Some Extensions of Tukey's Depth Function

Jian Zhang
EURANDOM, Eindhoven and
The Chinese Academy of Sciences, Beijing

revised August 27, 2000

Abstract

As the extensions of Tukey's depth, a family of affine invariant depth functions are introduced for multivariate location and dispersion. The location depth functions can be used for the purpose of multivariate ordering. Such a kind of ordering can retain more information from the original data than that based on Tukey's depth. The dispersion depth functions provide some additional view of the dispersion of the data set. It is shown that these sample depth functions converge to their population versions uniformly on any compact subset of the parameter space. The deepest points of these depth functions are affine equivariant estimators of multivariate location and dispersion. Under some general conditions these estimators are proved to have asymptotic breakdown points at least 1/3 and convergence rates of $1/\sqrt{n}$. Their asymptotic distributions are also obtained under some regularity conditions. A new algorithm based on the idea of thresholding is presented for computing these kinds of estimators and realized in the bivariate case. Simulations indicate that some of them could have the empirical mean squared errors smaller than those based on Tukey's depth function or Donoho's depth function.

Key words and Phrases: Depth and outlyingness, location and dispersion estimators, breakdown point, multivariate ordering.
AMS 1991 subject classification: 60E99, 62H99.
Short running title: Depth Functions

1 Introduction

In the last decades much research has been made on how to rank a point relative to a multivariate data set and on the way of multivariate ordering (see Tukey, 1975; Barnett, 1976; Eddy, 1985; Reiss, 1989; Liu, 1990; Donoho and Gasko, 1992; Liu and Singh, 1993; He and Wang, 1997; Koltchinskii, 1997; Small, 1997; Liu, Parelius and Singh, 1999; Rousseeuw and Hubert, 1999; Zuo and Serfling, 2000, and the references therein). An important notion, now called depth, forms the core of these works. The depth of a point in a parameter space is its rank relative to a data set. One of the incentives to these developments is that the depth notion could be a powerful tool for developing affine equivariant robust estimators for multivariate location and dispersion (see, for example, Tukey, 1975; Donoho and Gasko, 1992; Maronna, Stahel and Yohai, 1992; and Tyler, 1994). Tukey's depth is very simple but it is a step function and uses only the qualitative information in the data. As a result the related deepest point is quite less efficient than the sample mean and sometimes can be locally unstable, jumping suddenly to another value when slight changes are made in the data (see Hettmansperger and Sheather, 1992). The goal of this paper is to improve the performance of Tukey's depth by incorporating some auxiliary functions into Tukey's depth notion. A similar technique was adapted by Plackett (1976) for some multivariate ordering problems.

To highlight the basic idea behind our extensions, we consider a univariate data set $Z = (z_1, \cdots, z_n)$ with empirical distribution F_{nZ}. Let $\mu(\cdot)$ and $s(\cdot)$ represent translation and scale equivariant location and scale functional on the family of all univariate distributions. Then the M-estimators of location and scale are, respectively, defined as the solutions (β, σ) of the equations

$$\frac{1}{n}\sum_{i=1}^{n} \psi((z_i - \beta)/s(F_{nZ})) = 0, \quad \frac{1}{n}\sum_{i=1}^{n} \chi((z_i - \mu(F_{nZ}))/\sigma) = 0$$

where ψ and χ are commonly taken to be an odd function and an even function, respectively. It is obvious that $O_{\psi n}(\beta, Z) = |\sum_{i=1}^{n} \psi((z_i - \beta)/s(F_{nZ}))/n|$ is a measure of the outlyingness of β relative to Z. $(1 + O_{\psi n}(\beta, Z))^{-1}$ describes the depth of β relative to Z. Similarly $(1 + 0_{\mu\chi n}(\sigma, Z))^{-1}$ with $0_{\mu\chi n}(\sigma, Z) = |\sum_{i=1}^{n} \chi((z_i - \mu(F_{nZ}))/\sigma))/n|$ gives the depth of σ relative to $(|z_1 - \mu(F_{nZ})|, \cdots, |z_n - \mu(F_{nZ})|)$. The M-estimators are just the deepest points of these depth functions. For a multivariate data set X, and for parameters $\theta \in R^p$ and positive $p \times p$ matrix Σ, the depths of θ and Σ can be easily defined by applying the above idea to the worst one-dimensional projections, namely, $(a^\tau X, a^\tau \theta, a^\tau \Sigma a)$ (see Section 2). The dispersion depth can be established even without the location functional $\mu(\cdot)$. Like Tukey's depth, the location depths can be applied for the ordering of X through its depth contours, while the dispersion depths provide some additional view of the dispersion of X.

We show that the new depth functions have the basic properties of Tukey's depth: affine invariance, monotonicity relative to the deepest point, strong and uniform consistence with respect to the compact subset of the parameter space, and weak convergency. We take the deepest points of these depth functions to construct some new affine equivariant estimates for multivariate location and dispersion. Most of these estimates are proved to have at least 1/3 (sometimes approximate 1/2) asymptotic breakdown points under elliptic symmetry. So their resistance to outliers can be better than Tukey's median whose asymptotic breakdown point is 1/3. Our simulations indicate at least in the bivariate case these estimates can be more efficient than Tukey's median and the location estimates studied in Tyler (1994). Furthermore, we show that these estimates converge weakly to certain functionals of some Gaussian processes and have converges rates of $1/\sqrt{n}$. In one word, both theoretical and empirical results indicate that the ordering based on our generalized Tukey depths can retain much more information from the original data than that using the Tukey depth.

Note that depth functions are semiparametric or nonparametric in nature. Another well-known tool to rank points in a parameter space is the likelihood. It is natural to ask whether the depth contours (see He and Wang, 1997) can be interpreted as some contours based on some semiparametric likelihood. In this paper we show that it is true for Tukey's depth. In fact, we show that there exists a strictly increasing transformation between Tukey's depth and an empirical likelihood of location.

One of our computational algorithms for implementing our new estimators is based on the idea of thresholding which aims at reducing the double optimization problem to a few of single optimization problems. Another is the direct application of the simulated annealing of Vetterling, Teukolsky, Press and Flannery (1992).

The rest of paper is arranged as follows. In Section 2, we give a uniform definition of the new depth functions. Then we unveil some of their properties. In Section 3, we investigate the breakdown behavior and asymptotic properties of the multivariate estimators based on these new depth functions. In Section 4, we present some simulation results. In Section 5, we establish the relationship between Tukey's depth and the projection based empirical likelihood. Technical proofs of the main results are deferred to the last section. Throughtout the paper we denote by $X = (x_1, \cdots, x_n)$ a sample of size n from the p-dimensional distribution F. Let F_n be the empirical distribution function of X. Denote by F^a and F_n^a the theoretical and empirical distribution functions of the projected sample in direction a. Let P, P_n, P^a, and P_n^a denote the probability measures induced by F, F_n, F^a and F_n^a respectively. For simplicity, we write the expectation of h under P as Ph. "$\xrightarrow{\mathcal{L}}$" means convergence in distribution. Denote by I_A the indicator function of set A.

2 A family of depth functions

Following the basic idea illustrated in the last section, first we choose a univariate function g_a for each direction a. Here g_a is allowed to depend on X and θ. Set $z_i(a,\theta) = a^\tau(x_i - \theta)/s(F_n^a)$ for each i and $s(\cdot)$ in the last section. Then the projection based outlyingness function of location induced by g_a is defined by

$$O_{gn}(\theta) = O_{gn}(X,\theta) = \max_{||a||=1} |\frac{1}{n}\sum_{i=1}^{n} g_a(z_i(a,\theta))|. \tag{2.1}$$

The corresponding depth function, namely $D_{gn}(\theta)$ is

$$D_{gn}(\theta) = D_{gn}(X,\theta) = \frac{1}{1 + O_{gn}(\theta)}.$$

Let $x_{(1)}, \cdots, x_{(n)}$ be order statistics according to the corresponding depths $D_{gn}(x_{(1)}) \geq \cdots \geq D_{gn}(x_{(n)})$. Then, for $0 < \alpha < 1$, the $\alpha-$th depth contour can be constructed by

$$C_n(\alpha) = \{x : D_{gn}(x) = D_{gn}(x_{[\alpha n]})\}$$

where $[\alpha n]$ stands for the integer part of αn.

If we let $g_a(z) = 1$ when $z \geq 0$; $g_a(z) = -1$ when $z < 0$, then the scale $s(\cdot)$ is not required, since D_{gn} is scale-free in this case. Moreover, if we let $D_{Tn}(\theta) = \min_{||a||=1} \sum_{i=1}^{n} I_{[a^\tau(x_i-\theta)<0]}$, then an equivalent relationship between D_{gn} and D_{Tn} is recovered in the sense that there exists a strictly increasing transformation:

$$D_{gn}(\theta) = \frac{1}{2 - 2D_{Tn}(\theta)/n}. \tag{2.2}$$

We call them equivalent because the depth contours (see He and Wang, 1997) are invariant when the depth function is subject to a strictly increasing transformation. D_{Tn} is slightly different from the traditional form of the Tukey depth function, which is defined by

$$\min_{||a||=1} \sum_{i=1}^{n} I_{[a^\tau(x_i-\theta)\leq 0]} = \min_{||a||=1} \min\{\sum_{i=1}^{n} I_{[a^\tau(x_i-\theta)\leq 0]}, \sum_{i=1}^{n} I_{[a^\tau(x_i-\theta)\geq 0]}\}. \tag{2.3}$$

We call D_{Tn} the left continuous version of the Tukey depth function, since D_{Tn} becomes the traditional form if we only replace the left continuous distribution $\sum_{i=1}^{n} I_{[a^\tau(x_i-\theta)<0]}/n$ by its right continuous version.

There are other ways to generalize Tukey's depth function. For example, if we view $I_{[a^\tau(x_i-\theta)\leq 0]}$ in (2.3) as the absolute score to quantify the contribution of $a^\tau(x_i-\theta)$ to the depth of θ, then we have the following extension:

$$\begin{aligned} D_{\psi 1n} &= \min_{||a||=1} \min\{\textstyle\sum_{i=1}^{n} |\psi(a^\tau(x_i-\theta)/s(F_n^a))| I_{[a^\tau(x_i-\theta)\leq 0]}, \\ &\qquad \textstyle\sum_{i=1}^{n} |\psi(a^\tau(x_i-\theta)/s(F_n^a))| I_{[a^\tau(x_i-\theta)\geq 0]}\} \end{aligned} \tag{2.4}$$

where ψ is defined in the last section.

