Projection based scatter depth functions and associated scatter estimators

Weihua Zhou[1]

University of North Carolina at Charlotte

and

Xin Dang[2]

University of Mississippi

June 26, 2008

[1]Department of Mathematics and Statistics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA. Telephone: (704)687-4562. Fax: (704)687-6415. Email: wzhou2@uncc.edu.

[2]Department of Mathematics, University of Mississippi, University, MS 38677, USA. Telephone: (662)915-7409. Fax: (662)915-2361. Email: xdang@olemiss.edu.

Abstract

A class of projection based scatter depth functions is introduced and studied. In order to use the depth function effectively, some favorable properties are suggested for the common scatter depth functions. We show the proposed scatter depth totally satisfies these desirable properties and its sample version possess strong and $\sqrt{n}$ uniform consistency. Under some regularity conditions, the limiting distribution of the empirical process of the scatter depth function is derived. We also found that the aforementioned depth functions assess the bounded influence functions.

A maximum depth-based affine equivariant scatter estimator is induced. The limiting distributions as well as the strong and $\sqrt{n}$ consistency of the sample scatter estimators are establish. The finite sample performance of the related scatter estimator shows it has a very high breakdown point and appropriate efficiency.

AMS 2000 Subject Classification: Primary 62G99 Secondary 60F15.

Key words and phrases: Scatter depth function; scatter estimator; empirical depth process; limiting distribution; breakdown point; influence functions.

1 Introduction

As an extension of univariate rank into multivariate data setting, the *location* depth function provides a *center-outward* ordering of multivariate observations. Points deep inside a data cloud get higher depth and those on the outskirts receive lower depth. There are a number of notions of *location* depth, including Mahalanobis depth, halfspace depth [14], simpilicial depth [5] and projection depth [19]. Depth-induced ordering provides promising new tools in multivariate data analysis and inference, especially in developing affine equivariant robust estimates of multivariate location and dispersion.

In this paper, we concentrate on another type of depth functions: *scatter* depth. The notions of *scatter* depth provide a *center-outward* order within the all positive definite symmetric matrices relative to a given distribution or a sample of data. We may use it as a selection criterion for different covariance estimators. Furthermore, the scatter depth can induce the maximum depth covariance estimator, which could be regarded as an extension of the univariate median type estimator to the multivariate data setting. The literature is replete with discussions of *location* depth and its applications; see, e.g., Liu, Parelius and Singh [6], Zuo and Serfling [18], Zhang [17], Zuo, Cui and He [20], and Zuo and Cui [21]. There are, however, very few discussions on *scatter* depth (exceptions are made in Zhang [17] and Zuo [22]. To fill this gap, in this research we study a class of projection based *scatter* depth functions proposed by Zuo [22], as a complementarily to the *location* depth.

In Section 2, the projection based scatter depth functions are defined. In order to obtain the effective usage of the *location* depth functions, it is argued in Zuo and Serfling [18] that the following four properties are favorable: *Affine Invariance, Maximality at Center, Monotonicity relative to Deepest point, Vanishing at Infinity.* Here we suggest some similar properties for the common scatter depth functions. It is shown that the projection based scatter depth functions possess all those properties and their sample versions of this depth function are strongly and $\sqrt{n}$ uniformly consistent. We also study the asymptotic behavior of the empirical process of the projection based scatter depth functions. Our results show that in general the empirical process does not converge uniformly over the whole set $\mathcal{M}$ = all positive definite symmetric matrices. However, with some restrictions, we do obtain the weak convergence for the empirical process over some subsets. In order to assess the robustness of the proposed depth functions, we derive its influence function which is bounded. Section 2 ends with an illustration.

Section 3 is devoted to the study of the affine equivariant scatter estimators induced from projection based scatter depth functions. Large and finite sample behavior of sample projection induced scatter estimators are investigated. Strong consistency and limiting distributions of sample projection induced scatter estimators are obtained. Study of the finite sample behavior indicates that with appropriate choice of univariate scale estimator the proposed scatter estimator can have very high breakdown point and good efficiency. Section 4 ends the paper with some concluding remarks.

2 Projection based scatter depth functions and associated properties

In this section we study a class of projection based scatter depth functions and associated properties. We first introduce the definition.

2.1 Definition

Assume $X \in \mathbb{R}^d$, σ be *scale equivariant* and *translation invariant* scale, that is, $\sigma(F_{sY+c}) = |s|\sigma(F_Y)$, respectively, for any scalars s and c and random variable $Y \in \mathbb{R}^1$. Maronna, Stahel and Yohai [8] proposed a scatter estimator

$$V(X) = \mathrm{arginf}_{\Sigma}\Big[\sup_{\|u\|=1}\Big\{\big|1 - \frac{\sigma(F_u)}{\sqrt{u^T \Sigma u}}\big|\Big\}\Big], \tag{1}$$

where the infimum is over $V \in \mathcal{M}$ with $\mathcal{M}$ the set of all $d \times d$ positive definite matrices and F_u the distribution of $u^T X$, $u \in \mathbb{R}^d$.

Note that if $\sigma(\cdot)$ is taken to be the sample standard deviation, then $V(X)$ is the sample covariance matrix. Tyler [16] modified and generalized it to the following version

$$V(X) = \text{arginf}_\Sigma \Big[\sup_{\|u\|=1} g(\frac{\sigma^2(F_u)}{u^T \Sigma u}) \Big], \tag{2}$$

and he pointed out that the solutions to (1) and (2) differ only up to a scalar if for some $x_0 > 0$ the function $g(x)$ decreases for $x < x_0$ and increases for $x > x_0$.

Based on the above argument, it is very natural to define the *outlyingness* of a positive definite symmetric matrix $\Sigma \in \mathbb{R}^{d\times d}$ with respect to a given distribution function F in $\mathbb{R}^d$ as

$$O(\Sigma, F) = \sup_{\|u\|=1} g(\frac{\sigma(F_u)}{\sqrt{u^T \Sigma u}}). \tag{3}$$

This definition was first mentioned by Zuo [22].

The *projection based scatter depth* of a matrix $\Sigma \in \mathcal{M}$ w.r.t. the given F, $D(\Sigma, F)$, is then defined as

$$D(\Sigma, F) = 1/(1 + O(\Sigma, F)). \tag{4}$$

Sample versions of $O(\Sigma, F)$ and $D(\Sigma, F)$ denoted by $O(\Sigma, F_n)$ and $D(\Sigma, F_n)$ are obtained by replacing F with its empirical distribution F_n respectively.

Remark 2.1 *(i) For a given $g(\cdot)$ function, a specific choice of σ results in a specific $D(\Sigma, F)$. In this paper, we recommend to choose σ = the median absolute deviation (MAD) for the robustness property of the scatter depth.*
(ii) In the literature, a scatter depth defined in the same spirit was also given in Zhang [17]. According to the simultaneous univariate M-estimates of the location and scale, he defined

$$O(\sigma, F_n) = \frac{1}{n}\Sigma_{i=1}^n g(\frac{x_i - \mu(F_n)}{\sigma})$$

as a measure of the outlyingness of σ relative to $(|x_1 - \mu(F_n)|, \cdots, |x_n - \mu(F_n)|)$. Then a natural extension to the multivariate data is

$$O(\Sigma, F) = \sup_{\|u=1\|} \int g(\frac{u^T x - \mu(F_u)}{\sqrt{u^T \Sigma u}}) F(dx).$$

(iii) The projection based scatter depth $D(\Sigma, F)$ enjoys desirable properties of depth functions (Section 2.2) and thus provides a center-outward ordering of covariance matrices. One possible application of this aforementioned depth is to serve as a selection criterion for those covariance estimators via different methods such as M-estimator, Minimum Covariance Determinant (MCD) and S-estimator.
(iv) An important application of this depth functions is that it induces the deepest scatter estimator which can achieve high breakdown point.

Now we explore various properties of the projection based scatter depth functions.

2.2 Some properties

In order to use the *location* depth functions effectively, Zuo and Serfling [18] provided four favorable properties: affine invariance, maximality at a center, quasi-concave and vanishing at infinity. Similar with the *location* depth functions, here we suggest that the following properties may be favorable for the *scatter* depth functions:

P1. Affine invariance. That is say, $D(A\Sigma A^T, F_{AX+b}) = D(\Sigma, F)$ for any random vector $X \in \mathbb{R}^d$, any $d \times d$ nonsingular matrix A and any vector $b \in \mathbb{R}^d$. The depth of a matrix $\Sigma \in \mathcal{M}$ should not depend on the underlying coordinate system or, in particular, on the scales of the underlying measurements.

P2. Maximality at some matrix denoted by Σ_0. For a distribution having a uniquely defined covariance matrix Σ_F, the depth function should attain the maximum value at some matrix $\Sigma_0 = c\Sigma_F$ for some constant c.

P3. Monotonicity relative to Σ_0. That is, $D(\Sigma_1, F) \geq D(\Sigma_2, F)$ if $\Sigma_1 - \Sigma_0$ and $\Sigma_2 - \Sigma_1$ are positive definite and $D(\Sigma_1, F) \leq D(\Sigma_2, F)$ if $\Sigma_0 - \Sigma_2$ and $\Sigma_2 - \Sigma_1$ and are positive definite. As a matrix $\Sigma \in \mathcal{M}$ moves away from the "deepest point" (Σ_0), the depth of Σ should decrease.

