}(\underset{\sim 1}{x} + \underset{\sim}{b}, ..., \underset{\sim n}{x} + \underset{\sim}{b}) = \boldsymbol{T}(\underset{\sim 1}{x}, ..., \underset{\sim n}{x}) + \underset{\sim}{b}$ for all $\underset{\sim}{b} \in I\!R^d$. There are two other groups of transformations which play pivotal roles in this context. These are the groups of orthogonal and non-singular transformations respectively. If a statistic is translation equivariant and equivariant under orthogonal transformations (non-singular transformations),then the statistic is *orthogonally* (*affine*) equivariant.

For *finite sample breakdown point* we use the definition introduced by Donoho and Huber (1983), i.e.,

$$BD(\boldsymbol{T}, \boldsymbol{X}) = \inf_m \{\frac{m}{n} : \sup_{\boldsymbol{Y}_m} \|\boldsymbol{T}(\boldsymbol{Y}_m) - \boldsymbol{T}(\boldsymbol{X})\| = \infty\} \quad \ldots (1.1)$$

where, $\boldsymbol{X} = \{\underset{\sim 1}{x}, ..., \underset{\sim n}{x}\}$ and $\boldsymbol{Y}_m$ is another set of n points satisfying $|\boldsymbol{Y}_m \cap \boldsymbol{X}| = n - m$.

Among orthogonally equivariant measures the most studied one is the L_1-median (see, Small (1990)). This statistic is a natural extension of sample median in the univariate case and has a breakdown point about $\frac{1}{2}$. There is a host of other procedures which are affine equivariant. Among them the minimum volume ellipsoid (MVE) statistic introduced by Rousseeuw (1985), efficient multivariate M-estimators by Lopuhaä (1992) are worth mentioning. Lopuhaä (1990) gives a detailed study of the problem of finding robust covariance matrices. These procedures are classical in the sense that they lead to ellipsoidal outward ranking. A general technique introduced by Tukey (1975), called 'data depth' works quite well for the problem of constructing affine equivariant multivariate median (with an associated centre outward ranking). Liu (1990) introduced a notion called 'simplicial depth' and related it to Oja's simplicial median (Oja (1983)). Small (1990) did a thorough review of the literature on medians in higher dimensions. As far as the computation of finite sample breakdown point of various measures of location is concerned we refer to a couple of excellent papers in this direction, namely, Lopuhaä and Rousseeuw (1991) and Donoho and Gasko (1992).

The purpose of this article is to introduce a new scheme for robust multivariate ranking by making use of a not so familiar notion called *monotonicity.* Under this scheme, as in the case of classical outward ranking, we get an increasing sequence of regions diverging away from a central region (may be a single point) as nucleus. The nuclear region may be defined as the *median region.* According to Bassett (1991), the univariate sample median is the only monotonic, affine equivariant statistic with breakdown point $\frac{1}{2}$. Such a characterization of sample median is indeed interesting. We look into the problem of extending the above fact to higher dimensions. The monotonicity property is a natural requirement in many applications (for example, in case of income/expenditure economic data). It is also worth mentioning that there are measures of location (for example, the 'shorth') which are sometimes used in practice and which show anti-monotonicity property. In higher dimensions the problem becomes more involved as there is no straightforward extension of univariate monotonicity.

In section 2 we define some notions of multivariate monotonicity via contractions. Several monotonicity properties are discussed. We study these properties with respect to some standard measures of location such as the coordinatewise median and the sample mean in section 3. In the next section the emphasis is given to the problem of constructing monotonic measures of location with specified equivariance properties. In section 5 we discuss the breakdown (or, descriptive robustness) properties of these measures together with other issues and concluding remarks.

2. Monotonicity in General Euclidean Spaces

A vector valued function $g : \mathbb{R}^d \to \mathbb{R}^d$ is a *contraction* towards $\underset{\sim}{\mu} \in \mathbb{R}^d$ if it satisfies $\|g(\underset{\sim}{x}) - \underset{\sim}{\mu}\| \leq \|\underset{\sim}{x} - \underset{\sim}{\mu}\|$ for every $\underset{\sim}{x} \in \mathbb{R}^d$. Any geometric notion of monotonicity is intrinsically related to the concept of contraction towards a point. In other words, given a set of points $\underset{\sim 1}{x}, ..., \underset{\sim n}{x} \in \mathbb{R}^d$ if we contract them towards a fixed point $\underset{\sim}{\mu}$ any monotonic measure of the center of this configuration of points should also move towards $\underset{\sim}{\mu}$. This is the key idea in this article as far as monotonicity is concerned. Because the class of contractions towards a point is quite large it is unlikely that the center of a data cloud would move towards the point of contraction for any kind of distortion of the original configuration. To avoid this problem, we restrict to linear convex combinations, i.e., $g(\underset{\sim}{x}) = \alpha x + (1-\alpha)\underset{\sim}{\mu}$ for some $0 \leq \alpha \leq 1$ and $\underset{\sim}{\mu} \in \mathbb{R}^d$. We shall denote this class by $\mathcal{C}(\underset{\sim}{\mu})$.

Definition 2.1 *A statistic $\boldsymbol{T}$ is monotonic at $\underset{\sim}{\mu} \in \mathbb{R}^d$ if for every $g_1, ..., g_n \in \mathcal{C}(\underset{\sim}{\mu})$*

$$\|\boldsymbol{T}(g_1(\underset{\sim 1}{x}), ..., g_n(\underset{\sim n}{x})) - \underset{\sim}{\mu}\| \leq \|\boldsymbol{T}(\underset{\sim 1}{x}, ..., \underset{\sim n}{x}) - \underset{\sim}{\mu})\| \qquad \ldots (2.1)$$

for every configuration $\boldsymbol{X} = \{\underset{\sim 1}{x}, ..., \underset{\sim n}{x}\}$.

Fact 2.1. If $\boldsymbol{T}$ is translation equivariant and monotonic at some $\underset{\sim 0}{\mu} \in \mathbb{R}^d$ then $\boldsymbol{T}$ is monotonic at every $\underset{\sim}{\mu} \in \mathbb{R}^d$.

Proof. Let $\underset{\sim}{\mu} \in I\!R^d$ and $g_i(\underset{\sim}{x}) = \alpha_i \underset{\sim}{x} + (1-\alpha_i)\underset{\sim}{\mu}, 1 \le i \le n$. Then,

$$\begin{aligned} & \|\boldsymbol{T}(g_1(\underset{\sim 1}{x}),...,g_n(\underset{\sim n}{x})) - \underset{\sim}{\mu}\| \\ = & \|\boldsymbol{T}(\underset{\sim}{\mu} + \alpha_1(\underset{\sim 1}{x} - \underset{\sim}{\mu}),...,\underset{\sim}{\mu} + \alpha_n(\underset{\sim n}{x} - \underset{\sim}{\mu})) - \underset{\sim}{\mu}\| \\ = & \|\boldsymbol{T}(\alpha_1(\underset{\sim 1}{x} - \underset{\sim}{\mu}),...,\alpha_n(\underset{\sim n}{x} - \underset{\sim}{\mu}))\| \text{ (by translation equivariance)} \\ = & \|\boldsymbol{T}(\alpha_1 \underset{\sim 1}{y} + (1-\alpha_1)\underset{\sim 0}{\mu},...,\alpha_n \underset{\sim n}{y} + (1-\alpha_n)\underset{\sim 0}{\mu}) - \underset{\sim 0}{\mu}\|, \end{aligned}$$

where $\underset{\sim i}{y} = \underset{\sim i}{x} - (\underset{\sim}{\mu} - \underset{\sim 0}{\mu}), 1 \le i \le n$. The above step again requires translation equivariance. Now using monotonicity of $\boldsymbol{T}$ at $\underset{\sim 0}{\mu}$ we have

$$\begin{aligned} \|\boldsymbol{T}(g_1(\underset{\sim 1}{x}),...,g_n(\underset{\sim n}{x})) - \underset{\sim}{\mu}\| & \le \|\boldsymbol{T}(\underset{\sim 1}{y},...,\underset{\sim n}{y}) - \underset{\sim 0}{\mu}\| \\ & = \|\boldsymbol{T}(\underset{\sim 1}{x},...,\underset{\sim n}{x}) - \underset{\sim}{\mu}\|. \end{aligned}$$

In view of Fact 2.1 we can say that a translation equivariant statistic is simply *monotonic* if it is monotonic at $\underset{\sim}{\mu} = \underset{\sim}{0}$, which is equivalent to saying that $\|\boldsymbol{T}(\alpha_1\underset{\sim 1}{x},...,\alpha_n\underset{\sim n}{x})\| \le \|\boldsymbol{T}(\underset{\sim 1}{x},...,\underset{\sim n}{x})\|$ for every $\underset{\sim 1}{x},...,\underset{\sim n}{x} \in I\!R^d$ and $0 \le \alpha_1,...,\alpha_n \le 1$. Also note that while actually verifying monotonicity of a translation equivariant statistic it is necessary and sufficient to verify it for a single coordinate.