Donoho and Gasko (1992) introduced a notion of outlyingness of θ relative to X. It is defined as the distance between θ and the center of X in the worst one-dimensional projection:

$$O_{dn}(\theta) = \max_{||a||=1} |Med(\frac{a^\tau(X-\theta)}{MAD(a^\tau X)})|$$

where Med denotes median and MAD denotes median absolute deviation about Med. We call $D_{dn} = \{1 + O_{dn}(\theta)\}^{-1}$ Donoho's depth function. Tyler (1994) showed that the deepest point of Donoho's depth function has the finite sample breakdown point close to 1/2. However, our simulations show that although, as an estimator of location, it has a high breakdown point, its efficiency can be significantly lower than some of the deepest points of the our generalized Tukey depth functions (see Section 4). We use the weighted sum of the outlyingness functions O_{dn} and O_{gn} to construct a new depth function $\{1 + O_{dn} + wO_{gn}\}^{-1}$ where w is a positive constant and O_{gn} is defined by (2.1). We demonstrate that under some general conditions the deepest point of the new depth function has the same finite breakdown point as that of D_{dn}, however, it could be more efficient.

The following are some *special cases:*

(1) $D_{\psi 1n}$ in (2.4) with ψ being bounded and odd, is equivalent to D_{gn} with $g_a(z) = \psi(\infty) - |\psi(z)| I_{[z\leq 0]}$. So the traditional form of Tukey's depth is recovered by D_{gn} if letting $g_a(z) = 1 - I_{[z\leq 0]}$.

(2) If

$$g_a(z) = \begin{cases} z, & |z| \leq c; \\ c, & z > c; \\ -c, & z < -c. \end{cases}$$

with a tuning constant $c > 0$, we obtain the generalized Tukey depth function based on Huber's function. If let $g_a(z) = (1 - \exp(-cz))/(1 + exp(-cz))$ with a large tuning constant $c > 0$, then we obtain a smooth version of Tukey's depth function. Such kind of smooth can avoid the local instability of Tukey's depth.

(3) *Depth functions associated with quantile functions.* For $0 < \alpha < 1$, let $g_a(z) = -\alpha I_{[z \le 0]} + (1-\alpha) I_{[z>0]}$ or $g_a(z) = (1+\alpha) I_{[z \le 0]} + \alpha I_{[z>0]}$. Then we have a depth function in which a prelimiary univariate scale estimator is not required.

(4) *Depth functions based on trimmed statistics.* Let $z_{(1)}(a,\theta) \le \cdots \le z_{(n)}(a,\theta)$ denote the order statistics of $z_1(a,\theta), \cdots, z_n(a,\theta)$. For a fraction α such that αn is a nonnegative integer and $2\alpha n \le n-1$, let

$$g_a(z) = z I_{[z_{(n\alpha+1)}(a,\theta) \le z \le z_{(n-n\alpha)}(a,\theta)]}.$$

Then the depth functions based on the trimmed statistics are obtained. In particular, let $\alpha = 1/2 - 1/n$ when n is even and $\alpha = 1/2 - 1/(2n)$ when n is odd, we recover Donoho's depth function.

(5) *Depth functions based on some discrepancy function.* Note that the M-estimator can be defined as the solution of the optimization of a certain discrepancy function, namely $g-$function. So we can let $g_a(z) = g(z)$. For bounded g-function, without loss of generality, we assume that $0 \le g \le 1$ because the solution of the above optimization is invariant when $g(z)$ is multiplied by a positive constant or added by some constant.

In what follows we focus on the case when $g_a = g$ is independent of a, θ and X. The results can be readily extended to the other depth functions like those based on some trimmed statistics. For completeness, the following propositions or theorems will include the results of Donoho and Gasko (1992), Nolan (1992), Chen (1995) and He and Wang (1997) in the case when g is the sign function.

Set $z(a,\theta) = a^\tau(x-\theta)/s(F^a)$ with $s(\cdot)$ in the last section. Then the population version of $D_{gn}(\theta)$ can be expressed as

$$D_g(\theta) = D_g(F,\theta) = \frac{1}{1+O_g(\theta)}$$

where $O_g(\theta) = \max_{||a||=1} |Pg(z(a,\theta))|$.

It is easy to see that D_{gn} is affine invariant. Here we call a depth function $D(X,\theta)$ affine invariant if for any nonsingular $p \times p$ matrix A and $p-$vector v,

$$D(AX+v, A\theta+v) = D(X,\theta).$$

The following proposition shows D_{gn} also has nice asymptotic behavior in which we need the conditions:

(S0). $\max_{||a||=1} |s(F_n^a) - s(F^a)| \to 0, \quad 0 < \min_{||a||=1} s(F^a) \le \max_{||a||=1} s(F^a) < \infty.$

(G0). $P|g(s_1||x||+s_2)| < \infty, \quad$ for any $s_1 \ne 0, s_2 \ne 0$.

(G0'). $P|g(s_1||x||+s_2)|^2 < \infty, \quad$ for any $s_1 \ne 0, s_2 \ne 0$, and as $\quad ||a-b|| + |s-t| \to 0$,

$$P\{g(a^\tau(x-\theta)/s) - g(b^\tau(x-\theta)/t)\}^2 \to 0.$$

The commonly used g function satisfies (B) or (U) below. See Zhang and Li (1998) for some examples.

(B). $g(z)$ attains its minimum 0 at $z=0$; g is nonincreasing for $z<0$ and nondecreasing for $z>0$; furthermore, $g(z) \to 1$ as $|z| \to \infty$.

(U). $g(z)$ attains its minimum 0 at $z=0$; g is even and is nondecreasing for $z>0$; $\lim_{|z|\to\infty} g(z) = \infty$; $\psi = g'$ is continuous in R^1, and there exists $z_o \ge 0$ such that ψ is nondecreasing in $(0, z_o]$ and nonincreasing in (z_o, ∞).

For each fixed θ, set $h_a(x) = g(a^\tau(x-\theta)/s(F^a))$,

$$\begin{aligned} A_+ &= \{a \in R^p : ||a|| = 1, Ph_a = \max_{||b||=1} |Ph_b|\}, \\ A_- &= \{a \in R^p : ||a|| = 1, Ph_a = -\max_{||b||=1} |Ph_b|\}. \end{aligned}$$

Note that for simplicity, we suppress the notation θ in $h_a(x)$, A_+ and A_- above.

Proposition 2.1 *Suppose that g is a monotone function or satisfies one of conditions* **(B)** *and* **(U)**. *Suppose that Conditions* **(S0)** *and* **(G0)** *hold. Then $D_{gn}(\theta)$ converges to $D_g(\theta)$ almost surely and uniformly on any compact subset of R^p (on R^p when g is bounded). Furthermore, if* **(G0)** *is replaced by* **(G0')** *above, then for each θ,*

$$\sqrt{n}(D_{gn}(\theta) - D_g(\theta)) \xrightarrow{\mathcal{L}} \frac{1}{(1+O_g(\theta))^2} \max\{\max_{a\in A_+} W(h_a), -\min_{a\in A_-} W(h_a)\}.$$

where $\{W(h_a) : ||a|| = 1\}$ is a centered Gaussian process with continuous sample paths and covariance

$$EW(h_a)W(h_b) = Ph_ah_b - Ph_aPh_b.$$

Here, we define $\max_{a\in\emptyset}\{\cdot\} = -\infty$ when A_+ or A_- is empty.

In the following corollary, we assume that F has an elliptic density

$$det(\Sigma_o)^{-1/2} f_o((x-\theta_o)^\tau \Sigma_o^{-1}(x-\theta_o)) > 0, \quad x \in R^p$$

where Σ_o is positive definite. Note that the distribution of $a^\tau \Sigma_o^{-1/2}(x-\theta_o)$ is same for all a, $||a|| = 1$. So we let F_{mo} denote this common projection distribution.

Corollary 2.1 *Suppose that* **(S0)** *holds. Let g be a bounded, odd function such that $|\int g(z+v)dF_{mo}(z)|$ is strictly increasing with respect to $v \geq 0$. Assume* **(G0)** *holds. Then, under the Hausdorff distance, for $0 < \alpha < 1$, the $[n\alpha]$-th depth contour $C_n(\alpha)$ converges almost surely to an ellipsoid of the form $\{x \in R^p : (x-\theta_o)^\tau \Sigma_o^{-1}(x-\theta_o) = q(\alpha)\}$ where constant $q(\alpha)$ depends on α, f_o and g. Here the Hausdorff distance of two sets A and B is defined as $\max\{max_{x_B\in B}\min_{x_A\in A}||x_A - x_B||, max_{x_A\in A}\min_{x_B\in B}||x_B - x_A||\}$.*

Remark 2.1 *Suppose that g is a nondecreasing odd function and the underlying distribution is spherical. Then $D_g(\theta)$ is monotonically decreasing along any fixed ray stemming from the center of the distribution (see Liu and Singh (1993) for the definition of monotonicity for depth function). Therefore $D_g(\theta)$ inherits the monotonicity property of Tukey's depth function.*

We conclude this section by developing a family of depth functions for dispersion. The basic idea hehind these depths has already been illustrated in the last section. First, we choose an even function g_a for each direction a. To measure the scale alone, we introduce two methods to filter out the location effect. One is based on the external location functional $\mu(\cdot)$ defined in the last section, another is based on U-statistics.

With $\mu(\cdot)$, for each positive $p \times p$ matrix Σ, its outlyingness relative to X is defined by

$$O_{\mu gn}(\Sigma) = O_{\mu gn}(X, \Sigma) = \max_{||a||=1} |\frac{1}{n}\sum_{i=1}^n g_a((a^\tau x_i - \mu(F_n^a))/\sqrt{a^\tau \Sigma a})|.$$

Then the depth function of Σ is defined by

$$D_{\mu gn}(\Sigma) = \frac{1}{1+O_{\mu gn}(\Sigma)}.$$

The depth concept can be established even without the help of the external location functional. For this purpose, we note that for each i, the contribution of $a^\tau x_i$ to the outlyingness of $\sqrt{a^\tau \Sigma a}$ can be measured by

$$\frac{1}{n-1}\sum_{j=1, j\neq i}^n g_a(a^\tau(x_i - x_j)/\sqrt{a^\tau \Sigma a}).$$

Averaging these contributions, we have the following definition of the outlyingness of Σ relative to X by $O_{ugn}(\Sigma) = \max_{||a||=1} |\frac{2}{n(n-1)} \sum_{i<j} g_a(a^\tau(x_i - x_j)/\sqrt{a^\tau \Sigma a})|$. The corresponding depth function of Σ is defined by

$$D_{ugn}(\Sigma) = \frac{1}{1 + O_{ugn}(\Sigma)}.$$

Note that similar to the location case, we can define the population versions, namely $D_{\mu g}(\Sigma)$ and $D_{ug}(\Sigma)$ for $D_{\mu gn}(\Sigma)$ and $D_{ugn}(\Sigma)$, respectively. It is obvious that the above depth functions are affine invariant.