P4. Vanishing at infinity. $D(\Sigma, F) \to 0$ as $\lambda_1 \to 0$ or $\lambda_d \to \infty$ where λ_1 and λ_d are the largest and smallest eigenvalue of Σ.

Since the P2 requires the maximality at some center, in the sequel, we will assume that the distribution F has a uniquely defined covariance matrix. The theorem below shows that the projection based scatter depth possess all the above desirable properties under the following assumptions for function $g(\cdot)$:

(A1) g is a nonnegative, differentiable function on $[0, \infty)$ with $g(0) = g(\infty) = \infty$.

(A2) For some $x_0 > 0$ the function $g(x)$ decreases for $x < x_0$ and increases for $x > x_0$.

A simple example of $g(\cdot)$ satisfied the above conditions is $g(x) = e^x/x$ with $x_0 = 1$.

Theorem 2.1 *For a fixed F in $\mathbb{R}^d$, assume (A1)-(A2) hold and $\sigma(\cdot)$ is Fisher consistent for a given scaler parameter, then $D(\Sigma, F)$ possesses P1-P4.*

PROOF OF THEOREM 2.1. For P1, since σ is *scale equivariant* and *translation invariant*, we have

$$\sigma(u^T(AX + b)) = \sigma(u^T AX) = \|u^T A\| \sigma(\frac{u^T A}{\|u^T A\|} X).$$

Also we have

$$\sqrt{u^T A \Sigma A^T u} = \|u^T A\| \sqrt{\frac{u^T A}{\|u^T A\|} \Sigma \frac{u A^T}{\|u A^T\|}}.$$

If we choose $v = uA^T/\|uA^T\|$, straightforwardly we can obtain $O(A\Sigma A^T, F_{AX+b}) = O(\Sigma, F)$ and P1 follows.

Let $\sigma(\cdot)$ be a Fisher consistent scale estimator, then $\sigma(F_u) = m_0(u^T \Sigma_F u)^{1/2}$ with some constant m_0 provided that F has covariance matrix Σ_F. Hence

$$\frac{\sigma(F_u)}{\sqrt{u^T \Sigma u}} = m_0 \sqrt{\frac{u^T \Sigma_F u}{u^T \Sigma u}} = m_0 \sqrt{\frac{x^T \Sigma^{-1/2} \Sigma_F \Sigma^{-1/2} x}{x^T x}} = m_0 \sqrt{R(x)}$$

Where $x = \Sigma^{1/2} u$ and $R(x)$ is called the *Rayleigh quotient* of matrix $H = \Sigma^{-1/2} \Sigma_F \Sigma^{-1/2}$ (see Lancaster and Tismenetsky [10]). In particular, $R(x)$ is a continuous function in x where x is on the unit sphere in $\mathbb{R}^d$. Further, from results of matrix theory, we have $\lambda_d \leq R(x) \leq \lambda_1$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ are the eigenvalues of H and the minima λ_d and maxima λ_1 are attained at the corresponding eigenvectors v_d, v_1 of H. Note that v_1 or v_d may not be a singleton vector. By conditions A1-A2, the superimum will be taken at end points of the interval, i.e.

$$O(\Sigma, F) = \max(g(m_0\lambda_1^{1/2}), g(m_0\lambda_d^{1/2})). \tag{5}$$

Let Σ_0 be a solution which results in the maximum depth of $D(\Sigma, F)$, from A2, we obtain

$$\begin{aligned} \sup_{\|u\|=1} g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma_0 u}}) &= \sup_{\|u\|=1} g\Big(m_0\sqrt{\frac{u^T\Sigma_F u}{u^T\Sigma_0 u}}\Big) \\ &= g(x_0). \end{aligned} \tag{6}$$

Note that $g(x_0)$ is the minima of $g(\cdot)$ and combined with equation (5), it is straightforward to see that equation (6) will only hold when $\lambda_1 = \lambda_d$. This results in $\Sigma_0 = (\frac{m_0}{x_0})^2\Sigma_F$ and any other Σ differing with Σ_0 will lower the corresponding depth. P2 follows.

For P3, we only prove one side and the other side is similar and omitted. Suppose $\Sigma_1 - \Sigma_0$ and $\Sigma_2 - \Sigma_1$ are positive definite. From the proof of P2, it is easy to see

$$g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma_1 u}}) = g\Big(m_0\sqrt{\frac{u^T\Sigma_F u}{u^T\Sigma_1 u}}\Big) = g\Big(x_0\sqrt{\frac{u^T\Sigma_0 u}{u^T\Sigma_1 u}}\Big)$$

and

$$g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma_2 u}}) = g\Big(x_0\sqrt{\frac{u^T\Sigma_0 u}{u^T\Sigma_2 u}}\Big).$$

Since $u^T\Sigma_0 u \leq u^T\Sigma_1 u \leq u^T\Sigma_2 u$ for any u, according to A2, $g(\cdot)$ is a decreasing function in this domain. So for any u, we have $g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma_1 u}}) \leq g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma_2 u}})$, which means $O(\Sigma_1, F) \leq O(\Sigma_2, F)$. This concludes the proof of P3. P4 is straightforward and omitted. □

2.3 Continuity of the projection based scatter depth

For any $\Sigma \in \mathcal{M}$, we define its norm $\|\Sigma\| = \sup_{\|u\|=1} |u^T\Sigma u|$. Since the projection scatter depth functions are based on a univariate scale functional, conditions on σ are given first. We use F_{nu} as the empirical distribution of $\{u^T X_i, i = 1, \ldots, n\}$ for any $u \in \mathbb{R}^d$. Define

(C1) $\sup_{\|u\|=1} \sigma(F_u) < \infty, \inf_{\|u\|=1} \sigma(F_u) > 0$.

(C2) $\sup_{\|u\|=1} |\sigma(F_{nu}) - \sigma(F_u)| = o_p(1)$.

(C3) $\sup_{\|u\|=1} |\sigma(F_{nu}) - \sigma(F_u)| = o(1)$ a.s..

(C4) $\sup_{\|u\|=1} |\sigma(F_{nu}) - \sigma(F_u)| = O_p(n^{-1/2})$.

(C5) $\{\sqrt{n}(\sigma(F_{nu}) - \sigma(F_u)) : \|u\| = 1\} \xrightarrow{d} \{Z_\sigma(u) : \|u\| = 1\}$ with $Z_\sigma(u)$ having continuous sample path and "$\xrightarrow{d}$" representing convergence in distribution.

Under some mild conditions, the theorem below shows $D(\Sigma, F)$ is uniformly continuous in $\Sigma \in \mathcal{M}$ for fixed F.

Theorem 2.2 *Under (A1), (A2) and (C1), $D(\Sigma, F)$ is uniformly continuous in $\Sigma \in \mathcal{M}$.*

Proof of Theorem 2.2. Since for any Σ_1, $\Sigma_2 \in \mathcal{M}$,

$$
\begin{aligned}
|D(\Sigma_1, F) - D(\Sigma_2, F)| &= \frac{|O(\Sigma_1, F) - O(\Sigma_2, F)|}{(1 + O(\Sigma_1, F))(1 + O(\Sigma_2, F))} \\
&\le |O(\Sigma_1, F) - O(\Sigma_2, F)|
\end{aligned}
$$

the following proof thus are focused on the outlyingness functions. For any given $\epsilon > 0$, let Σ_1, $\Sigma_2 \in \mathcal{M}$ such that $\|\Sigma_1 - \Sigma_2\| = \sup_{\|u\|=1} |u^T(\Sigma_1 - \Sigma_2)u| \le c\epsilon$, where c is some constant. We have

$$
\begin{aligned}
|O(\Sigma_1, F) - O(\Sigma_2, F)| &\le \sup_{\|u\|=1} |g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma_1 u}}) - g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma_2 u}})| \\
&= \sup_{\|u\|=1} |g'(\theta)(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma_1 u}} - \frac{\sigma(F_u)}{\sqrt{u^T\Sigma_2 u}})| \\
&= \sup_{\|u\|=1} \Big| \frac{g'(\theta)\sigma(F_u)u^T(\Sigma_1 - \Sigma_2)u}{\sqrt{u^T\Sigma_1 u}\sqrt{u^T\Sigma_2 u}(\sqrt{u^T\Sigma_1 u} + \sqrt{u^T\Sigma_2 u})} \Big|,
\end{aligned}
$$

where θ is some value between $\frac{\sigma(F_u)}{\sqrt{u^T\Sigma_1 u}}$ and $\frac{\sigma(F_u)}{\sqrt{u^T\Sigma_2 u}}$. The uniform continuity of $O(\Sigma, F)$ in Σ follows for some suitable c. □

Furthermore, $D(\Sigma, F)$ is " continuous " in F uniformly relative to a given Σ. $D(\Sigma, F)$ is said to be continuous in F for fixed Σ if $D(\Sigma, F_n) \to D(\Sigma, F)$ when F_n converges to F in distribution as $n \to \infty$.