Fact 2.2. A translation equivariant statistic $\boldsymbol{T}$ is monotonic if and only if

$$\|\boldsymbol{T}(\alpha\underset{\sim 1}{x},\underset{\sim 2}{x},...,\underset{\sim n}{x})\| \le \|\boldsymbol{T}(\underset{\sim 1}{x},\underset{\sim 2}{x},...,\underset{\sim n}{x})\| \qquad \ldots(2.2)$$

for any $\alpha \in [0,1]$ *and* $\underset{\sim 1}{x},..,\underset{\sim n}{x} \in I\!R^d$.

Next notice that one implication of monotonicity of a translation equivariant statistic is the following. For $g_1,...,g_n \in \mathcal{C}(\underset{\sim}{\mu})$

$$\langle\, \boldsymbol{T}(g_1(\underset{\sim 1}{x}),...,g_n(\underset{\sim n}{x})) - \boldsymbol{T}(\underset{\sim 1}{x},...,\underset{\sim n}{x})\,,\, \boldsymbol{T}(\underset{\sim 1}{x},...,\underset{\sim n}{x}) - \underset{\sim}{\mu}\,\rangle \le 0, \qquad \ldots(2.3)$$

where $\langle\cdot,\cdot\rangle$ denotes the standard inner product on $I\!R$. Moreover, using translation equivariance, we can choose $\underset{\sim}{\mu} = \underset{\sim}{0}$ without loss of generality. In many cases of interest it is easier to verify (2.3) rather than (2.1) or (2.2). This property has nice geometric appeal on its own.

Definition 2.2. *A translation equivariant statistic* $\boldsymbol{T}$ *is weakly monotonic if*

$$\langle\, \boldsymbol{T}(\alpha_1\underset{\sim 1}{x},...,\alpha_n\underset{\sim n}{x}) - \boldsymbol{T}(\underset{\sim 1}{x},...,\underset{\sim n}{x})\,,\, \boldsymbol{T}(\underset{\sim 1}{x},...,\underset{\sim n}{x})\,\rangle \le 0 \qquad \ldots(2.4)$$

for $0 \le \alpha_1,...,\alpha_n \le 1$ *and* $\underset{\sim 1}{x},...,\underset{\sim n}{x} \in I\!R^d$. The following is an easy consequence of the above discussions.

Fact 2.3. A monotonic translation equivariant statistic is also weakly monotonic.

The notions of monotonicity introduced so far are quite natural and intuitively plausible. Next we consider another notion of monotonicity which reduces to the usual coordinatewise definition when $d = 1$. To fix the idea let us consider a real valued function $h(x_1, ..., x_n)$, $x_1, ..., x_n \in \mathbb{R}$ which is symmetric in its arguments. The function h is said to be coordinatewise monotonic if $h(x_1 + u, x_2, ..., x_n) - h(x_1, ..., x_n) \geq 0$ (≤ 0) whenever $u \geq 0$ (≤ 0). If we think in terms of the configuration of the set of points $\{x_1, ..., x_n\}$, the geometric interpretation of shifting x_1 by u amounts to saying that we are contracting the configuration towards (sign $(u))\infty$. For the real line the points at infinity are characterized by $\{-1, 1\}$. Analogously in $\mathbb{R}^d$ the points at ∞ are characterized by various unit directions, i.e., by the points on the unit sphere, $S^{(d-1)}$ to be more precise. Next fix some unit direction $\underset{\sim}{\mu} \in \mathbb{R}^d$ and denote the point at infinity in that direction by $\infty(\underset{\sim}{\mu})$. Also let $\underset{\sim}{H}(t, \underset{\sim}{\mu}), t \in \mathbb{R}$ denote the family of hyperplanes orthogonal to $\underset{\sim}{\mu}$ and shifted to the point $t\underset{\sim}{\mu}$. The half spaces formed by this family in the direction of $\underset{\sim}{\mu}$ can be interpreted as the family of concentric spheres centered at $\infty(\underset{\sim}{\mu})$. Therefore a natural notion of monotonicity at ∞ can be defined as follows.

DEFINITION 2.3. *A translation equivariant statistic* $\boldsymbol{T}$ *is directional monotonic if for any* $\underset{\sim}{\mu} \in S^{(d-1)}$ *and* $\alpha_1, ..., \alpha_n \geq 0$,

$$\langle \underset{\sim}{\mu}, \boldsymbol{T}(\underset{\sim 1}{x} + \alpha_1 \underset{\sim}{\mu}, ..., \underset{\sim n}{x} + \alpha_n \underset{\sim}{\mu}) - \boldsymbol{T}(\underset{\sim 1}{x}, ..., \underset{\sim n}{x}) \rangle \geq 0 \qquad \ldots (2.5)$$

for every $\underset{\sim 1}{x}, ..., \underset{\sim n}{x} \in \mathbb{R}^d$.

REMARK 2.1. Note that because $\boldsymbol{T}$ is symmetric in its arguments it is enough to take $\alpha_1 \geq 0, \alpha_2 = ... = \alpha_n = 0$. The concept of directional monotonicity reduces to usual monotonicity in each coordinate when $d = 1$. Although the definition 2.3 is a direct extension of the univariate monotonicity, definitions 2.1 and 2.2 are also equally appealing in higher dimensions.

REMARK 2.2 There is another popular notion of multivariate ordering, namely the coordinatewise ordering. This concept can be used to define monotonicity. The major drawback of this ordering is that it is only a partial order. Secondly, it is not quite compatible with orthogonal and affine group operations where the coordinates get mixed up after transformation. The coordinatewise ordering of the transformed data does not seem to carry any meaning. Finally, the concept of directional monotonicity implies this sort of monotonicity. To see this, apply (2.5) with $\underset{\sim}{\mu} = \underset{\sim 1}{e}, \underset{\sim}{\mu} = \underset{\sim 2}{e}, ..., \underset{\sim}{\mu} = \underset{\sim d}{e}$ sequentially, where $\underset{\sim 1}{e}, ..., \underset{\sim d}{e}$ are standard basis vectors. This will imply if $\underset{\sim 1}{x} \preceq \underset{\sim 1}{y}, ..., \underset{\sim n}{x} \preceq \underset{\sim n}{y}$ then $\boldsymbol{T}(\underset{\sim 1}{x}, ..., \underset{\sim n}{x}) \preceq \boldsymbol{T}(\underset{\sim 1}{y}, ..., \underset{\sim n}{y})$. Here $\preceq$ stands for the coordinatewise ordering of multivariate vectors. See, Barnett (1976) for an excellent discussion on various aspects of multivariate ordering.

3. Monotonicity Properties of Some Standard Statistics

In this section we study the monotonicity properties of some standard translation equivariant statistics which are commonly used as measures of central tendency of a data cloud. First we consider the example of weighted mean. Let

$$\boldsymbol{T}_w(\underset{\sim}{x}_1, ..., \underset{\sim}{x}_n) = \sum_{i=1}^{n} w_i \underset{\sim}{x}_i \qquad \ldots (3.1)$$

where $w_1, ..., w_n$ is a set of nonnegative weights with $\sum w_i = 1$.