The counterpart of Tukey's depth function in the dispersion setting is obtained by letting $g_a(z) = sign(|z| - 1)$. Importantly, $O_{ugn}(\Sigma)$ requires no external location estimator, making it particularly suitable for problems such as multivariate dispersion estimation, where robust estimation of location is difficult.

Let $\hat{\Sigma}_\mu$ be the deepest point of $D_{\mu gn}$, and $\hat{a}$ the worst direction, which gives the "deepest" view of dispersion, that is,

$$D_{\mu gn}(\hat{\Sigma}) = |\frac{1}{n} \sum_{i=1}^{n} g_{\hat{a}}((\hat{a}^\tau x_i - \mu(F_n^{\hat{a}}))/\sqrt{\hat{a}^\tau \hat{\Sigma} \hat{a}})|.$$

We view the dispersion of X by using the depths of $\{|\hat{a}^\tau x_i - \mu(F_n^{\hat{a}})|\}$, say $\{|r_i|\}$, with

$$r_i = \frac{1}{n} \sum_{j=1}^{n} g_{\hat{a}}((\hat{a}^\tau x_j - \mu(F_n^{\hat{a}}))/|\hat{a}^\tau x_i - \mu(F_n^{\hat{a}})|), \quad 1 \le i \le n,$$

where we let $0/0 = 1$ and $c/0 = \infty$ for $c > 0$.

The next proposition concerns the consistency of the proposed depth functions.

Proposition 2.2 *Let $g_a(z) = g(z)$ be bounded, even and be nondecreasing in $z > 0$. Then*

(i) $D_{ugn}(\Sigma)$ converges to $D_{ug}(\Sigma)$ uniformly for all positive definite $p \times p$ matrices Σ.

(ii) Furthermore, suppose that there exists a location functional $\mu(F^a)$ such that $\max_{||a||=1} |\mu(F_n^a) - \mu(F^a)|$ tends to zero almost surely. Then $D_{\mu gn}(\Sigma)$ converges to $D_{\mu g}(\Sigma)$ almost surely and uniformly for all positive definite $p \times p$ matrices Σ.

Remark 2.2 *$D_{ug}(\Sigma)$ and $D_{\mu g}(\Sigma)$ also have certain monotonicity property. Assume F has an elliptic density and a marginal distribution P_{mo} defined in Corollary 2.1. Suppose that g is a bounded even function and that $|(P_{mo} \times P_{mo})g((z_1 - z_2)/s)|$ is increasing for $s \ge 1$ and is decreasing for $0 < s \le 1$. Then $D_{ug}(\Sigma_2) \le D_{ug}(\Sigma_1)$ when $\Sigma_1 - I_p$ and $\Sigma_2 - \Sigma_1$ are positive definite; $D_{ug}(\Sigma_2) \ge D_{ug}(\Sigma_1)$ when $I_p - \Sigma_2$ and $\Sigma_2 - \Sigma_1$ are positive definite.*

3 Some properties of the deepest points

In this section we assume that $F = F(\cdot; \theta, \Sigma)$ is an elliptic distrubution with unknown parameters θ and Σ. There are several kinds of affine estimators of θ and Σ (see Huber, 1981; Stahel, 1981; Rousseeuw, 1985; Maronna, Stahel and Yohai, 1992; Donoho and Gasko, 1992; Tyler, 1994 and Zhang and Li, 1995). The deepest points of our new depth functions give some alternative estimators.

Location estimator denoted by $\hat{\theta} = \hat{\theta}(g, X)$ will be any element of $\{argmax_\theta D_{gn}(\theta)\}$. Similarly, its population version will be any element from $\{argmax_\theta D_g(\theta)\}$. Usually the population version is unique.

Dispersion estimator with μ, denoted by $\hat{\Sigma}_\mu = \hat{\Sigma}(\mu, g, X)$, will be any element of $\{argmax_\Sigma D_{\mu gn}(\Sigma)\}$. Its population version $\Sigma_\mu(F)$ will be any element of $\{argmax_\Sigma D_{\mu g}(\Sigma)\}$. Here Σ runs over all the $p \times p$ positive definite matrices.

Analogously, we define the dispersion estimator, denoted by $\hat{\Sigma} = \hat{\Sigma}(g, X)$, without the location estimator μ.

In practice, we select the element according to some fixed rule. For example, for the location estimator, it should be the nearest one to the coordinate-wise median or Tukey's median, given for example by Rousseeuw and Ruts (1998). For the large sample, the pair-wise differences for the elements of $\{argmax_\theta D_{gn}(\theta)\}$ are small when the population version is unique.

It is obvious that all these estimators are affine equivariant. That is, for any $p \times p$ nonsingular matrix A and $p-$vector v,

$$\begin{aligned}\hat{\theta}(g, AX + v) &= A\hat{\theta}(g, X) + v;\\ \hat{\Sigma}_\mu(g, AX + v) &= A^\tau \hat{\Sigma}_\mu(g, X)A;\\ \hat{\Sigma}(g, AX + v) &= A^\tau \hat{\Sigma}(g, X)A.\end{aligned}$$

So, without loss of generality, we assume that the underlying distribution function F_o is spherical with marginal distribution F_{mo}. Let P_o and P_{mo} denote the probability measure induced by F_o and F_{mo}, respectively.

3.1 Asymptotics

The above estimators have good asymptotic behavior. First, they are consistent under some regularity conditions. Secondly, some of them have convergence rates of $n^{-1/2}$. For presenting these theorems, the following additional conditions are needed:

(G1): $\int g(z - \beta)dP_{mo}(z) = 0$ if and only if $\beta = 0$, $\int g(z - \beta)dP_{m_o}(z)$ is continuous in β, and $\lim_{\delta\to\infty} \inf_{||\beta||\geq\delta} |\int g(z - \beta)dP_{mo}(z)| > 0$.

(G2): $|\int g(z - \beta)dP_{mo}(z)|$ attains the minimum only at $\beta = 0$.

(G3): $\int g((z_1 - z_2)/s)dP_{mo}(z_1)dP_{m_o}(z_2) = 0$ if and only if $s = 1$, $\int g((z_1 - z_2)/s)dP_{mo}(z_1)dP_{mo}(z_2)$ is continuous in s, and $\lim_{\delta\to\infty} \inf_{s\notin[1/\delta,\delta]} |\int g((z_1 - z_2)/s)dP_{mo}(z_1)dP_{mo}(z_2)| > 0$.

(G4): $\int g(z/s)dP_{mo}(z) = 0$ if and only if $s = 1$, $\int g(z/s)dP_{mo}(z)$ is continuous in s, and

$$\lim_{\delta\to\infty} \inf_{s\notin[1/\delta,\delta]} |\int g(z/s)dP_{mo}(z)| > 0.$$

(G5): $\max_{||a||=1} |s(F_n^a) - 1| \to 0$ almost surely as $n \to \infty$.

(G6): $\max_{||a||=1} |\mu(F_n^a)| \to 0$ almost surely as $n \to \infty$.

In the following we say that $\hat{\theta}(g, X)$ is consistent (with the true value, θ_o, of parameter) if $\sup\{|v - \theta_o| : v \in \{argmax_\theta D_{gn}(\theta)\}\} \to 0$ almost surely. Similar notions for $\hat{\Sigma}(g, X)$ and $\hat{\Sigma}(\mu, g, X)$ can be defined.

Theorem 3.1 *(i) Suppose* **(G5)** *holds. If g is monotone and satisfies* **(G0)** *and* **(G1)** *or g satisfies* **(B)** *and* **(G2)** *or g satisfies* **(G0), (G2)** *and* **(U)**, *then $\hat{\theta}(g, X)$ is consistent.*

(ii) If g is bounded and monotone and satisfies **(G3)**, *then $\hat{\Sigma}(g, X)$ is a consistent.*

(iii) If g is bounded and monotone and satisfies **(G4)** *and* **(G6)**, *then $\hat{\Sigma}(\mu, g, X)$ is also consistent.*

Remark 3.1 *Suppose that F_{mo} has a density with respect to the Lebesgue measure which is even and strictly decreasing in $z \geq 0$. Then* **(G2)** *holds if g satisfies* **(B)** *or $g = \psi(\infty) - |\psi(z)|I_{[z\leq 0]}$ with $\psi(z)$ being bounded, odd and nondecreasing.*

To derive the asymptotic distributions of our estimators, we need the further conditions:

(S1) $\sqrt{n}(s(F_n^a)-1) = \sum_{i=1}^n s_I(a^\tau x_i)/\sqrt{n} + o_p(1) \xrightarrow{\mathcal{L}} W_s(a)$ where $\{W_s(a) : ||a|| = 1\}$ is a Gaussian process.

(G7) g is nondecreasing and satisfies:

$$P_{mo}g(z) = 0; \quad P_o\{g(s_1||x|| + s_2)\}^2 < \infty, s_1 \neq 0, s_2 \neq 0;$$
$$P_o\{g(a^\tau(x-\theta)/s) - g(b^\tau(x-\theta_1)/t)\}^2 \to 0, \quad \text{as } ||a-b|| + ||\theta-\theta_1|| + |s-t| \to 0;$$

and $\Psi_1^a(\theta, s) = \partial P_o g(a^\tau(x-\theta)/s)/\partial\theta$ and $\Psi_2^a(\theta, s) = \partial P_o g(a^\tau(x-\theta)/s)/\partial s$ are continuous with $\Psi_1^a(0,1) = d(a)a, \quad \min_{||a||=1}|d(a)| > 0$.

Let $\{W(g,a) : ||a|| = 1\}$ be a Gaussian process with zero means and covariance

$$EW(g,a)W(g,b) = P_o g(a^\tau x)g(b^\tau x) - P_o g(a^\tau x)P_o g(a^\tau x).$$

Theorem 3.2 *Under Conditions* **(S1)** *and* **(G7)**, *we have*

$$\hat{\theta}(g, X) = O_p(1/\sqrt{n}).$$

Furthermore,

$$\sqrt{n}\hat{\theta}(g,X) \xrightarrow{\mathcal{L}} argmin_u \max_{||a||=1} |W(g,a) + u^\tau \Psi_1^a(0,1) + W_s(a)\Psi_2^a(0,1)|$$

provided the argmin is unique with probability 1.