Theorem 2.3 *Under (A1), (A2) and (C1), we have*

(1) $\sup_{\Sigma\in\mathcal{M}}|D(\Sigma,F_n)-D(\Sigma,F)|=o_p(1)$ *if (C2) holds;*

(2) $\sup_{\Sigma\in\mathcal{M}}|D(\Sigma,F_n)-D(\Sigma,F)|=o(1)$ *a.s. if (C3) holds;*

Proof of Theorem 2.3. Now we show part (2), the proof of part (1) is similar and omitted. Note that

$$\begin{aligned}|O(\Sigma,F_n)-O(\Sigma,F)| &\le \sup_{\|u\|=1}|g(\frac{\sigma(F_{nu})}{\sqrt{u^T\Sigma u}})-g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}})| \\ &= \sup_{\|u\|=1}\Big|g'(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}})\frac{\sigma(F_{nu})-\sigma(F_u)}{\sqrt{u^T\Sigma u}}\Big| \\ &\quad +O((\frac{\sigma(F_{nu})-\sigma(F_u)}{\sqrt{u^T\Sigma u}})^2). \qquad (7)\end{aligned}$$

Under the condition (C3), it follows that for any fixed $c>0$ and $\Sigma\in\mathcal{M}$ such that $\inf_{\|u\|=1}u^T\Sigma u\ge c$,

$$\sup_{\inf_{\|u\|=1}u^T\Sigma u\ge c}|D(\Sigma,F_n)-D(\Sigma,F)|\to 0,\quad \text{as}\quad n\to\infty.$$

Part (2) follows if we can show that the above is also true for any $\Sigma\in\mathcal{M}$ such that $\|\Sigma\|<c$. By Theorem 2.1, $D(\Sigma,F)\to 0$ as $\|\Sigma\|\to 0$. So we need only show that $D(\Sigma,F_n)\to 0$ as $u^T\Sigma u\to 0$ and $n\to\infty$. By (C1), $\sigma(F_u)$ is uniformly bounded below from 0. Thus, if we can show that $\sigma(F_{nu})$ is uniformly bounded below from 0, then

$$O(\Sigma,F_n)=\sup_{\|u\|=1}g(\frac{\sigma(F_{nu})}{\sqrt{u^T\Sigma u}})\to\infty \text{ as } n\to\infty.$$

Since

$$|\inf_{\|u\|=1}\sigma(F_{nu})-\inf_{\|u\|=1}\sigma(F_u)|\le\sup_{\|u\|=1}|\sigma(F_{nu})-\sigma(F_u)|,$$

thus $\inf_{\|u\|=1}\sigma(F_{nu})\to\inf_{\|u\|=1}\sigma(F_u)$ as $n\to\infty$ and consequently $\sigma(F_{nu})$ is uniformly bounded below from 0 for sufficiently large n. Part (2) follows. □

Theorem 2.3 can be strengthened to $\sqrt{n}$ consistency under some conditions.

Theorem 2.4 *Under (A1)-(A2),(C1) and (C4),* $\sup_{\Sigma\in\mathcal{M}}|D(\Sigma,F_n)-D(\Sigma,F)|=O_p(n^{-1/2})$.

Proof of Theorem 2.4. From equation (7), we see that

$$\sqrt{n}|O(\Sigma, F_n) - O(\Sigma, F)| \le \frac{Q_n}{\inf_{\|u\|=1}\sqrt{u^T\Sigma u}} + \frac{R_n}{\{\inf_{\|u\|=1}\sqrt{u^T\Sigma u}\}^2},$$

where

$$Q_n = \sup_{\|u\|=1}\Big|g'(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}})\sqrt{n}(\sigma(F_{nu}) - \sigma(F_u))\Big|$$

and

$$R_n = \sqrt{n}O((\sigma(F_{nu}) - \sigma(F_u))^2).$$

By the given conditions, Q_n is bounded and $R_n = o_p(1)$. Thus, for $\Sigma \in \mathcal{M}$ and any fixed $c > 0$, we have

$$\sqrt{n}\sup_{\inf_{\|u\|=1} u^T\Sigma u \ge c}|D(\Sigma, F_n) - D(\Sigma, F)| = O_p(1).$$

Denote Σ_1 as a solution for $O(\Sigma_1, F) = \inf_\Sigma O(\Sigma, F)$. From the proof of Theorem 2.2, we have

$$O(\Sigma, F_n) = \sup_{\|u\|=1}\Big\{g(\frac{\sigma(F_{nu})}{\sqrt{u^T\Sigma_1 u}}) + \frac{g'(\theta_1)\sigma(F_{nu})u^T(\Sigma_1 - \Sigma)u}{\sqrt{u^T\Sigma_1 u}\sqrt{u^T\Sigma u}(\sqrt{u^T\Sigma_1 u} + \sqrt{u^T\Sigma u})}\Big\} \quad (8)$$

where θ_1 is some value between $\frac{\sigma(F_{nu})}{\sqrt{u^T\Sigma_1 u}}$ and $\frac{\sigma(F_{nu})}{\sqrt{u^T\Sigma u}}$. Notice that $\Sigma_1 - \Sigma$ is positive definite and $g'(\theta_1)$ is positive when $\|\Sigma\| \to 0$. Using the fact that

$$\frac{\inf f(x)}{\inf g(x)} \le \sup\frac{f(x)}{g(x)} \le \frac{\sup f(x)}{\inf g(x)}$$

for any positive function $f(x)$ and $g(x)$, we obtain

$$O(\Sigma, F_n) \ge \frac{A_n}{\inf_{\|u\|=1}\sqrt{u^T\Sigma u}}, \quad (9)$$

where

$$A_n = \frac{\inf_{t\in(x_0,\infty)} g'(t)\inf_{\|u\|=1}[\sigma(F_{nu})u^T(\Sigma_1 - \Sigma)u]}{\sup_{\|u\|=1}\sqrt{u^T\Sigma_1 u}(\sqrt{u^T\Sigma_1 u} + \sqrt{u^T\Sigma u})}.$$

Similarly, we can get

$$O(\Sigma, F) \ge \frac{B_n}{\inf_{\|u\|=1}\sqrt{u^T\Sigma u}}, \quad (10)$$

where

$$B_n = \frac{\inf_{t\in(x_0,\infty)} g'(t)\inf_{\|u\|=1}[\sigma(F_u)u^T(\Sigma_1 - \Sigma)u]}{\sup_{\|u\|=1}\sqrt{u^T\Sigma_1 u}(\sqrt{u^T\Sigma_1 u} + \sqrt{u^T\Sigma u})}.$$

Note that

$$
\begin{aligned}
|O(\Sigma, F_n) - O(\Sigma, F_n)| &\leq \sup_{\|u\|=1} |g(\frac{\sigma(F_{nu})}{\sqrt{u^T \Sigma u}}) - g(\frac{\sigma(F_u)}{\sqrt{u^T \Sigma u}})| \\
&= \sup_{\|u\|=1} \left| g'(\theta_2) \frac{\sigma(F_{nu}) - \sigma(F_u)}{\sqrt{u^T \Sigma u}} \right| \\
&\leq \frac{C_n}{\inf_{\|u\|=1} \sqrt{u^T \Sigma u}}, \qquad (11)
\end{aligned}
$$

where θ_2 is some value between $\frac{\sigma(F_{nu})}{\sqrt{u^T \Sigma u}}$ and $\frac{\sigma(F_u)}{\sqrt{u^T \Sigma u}}$ and

$$C_n = \sup_{t \in [0,\infty)} g'(t) \sup_{\|u\|=1} |\sigma(F_{nu}) - \sigma(F_u)|.$$

Under the conditions, A_n, B_n and C_n are bounded. Based on the above argument, for any $\Sigma \in \mathcal{M}$ such that $\|\Sigma\| < c$ (c sufficiently small), we see that for sufficiently large n

$$
\begin{aligned}
\sqrt{n}|D(\Sigma, F_n) - D(\Sigma, F)| &\leq \frac{|O(\Sigma, F_n) - O(\Sigma, F)|}{O(\Sigma, F_n) O(\Sigma, F)} \\
&\leq \frac{C_n \inf_{\|u\|=1} \sqrt{u^T \Sigma u}}{A_n B_n} \\
&\to 0. \qquad (12)
\end{aligned}
$$

This completes the proof. □

2.4 Empirical depth process

Since we already get the strong and $\sqrt{n}$ consistency for the sample projection scatter depth $D(\Sigma, F_n)$. A natural question is: does the empirical process of the projection scatter depth $\sqrt{n}(D(\Sigma, F_n) - D(\Sigma, F)$ possess a limiting distribution? The next theorem answers the question. In the following, $u(\Sigma)$ is the set of u satisfying $O(\Sigma, F) = g(\frac{\sigma(F_u)}{\sqrt{u^T \Sigma u}})$ for a given Σ.