THEOREM 3.1. *The statistic $\boldsymbol{T}_w$ is directional monotonic but neither monotonic nor weakly monotonic.*

PROOF. Take any $\underset{\sim}{\mu} \in S^{(d-1)}$ and $\alpha_1, ..., \alpha_n \geq 0$. Then

$$\begin{aligned} \langle\, \boldsymbol{T}_w(\underset{\sim}{x}_1 + \alpha_1\underset{\sim}{\mu}, ..., \underset{\sim}{x}_n + \alpha_n\underset{\sim}{\mu}) - \boldsymbol{T}_w(\underset{\sim}{x}_1, ..., \underset{\sim}{x}_n)\,,\, \underset{\sim}{\mu}\,\rangle &= \langle\, (\textstyle\sum \alpha_i w_i)\underset{\sim}{\mu}, \underset{\sim}{\mu}\,\rangle \\ &= \textstyle\sum \alpha_i w_i \geq 0. \end{aligned}$$

To show that $\boldsymbol{T}_w$ is not weakly monotonic take any configuration of points $\underset{\sim}{x}_1, ..., \underset{\sim}{x}_n$ satisfying $\langle \underset{\sim}{x}_1, \boldsymbol{T}_w(\underset{\sim}{x}_1, ..., \underset{\sim}{x}_n)\rangle < 0$. Also assume w.l.o.g. that $w_1 > 0$. Next choose $\alpha_1 = 0, \alpha_2 = \cdots = \alpha_n = 1$. Then

$$\boldsymbol{T}_w(\alpha_1\underset{\sim}{x}_1, ..., \alpha_n\underset{\sim}{x}_n) - \boldsymbol{T}_w(\underset{\sim}{x}_1, ..., \underset{\sim}{x}_n) = -w_1\underset{\sim}{x}_1.$$

Hence we have $\langle \boldsymbol{T}_w(\alpha_1\underset{\sim}{x}_1, ..., \alpha_n\underset{\sim}{x}_n) - \boldsymbol{T}_w(\underset{\sim}{x}_1, ..., \underset{\sim}{x}_n), \boldsymbol{T}_w(\underset{\sim}{x}_1, ..., \underset{\sim}{x}_n)\rangle = -w_1\langle \underset{\sim}{x}_1, \boldsymbol{T}_w(\underset{\sim}{x}_1, ..., \underset{\sim}{x}_n)\rangle > 0$.
This is a contradiction to the weakly monotonic property. Now, by fact 2.3 it is also clear that $\boldsymbol{T}_w$ cannot be monotonic. □

Next we consider the example of sample median for $d = 1$. Let us also assume for the sake of simplicity that n is odd and $x_1, ..., x_n$ are random samples from a continuous distribution so that there is no tie. This will uniquely define the sample median as $\frac{1}{2}(n+1)$th order statistic, namely, $x_{(\frac{n+1}{2})}$.

THEOREM 3.2 *The sample median is both monotonic and directional monotonic.*

PROOF. To prove the theorem we shall first verify the condition (2.2) of the fact 2.2. This will show that the sample median is monotonic. In order to do so we consider the following cases. First assume that $x_{(\frac{n+1}{2})} > 0$.

Case (i) $x_1 < x_{(\frac{n+1}{2})}$. In this situation $\alpha x_1 < x_{(\frac{n+1}{2})}$ for $0 \le \alpha \le 1$. Hence in the new configuration the position of the $\frac{n+1}{2}$th order statistic is not altered. Therefore (2.2) is verified.

Case (ii) $x_1 = x_{(\frac{n+1}{2})}$. Since $x_{(\frac{n+1}{2})} > 0$, in the new configuration $\{\alpha x_1, x_2, ..., x_n\}$ the total number of nonnegative observations remains same. If $0 \le \alpha x_1 \le x_{(\frac{n-1}{2})}$ then the new median is located at $x_{(\frac{n+1}{2})}$. Therefore we have $0 \le$ median $\{\alpha x_1, x_2, ..., x_n\} = x_{(\frac{n-1}{2})} < x_{(\frac{n+1}{2})}$. Thus (2.2) is verified. Otherwise if $x_{(\frac{n-1}{2})} < \alpha x_1 \le x_{(\frac{n+1}{2})}$, the median of the new configuration is $\alpha x_{(\frac{n+1}{2})}$ and (2.2) is again verified.

Case (iii) $x_1 > x_{(\frac{n+1}{2})}$. Using similar arguments we observe that if $\alpha x_1 < x_{(\frac{n+1}{2})}$, the median of the new configuration will move towards 0. It will remain unaltered otherwise. The condition (2.2) is satisfied in either case.

Next consider the case when $x_{(\frac{n+1}{2})} < 0$. The same proof as in the earlier case goes through by symmetry of the configuration with respect to reflection around 0.

The remaining case is when $x_{(\frac{n+1}{2})} = 0$. In this case the number of positive and negative data points remain same in the new configuration $\{\alpha x_1, ..., x_n\}$ regardless of the position of x_1 with respect to the data set.

Hence it follows that the sample median is monotonic. Since we have already observed that the notion of directional monotonicity reduces to usual coordinatewise monotonicity for $d = 1$, the remaining part of the result follows from observations made by Bassett (1991). □

The assumptions that n is odd and there are no ties can be assumed without loss of any generality. Also, the same is true for any quantile (not necessarily median). If we define the qth quantile $T_q(x_1, ..., x_n) = F_n^{-1}(q)$ for $0 < q < 1$ where F_n is the empirical cumulative distribution function then we have the following.

COROLLARY 3.3. *The family of quantiles $T_q, 0 < q < 1$ are both monotonic and directional monotonic.*

The argument used to prove theorem 3.2 has other interesting implications. For example, the same argument when applied coordinatewise works for coordinatewise multivariate median. In general suppose $T_1, ..., T_d$ are d univariate translation equivariant statistics defined on sets of samples of size n. Given n points $\underset{\sim}{x}_1, ..., \underset{\sim}{x}_n \in \mathbb{R}^d$ let x_{ij} denote the jth coordinate of $\underset{\sim}{x}_i, 1 \le i \le n$ and $1 \le j \le d$. Define

$$\boldsymbol{T}_{0n}(\underset{\sim}{x}_1, ..., \underset{\sim}{x}_n) = (T_1(\underset{\sim}{z}_1), ..., T_d(\underset{\sim}{z}_d))', \qquad \dots (3.2)$$

where $\underset{\sim}{z}_j = (x_{1j}, ..., x_{nj}) \quad \text{for} \quad 1 \le j \le d.$

THEOREM 3.4. *If each of $T_1, ..., T_d$ is monotonic (directional monotonic) then $\boldsymbol{T}_{0n}$, defined by (3.2), is also monotonic (directional monotonic).*

We have discussed so far, certain features of monotonicity through different examples. There are other commonly used orthogonal and affine equivariant multivariate medians such as the L_1-median and Tukey's halfspace median. These measures (so called geometric medians, cf., Small (1990)) are highly nonlinear in nature and we are unable to verify their monotonic status directly. Intuitively it seems they should possess some of the monotonicity properties. We look into this issue at length in the next section and obtain partial results for some of these highly nonlinear estimators.

REMARK 3.1 It should be noted that theorem 3.1 and theorem 3.2 combined, have a striking implication. The notion of monotonicity seems to act as a line of demarcation between the 'mean type' and 'median type' measures of location. There has been a long standing debate regarding which 'type' actually serves as a more efficient estimator of location. See Huber (1981) for some useful comments on this issue. Also Chaudhuri and Sengupta (1993a) established certain property of 'median type' measures which gives the sample median a unique status. We also refer to Bassett (1991) in this context which acted as a major motivation for the current investigation.

4. The Geometry of Constructing Monotonic Multivariate Measures of Location

As remarked earlier it is difficult to verify monotonicity properties for general orthogonally or affine equivariant estimators such as L_1-median and Tukey's half space median. We shall verify a weaker form of monotonicity for the L_1-median next.

DEFINITION 4.1. *A translation equivariant estimator $\boldsymbol{T}$ is locally weakly monotonic (directional monotonic) at a set of points $\boldsymbol{X} = \{\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n\} \in I\!R^d$ with respect to an inner product $< \cdot, \cdot >$ if (2.4) holds for almost all $\alpha_1, ..., \alpha_n$ sufficiently close to 1 ((2.5) holds for almost all $\underset{\sim}{\alpha} = (\alpha_1, \cdots, \alpha_n)$ in a neighborhood of $\underset{\sim}{0}$ for all $\underset{\sim}{\mu} \in S^{(d-1)}$).*

Next we study the local monotonicity properties of L_1-median to get a general insight into the geometry of multivariate monotonicity.