We now consider the asymptotic behavior of the dispersion estimators. To begin with, we introduce a few more notations. Denote

$$\mathcal{F} = \left\{ g(\frac{a^\tau(x_1-x_2)}{\sqrt{a^\tau\Sigma a}}) : ||a|| = 1, \quad \Sigma \quad \text{is any } p\times p \quad \text{positive matrix} \right\},$$
$$P_o\mathcal{F} = \left\{ \int g(\frac{a^\tau(x_1-x)}{\sqrt{a^\tau\Sigma a}})dP_o(x_1) : ||a|| = 1, \quad \Sigma \text{ is any} \quad p\times p \quad \text{positive matrix} \right\}.$$

Denote by G_p the centered Gaussian process indexed by $L^2(R^p)$ with covariance

$$EG_p(h_1)G_p(h_2) = P_o h_1 h_2 - P_o h_1 P_o h_2, \quad h_1, h_2 \in L^2(R^p).$$

Denote by W_l the centered Gaussian process with covariance

$$EW_l(a_1)W_l(a_2) = P_o l(a_1^\tau x)l(a_2^\tau x) - P_o l(a_1^\tau x)P_o l(a_2^\tau x)$$

where l is a univariate function such that $P_{m_o}l^2(z) < \infty$.

Set

$$U_2^n(f) = \frac{2}{n(n-1)}\sum_{i<j} f(x_i, x_j), \quad f \in \mathcal{F}.$$

The following technical conditions are needed in the next theorem:

(G8): $P_o \times P_o g(a^\tau(x_1-x_2)/s)$ has continuous derivative $u(a,s)$ in s and $\min_{||a||=1}|u(a,1)| > 0$. As $||a-b|| + trace(\Sigma_1 - \Sigma_2) \to 0$,

$$P_o \times P_o \left\{ g(\frac{a^\tau(x_1-x_2)}{\sqrt{a^\tau\Sigma_1 a}}) - g(\frac{b^\tau(x_1-x_2)}{\sqrt{b^\tau\Sigma_1 b}}) \right\}^2 \to 0.$$

(L1): Under P_o,

$$\mu(F_n^a) = \frac{1}{n}\sum_{i=1}^n l(a^\tau x_i) + o_p(n^{-1/2}).$$

Theorem 3.3 *(i) Assume that g is bounded and nondecreasing and satisfies Conditions* **(G3)** *and* **(G8)**. *Then*

$$\hat{\Sigma}(g, X) = I_p + O_p(n^{-1/2}),$$

and

$$\sqrt{n}(\hat{\Sigma}(g, X) - I_p) \xrightarrow{\mathcal{L}} argmin_{\Delta} \max_{||a||=1} |G_p(\int g(a^\tau(x_1 - \cdot))dP_o(x_1)) + u(a, 1)a^\tau \Delta a/2|$$

provided the argmin is unique with probability 1, *where* Δ *runs over all* $p \times p$ *positive definite matrices.*

(ii) Assume that Condition **(L1)** *holds and that g is bounded and nondecreasing and satisfies Condition* **(G7)**. *Then*

$$\hat{\Sigma}(\mu, g, X) = I_p + O_p(n^{-1/2})$$

and

$$\sqrt{n}(\hat{\Sigma}(\mu, g, X) - I_p) \xrightarrow{\mathcal{L}} argmin_{\Delta} \max_{||a||=1} |W(g, a) + \Psi_1^a(0, 1)^\tau W_l(a) + \Psi_2^a(0, 1)a^\tau \Delta a/2|$$

provided that the argmin is unique with probability 1, *where* Δ *runs over all* $p \times p$ *positive definite matrices and* $W(g, a)$ *is the process defined in Theorem 3.2.*

Remark 3.2 *The assumption that the* argmin *is unique in the above theorems seems reasonable but difficult to check. He and Portnoy (1998) provided a way to check such assumption.*

3.2 Breakdown behavior

Let $\mathcal{H}$ be the set of all distributions on R^p. Recall that $\theta(\cdot)$, $\Sigma_\mu(\cdot)$ and $\Sigma(\cdot)$ are the population versions of $\hat{\theta}$, $\hat{\Sigma}_\mu$ and $\hat{\Sigma}$ presented in Section 2. Then the asymptotic breakdown point of location estimator $\hat{\theta}$ at the assumed distribution F_o is

$$\varepsilon(\theta(F_o)) = \inf\{\varepsilon \geq 0 : \sup_{H \in \mathcal{H}} |\theta((1 - \varepsilon)F_o + \varepsilon H)| = \infty\},$$

and the asymptotic breakdown point of dispersion estimator $\hat{\Sigma}_\mu$ at the assumed distribution F_o is

$$\begin{aligned}\varepsilon(\Sigma_\mu(F_o)) \quad = \quad & \inf\{\varepsilon \geq 0 : \inf_{H \in \mathcal{H}} \lambda_{min}(\Sigma_\mu((1 - \varepsilon)F_o + \varepsilon H)) = 0, \\ & \sup_{H \in \mathcal{H}} \lambda_{max}(\Sigma_\mu((1 - \varepsilon)F_o + \varepsilon H) = \infty\}\end{aligned}$$

where $\lambda_{min}(\cdot)$ and $\lambda_{max}(\cdot)$ stand for the minimum and maximum eigenvalues. Similarly, we define $\varepsilon(\Sigma(F_o))$ for the dispersion estimator $\hat{\Sigma}$. For the rest of this section we restrict ourselves to the case when $F_o = F_o(\cdot; \theta, \Sigma)$ is an elliptic distribution with unknown parameters θ and Σ. Let P_o denote the probability measure induced by F_o. Note that Lopuhaä and Rousseeuw (1991) showed that the highest asymptotic breakdown point that the affine equivariant estimators of location and dispersion can attain is 1/2.

Theorem 3.4 *Suppose that g is bounded, nondecreasing and odd. Assume that for any* $0 \leq \varepsilon < 1/3$,

$$0 < \inf_{||a||=1, H} s((1 - \varepsilon)F_o^a + \varepsilon H^a) \leq \sup_{||a||=1, H} s((1 - \varepsilon)F_o^a + \varepsilon H^a) < \infty.$$

Then $\varepsilon(\theta(F_o)) \geq 1/3$.

Remark 3.3 *Theorem 3.4 can be extended to the case when* g *is bounded and nondecreasing but not odd. For this pourpose, for* $0 \le \varepsilon < 1/2$, *we define*

$$\begin{aligned} C(\varepsilon) &= \sup_{H\in\mathcal{H}} \min_{\theta} \max_{||a||=1} |P_o g(a^\tau(x-\theta)/s((1-\varepsilon)F_o^a+\varepsilon H^a))|, \\ g_u &= \max\{g(\infty), -g(-\infty)\}, \quad g_l = \min\{g(\infty), -g(-\infty)\}. \end{aligned}$$

Then

$$\varepsilon(\theta(F_o)) \ge \frac{g_u - C(\varepsilon(\theta(F_o)))}{2g_u + g_l - C(\varepsilon(\theta(F_o)))}.$$

For example, if $g(z) = -\alpha I_{[z\le 0]} + (1-\alpha)I_{[z>0]}$, *then under the same condition of Theorem 3.4,* $\varepsilon(\theta(F_o)) \ge 1/3$.

Remark 3.4 *Chen (1995) and He and Wang (1997) proved that the asymptotic breakdown point of Tukey's median is* $1/3$. *So Theorem 3.4 implies that the asymptotic breakdown points of our new location estimators are not lower than that of Tukey's median. Generally it remains to see whether our extensions can improve the breakdown property of Tukey's median. However, the next theorem indicates that the answer is positive if* g *satisfies Condition* **(B)** *and with a suitably chosen tuning constant. See Zhang and Li (1998) for several commonly used* g*-functions.*

For $0 < \varepsilon < 1/2$, define

$$\begin{aligned} s_*(\varepsilon) &= \inf\{s((1-\varepsilon)F_o^a+\varepsilon H^a)) : ||a||=1, \quad H\in\mathcal{H}\}, \\ A_*(\varepsilon, a) &= \int [1-g(a^\tau x/s_*(\varepsilon))]dF_o(x), \quad A_*(\varepsilon) = \min_{||a||=1} A_*(\varepsilon, a). \end{aligned}$$

Theorem 3.5 *Suppose that* g *satisfies Condition* **(B)**. *Then*

$$\varepsilon(\theta(F_o)) \ge \sup\{\varepsilon \ge 0 : \varepsilon < A_*(\varepsilon)/[1+A_*(\varepsilon)]\}.$$

Remark 3.5 *It follows from Zhang and Li (1998, p. 1177) that if* g *is even with a derivative function* $\psi(x)$ *satisfying*

$$\left(\int g(z-t)dF_{mo}(z)\right)' = -\int \psi(z-t)dF_{mo}(z), \quad \int |\psi(z-t)|dF_{mo}(z) < \infty;$$

$\psi(z) \ge 0$ *for* $z \ge 0$, *and* F_o *is spherical with density* $f_o(||x-\theta||^2)$ *and* $f_o(z)$ *is even, strictly decreasing for* $z \ge 0$, *then*

$$A_*(\varepsilon) = \sup_t \int \{1 - g((z-t)/s_*(\varepsilon))\}\, dF_{mo}(z).$$

For the fixed F_{mo}, *we can adjust the tuning constant in* g *(see Zhang and Li, 1998, p. 1179) so that* $\varepsilon(\theta(F_o))$ *is close to* $1/2$.

We now turn to the breakdown behavior of the dispersion estimator $\hat{\Sigma}$.

Theorem 3.6 *Assume that* $g(z)$ *is nonincreasing for* $z < 0$ *and nondecreasing for* $z \ge 0$, *and* $0 < g(\infty) = g(-\infty) < \infty$. *Assume that*

$$\int g(z_1 - z_2)dP_{mo}(z_1)dP_{mo}(z_2) = 0.$$

Then

$$\varepsilon(\Sigma(F_o)) \ge 1 - \sqrt{\frac{1+c_o}{1+2c_o}} > 0$$

where $c_o = \min\{|g(0)|, g(\infty)\}/\max\{|g(0)|, g(\infty)\}$. *Especially, when* $|g(0)| = g(\infty)$, $\varepsilon(\Sigma(F_o)) \ge 0.1835$.