Theorem 2.5 *Assume (A1)-(A2), (C1) and (C5) hold. Then we have*

$$\sqrt{n}(D(\Sigma, F_n) - D(\Sigma, F)) \xrightarrow{d} - \sup_{u \in u(\Sigma)} Z(u, \Sigma),$$

where

$$Z(u, \Sigma) = g'(\frac{\sigma(F_u)}{\sqrt{u^T \Sigma u}}) Z_\sigma(u) / [\sqrt{u^T \Sigma u}(1 + O(\Sigma, F))^2].$$

PROOF OF THEOREM 2.5. Since for $\Sigma \in \mathcal{M}$,

$$\sqrt{n}\big(g(\frac{\sigma(F_{nu})}{\sqrt{u^T\Sigma u}}) - g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}})\big) = \sqrt{n}g'(\theta_3)\frac{\sigma(F_{nu}) - \sigma(F_u)}{\sqrt{u^T\Sigma u}},$$

where θ_3 is some value between $\frac{\sigma(F_{nu})}{\sqrt{u^T\Sigma u}}$ and $\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}}$ and $|\theta - \frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}}| = O_p(n^{-1/2})$, under the given conditions and the continuous mapping theorem, we have

$$\Big\{\sqrt{n}\big(g(\frac{\sigma(F_{nu})}{\sqrt{u^T\Sigma u}}) - g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}})\big) : \|u\| = 1\Big\} \xrightarrow{w} \Big\{g'(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}})Z_\sigma(u)/\sqrt{u^T\Sigma u} : \|u\| = 1\Big\},$$

where "$\xrightarrow{w}$" represents weak convergence. By the representation theorem (see for example Theorem 3.5.1 in Dudley [4]), for a fixed Σ, there exist two processes $\{S_n(u) : \|u\| = 1\}$ and $\{S(u) : \|u\| = 1\}$ which follow the same joint distributions as those of

$$\Big\{\sqrt{n}\big(g(\frac{\sigma(F_{nu})}{\sqrt{u^T\Sigma u}}) - g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}})\big) : \|u\| = 1\Big\}$$

and

$$\Big\{g'(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}})Z_\sigma(u)/\sqrt{u^T\Sigma u} : \|u\| = 1\Big\},$$

and satisfy

$$\sup_{\|u\|=1} |S_n(u) - S(u)| \longrightarrow 0, \quad \text{a.s..}$$

Letting $V_{1n} = g(\frac{\sigma(F_{nu})}{\sqrt{u^T\Sigma u}})$, $V_1 = g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}})$, $S_{1n} = S_n$, $S_1 = S$, $\Delta = \{u : \|u\| = 1, u \in \mathbb{R}^d\}$, $B_1 = u(\Sigma)$ and invoking Lemma 6.1 in Zhang [17], we obtain that as $n \to \infty$,

$$\sqrt{n}(O(\Sigma, F_n) - O(\Sigma, F)) \xrightarrow{d} \sup_{u \in u(\Sigma)} S(u).$$

Via the definition of the depth function, we obtain the desired result. □

For a given Σ where $u(\Sigma)$ is a singleton, the distribution of $Z(u(\Sigma), \Sigma)$ is typically Gaussian. It is impossible to get the weak convergence of the projection scatter depth process over the whole $\mathcal{M}$. Matrices Σ with $u(\Sigma)$ different from a singleton present a problem. The following theorem provide sufficient conditions for the weak convergence of the projection scatter depth process over certain subsets of $\mathcal{M}$.

Theorem 2.6 *Assume (A1)-(A2), (C1) and (C5) hold. Let $K \subset \mathcal{M}$ be a set such that for each $c > 0$ and each $\delta > 0$,*

$$\inf_{\inf u^T \Sigma u \geq c, \Sigma \in K} \inf_{u \notin u(x,\delta)} (O(\Sigma, F) - g(\Sigma, u, F)) > 0, \tag{13}$$

where $u(\Sigma, \delta) = \{u \in \mathbb{R}^d : \|u\| = 1, d(u, u(\Sigma)) \leq \delta\}$. Then we have

$$\{\sqrt{n}(D(\Sigma, F_n) - D(\Sigma, F)) : \Sigma \in K\} \xrightarrow{w} \{-\sup_{u \in u(\Sigma)} Z(u, \Sigma) : \Sigma \in K\}.$$

Proof of Theorem 2.6. By equation (12), for each $\epsilon > 0$,

$$\lim_{c \to 0} \lim_{n \to \infty} P\{\sup_{\|\Sigma\| < c} |\sqrt{n}(D(\Sigma, F_n) - D(\Sigma, F))| > \epsilon\} = 0.$$

So we only need to show that for any $c > 0$,

$$\begin{aligned} \{\sqrt{n}(D(\Sigma, F_n) - D(\Sigma, F)) \quad &: \quad \Sigma \in K, \inf_{\|u\|=1} u^T \Sigma u \geq c\} \\ &\xrightarrow{w} \{-\sup_{u \in u(\Sigma)} Z(u, \Sigma) : \Sigma \in K, \inf_{\|u\|=1} u^T \Sigma u \geq c\}. \end{aligned}$$

This follows from the fact that for each $c > 0$,

$$\begin{aligned} &\{\sqrt{n}(O(\Sigma, F_n) - O(\Sigma, F)) : \Sigma \in K, \inf_{\|u\|=1} u^T \Sigma u \geq c\} \\ &\xrightarrow{w} \{-\sup_{u \in u(\Sigma)} g'(\frac{\sigma(F_u)}{\sqrt{u^T \Sigma u}}) Z_\sigma(u) / \sqrt{u^T \Sigma u} : \Sigma \in K, \inf_{\|u\|=1} u^T \Sigma u \geq c\}. \end{aligned}$$

In order to get this, we will use Theorem 2.1 in Arcones, Cui and Zuo [1]. Condition (i), (iii) hold trivially. Condition (ii) comes from the proof of Theorem 2.3. Condition (iv) is straightforward to verify. Thus invoking Theorem 2.1 in Arcones, Cui and Zuo [1], we finish the proof. □

Corollary 2.1 *Assume (A1)-(A2), (C1) and (C5) hold. Suppose that $u(\Sigma)$ consists of a singleton except for finitely many points $\{\Sigma_1, \cdots, \Sigma_m\}$. Then, for each $\delta > 0$,*

$$\begin{aligned} &\Big\{\sqrt{n}(D(\Sigma, F_n) - D(\Sigma, F)) : \Sigma \in \mathcal{M} - \cup_{j=1}^m B(\Sigma_j, \delta)\Big\} \\ &\xrightarrow{w} \Big\{ - Z(u, \Sigma) : \Sigma \in \mathcal{M} - \cup_{j=1}^m B(\Sigma_j, \delta)\Big\}, \end{aligned}$$

where $B(\Sigma_j, \delta) = \{\Sigma \in \mathcal{M} : \|\Sigma_j - \Sigma\| < \delta\}$.

PROOF OF COROLLARY 2.1. From Theorem 2.2, $O(\Sigma, F)$ is continuous in Σ. In the following we claim that $u(\Sigma)$ is a continuous function in $\mathcal{M} - \cup_{j=1}^m B(\Sigma_j, \delta)$. Take $\Sigma \in \mathcal{M} - \cup_{j=1}^m B(\Sigma_j, \delta)$. If $u(\Sigma)$ is not continuous at Σ, then there exists a sequence $\Sigma_n \to \Sigma$ such that $u(\Sigma_n) \nrightarrow u(\Sigma)$. We may assume that $u(\Sigma_n) \to \Sigma_0 \neq u(\Sigma)$. Since $O(\Sigma_n, F) \to O(\Sigma, F)$,

$$g(\frac{\sigma(F_{u(\Sigma_n)})}{\sqrt{u(\Sigma_n)^T \Sigma_n u(\Sigma_n)}}) \to g(\frac{\sigma(F_{u(\Sigma)})}{\sqrt{u(\Sigma)^T \Sigma u(\Sigma)}}).$$

But,

$$g(\frac{\sigma(F_{u(\Sigma_n)})}{\sqrt{u(\Sigma_n)^T \Sigma_n u(\Sigma_n)}}) \to g(\frac{\sigma(F_{u(\Sigma_0)})}{\sqrt{u(\Sigma_0)^T \Sigma_0 u(\Sigma_0)}}),$$

in contradiction. The continuity of $u(\Sigma)$ implies that condition (13) holds. Applying Theorem 2.6, we obtain the desired result. □

2.5 Influence function of the scatter depth

In the next, we will explore the influence function of the scatter depth. The influence function of T at a given point $z \in \mathbb{R}^d$ is defined as

$$IF(z; T, F) = \lim_{\epsilon \to 0^+} \frac{T((1-\epsilon)F + \epsilon\delta_z) - T(F)}{\epsilon},$$

where δ_z denotes the point mass distribution at z. The influence function measures the local robustness of T.

Theorem 2.7 *Assume that the conditions in Theorem 2.3 hold. Then for the given $z \in \mathbb{R}^d$ and $\Sigma \in \mathcal{M}$,*

$$IF(z; O(\Sigma, F), F) = \sup_{u \in u(\Sigma)} g'\Big(\frac{\sigma(F_u)}{\sqrt{u^T \Sigma u}}\Big) IF(u^T z; \sigma(F_u), F)/\sqrt{u^T \Sigma u} \quad (14)$$

Accordingly, $IF(z; D(\Sigma, F), F) = -IF(z; O(\Sigma, F), F)/(1 + O(\Sigma, F))^2$.

The theorem shows how the influence function of the depth of a positive definite matrix depends on the influence function of the scale estimator σ. If we choose a robust scale estimator σ, say, the median absolute deviation about the median (MAD), then the influence function of the depth for Σ is bounded regardless of the choice of function g, though its value does depend on g.