Let $h : [0, 1] \times I\!R^d \to I\!R^d$ be a differentiable vector valued function with the property that $h(0, \underset{\sim}{x}) = \underset{\sim}{x}$ for any $\underset{\sim}{x} \in I\!R^d$. The function h can be thought of as a smooth deformation of $I\!R^d$. A given set of points $\boldsymbol{X} = \{\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n\} \in I\!R^d$ is a *regular* for L_1-median if the solution $\hat{\underset{\sim}{\theta}}$ of the equation

$$\sum_1^n \frac{\underset{\sim i}{x} - \underset{\sim}{\theta}}{\|\underset{\sim i}{x} - \underset{\sim}{\theta}\|} = \underset{\sim}{0} \qquad \ldots (4.1)$$

is unique and is not one of the $\underset{\sim i}{x}'$s. If $\{\underset{\sim 1}{x}, \cdots, \underset{\sim n}{x}\}$ is a random sample from a continuous density in $\mathbb{R}^d, d \geq 2$ then it is easy to see that $\{\underset{\sim 1}{x}, \cdots, \underset{\sim n}{x}\}$ is *regular* with probability one. Next let us define a family of transformed data points $\underset{\sim i}{y}(\alpha_i) = h(\alpha_i, \underset{\sim i}{x}), 1 \leq i \leq n$. Also, let $\underset{\sim}{\theta}(\alpha_1, \cdots, \alpha_n)$ denote the L_1-median of the set of points $\{\underset{\sim 1}{y}(\alpha_1), \cdots, \underset{\sim n}{y}(\alpha_n)\}$. Then we have the following facts: (i) $\hat{\underset{\sim}{\theta}}(0, \cdots, 0) = \hat{\underset{\sim}{\theta}}$ (ii) The set of points $\{\underset{\sim 1}{y}(\alpha_1), \cdots, \underset{\sim n}{y}(\alpha_n)\}$ is *regular* for $\underset{\sim}{\alpha} = (\alpha_1, \cdots, \alpha_n)$ belonging to a sufficiently small neighborhood of $(0, \cdots, 0)$. (This is true because the deformation h is continuous and also the left hand side of (4.1) is a continuous function at regular points $\{\underset{\sim 1}{x}, \cdots, \underset{\sim n}{x}\}$. Also note that by smoothness of h and the estimating equation (4.1), $\hat{\underset{\sim}{\theta}}(\alpha_1, \cdots, \alpha_n)$ is differentiable at $(0, \cdots, 0)$ by implicit function theorem (cf. Apostol(1974)). Let us next define

$$\hat{\boldsymbol{U}}_i = \|\underset{\sim i}{x} - \hat{\underset{\sim}{\theta}}\|^{-1}(\underset{\sim i}{x} - \hat{\underset{\sim}{\theta}}),\ 1 \leq i \leq n.$$

LEMMA 4.1 *Suppose $\{\underset{\sim 1}{x}, \cdots, \underset{\sim n}{x}\}$ is a set of regular points for the L_1-median (defined by (4.1)). Then*

$$\Gamma(\boldsymbol{X})\, \frac{\partial \hat{\underset{\sim}{\theta}}}{\partial \alpha_k}(\underset{\sim}{0}) = \frac{1}{\|\underset{\sim k}{x} - \hat{\underset{\sim}{\theta}}\|}\,(I_d - \hat{\boldsymbol{U}}_k \hat{\boldsymbol{U}}_k')\, \frac{\partial h}{\partial \alpha_k}(0, \underset{\sim k}{x}) \qquad \ldots (4.2)$$

for $1 \leq k \leq n$, where

$$\Gamma(\boldsymbol{X}) = \sum_{i=1}^n \frac{1}{\|\underset{\sim i}{x} - \hat{\theta}\|}\,(I_d - \hat{\boldsymbol{U}}_i \hat{\boldsymbol{U}}_i') \qquad \ldots (4.3)$$

and I_d is the $d \times d$ identity matrix. PROOF. We start by differentiating the relation

$$\sum_1^n \frac{\underset{\sim i}{y}(\alpha_i) - \hat{\underset{\sim}{\theta}}(\underset{\sim}{\alpha})}{\|\underset{\sim i}{y}(\alpha_i) - \hat{\underset{\sim}{\theta}}(\underset{\sim}{\alpha})\|} = \underset{\sim}{0}. \qquad \ldots (4.4)$$

While differentiating with respect to α_k, the terms for which $i \neq k$ are to be treated separately from the term $i = k$. Now differentiating by product rule we get

$$\frac{\partial}{\partial \alpha_k}\left[\frac{\underset{\sim k}{y}(\alpha_k) - \hat{\underset{\sim}{\theta}}(\underset{\sim}{\alpha})}{\|\underset{\sim k}{y}(\alpha_k) - \hat{\underset{\sim}{\theta}}(\underset{\sim}{\alpha})\|}\right](\underset{\sim}{0}) = \frac{1}{\|\underset{\sim k}{x} - \hat{\underset{\sim}{\theta}}\|}\left[\frac{\partial \underset{\sim k}{y}}{\partial \alpha_k}(0) - \frac{\partial \hat{\underset{\sim}{\theta}}}{\partial \alpha_k}(\underset{\sim}{0})\right] - \frac{1}{\|\underset{\sim k}{x} - \hat{\underset{\sim}{\theta}}\|}\hat{\boldsymbol{U}}_k\hat{\boldsymbol{U}}_k'\left[\frac{\partial \underset{\sim k}{y}}{\partial \alpha_k}(0) - \frac{\partial \hat{\underset{\sim}{\theta}}}{\partial \alpha_k}(\underset{\sim}{0})\right]. \qquad \ldots (4.5)$$

Note that by definition $\frac{\partial \underset{\sim k}{y}}{\partial \alpha_k}(0) = \frac{\partial h}{\partial \alpha_k}(0, \underset{\sim k}{x})$. Next for $i \neq k$

$$\frac{\partial}{\partial \alpha_k}\left[\frac{\underset{\sim i}{y}(\alpha_i) - \hat{\underset{\sim}{\theta}}(\underset{\sim}{\alpha})}{\|\underset{\sim i}{y}(\alpha_i) - \hat{\underset{\sim}{\theta}}(\underset{\sim}{\alpha})\|}\right](\underset{\sim}{0}) = -\frac{1}{\|\underset{\sim k}{x} - \hat{\underset{\sim}{\theta}}\|}\left[(I_d - \hat{\boldsymbol{U}}_i\hat{\boldsymbol{U}}_i')\frac{\partial \hat{\underset{\sim}{\theta}}}{\partial \alpha_k}(\underset{\sim}{0})\right]. \qquad \ldots (4.6)$$

The lemma follows after combining (4.5) and (4.6). □

The above lemma gives some insight into the geometry of L_1-median. The matrix $n^{-1}\Gamma(\boldsymbol{X})$ is an estimator of the inverse of asymptotic covariance matrix of $\hat{\underset{\sim}{\theta}}$ when samples are generated from a spherically symmetric distribution.

THEOREM 4.2. *Suppose $\underset{\sim 1}{x}, \cdots, \underset{\sim n}{x} \in \mathbb{R}^d, n \geq 3$ are i.i.d. samples from a continuous population. Then the L_1-median is locally weakly and directional monotonic with probability one with respect to the inner product generated by $\Gamma(\boldsymbol{X})$.*

PROOF. First consider the case of weak monotonicity. Because the samples are drawn from a continuous distribution the data will be regular with probability one. Also note that the matrix Γ is positive definite whenever there will be at least two distinct $\hat{U}_i$'s. This event occurs with probability one too. Now without loss of generality we can change α to $(1-\alpha)$ so that we can apply lemma 4.1 as it is. In the case of weak monotonicity we apply the lemma for $h(\alpha, \underset{\sim}{x}) = \underset{\sim}{x} - \alpha\underset{\sim}{x}$. Let $J(\boldsymbol{X})$ denote the Jacobian of $\hat{\underset{\sim}{\theta}}$ at $\{\underset{\sim 1}{x}, \cdots, \underset{\sim n}{x}\}$. Then for a small perturbation $\underset{\sim}{\alpha} = (\alpha_1, \cdots, \alpha_n)'$,

$$\Delta\hat{\underset{\sim}{\theta}}(\underset{\sim}{\alpha}) := \hat{\underset{\sim}{\theta}}(\alpha_1\underset{\sim 1}{x}, \cdots, \alpha_n\underset{\sim n}{x}) - \hat{\underset{\sim}{\theta}}(\underset{\sim 1}{x}, \cdots, \underset{\sim n}{x}) = J(\boldsymbol{X})\underset{\sim}{\alpha} + o(\|\underset{\sim}{\alpha}\|). \qquad \ldots (4.7)$$