Remark 3.6 *For $\hat{\Sigma}_\mu$, if g is nonincreasing for $z < 0$ and nondecreasing for $z \geq 0$, and $0 < g(\infty) = g(-\infty) < \infty$, then*

$$\varepsilon(\Sigma_\mu(F_o)) \geq \min\{\varepsilon \geq 0 : B(\varepsilon) \geq |(1-\varepsilon)\min\{|g(0)|, g(\infty)\} - \varepsilon \max\{|g(0)|, g(\infty)\}|\}$$

with

$$B(\varepsilon) = \sup_{H \in \mathcal{H}} \max_{||a||=1} |\int g(a^\tau x - \mu(F_\varepsilon^a))dF_\varepsilon(x)|, \quad F_\varepsilon = (1-\varepsilon)F_o + \varepsilon H.$$

Furthermore if $P_{mo}g(z) = 0$ and location is known, then

$$\varepsilon(\Sigma_\mu(F_o)) \geq \frac{\min\{|g(0)|, g(\infty)\}}{\max\{|g(0)|, g(\infty)\} + |g(0)| + g(\infty)}.$$

In particular, if $|g(0)| = g(\infty)$, then $\varepsilon(\Sigma_\mu(F_o)) \geq 1/3$.

4 A simulation study

In this section we undertake an extensive simulation study to assess the performance of the deepest points of our generalized Tukey depth functions in the bivariate location setting. The computation seems awkward because of the double optimization and of many local optimal points when g is not smooth. We first present what is called the thresholding algorithm to reduce the double optimization to a few of single optimizations. For simplicity, we focus on the bivariate location case with

$$g(z) = \begin{cases} z, & |z| \leq c; \\ c, & z > c; \\ -c, & z < -c. \end{cases} \tag{4.1}$$

Recall that $\hat{\theta}$ is any element of set $\{argmin_\theta O_{gn}(\theta)\}$. Assume that we devise a sieve $\Theta_n = \{\theta_i, 1 \leq i \leq N_1\}$ and use $argmin_{\theta \in \Theta_n} O_{gn}(\theta)$ to approximate $argmin_\theta O_{gn}(\theta)$. We take the coordinate-wise median θ_{co} as the initial vector. Note that by the expression (2.1), $O_{gn}(\theta_{co}) = \max_{||a||=1} |\frac{1}{n}\sum_{i=1}^n g(z_i(a, \theta_{co}))|$. So we can calculate $O_{gn}(\theta_{co})$ by the direct approximation:

$$\max_{a \in U_o} |\frac{1}{n}\sum_{i=1}^n g(z_i(a, \theta_{co}))|$$

where U_o is a set of the grid points of $\{a : ||a|| = 1\}$, and can be made via the expression $a = (\cos\phi, \sin\phi), \phi \in [-\pi, \pi]$ and the grid points of $[-\pi, \pi]$. Then we choose the smallest among $O_{gn}(\theta), \theta \in \Theta_n$ and $O_{gn}(\theta_{co})$. The basic idea here is to avoid full evaluation of depth function for all $\theta \in \Theta_n$. For instance, we don't need to evalate the depth function at θ_1 fully if we find some $a, ||a|| = 1$ such that

$$|\frac{1}{n}\sum_{i=1}^n g(z_i(a, \theta_1))| > O_{gn}(\theta_{co})$$

which implies $O_{gn}(\theta)$ can not attain the minimum at θ_1. It turns out that a large number of candidates in Θ_n can be filtered out by calculating $|\frac{1}{n}\sum_{i=1}^n g(z_i(a, \theta))|$ only for a in some nested finite subsets of $\{a : ||a|| = 1\}$. We note that a similar idea was used independently by He (1999) for the regression depth.

In summary, the **thresholding algorithm** consists of the following steps:

1. Calculate O_{gn} at the coordinate-wise median θ_{co}. Set $\theta_f = \theta_{co}$. Calculate $O_{gn}(\theta_{co})$ by the direct optimization mentioned before.

2. Choose three nested finite subsets of $\{a : ||a|| = 1\}$: $U_1 \subset U_2 \subset U_3$ of sizes m_1, m_2 and m_3, respectively. Our numerical experience indicates that for the bivariate case, we can choose $m_1 = 25$, $m_2 = 324$ and $m_3 = 924$. The first two sets act as "filters" while the last one is used to evaluate the depth approximately for $\theta \in \Theta_n$ which has not been filtered out. For this purpose, set $r = 1$.

3. For θ_r, we first calculate $\frac{1}{n}\sum_{i=1}^n g(z_i(a, \theta_r))$ for $a \in U_1$. Observe that if

$$\max_{a\in U_1} |\frac{1}{n}\sum_{i=1}^n g(z_i(a, \theta_r))| > O_{gn}(\theta_f)$$

then O_{gn} can not reach the minimum at θ_r. In this case we need not calculate the values of O_{gn} on the large set U_3. Set $r \Leftarrow r + 1$ and if $r \leq N_1$, then back to the beginning of Step 3. If $r > N_1$, then go to Step 6.

4. If

$$\max_{a\in U_1} |\frac{1}{n}\sum_{i=1}^n g(z_i(a, \theta_r))| \leq O_{gn}(\theta_f)$$

then we calculate O_{gn} on the set U_2. Similarly if

$$\max_{a\in U_2} |\frac{1}{n}\sum_{i=1}^n g(z_i(a, \theta_r))| > O_{gn}(\theta_f)$$

then O_{gn} can not reach the minimum at θ_r. Set $r \Leftarrow r + 1$ and if $r \leq N_1$, then go to Step 3. If $r > N_1$, then go to Step 6.

5. If

$$\max_{a\in U_2} |\frac{1}{n}\sum_{i=1}^n g(z_i(a, \theta_r))| \leq O_{gn}(\theta_f)$$

we calculate the values of O_{gn} on the set U_3. If

$$\max_{a\in U_3} |\frac{1}{n}\sum_{i=1}^n g(z_i(a, \theta_r))| < O_{gn}(\theta_f)$$

then replace θ_f by θ_r and $O_{gn}(\theta_f)$ by $O_{gn}(\theta_r)$. Set $r \Leftarrow r + 1$ and if $r \leq N_1$, then go to Step 3. If $r > N_1$ then go to Step 6.

6. Let $\theta_{fi}, i = 1, \cdots, p$ be the components of θ_f. Choose a sieve of the interval $[\theta_f - \delta, \theta_f + \delta] = [\theta_{f1} - \delta, \theta_{f1} + \delta] \times \cdots \times [\theta_{fp} - \delta, \theta_{fp} + \delta]$ ($\delta = 0.01$ in our code) and repeat Steps 3 to 5 but the sentence " If $r > N_1$, then go to Step 6" is replaced "If $r > N_1$, then go to Step 7."

7. Take the current θ_f as an approximation of $\hat{\theta}$.

The above algorithm can be further refined. The code is available from the author.

The similated annealing algorithm of Vetterling, Teukolsky, Press and Flannery (1992) is also applied to calculate $\hat{\theta}$. This algorithm could be faster than the above thresholding algorithm when O_{gn} has many local minimum points, whereas the idea of thresholding is safer. In our code, we run the subroutine AMEBSA of Vetterling, Teukolsky, Press and Flannery (1992) with temperature schedule: 0.1, 0.01, 0.01/11, 0.01/21. At each temperature we run AMEBSA 20 times.

We now use these algorithms to simulate the mean squared errors of $\hat{\theta}(g, X)$. For the sample sizes $n = 30$ and 60, we respectively generate $m = 1000$ samples from the bivariate standard normal distribution, and apply the thresholding algorithm and simulated annealing algorithm to each sample. From the m estimators $\hat{\theta}_1, \cdots, \hat{\theta}_m$ we compute the empirical mean squared error:

$$ERR(c) = \frac{1}{m} \sum_{i=1}^{m} ||\hat{\theta}_i||^2.$$

The results are shown in Table 1. In it, $ERR(c)_t$ and $ERR(c)_s$, respectively, stand for the corresponding empirical mean squared errors of $\hat{\theta}$ when the thresholding algorithm and the simulated annealing algorithm are applied. Comparing the values of $ERR(c)_t$ and $ERR(c)_s$ in Table 1, we see that in general the above temperature schedule is suitable for the cases $n = 60$, and $n = 30$.

At the same time, we calculate the empirical mean squared errors of the sample mean $\overline{x}$, coordinate-wise median θ_{co}, Donoho's depth based deepest point θ_D, the deepest point θ_{tr} of the generalized Tukey depth based the trimmed function with $\alpha = 1/3$. The results are presented in Table 2.

From Tables 1 and 2, we see that the empirical efficiency of our new estimators can be significantly higher than those of the deepest points of Tukey's depth and Donoho's depth. For example, for simple size 60 and $c = 1.2$, $ERR(1.2)_t/ERR_T = 0.80$ and $ERR(1.2)_t/ERR_D = 0.80$ where ERR_T and ERR_D denote the empirical mean squared errors of the deepest points of Tukey's depth and Donoho's depth, respectively.

To improve the efficiency of the estimator based on Donoho's depth, we combine Donoho's depth with the generalized Tukey depth as pointed out in Section 1:

$$D_{dT} = \{1 + O_{dn} + wO_{gn}\}^{-1}$$

where w is a positive constant and g is defined in (4.1). Here we choose $w = 5$.

Applying Tyler's technique, we can easily show that the asymptotic breakdown point of the deepest point $\theta(w, c)$ of D_{dT} is 1/2 under symmetry. We also simulated the mean squared errors for the sample sizes 60 and 30 with $m = 1000$. The results are presented in Table 2.

Table 1. The empirical mean squared errors of $\hat{\theta}$ based on Huber's function with tuning constant c.

$n = 60, m = 1000$						
c	0.0000001	0.00001	0.5	1.2	2.	3.
$ERR(c)_t$	0.0454	0.0453	0.0382	0.0341	0.0329	0.0327
$ERR(c)_s$	0.0457	0.0441	0.0403	0.0355	0.0343	0.0336
$n = 30, m = 1000$						
$ERR(c)_t$	0.0958	0.0956	0.0775	0.0680	0.0653	0.0646
$ERR(c)_s$	0.0958	0.0885	0.0850	0.0699	0.0670	0.0681

Table 2. The empirical mean squared errors of the sample mean, coordinate-wise median, deepest points based on Donoho's depth and $\hat{\theta}$ based on the trimmed with $\alpha = 1/3$.

