PROJECTION-BASED DEPTH FUNCTIONS
AND ASSOCIATED MEDIANS

BY YIJUN ZUO

Arizona State University and Michigan State University

A class of projection-based depth functions is introduced and studied. These projection-based depth functions possess desirable properties of statistical depth functions and their sample versions possess strong and order \( \sqrt{n} \) uniform consistency. Depth regions and contours induced from projection-based depth functions are investigated. Structural properties of depth regions and contours and general continuity and convergence results of sample depth regions are obtained.

Affine equivariant multivariate medians induced from projection-based depth functions are probed. The limiting distributions as well as the strong and order \( \sqrt{n} \) consistency of the sample projection medians are established. The finite sample performance of projection medians is compared with that of a leading depth-induced median, the Tukey halfspace median (induced from the Tukey halfspace depth function). It turns out that, with appropriate choices of univariate location and scale estimators, the projection medians have a very high finite sample breakdown point and relative efficiency, much higher than those of the halfspace median.

Based on the results obtained, it is found that projection depth functions and projection medians behave very well overall compared with their competitors and consequently are good alternatives to statistical depth functions and affine equivariant multivariate location estimators, respectively.

1. Introduction. Depth functions for multivariate data have been pursued in robust and nonparametric data analysis and inference. Among existing notions of depth are Tukey (1975) “halfspace depth,” Liu (1990) “simplicial depth,” and Rousseeuw and Hubert (1999) “regression depth.” The main idea of location depth is to provide a center-outward ordering of points in high dimension relative to a given data set or distribution. Broad treatments of location depth functions are given in Liu, Parelius and Singh (1999) and Zuo and Serfling (2000a). Other studies of depth functions and applications could be found in, for example, Donoho and Gasko (1992), Liu (1995), Liu and Singh (1993, 1997), He and Wang (1997), Rousseeuw and Ruts (1999), Zuo and Serfling (2000b, c, d) and Zhang (2001).

In Zuo and Serfling (2000a), a projection depth function and several other types of depth functions are investigated. It is found that the halfspace and projection...
depth functions (both are implementations of the projection pursuit methodology) appear to represent two very favorable choices among all those examined there. The halfspace depth function and its associated median have received tremendous attention in the literature, whereas not much attention has been paid to the projection depth function. To fill the gap, this paper introduces and studies a class of projection-based depth functions and associated medians, complementing Zuo and Serfling (2000a).

In Section 2, projection-based depth functions and associated depth regions and contours are defined and examples are presented. It is shown that these functions possess the four desirable properties of statistical depth functions introduced by Liu (1990) and Zuo and Serfling (2000a, b) and their sample versions are strongly and \( \sqrt{n} \) uniformly consistent. Depth regions and contours induced from projection depth functions are shown to possess nice structural properties and sample depth contours are proved to converge to their population counterparts.

Section 3 is devoted to the study of the affine equivariant multivariate medians induced from projection-based depth functions. Large and finite sample behavior of sample projection medians are investigated. Strong consistency and limiting distributions of sample projection medians are obtained. Study of the finite sample behavior indicates that, with appropriate choices of univariate location and scale estimators, the sample projection medians can have (simultaneously) a very high breakdown point and relative efficiency, which are much higher than those of the most prevalent depth-based multivariate median, the Tukey halfspace median. (In fact, the breakdown points obtained for the sample projection medians are the highest among all existing affine equivariant multivariate location estimators.) These findings suggest that projection medians are good alternatives of affine equivariant multivariate location estimators to the Tukey halfspace median.

Section 4 ends the paper with some concluding remarks. Selected proofs and auxiliary lemmas are saved for the Appendix.

2. Projection-based depth functions and contours. In this section we study a class of projection-based depth functions. It is a broader generalization of the projection idea behind the Stahel–Donoho (S–D) estimator [Stahel (1981) and Donoho (1982)]. The earlier generalizations were given in Liu (1992) and Zuo and Serfling (2000a).