Next notice that $\frac{\partial h}{\partial \alpha_k}(0, \underset{\sim k}{x}) = -\underset{\sim k}{x}$. Because $(I_d - \hat{U}_k\hat{U}_k')(\underset{\sim k}{x} - \hat{\underset{\sim}{\theta}}) = \underset{\sim}{0}$, by lemma

4.1 we have

$$\Gamma(\boldsymbol{X})\frac{\partial \hat{\underset{\sim}{\theta}}}{\partial \alpha_k}(\underset{\sim}{0}) = -\frac{1}{\|\underset{\sim}{x}_k - \hat{\underset{\sim}{\theta}}\|}\,(I_d - \hat{U}_k\hat{U}_k')\hat{\underset{\sim}{\theta}}. \qquad \ldots(4.8)$$

Therefore in view of (4.7) and (4.8)

$$\begin{aligned}\Gamma(\boldsymbol{X})\Delta\hat{\underset{\sim}{\theta}}(\underset{\sim}{\alpha}) &= \Gamma(\boldsymbol{X})J(\boldsymbol{X})\underset{\sim}{\alpha} + o(\|\underset{\sim}{\alpha}\|)\\ &= -\left[\sum_1^n \frac{\alpha_k}{\|\underset{\sim}{x}_k - \hat{\underset{\sim}{\theta}}\|}(I_d - \hat{U}_k\hat{U}_k')\right]\hat{\underset{\sim}{\theta}} + o(\|\underset{\sim}{\alpha}\|).\end{aligned}$$

Thus

$$\langle\Delta\hat{\underset{\sim}{\theta}}(\underset{\sim}{\alpha}), \hat{\underset{\sim}{\theta}}\rangle_\Gamma = -\hat{\underset{\sim}{\theta}}'\left[\sum_1^n \frac{\alpha_k}{\|\underset{\sim}{x}_k - \hat{\underset{\sim}{\theta}}\|}(I_d - \hat{U}_k\hat{U}_k')\right]\hat{\underset{\sim}{\theta}} + o(\|\underset{\sim}{\alpha}\|).$$

The matrix on the right hand is positive definite as long as more than two α_k's are strictly positive. The collection of directions having less than or equal to two nonzero coordinates have zero measure and hence the weak monotonicity follows, once we note that the property is trivially true when $\hat{\theta} = \underset{\sim}{0}$.

The local directional monotonicity follows exactly the same way. The only difference is that now for given $\underset{\sim}{\mu} \in S^{(d-1)}$ we have to choose $h(\alpha, \underset{\sim}{x}) = \underset{\sim}{x} + \alpha\underset{\sim}{\mu}$. Hence the theorem. □

Remark 4.1. Theorem 4.2 points out where the actual difficulty lies in handling highly nonlinear measures like the L_1-median. The main trouble here is that the inner product under which the monotonicity property is to be studied depends on the local geometry of the configuration of points. It is interesting to see that the driving inner product matrix is the 'observed' precision matrix of the L_1-median if the population is spherically symmetric. While studying the geometry of the maximum likelihood estimators the Fisher information matrix (which is the asymptotic precision matrix of the m.l.e) becomes a natural inner product matrix. We feel that the connection between these two apparently unrelated ideas should be studied further. We refer to Efron (1978), Barndorff-Nielsen (1978) in this regard.

The analysis of L_1-median gave us useful insight into the relationship between monotonicity in local sense and the 'Δ-method geometry' (as we are tempted to call it) of such nonlinear measures. Apart from this one can comprehend the issue of monotonicity from a purely geometric point of view.

The problem of constructing affine equivariant statistics with monotonicity or even orthogonally equivariant statistics with directional monotonicity is a much harder problem. One method of construction can be conceived of by making use of an auxiliary data on the same set of variables $X_1, \cdots, X_d$. Think of a set of *carefully* collected observations $\underset{\sim}{z}_1, \cdots, \underset{\sim}{z}_m$ on a set of attributes (say, incomes of various individuals from various sources) with higher sampling cost. Suppose the current data is collected *less carefully* for the same set of attributes and may have some systematic bias (such as under-reporting or other directional biases). In some applications, one can even split the available data into two parts. The first part would take the role of auxiliary data, $\underset{\sim}{z}_1, \cdots, \underset{\sim}{z}_m$, while the overall measure of location will be monotonic only with respect to the second part. Any reasonable measure of the center of the current data should reflect the pattern of bias relative to the center of the auxiliary data $\underset{\sim}{z}_1, \cdots, \underset{\sim}{z}_m$. By equivariance, if we transform the current data $\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n$, the auxiliary data $\underset{\sim}{z}_1, \cdots, \underset{\sim}{z}_m$ should also be transformed accordingly. Therefore exploting the idea described above we can construct a measure of the center of the current data where the coordinate system is chosen on the basis of the auxiliary data $\underset{\sim}{z}_1, \cdots, \underset{\sim}{z}_m$. This is an appropriate thing to do in this framework because, we are equating the requirement for monotonicity with the detection of any severe systematic bias in the current data; which is assumed to be absent in the auxiliary data.

First let us consider orthogonally equivariant monotonic statistics. Let $\boldsymbol{X}$ denote the $d \times n$ data matrix whose columns are $\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n$ respectively. We are interested here in transformations of the from $\boldsymbol{X} \to \boldsymbol{Y} = P\boldsymbol{X}A + \underset{\sim}{b}\,\underset{\sim}{1}_n'$ where P is an orthogonal matrix, $A = \text{diag}(\alpha_1, \cdots, \alpha_n)$ with $0 \le \alpha_1, \cdots, \alpha_n \le 1$, $\underset{\sim}{b} \in I\!R^d$ and $\underset{\sim}{1}_n$ the $n \times 1$ vector with all components equal to 1. The problem of constructing monotonic orthogonally equivariant statistics is in a sense equivalent to producing a data dependent orthonormal reference frame say, $\underset{\sim}{\hat{\eta}}_1, ..., \underset{\sim}{\hat{\eta}}_d$ which is (i) equivariant under orthogonal transformations (P) and (ii) invariant under the joint action of the set transformations produced by $(A, \underset{\sim}{b})$.

By studying the latter set of transformations we see that the invariant functions under $(A, \underset{\sim}{b})$ are not orthogonally equivariant. Notice that if we do away with the requirement of monotonicity with respect to the full data set (*i.e*, with respect to both $\boldsymbol{X}$ and the auxiliary data $\underset{\sim}{z}_1, \cdots, \underset{\sim}{z}_m$), we can make use of the eigenvectors of sample dispersion matrix such as

$$\boldsymbol{R} = \sum_{i=1}^{m} (\underset{\sim}{z}_i - \underset{\sim}{\hat{\mu}})(\underset{\sim}{z}_i - \underset{\sim}{\hat{\mu}})'$$

to construct the basic reference frame $\underset{\sim}{\hat{\eta}}_1, ..., \underset{\sim}{\hat{\eta}}_d$, where $\underset{\sim}{\hat{\mu}}$ denotes some orthogonally equivariant directional measure of location of the set of points $\underset{\sim}{z}_i, 1 \le i \le n\,\}$. We can take $\underset{\sim}{\hat{\mu}}$ to be the usual mean for example. A host of other techniques for circular and spherical data can be found in Mardia (1972). Let us rewrite

$$\boldsymbol{R} = \sum_1^d \hat{\lambda}_i \hat{\underset{\sim}{\eta}}_i \hat{\underset{\sim}{\eta}}_i' \qquad \ldots (4.9)$$

where $\hat{\underset{\sim}{\eta}}_1, \cdots, \hat{\underset{\sim}{\eta}}_d$ constitute an orthonormal basis for $I\!R^d$. There might be some ambiguity in the choice of $\hat{\underset{\sim}{\eta}}_1, \cdots, \hat{\underset{\sim}{\eta}}_d$. Because they are orthogonally equivariant, once we define them for a fixed point in each orbit of the orthogonal group the choice is unique. Also notice that we can actually choose $\hat{\underset{\sim}{\eta}}_1, \cdots, \hat{\underset{\sim}{\eta}}_d$ as smooth functions of the data. Next let

$$\hat{t}_{in} = t(\hat{\underset{\sim}{\eta}}_i' \underset{\sim}{x}_1, \cdots, \hat{\underset{\sim}{\eta}}_i' \underset{\sim}{x}_n),\ 1 \le i \le d,$$

where t is some univariate affine equivariant statistic. Finally let

$$\boldsymbol{T}_n^* = \sum_1^d \hat{t}_{in} \hat{\underset{\sim}{\eta}}_i. \qquad \ldots (4.10)$$

If t is chosen to be the sample median the corresponding $\boldsymbol{T}_n^*$ in (4.12) may be thought of as a multivariate median. It is clear from the definition that $\boldsymbol{T}_n^*$ defined this way is translation equivariant because the orthonormal system obtained from $\{\underset{\sim}{z}_i, 1 \le i \le n\}$ is translation invariant.