$n = 60, m = 1000$						
	$\overline{x}$	θ_D	θ_{tr}	θ_{co}	$\theta(5, 1.2)$	$\theta(5, 2.)$
ERR	0.0326	0.0421	0.0388	0.0491	0.0348	0.0332
$n = 30, m = 1000$						
ERR	0.0645	0.0985	0.0791	0.0984	0.070	0.066

5 Relation with empirical likelihood

If we know the parametric likelihood, we would prefer ordering the data set by the likelihood contours. When it is unknown, we often construct a nonparametric likelihood, for example, empirical likelihood, by using the auxiliary information (see Owen, 1988). In another word, the likelihood can be applied to construct some depth function. But the depth functions are often introduced in an adhoc way. So, as pointed out in Section 1, we hope to check whether some depth contours can be derived from a nonparametric likelihood. In the following we show that it is true for Tukey's depth contours. In fact, we find that Tukey's depth function is equivalent to a projection based empirical likelihood in the sense that there exists a strictly increasing transformation between them.

To begin with, we construct an empirical likelihood ratio $l(a, \theta)$ of $a^\tau \theta$ for each direction a as follows. Consider

$$\max \sum_{i=1}^{n} \log p_i$$

subject to

$$\begin{aligned} &\sum_{i=1}^{n} p_i = 1, \quad p_i \geq 0, \quad i \geq 1; \\ &\sum_{i=1}^{n} p_i I_{[a^\tau (x_i - \theta) \leq 0]} = 0. \end{aligned}$$

Let $p_i(a, \theta)$, $i = 1, 2, \cdots, n$ be the solution. Then

$$\begin{aligned} l(a, \theta) &= \sum_{i=1}^{n} \log p_i(a, \theta) + n \log n \\ &= -nR(F_n^a(a^\tau \theta)) \end{aligned}$$

where

$$R(z) = \log 2 + z \log z + (1 - z) \log(1 - z), \quad 0 < z < 1.$$

The projection based empirical likelihood (the least favorable empirical likelihood among all empirical likelihoods of one-dimensional projections of X), namely $\min_{||a||=1} l(a, \theta)$, is equal to $-nR(D_{rn}(\theta)/n)$ where D_{rn} is Tukey's depth function. Observe that $R(z)$ is strictly decreasing. We have the following proposition.

Proposition 5.1 *Tukey's depth function is equivalent to the projection based empirical likelihood.*

Analogously, we can show that the projection based depth function of α-quantile in Section 2 is equivalent to the corresponding projection based empirical likelihood. In this setting,

$$R(z) = z \log(z/(1 - \alpha)) + (1 - z) \log((1 - z)/\alpha).$$

However, in general the generalized Tukey depth functions are different from the semiparametric likelihood based on the corresponding estimation equations.

6 Technical proofs

Lemma 6.1 (Lemma 3.1 in Zhang and Li, 1993). Let Δ be a compact subset of a metric space with metric d. Let P_n be the empirical distribution of a probability distribution P. For each $t \in \Delta$, $V_1(t) = V_1(t, P)$

denotes the distribution functional of P and $V_{1n}(t) = V_{1n}(t, P_n)$ stands for the distribution functional of P_n. Suppose that for the fixed P, $V_1(t, P)$ is continuous in $t \in \Delta$. Suppose that $\sup_{t\in\Delta} V_{1n}(t)$ is measurable. Set $B_1 = \{t \in \Delta : V_1(t) = \sup_{s\in\Delta} V_1(s)\}$ and $S_{1n}(t) = \sqrt{n}(V_{1n}(t) - V_1(t)), t \in \Delta, n \geq 1$. If there is a process $\{S_1(t) : t \in \Delta\}$ with continuous sample paths such that

$$\sup_{t\in\Delta} |S_{1n}(t) - S_1(t)| \to 0, \text{a.s.},$$

then

$$\sqrt{n}(\sup_{t\in\Delta} V_{1n}(t) - \sup_{t\in\Delta} V_1(t)) \to \sup_{t\in B_1} S_1(t), \text{a.s.}.$$

Proof of Proposition 2.1. Set

$$s_{min} = \min_{||a||=1} s(F^a), \quad s_{max} = \max_{||a||=1} s(F^a).$$

Let Ω_1 be a compact subset of R^p bounded by constant c_o. For $0 < \delta < 1$, consider the empirical process

$$\{(P_n - P)h : \quad h \in \mathcal{F}_1\}$$

with

$$\mathcal{F}_1 = \{g(a^\tau(x-\theta)/s) : ||a|| = 1, a \in R^p, \theta \in \Omega_1, (1-\delta)s_{min} \leq s \leq (1+\delta)s_{max}\}$$

and with envelope $F_1(x) = |g((||x|| + c_o)/s_{min})|$. We see that the graphs (or subgraphs) of functions in $\mathcal{F}_1$ form a polynomial class (or a VC subgraph class) of sets (see Pollard, 1984, p. 17 for the definition). By Theorem 24 and Lemma 25 of Pollard (1984, p.25 and p.27), we deduce that as $n \to \infty$

$$\max_{\theta\in\Omega_1} |O_{gn}(\theta) - O_g(\theta)| \leq \max\{|(P_n - P)h| : h \in \mathcal{F}_1\} \to 0, \quad a.s..$$

Therefore

$$\max_{\theta\in\Omega_1} |D_{gn}(\theta) - D_g(\theta)| \to 0, \quad a.s.$$

Similarly we obtain

$$\{\sqrt{n}(P_n - P)h : h \in \mathcal{F}_1\} \xrightarrow{\mathcal{L}} \{W(h) : h \in \mathcal{F}_1\}$$

where W is a centered Gaussian process with covariance

$$EW(h_1)W(h_2) = Ph_1h_2 - Ph_1Ph_2.$$

Note that when g is bounded, the same result holds if we let $\Omega_1 = R^p$.

When $\max_{||a||=1} |Ph_a| = 0$, the asymptotic distribution of D_{gn} can follow directly from the functional central limit theorem of empirical processes. It remains to consider the case when $\max_{||a||=1} |Ph_a| > 0$. To this end, we first observe that the above Gaussian process has continuous sample paths in a, $||a|| = 1$ almost surely because of Condition **(G0')**. By the representation theorem of random elements (Pollard, 1984), for each fixed θ, there exist two processes $\{S_n(a) : ||a|| = 1\}$ and $\{S(a) : ||a|| = 1\}$ which follow the same joint distributions as those of $\{\sqrt{n}(P_n - P)h_a : ||a|| = 1\}$ and $\{W(h_a) : ||a|| = 1\}$, and satisfy

$$\max_{||a||=1} |S_n(a) - S(a)| \to 0, \quad a.s..$$

Set $V_n(a) = S_n(a)/\sqrt{n} + V(a), \quad V(a) = Ph_a, \quad a \in R^p, ||a|| = 1$. Without loss of generality, we assume that $A_+ \neq \emptyset$ and $A_- \neq \emptyset$. Letting $V_{1n} = V_n$, $V_1 = V$, $S_{1n} = S_n$, $S_1 = S$, $\Delta = A_+$, and $B_1 = A_+$ in Lemma 6.1, we obtain that as $n \to \infty$,

$$\sqrt{n}(\max_{a \in A_+} V_n(a) - \max_{a \in A_+} V(a)) \to \max_{a \in A_+} S(a), \text{a.}s..$$

Similarly, by using Lemma 6.1 we have

$$\sqrt{n}(\max_{a \in A_-}(-V_n(a)) - \max_{a \in A_-}(-V(a))) \to \max_{a \in A_-}(-S(a)), \text{a.}s..$$

Invoking the facts that $\max_{||a||=1} |V_n(a) - V(a)| \to 0$ almost surely, that for $a \in A_+$, $V(a) = \max_{||b||=1} |V(b)| > 0$, and that for $a \in A_-$, $-V(a) = \max_{||b||=1} |V(b)| > 0$, we have

$$\max_{a \in A_+} |V_n(a)| = \max_{a \in A_+} V_n(a), \quad \max_{a \in A_-} |V_n(a)| = \max_{a \in A_-}(-V_n(a)) \text{a.}s.$$

as $n \to \infty$. Thus, we have

$$\begin{aligned} &\sqrt{n}\left(\max_{||a||=1} |V_n(a)| - \max_{||a||=1} |V(a)|\right) \\ &= \sqrt{n} \max\left\{\max_{a \in A_+} V_n(a) - \max_{A_+} V(a), \quad \max_{a \in A_-}(-V_n(a)) - \max_{a \in A_-}(-V(a))\right\} \\ &\to \max\left\{\max_{a \in A_+} S(a), \quad -\min_{a \in A_-} S(a)\right\}, \quad a.s. \end{aligned}$$

which leads to

$$\sqrt{n}(O_{gn}(\theta) - O_g(\theta)) \xrightarrow{\mathcal{L}} \max\{\max_{a \in A_+} W(h_a), -\min_{a \in A_-} W(h_a)\}.$$

Now the results follow immediately from the definitions of D_{gn} and D_g.

Proof of Corollary 2.1. It is similar to the proof of Lemma 2.5 of Donoho and Gasko (1992).

Proof of Proposition 2.2. It is similar to the proof of Proposition 2.1 and relies on the result of Arcones and Giné (1993).

Proof of Theorem 3.1. It is a direct result of Propositions 2.1 and 2.2.

Proof of Remark 3.1. The first part is similar to Zhang and Li (1998) and thus omitted. To prove the second part, we denote by f_{mo} the density of F_{mo} and define

$$K(t) = \int_{-\infty}^{t} |\psi(z - t)| dF_{mo}(z).$$

When $t > 0$, it follows from the assumption on ψ that

$$\begin{aligned} K(t) &= \int_{-t}^{\infty} \psi(z + t) f_{mo}(z) dz \\ &> \int_{0}^{\infty} \psi(z) f_{mo}(z) dz = K(0). \end{aligned}$$

For $t_1 < t_2 \leq 0$, it follows from the assumption on f_{mo} that

$$\begin{aligned} K(t_1) &= \int_0^\infty \psi(z) f_{mo}(z - t_1) dz \\ &< \int_0^\infty \psi(z) f_{mo}(z - t_2) dz = K(t_2). \end{aligned}$$

Consequently, for any fixed $s > 0$,

$$\begin{aligned} &\max_{||a||=1} \int \left(\psi(\infty) - |\psi(a^\tau(x-\theta)/s)|\right) I_{[a^\tau(x-\theta)\leq 0]} dF_o(x) \\ &= \psi(\infty) - \min_{||a||=1} K(a^\tau\theta/s) \\ &= \psi(\infty) - K(-||\theta||/s) \end{aligned}$$

which attains the minimum only at $\theta = 0$. The proof is completed.