2.1. Definition and examples. Let \( \mu \) and \( \sigma \) be univariate location and scale measures, respectively. Define the outlyingness of a point \( x \in \mathbb{R}^d \) with respect to (w.r.t.) a given distribution function \( F \) of \( X \) in \( \mathbb{R}^d, d \geq 1 \), as

\[
O(x, F) = \sup_{\|u\|=1} g(x, u, F),
\]

(2.1)

where \( g(x, u, F) = |u'x - \mu(F_u)|/\sigma(F_u) \) and \( F_u \) is the distribution of \( u'X \). Then \( g(x, u, F) \) is defined to be 0 if \( u'x - \mu(F_u) = \sigma(F_u) = 0 \). The projection depth
(PD) of a point \( x \in \mathbb{R}^d \) w.r.t. the given \( F \), \( PD(x, F) \), is then defined as
\[
PD(x, F) = 1/(1 + O(x, F)).
\]

Sample versions of \( g(x, u, F) \), \( O(x, F) \) and \( PD(x, F) \), denoted by \( g_n(x, u) \), \( O_n(x) \) and \( PD_n(x) \), are obtained by replacing \( F \) with its empirical version \( \hat{F}_n \), respectively. Throughout our discussion we assume that \( \mu \) and \( \sigma \) exist uniquely. We also assume that \( \mu \) is translation and scale equivariant and \( \sigma \) is scale equivariant and translation invariant; that is, \( \mu(F_{sY + c}) = s\mu(F_Y) + c \) and \( \sigma(F_{sY + c}) = |s|\sigma(F_Y) \), respectively, for any scalars \( s \) and \( c \) and random variable \( Y \in \mathbb{R}^1 \).

**Remark 2.1.** (i) A specific pair \( (\mu, \sigma) \) results in a specific PD. The characteristics of PD and estimators induced from it thus depend on the choice of \( (\mu, \sigma) \). (ii) With the pair median (Med) and median absolute deviation (MAD), (2.1) has long been used as an outlyingness measure of points in \( \mathbb{R}^d \) \((d \geq 1)\); see Mosteller and Tukey (1977), Stahel (1981) and Donoho (1982). It is natural to characterize the depth of points in terms of their outlyingnesses, as (2.2) does. Of course any monotone decreasing function of \( O(x, F) \) can serve the purpose, but (2.2) ensures \( 0 \leq PD(x, F) \leq 1 \). (iii) PD enjoys desirable properties of depth functions (Section 2.2) and thus provides a center-outward ordering of multivariate points. It induces multivariate quantiles [Serfling (2001a, b)], medians (Section 3), depth-weighted means including as special cases the S–D estimator and multivariate trimmed and winsorized means [Zuo, Cui and He (2001)]. (iv) Moreover, estimators induced from PD can have a very high breakdown point while being extremely efficient whereas those induced from “rank-based” depth (e.g., the halfspace depth) have a relatively lower breakdown point and efficiency [see Section 3.2 and Zuo, Cui and He (2001)]. This is yet another motivation behind PD.

Call the set \( PD^\alpha(F) = \{ x : PD(x, F) \geq \alpha \} \) the \( \alpha \)th projection depth region for \( 0 \leq \alpha \leq 1 \). A sample version of \( PD^\alpha(F) \), \( PD^\alpha_n \), is obtained by replacing \( PD(x, F) \) with \( PD_n(x) \). \( PD^\alpha(F) \) is a multivariate analogue of the univariate \( \alpha \)th quantile region. Call the set \( \{ x : PD(x, F) = \alpha \} \) the \( \alpha \)th projection depth contour. Now let us see two examples of \( PD(x, F) \) and \( PD^\alpha(F) \) with symmetric and asymmetric \( F \) (in the usual sense).