THEOREM 4.3. *Suppose* $\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n$ *are i.i.d. samples from an angularly symmetric density about* $\underset{\sim}{\theta} \in I\!R^d$ *which is strictly positive in a neighborhood of* $\underset{\sim}{\theta}$*, and the univariate statistic* t *is the sample median. Let* $\underset{\sim}{z}_1, \ldots, \underset{\sim}{z}_n$ *be the auxiliary data. Then*

(i) $\boldsymbol{T}_n^*$ *is equivariant under orthogonal transformations of the data and monotonic.*

(ii) $\boldsymbol{T}_n^* \to \underset{\sim}{\theta}$ *almost surely as* $n \to \infty$*, for every choice of auxiliary data.*

PROOF. (i) Suppose we change $\underset{\sim}{x}_1 \to P\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n \to P\underset{\sim}{x}_n$ for some orthogonal matrix P. Therefore the matrix $\boldsymbol{R}$ changes to $P\boldsymbol{R}P'$ and thus $\hat{\underset{\sim}{\eta}}_1, \cdots, \hat{\underset{\sim}{\eta}}_d$ changes to $P\hat{\underset{\sim}{\eta}}_1, \cdots, P\hat{\underset{\sim}{\eta}}_d$ which is a new orthonormal system. On the other hand, $(P\hat{\underset{\sim}{\eta}}_i)'P\underset{\sim}{x}_i = \hat{\underset{\sim}{\eta}}_i' \underset{\sim}{x}_i$ for $1 \le i \le n$. Thus $\hat{t}_{1n}, \cdots, \hat{t}_{dn}$ remain invariant. Hence

$$\boldsymbol{T}_n^*(P\underset{\sim}{x}_1, \cdots, P\underset{\sim}{x}_n) = P\boldsymbol{T}_n^*(\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n),$$

so that $\boldsymbol{T}_n^*$ is equivariant under orthogonal transformations.

Next if we change $\underset{\sim}{x}_1 \to \alpha_1 \underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n \to \alpha_n \underset{\sim}{x}_n$ by construction the reference system $\hat{\underset{\sim}{\eta}}_1, \cdots, \hat{\underset{\sim}{\eta}}_d$ remains invariant. Let $\hat{t}_{1n}(\underset{\sim}{\alpha}), \cdots, \hat{t}_{dn}(\underset{\sim}{\alpha})$ denote the changed coordinates under the transformed data. Because t is a monotonic statistic, by virtue of theorem 3.4,

$$|\hat{t}_{in}(\underset{\sim}{\alpha})| \leq |\hat{t}_{in}| \quad \text{for} \quad 1 \leq i \leq n. \qquad \ldots (4.11)$$

Thus,

$$\begin{aligned} \|\boldsymbol{T}_n^*(\alpha_1 \underset{\sim}{x}_1, \cdots, \alpha_n \underset{\sim}{x}_n)\|^2 &= \sum_1^d \hat{t}_{in}^2(\underset{\sim}{\alpha}) \\ &\leq \sum_1^d \hat{t}_{in}^2 \\ &= \|\boldsymbol{T}_n^*(\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n)\|^2. \end{aligned} \qquad \ldots (4.12)$$

Therefore $\boldsymbol{T}_n^*$ is monotonic at $\underset{\sim}{0}$. Now making use of the fact 2.2, we can establish $\boldsymbol{T}_n^*$ is actually monotonic at each $\underset{\sim}{\mu} \in I\!R^d$.

(ii) First notice that by lemma 18 (p. 20) of Pollard (1984) the sets of the form $\{\underset{\sim}{\eta}' \underset{\sim}{x} \leq a\}$, $\underset{\sim}{\eta} \in S^{(d-1)}$ and $a \in I\!R$, have polynomial discrimination. Therefore, by theorem 14 (p. 18) of Pollard (1984)

$$\sup_{\underset{\sim}{\eta} \in S^{(d-1)}, a \in I\!R} \left| \frac{1}{n} \#\{\underset{\sim}{\eta}' \underset{\sim}{x}_i \leq a\} - P\{\underset{\sim}{\eta}' \underset{\sim}{x}_1 \leq a\} \right| \to 0 \qquad \ldots (4.13)$$

almost surely as $n \to \infty$. By the assumptions made $\eta' \underset{\sim}{\theta}$ is the unique solution of $P\{\underset{\sim}{\eta}' \underset{\sim}{x}_1 \leq a\} = \frac{1}{2}$ for every $\underset{\sim}{\eta} \in S^{(d-1)}$. Hence by (4.16)

$$D_n := \sup_{\underset{\sim}{\eta} \in S^{(d-1)}} \left| t(\underset{\sim}{\eta}' \underset{\sim}{x}_1, \cdots, \underset{\sim}{\eta}' \underset{\sim}{x}_n) - \underset{\sim}{\eta}' \underset{\sim}{\theta} \right| \to 0 \qquad \ldots (4.14)$$

almost surely as $n \to \infty$. Therefore

$$\begin{aligned} \|\boldsymbol{T}_n^* - \underset{\sim}{\theta}\|^2 &= \|\sum_1^d \hat{t}_{in} \hat{\underset{\sim}{\eta}}_i - \sum_1^d (\hat{\underset{\sim}{\eta}}_i' \underset{\sim}{\theta}) \hat{\underset{\sim}{\eta}}_i\|^2 \\ &= \sum_1^d (\hat{t}_{in} - \hat{\underset{\sim}{\eta}}_i' \underset{\sim}{\theta})^2 \\ &\leq dD_n^2 \end{aligned}$$

which converges to 0 almost surely as $n \to \infty$. □

The above theorem gives a partial solution to the problem of constructing multivariate medians with equivariance under orthogonal transformations and monotonicity. Moreover, the multivariate median statistics obtained in this manner remains strongly consistent for angularly symmetric distributions.

REMARK 4.2. Next we address the problem of affine equivariant medians. Employing the same logic used in constructing $\boldsymbol{T}_n^*$ in (4.10) we can construct an affine equivariant versionl of $\boldsymbol{T}_n^*$. In this case, we need to construct a suitable affine equivariant, 'data driven' coordinate system using $\underset{\sim}{z}_1, \cdots, \underset{\sim}{z}_n$. A general recipe for such constructions can be found in Chaudhuri and Sengupta (1993b). We shall denote the affine equivariant version of (4.18) by $\tilde{\boldsymbol{T}}_n^a$ for future references (with the corresponding affine equivariant reference frame $\underset{\sim}{\tilde{\eta}}_1, \cdots, \underset{\sim}{\tilde{\eta}}_d$).

5. Finite Sample Breakdown Points and Related Issues

As mentioned earlier, the main idea of Bassett (1991) can be stated as follows. For any $\alpha, 0 < \alpha \leq \frac{1}{2}$, let S_α be the range of all (univariate) affine equivariant, monotonic statistics with finite sample breakdown point at least α. The family $\{S_\alpha\}$ turns out to be a nested family of regions (actually intervals between certain order statistics) starting with the convex hull of the data and ending at the sample median (or, the median interval). In the multivariate case we developed certain classes of statistics, namely, $\boldsymbol{T}_{0n}, \boldsymbol{T}_n^*, \boldsymbol{T}_n^{*a}$ and $\tilde{\boldsymbol{T}}_n^a$. The key idea behind extending Bassett's result to higher dimension is to represent any measure of location as $\sum_1^d t_i \underset{\sim}{\eta}_i$. The quantities $t_1, \cdots, t_d$ are (univariate) affine equivariant, monotonic statistics. However they should be invariant under the group of transformations operating on the d-dimensional data. On the other hand the reference system $\{\underset{\sim}{\eta}_1, \cdots, \underset{\sim}{\eta}_d\}$ should be constructed in such a way that it is equivariant under the group of transformations but invariant under contractions of the data.