Proof of Theorem 2.7. For any $\epsilon > 0$ and the given $z \in \mathbb{R}^d$, let $\tilde{F} = (1-\epsilon)F + \epsilon\delta_z$, denote $u(\Sigma, \eta) = \{u : \|u\| = 1 \cap \|u - u(\Sigma)\| \leq \eta \text{ for any } \eta > 0\}$ and write

$$h(u, \Sigma, \tilde{F}) = -g\Big(\frac{\sigma(\tilde{F}_u)}{\sqrt{u^T\Sigma u}}\Big) + O(\Sigma, F).$$

Then

$$\begin{aligned}
-(O(\Sigma, \tilde{F}) - O(\Sigma, F))/\epsilon &= -(\sup_{\|u\|=1} g\Big(\frac{\sigma(\tilde{F}_u)}{\sqrt{u^T\Sigma u}}\Big) - O(\Sigma, F))/\epsilon \\
&= \inf_{\|u\|=1} (h(u, \Sigma, \tilde{F}))/\epsilon \\
&= \min\{\inf_{u\in u(\Sigma)} h(u, \tilde{F})/\epsilon, \inf_{u\in u(\Sigma)^C} h(u, \tilde{F})/\epsilon\} \\
&= \min\{\inf_{u\in u(\Sigma)} h(u, \tilde{F})/\epsilon, \inf_{u\in u(\Sigma)^C\cap u(\Sigma,\eta)} h(u, \tilde{F})/\epsilon, \\
&\qquad \inf_{u\in u(\Sigma)^C\cap u(\Sigma,\eta)^C} h(u, \tilde{F})/\epsilon\}, \qquad (15)
\end{aligned}$$

We know that as $\epsilon \to 0^+$,

$$(-g\Big(\frac{\sigma(\tilde{F}_u)}{\sqrt{u^T\Sigma u}}\Big) + g\Big(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}}\Big))/\epsilon \quad \to \quad g'\Big(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}}\Big)\frac{1}{\sqrt{u^T\Sigma u}}(-IF(u^Tz; \sigma, F_u)).$$

By the assumption, this is uniformly bounded u for the given Σ and z. Hence the third term in Equation (15) will be

$$\begin{aligned}
&\inf_{u\in u(\Sigma)^C\cap u(\Sigma,\eta)^C} h(u, \Sigma, \tilde{F})/\epsilon \\
&= \inf_{u\in u(\Sigma)^C\cap u(\Sigma,\eta)^C} [(-g\Big(\frac{\sigma(\tilde{F}_u)}{\sqrt{u^T\Sigma u}}\Big) + g\Big(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}}\Big))/\epsilon + (O(\Sigma, F) - g\Big(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}}\Big))/\epsilon] \\
&\geq \inf_{u\in u(\Sigma)^C\cap u(\Sigma,\eta)^C} (O(\Sigma, F) - g\Big(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}}\Big))/2\epsilon \qquad \text{(for sufficiently small } \epsilon) \\
&= O(1/\epsilon),
\end{aligned}$$

and the second term in Equation (15) satisfies

$$\inf_{u\in u(\Sigma)^C\cap u(\Sigma,\eta)} h(u,\Sigma,\tilde{F})/\epsilon$$

$$\begin{aligned}
&= \inf_{u\in u(\Sigma)^C\cap u(\Sigma,\eta)}((-g(\frac{\sigma(\tilde{F}_u)}{\sqrt{u^T\Sigma u}})+g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}}))/\epsilon+(O(\Sigma,F)-g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}}))/\epsilon)\\
&\geq \inf_{u\in u(\Sigma)^C\cap u(\Sigma,\eta)}(-g(\frac{\sigma(\tilde{F}_u)}{\sqrt{u^T\Sigma u}})+g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}}))/\epsilon\\
&\geq \inf_{u\in u(\Sigma,\eta)}(-g(\frac{\sigma(\tilde{F}_u)}{\sqrt{u^T\Sigma u}})+g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}}))/\epsilon
\end{aligned}$$

Thus for sufficiently small $\epsilon > 0$ and any given $\eta > 0$, we have,

$$\begin{aligned}
\inf_{u\in u(\Sigma,\eta)}(-g(\frac{\sigma(\tilde{F}_u)}{\sqrt{u^T\Sigma u}})+g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}}))/\epsilon &\leq \inf_{u\in u(\Sigma,\eta)} h(u,\Sigma,F)/\epsilon\\
&\leq (-O(\Sigma,\tilde{F})+O(\Sigma,F))/\epsilon\\
&\leq \inf_{u\in u(\Sigma)}(-g(\frac{\sigma(\tilde{F}_u)}{\sqrt{u^T\Sigma u}})+g(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}}))/\epsilon
\end{aligned}$$

Hence,

$$\begin{aligned}
&\inf_{u\in u(\Sigma,\eta)} g'(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}})\frac{1}{\sqrt{u^T\Sigma u}}(-IF(u^Tz;\sigma,F_u))\\
&\qquad\leq \lim_{\epsilon\to 0^+}(-O(\Sigma,\tilde{F})+O(\Sigma,F))/\epsilon\\
&\qquad\leq \inf_{u\in u(\Sigma)} g'(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}})\frac{1}{\sqrt{u^T\Sigma u}}(-IF(u^Tz;\sigma,F_u))
\end{aligned}$$

Let $\eta \to 0$, we obtain

$$\lim_{\epsilon\to 0^+} O(\Sigma,\tilde{F})-O(\Sigma,F)/\epsilon = \sup_{u\in u(\Sigma)} g'(\frac{\sigma(F_u)}{\sqrt{u^T\Sigma u}})\frac{1}{\sqrt{u^T\Sigma u}}IF(u^Tz;\sigma,F_u).$$

The desired result follows. □

As discussed previously, for a Fisher scale estimator σ, $u(\Sigma)$ is among the unit vectors of u_1 and u_2, where $u_1 = \Sigma^{-1/2}v_1$ and $u_2 = \Sigma^{-1/2}v_d$ with v_1 and v_d being the eigenvectors corresponding to the smallest and largest eigenvalues of $\Sigma^{-1/2}\Sigma_F\Sigma^{-1/2}$. Then the result of (14) becomes that

$$\begin{aligned}
IF(z;O(\Sigma,F),F) &= \max\Big(g'(m_0\lambda_1^{1/2})\frac{\|\Sigma^{-1/2}v_1\|}{\|v_1\|}IF(u_1^Tz;\sigma(F_{u_1}),F),\\
&\qquad g'(m_0\lambda_d^{1/2})\frac{\|\Sigma^{-1/2}v_d\|}{\|v_d\|}IF(u_2^Tz;\sigma(F_{u_2}),F)\Big). \qquad (16)
\end{aligned}$$

Further, for an elliptically symmetric distribution F with the covariance matrix Σ_F, without loss of generality, assuming symmetry about $\mathbf{0}$, i.e., the density function has the form $|\Sigma_F|^{-1/2}h(x^T\Sigma_F^{-1}x)$. With $\sigma = MAD$, Equation (16) has a following explicit form:

$$
\begin{aligned}
IF(z; O(\Sigma, F), F) \;=\; & \frac{1}{4h(1)} \max\Big(g'(m_0\lambda_1^{-1/2})\text{sign}(|v_1^T\Sigma^{-1/2}z| - \|(\Sigma^{-1}\Sigma_F)^{1/2}v_1\|), \\
& g'(m_0\lambda_d^{-1/2})\text{sign}(|v_d^T\Sigma^{-1/2}z| - \|(\Sigma^{-1}\Sigma_F)^{1/2}v_d\|)\Big).
\end{aligned}
$$

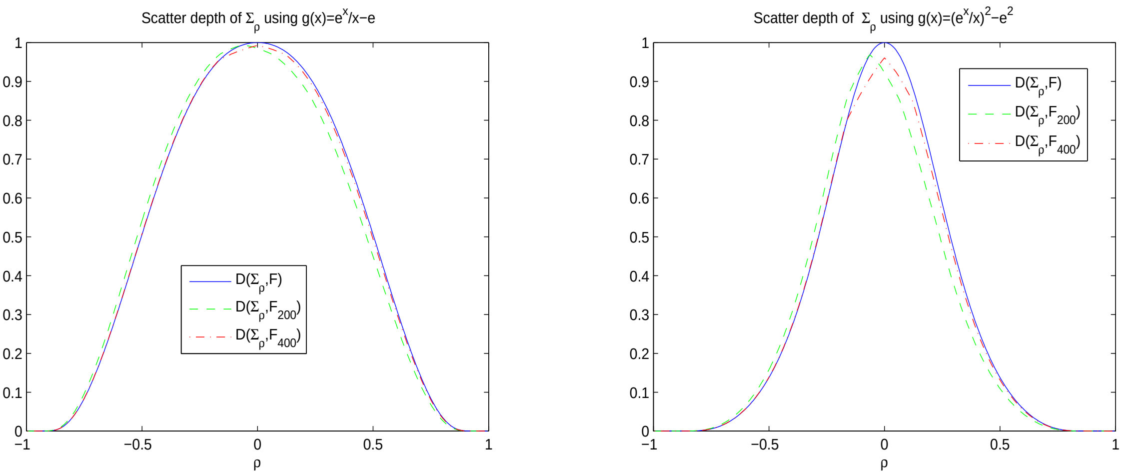
2.6 Examples for illustration

Maronna, Stahel and Yohai (1992) used $g(x) = |1-x|$ and Tyler (1994) suggested g function to be $|log(\cdot)|$. Both functions, however, are not differentiable due to involving the absolute function. Here we choose g function as $g_k(x) = (e^x/x)^k - e^k$, $k = 1, 2$. Substraction e^k is for the range of scatter depth function to be $(0, 1]$. The value k controls the shape of g function, for example, the sharpness or flatness at the neighborhood at the minimum point $x = 1$.