**Example 2.1.** Multivariate standard normal distribution \( F = N_d(0, I) \) in \( \mathbb{R}^d \), \( d \geq 1 \). Consider \((\mu, \sigma) = (\text{Med, MAD})\). It is seen that \( O(x, F) = \|x\|/c_N \), with \( c_N = \Phi^{-1}(\frac{\alpha}{2}) \approx 0.6744898 \), and
\[
PD(x, F) = c_N/(c_N + \|x\|),
\]
where \( \| \cdot \| \) denotes the Euclidean norm. The \( \alpha \)th depth region is then given by
\[
PD^\alpha(F) = \{ x : \|x\| \leq c_N(1 - \alpha)/\alpha \}.
\]
That is, the depth contours are circles in $\mathbb{R}^2$ and spheres in $\mathbb{R}^d$, $d > 2$. See Figure 1. If the underlying distribution $F$ is $N_d(\mu, \Sigma)$, then by affine invariance (see Section 2.2) $PD(x, F) = c_N / (c_N + \sqrt{(x - \mu)'\Sigma^{-1}(x - \mu)})$, and $PD^\alpha(F) = \{x: (x - \mu)'\Sigma^{-1}(x - \mu) \leq (c_N(1 - \alpha))^2/\alpha^2\}$. The depth contours are ellipses in $\mathbb{R}^2$ and ellipsoids in $\mathbb{R}^d$, $d > 2$.

**Example 2.2.** Uniform distribution $F$ over a triangle in $\mathbb{R}^2$. Since all triangles are affine images of a single triangle, we confine attention to the one with vertices $(0, 0)$, $(2, 0)$ and $(0, 2)$. The distribution $F$ in this case is asymmetric and the “center” for a center-outward ordering is not clear (in the usual sense). If we take the univariate mean and standard deviation as $\mu$ and $\sigma$, respectively, then the unique deepest point is $(2/3, 2/3)$, the mean of $F$. To see this, just assume, without loss of generality, that the mean of $F$ is at the origin (see Section 2.2 for affine invariance). It then can be shown that $O(x, F) = \sqrt{6(x_1^2 + x_1x_2 + x_2^2)}$ for any $x = (x_1, x_2)' \in \mathbb{R}^2$. Hence

$$PD(x, F) = 1/(1 + \sqrt{6(x_1^2 + x_1x_2 + x_2^2)}),$$

and the $\alpha$th depth region is given by

$$PD^\alpha(F) = \{x: x_1^2 + x_1x_2 + x_2^2 \leq (1 - \alpha)^2/(6\alpha^2)\},$$

implying that the depth contours are ellipses. See Figure 2.

Now we explore various properties of projection based depth functions and their induced depth regions and contours.

2.2. **Projection-based depth functions.** For a given distribution $F$ in $\mathbb{R}^d$, a functional $T(x, F)$ is said to be affine invariant if $T(Ax + b, F_{AX+b}) = T(x, F_X)$.
PROJECTION DEPTH AND MEDIANS

for any nonsingular $d \times d$ matrix $A$ and vector $b$ and $x$ in $\mathbb{R}^d$; $T(x, F)$ is said to be quasi-concave if $T(\lambda x_1 + (1 - \lambda)x_2, F) \geq \min\{T(x_1, F), T(x_2, F)\}$ for any $0 \leq \lambda \leq 1$ and points $x_1, x_2$ in $\mathbb{R}^d$. For a given univariate location (or “center”) measure $\mu$, a distribution function $F$ is called $\mu$-symmetric about point $\theta \in \mathbb{R}^d$ if $\mu(F_u) = u' \theta$ for any unit vector $u$ in $\mathbb{R}^d$. We have the following theorem.

THEOREM 2.1. For fixed $F$ in $\mathbb{R}^d$ ($d \geq 1$), $PD(x, F)$ is:

(i) affine invariant,
(ii) quasi-concave,
(iii) vanishing at infinity: $PD(x, F) \to 0$ as $\|x\| \to \infty$, and
(iv) maximized at the center of $\mu$-symmetric $F$.