In our method of construction the components, $t_1, \cdots, t_d$ are constructed on the basis of the projected data along $\underset{\sim}{\eta}_1, \cdots, \underset{\sim}{\eta}_d$ respectively. For coordinatewise measures like $\boldsymbol{T}_{0n}$ (defined by (3.2)) the reference system is fixed (the system consisting of the standard basis directions). For measures which are equivariant under orthogonal transformations such as $\boldsymbol{T}_n^*$ or $\boldsymbol{T}_n^{*a}$, the reference system is 'data driven' and is equivariant under orthogonal transformations. Notice that the L_1-median, $\underset{\sim}{\hat{\theta}}$ defined through (4.1) can be expressed in this fashion. Fix any orthonormal, reference system $\underset{\sim}{\hat{\eta}}_1, \cdots, \underset{\sim}{\hat{\eta}}_d$ which is equivariant under orthogonal transformations. Then,

$$\underset{\sim}{\hat{\theta}} = \sum_{1}^{d} \hat{t}_i \underset{\sim i}{\hat{\eta}} \qquad \ldots (5.1)$$

where $\hat{t}_i = \sum_{k=1}^{n} \hat{w}_k \hat{u}_{ki}$ with $\hat{u}_{ki} = \underset{\sim i}{\hat{\eta}}' \underset{\sim k}{x}$ and $\hat{w}_k = \|\underset{\sim k}{x} - \underset{\sim}{\hat{\theta}}\|^{-1} \left[\sum_{1}^{n} \|\underset{\sim \ell}{x} - \hat{\theta}\|^{-1}\right]^{-1}$ for $1 \leq k \leq n$. Because $\underset{\sim}{\hat{\theta}}$ is orthogonally equivariant, the components of $\hat{t}_1, \cdots, \hat{t}_d$ are also orthogonally equivariant. Next let us consider a reference system $\underset{\sim}{\eta}, \cdots, \underset{\sim d}{\eta}$ and fix some $0 < \alpha \leq \frac{1}{2}$. Consider the $d \times d$ matrix $\boldsymbol{E} = (\underset{\sim 1}{\eta}, \cdots, \underset{\sim d}{\eta})$ which is nonsingular by construction. The coordinates of $\underset{\sim 1}{x}, \cdots, \underset{\sim n}{x}$ under the new reference system are given by $\underset{\sim i}{y} = (y_{i1}, \cdots, y_{id})' = \boldsymbol{E}^{-1} \underset{\sim i}{x}, 1 \leq i \leq n$ respectively. If $\underset{\sim 1}{\eta}, \cdots, \underset{\sim d}{\eta}$ form an orthonormal system, $\boldsymbol{E}^{-1} = \boldsymbol{E}'$. Let $S_{j\alpha}$ denote the range of (univariate) affine equivariant, monotonic statistics with breakdown point $\geq \alpha$, based on the univariate data $\{y_{1j}, \cdots, y_{nj}\}$. Therefore a natural extension of Bassett's idea to the d-dimensional space would be

$$\boldsymbol{S}_\alpha = S_{1\alpha} \times \cdots \times S_{d\alpha}. \qquad \ldots (5.2)$$

Here '$\times$' denotes the Cartesian product. The sequence of regions $\boldsymbol{S}_\alpha$ for $0 < \alpha \leq \frac{1}{2}$ are rectangular and nested. The coordinatewise median with respect to the reference system constructed from the columns of $\boldsymbol{E}$ sits at the center.

Suppose $\boldsymbol{S}(\boldsymbol{X})$ denote a region in $\mathbb{R}^d$ for a given set of points $\boldsymbol{X} = \{\underset{\sim 1}{x}, \cdots, \underset{\sim n}{x}\}$. We shall define the breakdown point of $S(\boldsymbol{X})$ (just as in (1.1)) by

$$BD(\boldsymbol{S}, \boldsymbol{X}) = \min\{\frac{m}{n} : \boldsymbol{S}(\boldsymbol{Y}_m) \text{ is unbounded}\} \qquad \ldots (5.3)$$

where $\boldsymbol{Y}_m = \{\underset{\sim 1}{y}, \cdots, \underset{\sim n}{y}\}$ with $|\boldsymbol{Y}_m \cap \boldsymbol{X}| = n - m$.

Fact 5.1. When the basis matrix $\boldsymbol{E}$ is fixed or chosen in such a way that it is orthonormal (and equivariant under orthogonal transformations), the region $\boldsymbol{S}_\alpha$ (defined through (5.2)) has a breakdown point at least α (for $0 < \alpha \leq \frac{1}{2}$).

In order to see why the above result is true notice that the region $S(Y_m)$ becomes unbounded whenever one of the $S_{i\alpha}$ does so. From earlier univariate calculations we know that each $S_{i\alpha}$ has breakdown level α. Therefore, S_α must have breakdown α.

Fact 5.2. Under the assumptions of the theorem 4.3, the breakdown point of $\boldsymbol{T}_n^*$ (or, $\boldsymbol{T}_n^{*a}$) can be made as large as $(\frac{1}{2} - \frac{1}{2n})$.

Fact 5.2 can be obtained from Fact 5.1, by choosing each univariate t_i's as median. In view of the above facts we can construct monotonic statistics with breakdown point close to $\frac{1}{2}$ which are orthogonally equivariant in an asymptotic sense. If we are permitted to use auxiliary data on the same set of variates we can construct monotonic, orthogonally equivariant statistics with breakdown point as high as $\frac{1}{2}$ in an exact sense. One can also construct natural 'breakdown contours' in $\mathbb{R}^d$ (namely, $\boldsymbol{S}_\alpha$ described by (5.2)) using such statistics. Such contours would serve the same purpose as so called 'depth contours' (see, Liu (1990) or Small (1990) for example) introduced by Tukey (1975).

Finally we discuss the issue of coupling affine equivariance in $\mathbb{R}^d$ with monotonicity. The conclusions of the previous facts (5.1 and 5.2) are not true for affine equivariant statistics because the choice of the reference frame will affect the breakdown point of the statistic. If one considers the minimum volume ellipsoid (MVE) statistics, the breakdown point would be as high as $\frac{1}{2}$ but the statistic may show 'anti-monotonicity' behaviour for various configurations. There is a trade-off between monotonicity and breakdown point for affine equivariant statistics.

THEOREM 5.1. *Let $\boldsymbol{T}$ be an affine equivariant, directional monotonic statistic satisfying $\boldsymbol{T}(\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n) \in$ convex hull $(\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n)$. Then*

$$\inf_{\boldsymbol{X}} BD(\boldsymbol{T}, \boldsymbol{X}) \leq \frac{1}{3} + \frac{2}{3n} \qquad \ldots (5.4)$$

provided $d \geq 2$.

PROOF. First fix some $\underset{\sim}{\lambda} \in \mathbb{R}^d$, $\|\underset{\sim}{\lambda}\| = 1$ and $\underset{\sim}{z}_1, \cdots, \underset{\sim}{z}_n \in \mathbb{R}^d$ satisfying $\underset{\sim}{\lambda}' \underset{\sim}{z}_i = 0$ for $1 \leq i \leq n$. For fixed $\underset{\sim}{\lambda}$ and $\underset{\sim}{z}_1, \cdots, \underset{\sim}{z}_n$ define for $u_1, \cdots, u_n \in \mathbb{R}$

$$h(u_1, \cdots, u_n) = \underset{\sim}{\lambda}' \boldsymbol{T}(\underset{\sim}{z}_1 + u_1 \underset{\sim}{\lambda}, \cdots, \underset{\sim}{z}_n + u_n \underset{\sim}{\lambda}). \qquad \ldots (5.3)$$

Note that the definition depends on $\underset{\sim}{\lambda}$ and $\underset{\sim}{z}_1, \cdots, \underset{\sim}{z}_n$. We suppress this dependence for notational convenience. It is now easy to verify that h is affine equivariant. Because for $a \geq 0$, $b \in \mathbb{R}$

$$\begin{aligned} h(au_1 + b, \cdots, au_n + b) &= \underset{\sim}{\lambda}' \boldsymbol{T}(A\underset{\sim}{y}_1 + b\underset{\sim}{\lambda}, \cdots, A\underset{\sim}{y}_n + b\underset{\sim}{\lambda}) \\ &= \underset{\sim}{\lambda}' A \boldsymbol{T}(\underset{\sim}{y}_1, \cdots, \underset{\sim}{y}_n) + b\underset{\sim}{\lambda}'\underset{\sim}{\lambda} \qquad \ldots (5.6) \\ &= a\, h(u_1, \cdots, u_n) + b, \end{aligned}$$

where $\underset{\sim}{y}_i = \underset{\sim}{z}_i + u_i \underset{\sim}{\lambda}, 1 \leq i \leq n \text{and} A = (I_d - \underset{\sim}{\lambda}\underset{\sim}{\lambda}') + a\underset{\sim}{\lambda}\underset{\sim}{\lambda}'$ is nonsingular. Also, notice that $h(0, \cdots, 0) = 0$ because $\boldsymbol{T}(\underset{\sim}{z}_1, \cdots, \underset{\sim}{z}_n)$ is an element of the convex hull of $\underset{\sim}{z}_1, \cdots, \underset{\sim}{z}_n$, which is orthogonal to $\underset{\sim}{\lambda}$. Further by the assumption of directional monotonicity of $\boldsymbol{T}$, h is monotonic in each coordinate.