Proof of Theorem 3.2. To begin with, we consider the empirical process

$$\{\sqrt{n}(P_n - P_o)h : h \in \mathcal{F}_3\}$$

with

$$\mathcal{F}_3 = \{g(a^\tau(x-\theta)/s) : ||a|| = 1, \quad ||\theta|| \leq c, \quad 1 - \delta \leq s \leq 1 + \delta\}.$$

Set

$$W_n(a, \theta, s) = \sqrt{n}(P_n - P_o) g(a^\tau(x-\theta)/s).$$

Without loss of generality, we assume that P_o is the underlying distribution with zero location and unit dispersion. Note that the graphs (or subgraphs) of functions in $\mathcal{F}_3$ form a polynomial class of sets (or a VC-subgraph class) (see Pollard, 1984, p.17). Then it follows from the theorem in Pollard (1984) that the above empirical process is stochastic equicontinuous, which implies that for any $\theta_n = o_p(1)$,

$$\begin{aligned} &\max_{||a||=1} \sqrt{n}|P_n g(a^\tau(x-\theta_n)/s(F_n^a))| \\ &= \max_{||a||=1} |W_n(a, \theta_n, s(F_n^a)) + [\Psi_1^a(0,1)^\tau \sqrt{n}\theta_n + \Psi_2^a(0,1)\sqrt{n}(s(F_n^a) - 1)](1 + o_p(1))| \\ &= \max_{||a||=1} |\sqrt{n}(P_n - P_o)g(a^\tau x) + o_p(1) + [\Psi_1^a(0,1)^\tau \sqrt{n}\theta_n + \Psi_2^a(0,1)\sqrt{n}(s(F_n^a) - 1)](1 + o_p(1))| \end{aligned} \quad (6.1)$$

Since we have shown that $\hat{\theta} = o_p(1)$ in Theorem 3.1, by comparing the values of $\max_{||a||=1} \sqrt{n}|P_n g(a^\tau(x - \theta)/s(F_n^a))|$ at the points $\hat{\theta}$ and 0 and by using (6.1), we obtain

$$\begin{aligned} &\max_{||a||=1} |\sqrt{n}(P_n - P_o)g(a^\tau x) + o_p(1) + \Psi_1^a(0,1)^\tau \sqrt{n}\hat{\theta}(1 + o_p(1)) \\ &\quad + \Psi_2^a(0,1)\sqrt{n}(s(F_n^a) - 1)(1 + o_p(1))| \\ &\quad \leq \max_{||a||=1} |\sqrt{n}(P_n - P_o)g(a^\tau x) + o_p(1) + \Psi_2^a(0,1)\sqrt{n}(s(F_n^a) - 1)(1 + o_p(1))| \\ &= 0_p(1). \end{aligned}$$

This together with the assumptions on $\Psi_i^a(0,1), i = 1, 2$ yields that for some positive constant c_o, and the large n,

$$\begin{aligned} &\min_{||a||=1} |d(a)|||\sqrt{n}\hat{\theta}|| \leq c_o \max_{||a||=1} |d(a) a^\tau \sqrt{n}\hat{\theta}(1 + o_p(1))| \\ &\leq \max_{||a||=1} |\sqrt{n}(P_n - P_o)g(a^\tau x) + o_p(1) + \Psi_2^a(0,1)\sqrt{n}(s(F_n^a) - 1)(1 + o_p(1))| + 0_p(1) \end{aligned}$$

which implies that

$$||\sqrt{n}\hat{\theta}|| = 0_p(1). \tag{6.2}$$

To prove the second part of the theorem, we define $\theta_n(\xi) = \xi/\sqrt{n}$ and for any compact $\Theta_o \subset R^p$, consider the process

$$\{\sqrt{n}P_n g(a^\tau(x - \theta_n(\xi))/s(F_n^a)) : ||a|| = 1, \quad \xi \in \Theta_o\}.$$

Similar to the argument of (6.1), we show that the above empirical process converges weakly to the following Gaussian process

$$\{W(h) + \Psi_1^a(0,1)^\tau \xi + \Psi_2^a(0,1)W_s(a) : h(x) = g(a^\tau x), \quad ||a|| = 1, \quad \xi \in \Theta_o\}$$

where W is a centered Gausssian process indexed by $\mathcal{F}_3$ with covariance

$$EW(h_1)W(h_2) = P_o h_1 h_2 - P_o h_1 P_o h_2.$$

Now for $\xi \in R^p$, we define

$$\begin{aligned} Z_{1n}(\xi) &= \max_{||a||=1} \sqrt{n}|P_n g(a^\tau(x - \theta_n(\xi))/s(F_n^a)|, \\ Z_1(\xi) &= \max_{||a||=1} \{|W(h) + \Psi_1(0,1)^\tau \xi + \Psi_2(0,1)W_s(a)| : \quad h(x) = g(a^\tau x), \quad ||a|| = 1\}. \end{aligned}$$

Then, for any compact $\Theta_o \subset R^p$,

$$\{Z_{1n}(\xi) : \xi \in \Theta_o\} \xrightarrow{\mathcal{L}} \{Z_1(\xi) : \xi \in \Theta_o\} \tag{6.3}$$

by virtue of the continuous mapping theorem (see Pollard, 1984). Invoking Theorem 2.3 in Kim and Pollard (1990), we have

$$\{Z_{1n}(\xi) : \xi \in R^p\} \xrightarrow{\mathcal{L}} \{Z_1(\xi) : \xi \in R^p\} \tag{6.4}$$

Note that by (6.2),

$$\begin{aligned} O_p(1) &= \sqrt{n} argmin_{\theta \in R^p} \max_{||a||=1} |P_n g(a^\tau(x - \theta)/s(F_n^a))| \\ &= argmin_{\xi \in R^p} Z_{1n}(\xi). \end{aligned} \tag{6.5}$$

Combining this with (6.3), we easily show that

$$O_p(1) \quad \geq \quad arg \min_{\xi \in R^p} Z_1(\xi).$$

By the assumption $argmin_{\xi \in R^p} Z_1(\xi)$ is uniquely defined. We let $Z_n = Z_{1n}$, $t_n = argmin_{\xi \in R^p} Z_{1n}(\xi)$ and $\alpha_n = 0$ in Theorem 2.7 of Kim and Pollard (1990). Now the proof is completed by the direct application of that theorem, since the conditions in that theorem hold by (6.4) and (6.5).

Proof of Theorem 3.4. First we recall that F_o is an elliptic distribution and $s(\cdot)$ is scale equivariant by the assumption. So for $0 \leq \varepsilon < 1/3$, $||a|| = 1$,

$$P_o g(-a^\tau(x - \theta)/s((1 - \varepsilon)F_o^{-a} + \varepsilon H^{-a})) = P_o g(a^\tau(x - \theta)/s((1 - \varepsilon)F_o^a + \varepsilon H^a)).$$

By the assumption that g is odd,

$$P_o g(-a^\tau(x-\theta)/s((1-\varepsilon)F_o^{-a}+\varepsilon H^{-a})) = -P_o g(a^\tau(x-\theta)/s((1-\varepsilon)F_o^a+\varepsilon H^a)).$$

Thus,

$$P_o g(a^\tau(x-\theta)/s((1-\varepsilon)F_o^a+\varepsilon H^a)) = 0$$

which implies

$$\sup_{H\in\mathcal{H}} \min_{\Theta} \max_{||a||=1} |[(1-\varepsilon)P_o+\varepsilon H]g(a^\tau(x-\theta)/s((1-\varepsilon)F_o^a+\varepsilon H^a))| \le \varepsilon g(\infty). \tag{6.6}$$

On the other hand, for each $\varepsilon > 0$, if there exists $\{H_n\}$ such that the minimizer θ_n of

$$\max_{||a||=1} |[(1-\varepsilon)P_o+\varepsilon H_n]g(a^\tau(x-\theta)/s((1-\varepsilon)F_o^a+\varepsilon H_n^a))|$$

tends to ∞, then we show below that

$$\liminf \max_{||a||=1} |[(1-\varepsilon)P_o+\varepsilon H_n]g(a^\tau(x-\theta_n)/s((1-\varepsilon)F_o^a+\varepsilon H_n^a))| \ge (1-2\varepsilon)g(\infty). \tag{6.7}$$

To this end, for $r > 0$, set

$$B_{rn} = \{x\in R^p : |\theta_n^\tau x/||\theta_n|| - ||\theta_n||| \ge r\}.$$

For any $\delta > 0$ and $\delta_1 = \delta/(2(1-\varepsilon))$, choose r such that

$$g(r/s^*) \ge g(\infty) - \delta_1$$

where

$$s^* = \max_{||a||=1} \sup_{H\in\mathcal{H}} s((1-\varepsilon)F_o^a+\varepsilon H^a). \tag{6.8}$$

Then there exists $N(r,\delta)$, when $n \ge N(r,\delta)$,

$$P_o(B_{rn}^c) \le \delta/\left\{4(1-\varepsilon)g(\infty)\right\}.$$

Consequently, (6.7) follows from the following arguments:

$$\begin{aligned}
&\max_{||a||=1} |[(1-\varepsilon)P_o+\varepsilon H_n]g(a^\tau(x-\theta_n)/s((1-\varepsilon)F_o^a+\varepsilon H^a))| \\
&\quad\ge (1-\varepsilon)g(r/s^*)P_o(B_{rn}) - (1-\varepsilon)g(\infty)P_o(B_{rn}^c) - \varepsilon g(\infty) \\
&\quad\ge (1-\varepsilon)(g(\infty)-\delta_1)\left(1-\delta/(4(1-\varepsilon)g(\infty))\right) \\
&\qquad -(1-\varepsilon)g(\infty)\delta/\left\{4(1-\varepsilon)g(\infty)\right\} - \varepsilon g(\infty) \\
&\quad\ge (1-2\varepsilon)g(\infty) - \delta.
\end{aligned}$$

Combining (6.6) and (6.7), we have

$$\varepsilon g(\infty) \ge (1-2\varepsilon)g(\infty).$$

The proof is completed.