REMARK 2.2. (i) Affine invariance guarantees that $PD(x, F)$ does not depend on the underlying coordinate system and measurement scales while quasi-concavity ensures that $PD(x, F) \leq PD(\theta_0 + \beta(x - \theta_0), F)$ for $\beta \in [0, 1]$ and $\theta_0$ with $PD(\theta_0, F) = \sup_x PD(x, F)$. That is, $PD(x, F)$ decreases monotonically along any ray stemming from the deepest point $\theta_0$. Quasi-concavity also implies the convexity of depth regions (Section 2.2). (ii) A bounded nonnegative function with the four properties: affine invariance, maximality at center, monotonicity relative to deepest point and vanishing at infinity [see Liu (1990) and Zuo and Serfling (2000a, b)] is called a statistical depth function in the latter paper. In light of this, a general $PD(x, F)$ is a statistical depth function for $\mu$-symmetric $F$. Indeed, this was shown in Zuo and Serfling (2000a) for $(\mu, \sigma) = (\text{Med}, \text{MAD})$.

REMARK 2.3. (i) To shed light on $\mu$-symmetry, we consider two cases of $\mu$, the median and mean. If $\mu$ is the median functional, then $\mu$-symmetry is equivalent to halfspace symmetry, a notion, introduced in Zuo and Serfling (2000a, c), broadening spherical, elliptical, antipodal (central), and angular (directional)
symmetry. The latter four [see Liu (1990), Beran and Millar (1997), Liu, Parelius and Singh (1999), Randles (2000)] are increasingly less restrictive. If \( \mu \) is the mean functional, then any \( F \) is \( \mu \)-symmetric about its mean provided that the mean exists. So the choices of median and mean functionals represent two extreme cases of symmetry. (ii) The center \( \theta \) of \( \mu \)-symmetry is unique and \( PD(x, F) \) itself is symmetric about \( \theta \) in \( x \); that is, \( PD(\theta + x, F) = PD(\theta - x, F) \) for \( x \in \mathbb{R}^d \).

Under some mild conditions, \( PD(x, F) \) is uniformly continuous in \( x \) for fixed \( F \) and “continuous” in \( F \) uniformly relative to \( x \). \( PD(x, F) \) is said to be continuous in \( F \) for fixed \( x \) if \( PD(x, F_n) \rightarrow PD(x, F) \) when \( F_n \) converges to \( F \) in distribution \( (F_n \rightarrow_d F) \) as \( n \rightarrow \infty \). Throughout our discussion of convergence in the paper, the measurability of underlying objects is tacitly assumed. Define

\[
\begin{align*}
(C0) \quad & \sup_{\|u\|=1} |\mu(F_n) - \mu(F_u)| < \infty, \quad \sup_{\|u\|=1} \sigma(F_u) < \infty. \\
(C1) \quad & \inf_{\|u\|=1} \sigma(F_u) > 0. \\
(C2) \quad & \sup_{\|u\|=1} |\mu(F_nu) - \mu(F_u)| = o_P(1), \quad \sup_{\|u\|=1} |\sigma(F_nu) - \sigma(F_u)| = o_P(1). \\
(C3) \quad & \sup_{\|u\|=1} |\mu(F_nu) - \mu(F_u)| = O_P(1) \quad \text{a.s.,} \quad \sup_{\|u\|=1} |\sigma(F_nu) - \sigma(F_u)| = o(1) \quad \text{a.s.} \\
(C4) \quad & \sup_{\|u\|=1} |\mu(F_nu) - \mu(F_u)| = O_P(1/\sqrt{n}), \quad \sup_{\|u\|=1} |\sigma(F_nu) - \sigma(F_u)| = O_P(1/\sqrt{n}).
\end{align*}
\]

Here \( F_n \) is not necessarily \( \hat{F}_n \), the empirical version of \( F \).

**Remark 2.4.** (i) For (C0)–(C4) to hold, different choices of \( (\mu, \sigma) \) impose different restrictions on \( F \). (ii) For the pair (mean, standard deviation) and \( F_n = \hat{F}_n \), (C0)–(C4) hold for any \( F \) with a positive definite covariance matrix. (iii) For the pair (median, mean) and \( F_n = \hat{F}_n \), (C0)–(C4) hold for any \( F \) satisfying the conditions in Theorem 3.3 (but not necessarily \( \mu \)-symmetric about a point). (iv) For general \( M \)-functionals \( (\mu, \sigma) \), conditions for (C0)–(C4) to hold are addressed in Zuo, Cui and He (2001).