Let's consider a bivariate standard normal distribution $F = N_2(\mathbf{0}, \mathbf{I})$ and $\sigma =$ MAD as the scale estimator. Clearly $\text{MAD}(F_u) = \Phi^{-1}(3/4) = 0.6744898 = m_0$, where $\Phi(\cdot)$ is the cdf of the standard normal distribution. We consider the matrix $\Sigma_\rho = \left(\begin{smallmatrix}1 & \rho \\ \rho & 1\end{smallmatrix}\right)$ with ρ changing between -1 and 1.

We first compute the depth of Σ_ρ relative to F. From the discussion in Section 2.2, we know that the outlyingness of Σ_ρ with respect to F equals the maximum of $g(\sqrt{\lambda_1})$ and $g(\sqrt{\lambda_2})$, where λ_1, λ_2 are the eigenvalues of Σ_ρ^{-1}. We then compute the depth of Σ_ρ relative to F_n. In order to do so, we generate a random sample X of size n from F. For each unit direction $u = (\cos\theta, \sin\theta)^T$ with θ taking fine grids from 0 to π, we find the MAD of $u^TX_1, ..., u^TX_n$ and the ratio r between the MAD and $u^T\Sigma_\rho u$. Then the maximum of $g(r)$ over all directions will be considered as the sample projection *scatter* depth of Σ_ρ. The results based on the two g functions are showed in Fig1. The left panel corresponds to g_1 and g_2 for the right panel.

From Figure 1, for $\rho = 0$, Σ_ρ is the true covariance matrix of F and reaches the maximum depth 1. The farther the ρ away from 0, the smaller the depth. Hence scatter depth



(a) $g_1(x) = e^x/x - e$

(b) $g_2(x) = (e^x/x)^2 - e^2$

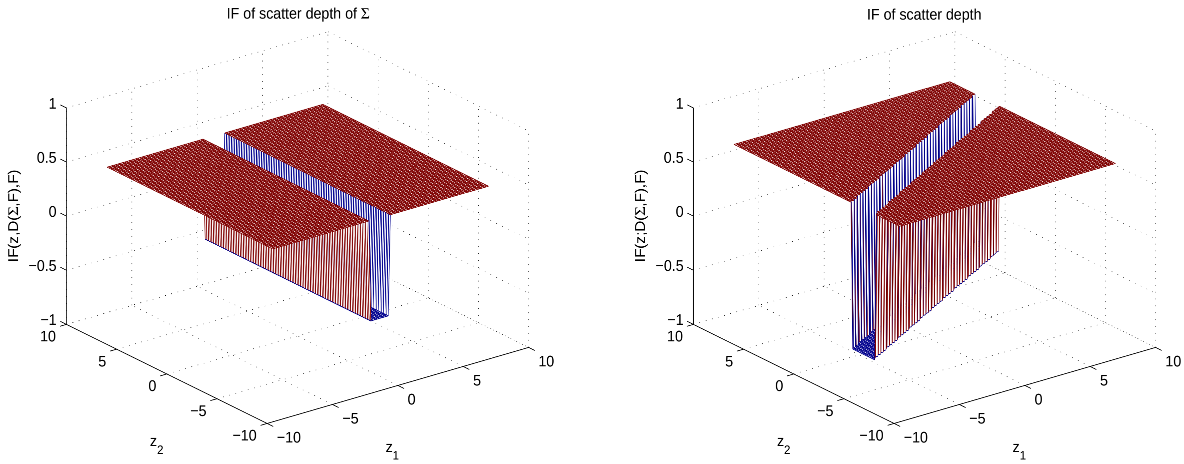
Figure 1: Depth of Σ_ρ w.r.t. F and F_n, F bivariate standard normal, F_n the empirical distribution with sample size n. g function is of the form $g_k(x) = (e^x/x)^k - e^k$ with $k = 1$ for the left panel and $k = 2$ for the right one.

provides center-outward ordering of a positive definite matrix with respect to a distribution. The sample depth estimates its population value very well, especially for ρ outside the neighborhood of 0. According to Theorem 2.5, the limiting distribution of sample scatter depth process is a stochastic process times a factor proportional to its squared population depth and the derivative of g function. Around $\rho = 0$, the depth attains a larger value close to 1, which explains a larger variability of sample depth around $\rho = 0$. Since $g_2'(x) \geq 2eg_1'(x)$, a larger deviation of sample depth from its population depth can be observed in the right panel than it in the left panel of Figure 1.

F	n	Mest	Mcd	Sest	Ccov
Normal	100	0.9452	0.9617	0.9568	0.9822
	200	0.9756	0.9813	0.9799	0.9911
	500	0.9913	0.9934	0.9925	0.9967
T_3	100	0.9158	0.8499	0.8991	0.4309
	200	0.9460	0.8922	0.9279	0.4985
	500	0.9643	0.9148	0.9460	0.5497

Table 1: Mean of depths for each scatter estimators.

Remark 2.1 (iii) mentioned an application of scatter depth, in which scatter depth may



(a) Σ is the identity matrix (b) $\Sigma_{11} = 2, \Sigma_{12} = \Sigma_{21} = -0.8, \Sigma_{22} = 1$

Figure 2: IF of $D(\Sigma, F)$ with F the bivariate standard normal distribution.

be served as a criteria to compare different scatter estimators. The estimator with largest depth value is the best one. Here we compare several scatter estimators M-estimator (Mest), Minimal Covariance Determinant (MCD), S-estimator (Sest) and sample covariance matrix (Ccov). More details and computations about those estimators see Section 3.2.2. We generated M samples of size n from the bivariate standard normal distribution or bivariate T_3 distribution. For each sample, each scatter estimators was obtained and the depth value with respect to F was calculated. Table 1 provided the mean of depths for each estimator with $M = 500$. Under normal models, sample covariance is the best estimator in terms of the largest depth value as well as efficiency. But it possesses the smallest depth among all estimators under the T_3 distribution.

Figure 2 provides the influence function of scatter depth $D(\Sigma, F)$, where F is bivariate standard normal distribution and $D(\Sigma, F)$ is based on g_1 with $\sigma(\cdot) = \text{MAD}$. The left panel is for Σ being the identity matrix, while $\Sigma = \begin{pmatrix} 2 & -0.8 \\ -0.8 & 1 \end{pmatrix}$ in the right panel. The influence function of scatter depth is a bounded step function due to the sign function in the influence function of univariate scale estimator MAD.

3 The maximal depth induced covariance estimator

For a given projection scatter depth, we define the symmetric positive definite matrix with the maximum depth as an estimator of the covariance matrix of F and denote it by V. Clearly, V can be defined as

$$V(F) = \arg\ \sup_{\Sigma \in \mathcal{M}} D(\Sigma, F).$$

Corollary 3.1 *Under the conditions in Theorem 2.1, the functional $V(F)$ has the following properties:*
(1) it is well defined and unique;
(2) it is Fisher consistent for the covariance matrix of F with some constant c which depends on the choices of $g(\cdot)$ and $\sigma(\cdot)$.
(3) it is affine equivariant. That is, $V(F_{AX+b}) = AV(F)A^T$ for any $d \times d$ nonsingular matrix and vector $b \in \mathbb{R}^d$.

This a direct result of Theorem 2.1. The sample analog of V denoted by V_n is obtained by replacing F with F_n. One need to take average if necessary to deal with non-uniqueness of V_n. V_n is affine equivariant and an unbiased estimator of V. Under some mild conditions, V_n is a consistent estimator of V and has a limiting distribution. Now we investigate the large and finite sample behavior of V_n.

3.1 Large sample behavior

In the following we first establish $\sqrt{n}$ consistency of V_n and then the limiting distribution of the maximal depth induced covariance estimator.

Theorem 3.1 *For a fixed F in $\mathbb{R}^d$, assume (A1),(A2),(C1) and (C4) hold and $\sigma(\cdot)$ is Fisher consistent for a given scaler parameter, then $\|V_n - V(F)\| = O_p(1/\sqrt{n})$ a.s..*

Proof of Theorem 3.1. Since V and V_n are the solutions to minimize the outlying functions $O(\Sigma, F)$ and $O(\Sigma, F_n)$, respectively and $O(V, F) = g(x_0)$, which is the smallest value of $g(\cdot)$, we have $O(V, F) \le O(V_n, F_n) \le O(V, F_n)$. We notice the following fact: for any u

$$g'(\frac{\sigma(F_u)}{\sqrt{u^T V u}}) = 0, \quad g''(\frac{\sigma(F_u)}{\sqrt{u^T V u}}) > 0.$$

Note that

$$\begin{aligned} O(V_n, F_n) &= \sup_{\|u\|=1} g(\frac{\sigma(F_{nu})}{\sqrt{u^T V_n u}}) \\ &= \sup_{\|u\|=1} \Big\{ g(\frac{\sigma(F_u)}{\sqrt{u^T V u}}) + \{ g''(\frac{\sigma(F_u)}{\sqrt{u^T V u}}) \frac{(\sigma(F_{nu}) - \sigma(F_u))^2}{u^T V u} + \\ &\quad \frac{1}{4} g''(\frac{\sigma(F_u)}{\sqrt{u^T V u}}) \frac{\sigma(F_u)^2}{(u^T V u)^3} u^T (V_n - V)^{\otimes 2} u \} (1 + o(1)) \Big\}, \end{aligned}$$

where $\Sigma^{\otimes 2} = \Sigma^T \Sigma$. It then follows that

$$\begin{aligned} n|O(V_n, F_n) - \quad O(V, F)| &\ge \sup_{\|u\|=1} \{ g''(\frac{\sigma(F_u)}{\sqrt{u^T V u}}) \frac{[\sqrt{n}(\sigma(F_{nu}) - \sigma(F_u))]^2}{u^T V u} + \\ &\quad \frac{1}{4} g''(\frac{\sigma(F_u)}{\sqrt{u^T V u}}) \frac{\sigma(F_u)^2}{(u^T V u)^3} u^T [\sqrt{n}(V_n - V)]^{\otimes 2} u \} (1 + o(1)). \end{aligned} \tag{17}$$