Proof of Theorem 3.5. The proof is similar to the second part of the proof of Theorem 3.2 in Zhang (1998). It suffices to prove that for any $0 < \varepsilon < 1/2$, $\varepsilon < A_*(\varepsilon)/[1 + A_*(\varepsilon)]$,

$$\varepsilon(\theta(F_o)) \geq \varepsilon.$$

To this end, we observe that for $0 < \varepsilon < A_*(\varepsilon)/[1 + A_*(\varepsilon)]$, there exists $\gamma_1 > 0$ such that

$$0 < \varepsilon < \frac{A_*(\varepsilon) - \gamma_1}{1 + A_*(\varepsilon) - \gamma_1}.$$

Then for any $||a|| = 1$,

$$\varepsilon < \frac{A_*(\varepsilon, a) - \gamma_1}{1 + A_*(\varepsilon, a) - \gamma_1}.$$

We choose $c_1 > 0$, $\delta > 0$ such that $1 - g(z) \leq \delta$ when $|z| > c_1$. Note that $\max_{||a||=1} |a^\tau x| = ||x||$. We can choose a compact subset K satisfying for any $||a|| = 1$,

$$\varepsilon < \frac{A_*(\varepsilon, a) - \gamma_1 - F_o^a(K)\delta - F_o^a(K^c)}{1 + A_*(\varepsilon, a) - \gamma_1 - F_o^a(K)\delta - F_o^a(K^c)} < \frac{A_*(\varepsilon, a)}{1 + A_*(\varepsilon, a)}$$

where $F_o^a(K) = \int_{a^\tau x \in K} dF_o(x)$ and $F_o^a(K^c) = \int_{a^\tau x \notin K} dF_o(x)$. Then, for any $||a|| = 1$,

$$\begin{aligned}
\int \{1 - g((a^\tau x - s)/s(F_\varepsilon^a))\} \, dF_\varepsilon(x) &\leq (1 - \varepsilon)[\delta F_o^a(K) + F_o^a(K^c)] + \varepsilon \\
&< (1 - \varepsilon)[A_*(\varepsilon, a) - \gamma_1] \\
&\leq \int [1 - g(a^\tau x / s_*(\varepsilon))] dF_\varepsilon(x) - (1 - \varepsilon)\gamma_1 \\
&\leq \int [1 - g(a^\tau x / s(F_\varepsilon^a))] dF_\varepsilon(x) - (1 - \varepsilon)\gamma_1
\end{aligned}$$

provided $d_2(s, K) = \inf_{z \in K} |s - z| > c_1 s^*$ (s^* is defined in (6.8)) and $F_\varepsilon = (1 - \varepsilon)F_o + \varepsilon H, H \in \mathcal{H}$. Thus as $d_2(s, K) > c_1 s^*$, we have

$$\begin{aligned}
\min_{||a||=1} \int \{1 - g((a^\tau x - s)/s(F_\varepsilon^a))\} \, dF_\varepsilon(x) &\leq \min_{||a||=1} \int \{1 - g(a^\tau x / s(F_\varepsilon^a))\} \, dF_\varepsilon(x) - (1 - \varepsilon)\gamma_1 \\
&\leq \sup_t \min_{||a||=1} \int \{1 - g((a^\tau x - t)/s(F_\varepsilon^a))\} \, dF_\varepsilon(x) - (1 - \varepsilon)\gamma_1.
\end{aligned}$$

This means that all the solutions of the following minimization problem with respect to t,

$$\max_{||a||=1} \int g((a^\tau x - t)/s(F_\varepsilon^a)) dF_\varepsilon^a(x) = \min!,$$

stay bounded. The proof is completed.

Proof of Theorem 3.6. To prove (ii), we first set for $H \in \mathcal{H}$ and dispersion matrix Σ,

$$T(H, \Sigma) = \max_{||a||=1} | \int \int g\left(a^\tau (x - y)/\sqrt{a^\tau \Sigma a}\right) ((1 - \varepsilon) dP_o(x) + \varepsilon dH(x)) ((1 - \varepsilon) dP_o(y) + \varepsilon dH(y)) |.$$

Note that

$$\begin{aligned}
\sup_{H \in \mathcal{H}} \min_{\Sigma} T(H, \Sigma) \quad &\leq \max_{||a||=1} |P_o^2 g(a^\tau (x - y))| + \sup_{H \in \mathcal{H}} \min_{\Sigma} \max_{||a||=1} | \int \int g(a^\tau (x - y)/\sqrt{a^\tau \Sigma a}) \\
&\quad \times \left(\varepsilon(1 - \varepsilon) dP_o(x) dH(y) + \varepsilon(1 - \varepsilon) dP_o(y) dH(x) + \varepsilon^2 dH(x) dH(y)\right) | \\
&\leq \varepsilon(2 - \varepsilon) \max\{g(\infty), |g(0)|\}. \qquad (6.9)
\end{aligned}$$

Assume that there exist $\{H_i\}$ such that the minimizer Σ_i of $T(H_i, \Sigma)$ is broken down, that is,

$$\max\{\lambda_{max})(\Sigma_i), 1/\lambda_{min}(\Sigma_i)\} \to \infty.$$

Then, analogous to the proof of Theorem 3.4, we obtain

$$\lim T(H_i, \Sigma_i) \geq ((1-\varepsilon)^2 - 2\varepsilon + \varepsilon^2) \min\{g(\infty), |g(0)|\}$$

which together with (6.9) yields

$$\varepsilon(2-\varepsilon) \max\{g(\infty), |g(0)|\} \geq ((1-\varepsilon)^2 - 2\varepsilon + \varepsilon^2) \min\{g(\infty), |g(0)|\}.$$

The desired result follows.

Proof of Remark 3.6. The proof is similar to that of Theorem 3.4 and thus omitted.

Acknowledgements

The work was partly supported by the NNSF of China and the contract "Projet d'Actions de Recherche Concertées" no. 93/98-164. The author would like to thank the referees, Editor and Professors I. Gijbels, L. Simar and Y. Zuo for their very valuable and constructive comments that helped to improve the presentation. The author is especially grateful to Professor Rousseeuw and his assistants for encouragememt and helpful discussions on the topic of depth functions. Some results of the paper were presented in the Annual Meeting of the Chinese Society of Probability and Statistics in 1994.

References

Arcones, M.A. and Giné, E. (1993). Limit theorems for U-processes. *Ann. Prob.* **21,** 1494-1542.

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

Chen, Z. (1995). Robustness of the half-space median. *J. Statist. Plann. Infer.* **46**, 175-181.

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

Eddy, W.F. (1985). Ordering of multivariate data. In *Computer Science and Statistics: the Interface* (L. Billard, ed.), pp. 25-30, North-Holland, Amsterdam.

He, X. (1999). Comment on "Regression Depth" by Rousseeuw and Hubert. *J. Amer. Statist. Assoc.* **94,** 403-404.

He, X. and Wang, G. (1997). Convergence of depth contours for multivariate data sets. *Ann. Statist.* **25**, 495-504.

He, X. and Portnoy, S. (1998). Asymptotics of the deepest line. In *Statistical Inference and Related Topics: A Festschrift in Honor of A.K. Md.E. Saleh,* Nova Science Publications Inc., New York (to appear).

Hettmansperger, T. P. and Sheather, S. J. (1992). A cautionary note on the method of least median squares. *American Statistician.* **46,** 79-83.

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

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Ann. Statist.* **18**, 191-219.

Koltchinskii, V.I. (1997). M-estimation, convexity and quantiles. *Ann. Statist.* **25,** 435-477.

Liu, R.Y. (1990). On a notion of data depth based on random simplices. *Ann. Statist.* **18**, 405-414.

Liu, R.Y. and Singh, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc* **88,** 257-260.

Liu, R.Y., Parelius, J.M. and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference (with discussions). *Ann. Statist.* **27**, 783-858.

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

Maronna, R.A., Stahel, W.A. and Yohai, V. (1992). Bias-robust estimators of multivariate scatter based on projections. *J. Multivariate Anal.* **42**, 141-161.

Nolan, D. (1992). Asymptotics for multivariate trimming. *Stochastic Process. Appl.* **42**, 157-169.

Owen, A.B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75,** 237-249.

Plackett, R.L.(1976). Discussion of Professor Barnett's paper. *J. Roy. Statist. Soc. Ser. A* **139,** 344-346.

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

Reiss, R.D.(1989). *Approximate Distribution of Order Statistics.* Springer, New York.

Rousseeuw, P.J.(1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications,* Vol. B, W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, eds, pp. 283-297, Dordrecht, Reidel.

Rousseeuw, P.J. and Hubert, M. (1999). Regression Depth. *J. Amer. Statist. Assoc.* **94,** 388-402.

Rousseeuw, P.J. and Ruts, I. (1998). Constructing the bivariate Tukey median. *Statistica Sinica* **8**, 827-839.

Small, C. G.(1997). Multidimensional medians arising from geodesics on graphs. *Ann. Statist.* **25**, 478-494.

Stahel, W.A.(1981). Robust estimation: infinitesimal optimality and covariance matrix estimators. Ph.D. thesis, ETH, Zurich.

Tukey, J.W.(1975). Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians* (R.D. James, ed.) **2**, 523-531, Vancouver.

Tyler, D.(1994). Finite sample breakdown points of projection based multivariate location and scatter statistics. *Ann. Statist.* **22**, 1024-1044.

Vetterling, W.T., Teukolsky, S.A., Press, W.H. and Flannery, B.P. (1992). *Numerical Recipes in Fortran 77: the Art of Scientific Computing, Second Edition.* Cambridge University Press, London.

Zhang, J. (1998). Some extensions of Tukey's depth function. *Discussion Paper* **9803,** Institut de Statistique, Universite Catholique de Louvain, Belgium.

Zhang, J. and Li, G. (1993). A new approach to asymptotic distributions of maximum likelihood ratio statistics. In *Statistical Science and Data Analysis: Proceedings of the Third Pacific Area Statistical Conference* (Matusita, K. et al. eds), pp. 325-336, VSP, Amsterdam.

Zhang, J. and Li, G.(1995). Some advances in robust estimates and tests (in Chinese). To appear in *Adv. Math.*.

Zhang, J. and Li, G. (1998). Breakdown properties of location M-estimators. *Ann. Statist.* **26,** 1170-1189.

Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Ann. Statist.*, to appear.

EURANDOM
Den Dolech 2
5612 AZ, Eindhoven
The Netherlands
E-mail: jzhang@euridice.tue.nl