**Theorem 2.2.** Under (C0) and (C1) we have:

(i) \( PD(x, F) \) is uniformly continuous in \( x \),

(ii) \( \sup_{x \in \mathbb{R}^d} |PD(x, F_n) - PD(x, F)| = o_P(1) \) if (C2) holds,

(iii) \( \sup_{x \in \mathbb{R}^d} |PD(x, F_n) - PD(x, F)| = o(1) \) a.s. if (C3) holds, and

(iv) \( \sup_{x \in \mathbb{R}^d} |PD(x, F_n) - PD(x, F)| = O_P(1/\sqrt{n}) \) if (C4) holds.

**Remark 2.5.** (i)–(iv) in the theorem can be strengthened to (i)*–(iv)*: Lipschitz continuous, \( \sup_{x \in \mathbb{R}^d} |PD(x, F_n) - PD(x, F)|(1 + \|x\|) = o_P(1) = o(1) \) a.s. and \( = O_P(1/\sqrt{n}) \), respectively. (iv)* is crucial in establishing asymptotic normality of \( PD \)-weighted means [see Zuo, Cui and He (2001)]. For the halfspace and simplicial depth, (i) holds for absolutely continuous \( F \) and (ii)–(iv) hold with no restriction on \( F \). (i)*–(iv)*, however, do not hold for them in general. (ii) Zuo and Serfling (2000b) established (iii) for \( (\mu, \sigma) = (\text{Med}, \text{MAD}) \) and \( F_n = \hat{F}_n \).
It is straightforward to verify that sample projection depth functions share many of the above properties of their population counterparts.

2.3. Projection depth induced regions and contours. \( PD^\alpha(F) \) is called affine equivariant if \( PD^\alpha(F_{AX+b}) = A(PD^\alpha(F)) + b \) for any nonsingular \( d \times d \) matrix \( A \) and vector \( b \in \mathbb{R}^d \). Let \( \alpha^* = \sup_{x \in \mathbb{R}^d} PD(x, F) \). For convenience we sometimes drop the \( F \) in \( PD^\alpha(F) \). Denote the interior of \( PD^\alpha \) by \( (PD^\alpha)^o \) and the boundary of \( PD^\alpha \) by \( \partial PD^\alpha \).

**Theorem 2.3.** Let \( PD^\alpha(F) \) be the \( \alpha \)th projection depth region for a given \( F \).

(i) \( PD^\alpha(F) \) is affine equivariant, nested, and convex,
(ii) \( PD^\alpha(F) \) is bounded, closed, and hence compact for any \( \alpha > 0 \) if (C0) and (C1) hold,
(iii) \( PD^\alpha(F) \) is symmetric about \( \theta \) if \( F \) is \( \mu \)-symmetric about \( \theta \),
(iv) \( \partial PD^\alpha(F) \) is \( \{x : PD(x, F) = \alpha\} \) if (C0) and (C1) hold and
(v) \( (PD^\alpha)^o \) is \( \emptyset \) under (C0), \( PD^\alpha \neq \emptyset \) if (C0) and (C1) hold.

Under the conditions given below, the projection depth regions are continuous in both \( \alpha \) and \( F \). For other related discussions of the continuity of depth regions, see Nolan (1992), Massé and Theodorescu (1994), He and Wang (1997), Kim (2000) and Zuo and Serfling (2000b). For a sequence \( A_1, A_2, \ldots \) of sets, define \( \limsup_n A_n = \bigcap_{n=1}^\infty \bigcup_{k=n}^\infty A_k \) and \( \liminf_n A_n = \bigcup_{n=1}^\infty \bigcap_{k=n}^\infty A_k \). Write \( A_n \rightarrow A \) as \( n \rightarrow \infty \) if \( \limsup_n A_n = \liminf_n A_n = A \). Write \( A_n \overset{a.s.}{\rightarrow} A \) with \( A = \limsup_n A_n \) if \( P(\{x : x \in \limsup_n A_n \text{ but } x \notin \liminf_n A_n\}) = 0 \).