Similarly, we have

$$\begin{aligned} n|O(V, F_n) - O(V, F)| &\le n \sup_{\|u\|=1} \Big| g(\frac{\sigma(F_{nu})}{\sqrt{u^T V u}}) - (\frac{\sigma(F_u)}{\sqrt{u^T V u}}) \Big| \\ &\le \sup_{\|u\|=1} \{ g''(\frac{\sigma(F_u)}{\sqrt{u^T V u}}) \frac{[\sqrt{n}(\sigma(F_{nu}) - \sigma(F_u))]^2}{u^T V u} \} (1 + o(1)) \\ &\le O_p(1). \end{aligned} \tag{18}$$

Based on the fact

$$n|O(V_n, F_n) - O(V, F)| \le n|O(V, F_n) - O(V, F)|$$

and combining equations (17) and (18), we obtain

$$\sup_{\|u\|=1} u^T [\sqrt{n}(V_n - V)]^{\otimes 2} u = O_p(1).$$

Then the desired result follows straightforwardly. □

A natural question raised after one which has $\sqrt{n}$ consistency of V_n is: does V_n possess a limiting distribution? The following theorem answers the question.

Theorem 3.2 *For a fixed F in $\mathbb{R}^d$, assume (A1),(A2),(C1) and (C5) hold and $\sigma(\cdot)$ is Fisher consistent for a given scaler parameter, then*
$\sqrt{n}(V_n - V)$

$$\xrightarrow{d} \textit{argmin}_{\Delta} \sup_{\|u=1\|} \Big\{ g''(\frac{\sigma(F_u)}{\sqrt{u^TVu}})\frac{Z_\sigma(u)^2}{u^TVu} + \frac{1}{4}g''(\frac{\sigma(F_u)}{\sqrt{u^TVu}})\frac{\sigma(F_u)^2}{(u^TVu)^3}u^T\Delta^{\otimes 2}u \Big\}$$

provided that the argmin is unique with probability 1, where Δ runs over all $d \times d$ symmetric matrices.

Proof of Theorem 3.2. Note that V and V_n are positive definite symmetric matrices. In order to prove theorem, for any $d \times d$ symmetric matrix Δ, we define

$$V_n(\Delta) = V + \Delta/\sqrt{n},$$

$$Q_n(\Delta) = \sup_{\|u=1\|} \sqrt{n}\Big\{ g(\frac{\sigma(F_{nu})}{\sqrt{u^TV_n(\Delta)u}}) - g(\frac{\sigma(F_u)}{\sqrt{u^TVu}} \Big\}$$

and

$$Q(\Delta) = \sup_{\|u\|=1} \Big\{ g''(\frac{\sigma(F_u)}{\sqrt{u^TVu}})\frac{Z_\sigma(u)^2}{u^TVu} + \frac{1}{4}g''(\frac{\sigma(F_u)}{\sqrt{u^TVu}})\frac{\sigma(F_u)^2}{(u^TVu)^3}u^T\Delta^{\otimes 2}u \Big\}$$

From the proof of Theorem 3.1 and by virtue of the given conditions and the continuous mapping theorem, for each finite set S over all $d \times d$ symmetric matrices, $\{Q_n(\Delta) : \Delta \in S\} \xrightarrow{d} \{Q(\Delta) : \Delta \in S\}$. On the other hand, it is straightforward to verify that (ii) of Theorem 2.3 of Kim and Pollard [7] holds. In light of Theorem 2.3 of Kim and Pollard [7], we conclude that (i) of Theorem 2.7 of Kim and Pollard [7] holds. Let $t_n = \sqrt{n}(V_n - V)$, then by Theorem 3.1, $t_n = O_p(1)$. Note that $V_n = \text{argmin}_\Sigma \sup g(\frac{\sigma(F_{nu})}{\sqrt{u^T\Sigma u}})$, we have

$$Q_n(t_n) = \sup_{\|u\|=1} \sqrt{n}\Big\{ g(\frac{\sigma(F_{nu})}{\sqrt{u^TV_nu}}) - g(\frac{\sigma(F_u)}{\sqrt{u^TVu}}) \Big\} \le \inf_t Q_n(t).$$

We let $Z_n = Q_n$ and $\alpha_n = 0$, invoking Theorem 2.7 of Kim and Pollard [7], we obtain the desired result. □

Remark 3.1 *The assumption that the* argmin *is unique in the above theorem seems reasonable but difficult to check. Bai and He [2] provided a way to check such assumptions for the location estimator. Here we will defer this problem to the future study.*

3.2 Finite sample behavior

In this section the finite sample robustness and relative efficiency of V_n are investigated.

3.2.1 Finite sample breakdown point

Let $X^n = \{X_1, \cdots, X_n\}$ be a sample of size n from X in $\mathbb{R}^d$. The replacement breakdown point denoted by ε [Donoho and Huber [3]] of a scatter estimator Σ at X^n is defined as

$$\varepsilon(\Sigma, X^n) = \min\Big\{\frac{m}{n} : \mathrm{trace}(\Sigma(X^n)\Sigma(X^n_m)^{-1} + \Sigma(X^n)^{-1}\Sigma(X^n_m)) = \infty\Big\},$$

where X^n_m is a contaminated sample resulting from replacing m points of X^n with arbitrary values.

In the following discussion, $\varepsilon(\sigma, X^n)$ represents the breakdown point for the scale estimator of $u^T X$. A random sample X^n is said to be in general position if there are no more than d sample points of X^n lying in any $(d-1)$-dimensional subspace. Tyler [16] [Theorem 3.2] already obtained that $\varepsilon(V_n, X^n) \geq \varepsilon(\sigma, X^n)$. On the other side, It is trivial to show $\varepsilon(V_n, X^n) \leq \varepsilon(\sigma, X^n)$. Immediately, we have the following result.

Theorem 3.3 $\varepsilon(V_n, X^n) = \varepsilon(\sigma, X^n)$.

Let $\lfloor \cdot \rfloor$ be the floor function.

Corollary 3.2 *For X^n in general position, if we choose σ=MAD, then*

$$\varepsilon(V_n, X^n) = \begin{cases} \dfrac{\lfloor (n+1)/2 \rfloor}{n}, & \text{for } d = 1, \\ \dfrac{\lfloor (n-2d+4)/2 \rfloor}{n}, & \text{for } d \geq 2. \end{cases}$$

Remark 3.2 *We can improve the breakdown point by replacing MAD_k from MAD, where MAD_k is a modified version of MAD. $MAD_k(X^n) = Med_k(\{|x_1 - Med(X^n)|, \ldots, |x_n - Med(X^n)|\})$, where*

$$Med_k(X^n) = \Big(x_{\lfloor (n+k)/2 \rfloor} + x_{\lfloor (n+k+1)/2 \rfloor}\Big)/2, \quad 1 \leq k \leq n,$$

and $x_{(1)} \leq \cdots \leq x_{(n)}$ being ordered values of $x_1, \cdots, x_n$ in $\mathbb{R}^1$. This idea was adopted by several authors in the literature, for example Tyler [16] and Zuo [19].

3.2.2 Computation and finite sample efficiency

As pointed out by Maronna, Stahel and Yohai [8], the exact computation of the maximal depth induced covariance estimator seems nearly impossible, because it involves two optimization procedures with constraints in the high dimension space $\mathbb{R}^{d\times d}$. We adopt the algorithm based on the idea of random subsampling. This algorithm proposed by Maronna, Stahel and Yohai [8] approximates the covariance estimator by calculating the extrema over a finite set of directions u's and a finite set of positive definite symmetric matrices Σ's. The computation complex of the algorithm is $L \log L$, where L is the number of random subsamples. Their numerical experiments suggested that $L = 200$ suffices for $d = 2$ and $L = 500$ for $p = 5$. That is remarkably saving and the approximation is good enough for a general practice use. More details see Maronna, Stahel and Yohai [8].

We ran a simulation to investigate the small sample behavior of the maximal depth induced covariance estimator (Pd) and compare with other well-known robust scatter estimators. The following scatter estimators were considered:

Mcd Minimum covariance determinant estimator of Rousseeuw and van Driessen [12] is computed by the R package rrcov. The MCD method looks for the $h(> n/2)$ observations (out of n) whose classical covariance matrix has the lowest possible determinant. Then MCD scatter estimator is the covariance matrix based on those h observations multiplied by a consistency factor and a finite sample correction factor.

Mest Constrained M-estimator is also obtained by the R package rrcov . Unlike the usual M-estimator which only has breakdown point $1/(d+1)$, the modified one based on the translated bi-weight function using a high breakdown point initial estimate like Mcd can achieve high robustness and efficiency.

Sest Re-weighted S-estimator (Sest) is calculated by R package riv using Tukey bi-weighted ρ function.

Cov Non-robust sample covariance matrix.