Next let $\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n$ be the given data and let us assume the hypothesis that $BD(\boldsymbol{T}, \boldsymbol{X}) \geq r/n$ for all $\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n$. For a given set of observations $\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n$

consider h with $\underset{\sim}{z}_i = (I_d - \underset{\sim}{\lambda}\underset{\sim}{\lambda}')\underset{\sim}{x}_i$ for some $\underset{\sim}{\lambda} \in S^{(d-1)}$. It is clear that the breakdown point of any such h is at least r/n. In view of Bassett (1991) we can now claim that

$$(\underset{\sim}{\lambda}'\underset{\sim}{x})_{(r)} \leq \underset{\sim}{\lambda}' T(\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n) \leq (\underset{\sim}{\lambda}'\underset{\sim}{x})_{(n-r+1)}, \qquad \ldots (5.7)$$

where $(\underset{\sim}{\lambda}'\underset{\sim}{x})_{(\cdot)}$ denotes the order statistics of $\underset{\sim}{\lambda}'\underset{\sim}{x}_1, \cdots, \underset{\sim}{\lambda}'\underset{\sim}{x}_n$. Next fix two unit vectors $\underset{\sim}{\lambda}_1, \underset{\sim}{\lambda}_2$ $(\underset{\sim}{\lambda}_1 \neq \underset{\sim}{\lambda}_2)$ and let $\underset{\sim}{\lambda}_3 = \|\underset{\sim}{\lambda}_1 + \underset{\sim}{\lambda}_2\|^{-1}(\underset{\sim}{\lambda}_1 + \underset{\sim}{\lambda}_2)$. Define $a_i = \underset{\sim}{\lambda}_1' \underset{\sim}{x}_i$ and $b_i = \underset{\sim}{\lambda}_2' \underset{\sim}{x}_i$ for $1 \leq i \leq n$. Also let $C = \underset{\sim}{\lambda}_1' \boldsymbol{T}(\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n)$ and $D = \underset{\sim}{\lambda}_2' \boldsymbol{T}(\underset{\sim}{x}_1, \cdots, \underset{\sim}{x}_n)$ respectively. By virtue of (5.7) the following inequalities are valid.

$$\begin{array}{ccccc} a_{(r)} & \leq & C & \leq & a_{(n-r+1)} \\ b_{(r)} & \leq & D & \leq & b_{(n-r+1)} \\ (a+b)_{(r)} & \leq & C+D & \leq & (a+b)_{(n-r+1)}, \end{array} \qquad \ldots (5.8)$$

where $(a+b)_{(\cdot)}$ are the order statistics of $(a_i + b_i), 1 \leq i \leq n$. Next vary the configuration of the set of observations arbitrarily. This way we shall be able to generate the sequences $\{a_i\}$ and $\{b_i\}$ independently of each other if $d \geq 2$. This follows by simply solving $\underset{\sim}{\lambda}_1' \underset{\sim}{x}_i = a_i$ and $\underset{\sim}{\lambda}_2' \underset{\sim}{x}_i = b_i$ for $1 \leq i \leq n$. Now by the first two inequalities in (5.8) we have $a_{(r)} + b_{(r)} \leq C + D \leq a_{(n-r+1)} + b_{(n-r+1)}$. Hence

$$(a+b)_{(r)} \leq a_{(n-r+1)} + b_{(n-r+1)}. \qquad \ldots (5.9)$$

However by matching $a_{(1)}, \cdots, a_{(r-1)}$ with $b_{(n-r+2)}, \cdots, b_{(n)}$ and $b_{(1)}, \cdots, b_{(r-1)}$ with $a_{(n-r+2)}, \cdots, a_{(n)}$ respectively and then making $a_{n-r+2)}, \cdots, a_{(n)}$, $b_{(n-r+2)}, \cdots, b_{(n)}$ sufficiently large while keeping the other order statistics moderate we can violate (5.9) unless $r \leq n - 2(r-1)$. In other words we must have

$$\begin{aligned} \frac{r}{n} &\leq \frac{n+2}{3n} \\ &= \frac{1}{3} + \frac{2}{3n}. \end{aligned}$$

Hence the theorem follows. □

It is noteworthy that Donoho and Gasko (1992) obtained a similar result from Tukey's halfspace median. Because the halfspace median is not a unique point (in general this is true for other affine equivariant medians such as Oja's simplex median Oja (1983)) it is virtually impossible to establish directional monotonicity for this median. However one can intuitively realize that an appropriate directional monotonicity holds for Tukey's halfspace median. Also by (5.7) we have an interesting relationship between any affine equivariant, directional monotonic statistic and the order statistics of the projected data in various directions. The halfspace median $\boldsymbol{T}$ is computed in such a way that r is maximized in a 'minimax' way. It seems that (5.7) should serve as the building block for constructing a large class of affine equivariant, directional monotonic measures of location with high breakdown.

Acknowledgment. The authors would like to thank the Co Editor for efficient handling of the manuscript and also to anonymous referees for helpful comments and suggestions.

References

Apostol, T.M. (1974). *Mathematical Analysis*, 2nd Ed., Addison- Wesley.

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*, John Wiley, N.Y.

Barnett, V. (1976). The ordering of multivariate data (with discussion), *J. Roy. Statist. Soc.* A, **139**, 318-354.

Bassett, G. (1991). Equivariant, monotonic, 50% breakdown estimators, *The Amer. Statist.*, **45**, 2, 135-137.

Chaudhuri, P. and Sengupta, D. (1993a). A note on robust estimation of location. To appear in *Stat. Probab. Letters.*

– – –– (1993b). Sign tests in multidimensions : Inference based on the geometry of data cloud. To appear in *Jour. Amer. Statist. Assoc.*.

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

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

Efron, B. (1978). The geometry of exponential families, *Ann. Statist.*, **6**, 2, 362-376.

Huber, P.J. (1981), *Robust Statistics*, Wiley, N.Y.

Liu, R.Y. (1990). On a notion of data depth based upon random simplices, *Ann. Statist.* **18**, 405-415.

Lopuhaä, H.P. (1990). Estimation of location and covariance with high breakdown point, *Ph.D. thesis*, Delft Univ. of Technology.

– – –– (1992). Highly efficient estimators of multivariate location with high breakdown point, *Ann. Statist.*, **20**, 1, 389-413.

Lopuhaä, H.P. and Rousseeuw, P.J. (1991). Breakdown properties of affine equivariant estimators of multivariate location and covariance matrices, *Ann. Statist.*, **19**, 1, 229-248.

Mardia, K.V. (1972). *Statistics of Directional Data*, Academic Press.

Oja, H. (1983). Descriptive statistics for multivariate distributions, *Stat. Probab. Letters.*, **1**, 327-332.

Pollard, D. (1984). *Convergence of Sstochastic Processes*, Springer-Verlag, N.Y.

Rousseeuw, P.J. (1985). Multivariate estimation with high breakdown point. In *Math. Statist. Appl.* (W. Grossmann, G. Pflug, I. Vincze and W. Wertz eds.), 283-297, Reidel Dordrecht.

Small, C.G. (1990). A survey of multidimensional medians. *Inter Statist. Rev.,* **58**, 3, 263-277.

Tukey, J.W. (1975). Mathematics and the picturing of data. In *Proc. International Congress of Mathematicians, Vancouver,* 1974, **2**, 523-531.

Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203
U.S.A.
e-mail: sghosh@stat.ncsu.edu

Stat-Math Unit
Indian Statistical Institute
203 BT Road, Calcutta-700 035
INDIA
e-mail: dps@isical.ac.in