**Theorem 2.4.** Let \( PD^\alpha(F) \) be the \( \alpha \)th depth region for a given \( F \).

(i) \( PD^\alpha_n(F) \overset{a.s.}{\rightarrow} PD^\alpha(F) \) if \( \alpha_n \rightarrow \alpha \) and \( P(\{x : PD(x, F) = \alpha\}) = 0 \),
(ii) \( PD^\alpha_n(F_n) \overset{a.s.}{\rightarrow} PD^\alpha(F) \) if \( \alpha_n \rightarrow \alpha \), \( P(\partial PD^\alpha) = 0 \), and (C0), (C1) and (C3) hold.

We say \( PD(x, F) \) decreases most slowly along a ray \( u \) stemming from the deepest point \( \theta \) with \( PD(\theta, F) = \alpha^* \) if for any \( x_{a_1}^u, x_{a_2}^u \) on \( u \) and \( x_{a_1}^v, x_{a_2}^v \) on any other ray \( v \) from \( \theta \) with \( PD(x_{a_1}^u, F) = PD(x_{a_1}^v, F) = \alpha_i \), \( i = 1, 2 \), \( \|x_{a_1}^u - x_{a_2}^u\| \geq \|x_{a_1}^v - x_{a_2}^v\| \). Such a direction exists in many cases, especially in the case that \( F \) is elliptically distributed and \( \mu = \text{Med} \) and \( \sigma = \text{MAD} \). For two sets, \( A \) and \( B \), the Hausdorff distance between them, \( \rho(A, B) \), is \( \inf\{\varepsilon : \varepsilon > 0, A \subseteq B+\varepsilon, B \subseteq A+\varepsilon\} \), where \( A^\varepsilon = \{x \mid d(x, A) < \varepsilon\} \) and \( d(x, A) = \inf\{\|x - y\| : y \in A\} \). The depth regions are continuous in \( \rho \) as well as in \( \alpha \) in the following sense.

**Theorem 2.5.** Let (C0) and (C1) hold and \( \rho \) be defined as above.
8 Y. Zuo

(i) \( \alpha_n \to \alpha \) if \( \rho(PD^{\alpha_n}, PD^\alpha) \to 0 \) as \( n \to \infty \),

(ii) \( \rho(PD^{\alpha_n}, PD^\alpha) \to 0 \) if \( \alpha_n \to \alpha(< \alpha^*) \) and \( PD(x, F) \) decreases strictly along any ray from \( \theta \) with \( PD(\theta, F) = \alpha^* \) and most slowly along a ray \( u \).

It is straightforward to verify that sample depth regions share many of the properties of their population versions.

3. Projection depth-induced medians. For a given PD, define the point with maximum depth as a multivariate analogue of the univariate median. That is, a median induced from PD, called projection median (PM), can be defined as

\[
PM(F) = \arg \sup_{x \in \mathbb{R}^d} PD(x, F).
\]

Tyler (1994) also obtained PM based on a slightly different approach. The nonuniqueness problem in the definition can be handled with a fixed rule (such as taking average). By Theorem 2.3, \( PM(F) \) is well defined if \((C0)\) and \((C1)\) hold. In \( \mathbb{R}^1 \), it reduces to the univariate median if \( \mu = \text{Med} \). Like its univariate counterpart, \( PM(F) \) is affine equivariant. That is, \( PM(AX + b) = A(PM(X)) + b \) for any sample \( X \), nonsingular \( d \times d \) matrix \( A \) and vector \( b \in \mathbb{R}^d \). \( PM(F) \) is able to identify the center of symmetry of any \( \mu \)-symmetric \( F \); see Zuo and Serfling (2000c) for a related discussion.