As performance criteria for matrices, two quantities are used to measure non-sphericity. One is the condition number of V, cond(V). It is the ratio of the largest eigenvalue to the smallest eigenvalue of V. The other one is the likelihood ratio test statistic $\varphi_0(V)$, that is,

$$\varphi_0(V) = (\text{trace}(V)/d)^d/\det(V).$$

Those two measures were also used in Maronna and Yohai (1995) and Zuo and Cui (2005).

For different choices of ε and the dimension d , we generate M data sets of size n from the bivariate normal mixture distribution, i.e., $(1-\varepsilon)N(\mathbf{0}, I_{d\times d}) + \varepsilon N((k,k)^t, I_{d\times d})$. Here $M = 200$, $n = 50$ and $k = 10$. The dimensions were taken as 2 and 5. The values $\varepsilon = 0, 0.1, 0.2, 0.3$ and 0.4 were chosen. For each scatter estimator, the mean of log(cond) (MLCN) and the mean of $\log(\varphi_0)$ (MLLRT) were calculated. They include both bias and variability, like MSE. Hence the finite sample relative efficiency (RE) of estimator V_n is obtained by the ratio of MLCN or MLLRT of sample covariance to that of V_n. Table 2 displays the results and values in the parentheses are RE's.

Although the values of RE are quiet different using different measures, the behavior of an estimator under two criteria is similar. Among all robust estimators, under the normal model ($\varepsilon = 0$), our maximal depth induced covariance estimator (Pd) has the highest relative efficiency for dimension $d = 2$, while for $d = 5$ Sest is the most efficient. The phenomenon that RE increases as the dimension increases under normal models and RE decreases as the dimension increases under contaminated normal models is observed with some exceptions on the our proposed Pd. We believe that the exception is mainly due to the computation algorithm of Pd. The approximation of the algorithm is less accurate in high dimensions than low dimensions. It is observed that Sest is less robust in high dimensions than in low dimensions. In dimension 2, it is robust even in the scenario with the fraction of contamination $\varepsilon = 0.4$. It, however, has totally broken down for $\varepsilon \geq 0.3$ in $d = 5$. This problem of S estimator was also pointed out by Rocke (1996). It is not surprising that Mest preforms very similar to Mcd because this modified M estimator is based on the initial Mcd estimate to achieve high robustness. In summary, Pd demonstrates competitive performance on robustness and

	ε	d	Pd(RE)	Mcd(RE)	Mest(RE)	Sest(RE)	Cov
MLCN	0	2	0.5846(0.62)	0.6512(0.56)	0.6644(0.55)	0.6332(0.58)	0.3649
		5	1.6571(0.58)	1.6970(0.57)	1.4397(0.67)	1.1551(0.83)	0.9611
	0.1	2	0.6628(4.46)	0.6371(4.64)	0.6955(4.25)	0.6496(4.55)	2.956
		5	1.6238(2.68)	1.6021(2.71)	1.4152(3.07)	1.1650(3.73)	4.3450
	0.2	2	0.5286(6.71)	0.5129(6.92)	0.5668(6.26)	0.5199(6.82)	3.5470
		5	1.6088(3.05)	1.4487(3.39)	1.3505(3.63)	1.1808(4.16)	4.9067
	0.3	2	0.5938(6.38)	0.4940(7.67)	0.5892(6.43)	0.5183(7.31)	3.7886
		5	1.5675(3.13)	1.3807(3.75)	1.3560(3.82)	4.3579(1.19)	5.1773
	0.4	2	0.6891(5.73)	0.5306(7.44)	0.5572(7.09)	0.5339(7.40)	3.9485
		5	1.4453(3.67)	1.3593(3.90)	1.3609(3.90)	5.6352(0.94)	5.3063
MLLRT	0	2	0.1063(0.40)	0.1309(0.32)	0.1389(0.31)	0.1244(0.34)	0.0424
		5	0.7776(0.38)	0.8179(0.36)	0.6310(0.47)	0.4188(0.71)	0.2955
	0.1	2	0.1310(12.9)	0.1258(13.4)	0.1440(11.7)	0.1277(13.2)	1.6890
		5	0.7652(10.5)	0.7415(10.8)	0.6034(13.3)	0.4250(18.9)	8.0345
	0.2	2	0.1181(18.9)	0.0840(26.6)	0.1074(20.8)	0.0908(24.6)	2.2321
		5	0.6936(14.6)	0.6319(16.0)	0.5552(18.2)	0.4699(21.5)	10.127
	0.3	2	0.1141(21.6)	0.0737(33.4)	0.0940(26.2)	0.0814(30.3)	2.4650
		5	0.6401(17.4)	0.5770(19.3)	0.5684(19.6)	8.9267(1.25)	11.131
	0.4	2	0.1037(25.1)	0.0848(30.7)	0.0924(28.1)	0.0857(30.4)	2.6012
		5	0.8845(13.2)	0.5810(20.1)	0.5821(20.1)	12.608(0.93)	11.675

Table 2: Mean of log condition number (MLCN), mean of log likelihood ratio test statistics (MLLRT) and finite sample relative efficiencies (RE) for various scatter estimators.

efficiency, but it does need a good algorithm, especially in high dimensions.

4 Concluding Remarks

This paper introduces and studies a class of projection based scatter depth functions and their associated scatter estimators. We first propose some properties which will be favorable for the common scatter depth functions. Then it is showed the projection based scatter depth functions enjoy these properties and their sample versions possess strong and $\sqrt{n}$ uniformly consistency. The limiting distribution of the empirical depth process is established and the influence function of the depth is derived. The induced scatter estimator share many desirable properties. For example they are affine equivariate and $\sqrt{n}$ consistent and possess very high breakdown point. Furthermore, under mild conditions limiting distributions of the sample scatter estimators exist. The finite sample relative efficiency is also competitive comparing with other robust scatter estimators. The main drawback of these estimators is the computation. Only the approximation algorithm exists and is less accurate in high dimensions.

References

[1] Arcones, M., Cui, H. and Zuo, Y. (2006). Empirical depth processes. *Test*, **15**, 151-177.

[2] Bai, Z. and He, X. (1999). Asymptotic distributions of the maximal depth estimators for regression and multivariate location. *Annals of Statistics*, **27**, 1616-1637.

[3] Donoho, D. L. and Huber, P. J. (1983). The notion of breakdown point. In *A Fretschrift foe Erich L. Lehman* (Bickel, P. J., Doksum, K. A. and Hodges, J. L., eds). 157-184. Wadsworth, Belmont, CA.

[4] Dudley, R. M. (1999). *Unifrom Central Limite Theorems.* Cambridge University Press, Cambridge.

[5] Liu, R. Y. (1990). On a notion of data depth based on random simplices. *Annals of Statistics*, **18**, 405-414.

[6] Liu, R. Y., Parelius, J. M. and Singh, K. (1999). Multivariate analysis by data depth: descriptive statistics, graphics and inference (with discussion). *Annals of Statistics*, **27**, 783-858.

[7] Kim, J. and Pollard, D. (1990). Cube root asymptotics. *Annals of Statistics*, **18**, 191-219.

[8] Maronna, R., Stahel, W. and Yohai, V. (1992). Bias-robust estimators of multivariate scatter based on projections. *Journal of Multivariate Analysis*, **42**, 141-161.

[9] Maronna, R. and Yohai, V. (1995). The behavior of the Stahel-Donoho robust multivariate estimator. *Journal of American Statistical Association*, **90**, 330-341.

[10] Lancaster, P. and Tismenetsky, M. (1989). *The Theory of Matrices.* 2nd Ed. Academic Press, Inc.

[11] Rocke, D. (1996). Robustness Properties of S-estimators of multivariate location and shape in high dimension. *Annals of Statistics*,**24**, 1327-1345.

[12] Rousseeuw, P. and van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, **41**, 212-223.

[13] Ruppert, D. (1992). Computing S estiamtors for regression and multivaraite location/dispersion. *Journal of comutational and graphical staistics*,**1**, 253-270.

[14] Tukey, J. W. (1975). Mathematics and picturing data. *Proceedings of Inernational Congress of Mathematicians*, Vancouver, 1974, **2**, 523-531.

[15] Tyler, D. (1983). Robustness and efficiency properties of scatter matrices. *Biometrika*, **70**, 411-420.

[16] Tyler, D. (1994). Finite sample breakdown points of projection based multivariate location and scatter statistics. *Annals of Statistics*, **22**, 1024-1044.

[17] Zhang, J. (2002). Some extensions of Tukey's depth functions. *Journal of Multivariate Analysis*, **82**, 134-165.

[18] Zuo, Y. and Serfling, R. (2000). General notions of statistical depth function. *Annals of Statistics*, **28**, 461-482.

[19] Zuo, Y. (2003). Projection-based depth functions and associated medians. *Annals of Statistics*, **31**, 1460-1490.

[20] Zuo, Y., Cui, H. and He, X. (2004). On the Stahel-Donoho estimator and depth-weighted means of multivariate data. *Annals of Statistics*, **32**, 167-188.

[21] Zuo, Y. and Cui, H. (2005). Depth weighted scatter estimators. *Annals of Statistics*, **33**, 381-413.

[22] Zuo, Y. (2005). Robust location and scatter estimators in multivariate analysis. *The Frontiers in Statistics*, Invited book chapter to honor Peter Bickel on his 65th Birthday, Imperial College Press.