For a given sample \( X_n = \{X_1, \ldots, X_n\} \) from \( F \), a sample version of \( PM(F) \), \( PM_n = PM(\hat{F}_n) \), is obtained (take on average if necessary to deal with the nonuniqueness problem). \( PM_n \) is affine equivariant; that is, \( PM_n(AX_n + b) = A(PM_n) + b \) for any sample \( X_n \) from \( X \), nonsingular \( d \times d \) matrix \( A \) and vector \( b \in \mathbb{R}^d \). If \( X \) is centrally symmetric about a point \( \theta \in \mathbb{R}^d \); that is, \( X - \theta \) and \( \theta - X \) have the same distribution, then the probability distribution of \( PM_n \) itself is also centrally symmetric about \( \theta \) [see Corollary 1.3.19 of Randles and Wolfe (1979)]. Further, if the expectation of this centrally symmetric \( X \) exists, then \( PM_n \) is an unbiased estimator of the location parameter \( \theta \). Under some mild conditions, \( PM_n \) is a consistent estimator of \( PM(F) \) and has a limiting distribution. Now we investigate the large and finite sample behavior of the sample projection medians.

3.1. Large sample behavior. In the following we establish first the strong and \( \sqrt{n} \) consistency and then limiting distributions of the sample projection medians. Lemma 3.3 of Bai and He (1999) turns out to be very important in establishing the limiting distributions.

**Theorem 3.1.** Assume that \((C0)\) and \((C1)\) hold and \( \theta \) is the unique point with \( \theta = \arg \sup_{x \in \mathbb{R}^d} PD(x, F) \). Let \( PM(F_n) = \arg \sup_{x \in \mathbb{R}^d} PD(x, F_n) \). Then:

(i) \( PM(F_n) - PM(F) = o(1) \) a.s. if \((C3)\) holds, and

(ii) \( PM(F_n) - PM(F) = O_P(1/\sqrt{n}) \) if \((C4)\) holds and \( F \) is \( \mu \)-symmetric.
Remark 3.1. (i) (C4) is more than we need for part (ii) of the theorem. The first part of (C4) and the second part of (C2) suffice. (ii) Consistency of $PM(F_n)$ can be established accordingly.

A natural question raised after one has the strong and $\sqrt{n}$ consistency of $PM(F_n)$ is: Does $PM(F_n)$ possess a limiting distribution? We answer the question for a general class of $\mu$ and $\sigma$ with $\mu_u = \mu(F_u)$ and $\sigma_u = \sigma(F_u)$ being the simultaneous $M$-functionals of location and scale [see Huber (1981)] and defined by $\lambda(\eta_u) = E_{F_u} \Psi((x - \mu_u)/\sigma_u) = 0$ for $x \in \mathbb{R}^1$, where $\Psi(\cdot) = (\psi(\cdot), \chi(\cdot))'$ and $\eta_u = (\mu_u, \sigma_u)'$. That is,

\begin{align}
\int \psi \left( \frac{x - \mu_u}{\sigma_u} \right) F_u(dx) &= 0, \\
\int \chi \left( \frac{x - \mu_u}{\sigma_u} \right) F_u(dx) &= 0.
\end{align}

It is readily seen that $\mu_u$ is translation and scale equivariant and $\sigma_u$ is scale equivariant and translation invariant. Then $\psi$ and $\chi$ in (3.1) and (3.2) are usually odd and even, respectively. Typical choices of them include $\psi(x) = \text{sign}(x)$ and $\chi(x) = \text{sign}(|x| - 1)$, which lead to the median (for $\mu$) and the median absolute deviation from 0 (for $\sigma$). Another choice is $\psi(x) = \text{Med}\{-k, k, x\}$ and $\chi(x) = (\psi(x))^2 - \int \min(k^2, x^2) \Phi(dx)$; see Huber (1981) for other popular choices of $\psi$ and $\chi$. If $\lambda$ has a nonsingular derivative matrix $\Lambda_u$ for each $u$, then

$$
\Lambda_u = \begin{pmatrix}
a_u & b_u \\
c_u & d_u
\end{pmatrix},
$$

with $a_u = -\int \psi'(y_u) F_u(dy_u)$, $b_u = -\int \psi'(y_u) y_u F_u(dy_u)$, $c_u = -\int \chi'(y_u) \times F_u(dy_u)$ and $d_u = -\int \chi'(y_u) y_u F_u(dy_u)$, where $y_u = (x - \mu_u)/\sigma_u$. Define for bounded $\mu_u$ and $\sigma_u$,

$$
F = \left\{ \frac{d_u \psi(h_u(\cdot)) - b_u \chi(h_u(\cdot))}{\det \Lambda_u} \|u\| = 1 \right\},
$$

where $h_u(x) = (u' x - \mu_u)/\sigma_u$ for any $x \in \mathbb{R}^d$. Under mild conditions, $F$ is a permissible class and the graphs of functions in $F$ form a polynomial class of sets [see Pollard (1984)]. Define

(C5) $\eta_{un} - \eta_u = -\frac{1}{n} \sum_{i=1}^{\eta} \Lambda_u^{-1} \Psi \left( \frac{u' X_i - \mu_u}{\sigma_u} \right) + o_p(1/\sqrt{n})$, uniformly in $u$ with $\eta_{un} = (\mu(F_{un}), \sigma(F_{un}))'$. Let $Y_u = (u' X - \mu_u)/\sigma_u$. Assume (w.l.o.g.) that $E \Psi(Y_u) = 0$.

Theorem 3.2. Let $\mu_u = \mu(F_u)$ and $\sigma_u = \sigma(F_u)$ be determined by (3.1) and (3.2). Assume that:

\begin{itemize}
\item [(i)] $\mu_u$ and $\sigma_u$ are translation and scale equivariant, respectively.
\item [(ii)] $\mu_u$ and $\sigma_u$ are translation invariant, respectively.
\item [(iii)] $\psi$ and $\chi$ are odd and even functions, respectively.
\item [(iv)] $\lambda(\eta_u) = E_{F_u} \Psi((x - \mu_u)/\sigma_u) = 0$ for $x \in \mathbb{R}^1$.
\item [(v)] $\Lambda_u$ has a nonsingular derivative matrix $\Lambda_u$ for each $u$.
\end{itemize}
(i) (C0) and (C1) hold, $F$ is $\mu$-symmetric, and the density $f$ of $F$ and the gradient $\mathbf{f}$ of $f$ exist,
(ii) $\inf_{u} \det \Lambda_{u} > 0$ and $\sup_{u} |d_{u}|$, $\sup_{u} |b_{u}|$, $E(\sup_{u} |\psi(Y_{u})|)^{2}$ and $E(\sup_{u} |\chi(Y_{u})|)^{2}$ exist,
(iii) $\mathcal{F}$ is a permissible class of functions whose graphs form a polynomial class of sets, and
(iv) (C5) holds.

Then
\[ \sqrt{n}(PM(\hat{F}_{n}) - PM(F)) \xrightarrow{d} \arg\inf_{u} \sup_{\|u\|=1} \left| (u' - Z(u))/\sigma(F_{u}) \right|, \]
where $Z(u)$ is a Gaussian process on the unit sphere with mean zero and covariance structure
\[ \text{Cov}(Z(u_{1}), Z(u_{2})) = \frac{E(\{d_{u_{1}}\psi(Y_{u_{1}}) - b_{u_{1}}\chi(Y_{u_{1}})\}[d_{u_{2}}\psi(Y_{u_{2}}) - b_{u_{2}}\chi(Y_{u_{2}})])}{(a_{u_{1}}d_{u_{1}} - b_{u_{1}}c_{u_{1}})(a_{u_{2}}d_{u_{2}} - b_{u_{2}}c_{u_{2}})}. \]

Remark 3.2. (i) Assume that $X \sim F$ is $\mu$-symmetric about 0 and $u'X =_{d} a(u)Y$ for each $\|u\| = 1$ with $a(u) \geq 0$ being continuous and even in $u$ and $Y$ having a density $f_{u}(y)$ continuous and even in $y$, where “$=_{d}$” stands for “equal in distribution.” Such $F$ includes elliptically symmetric $F$ with $a(u) = \sqrt{u'\Sigma u}$ for some positive definite matrix $\Sigma$ and $\alpha$-symmetric $F$ with $a(u) = (\sum_{i=1}^{d} a_{i}|u_{i}|^{2\alpha})^{1/\alpha}$, $\alpha > 0$ [Fang, Kotz an Ng (1990)]. The part (i) of the above theorem then holds trivially when such $F$ is smooth. If $\psi$ and $\chi$ are (almost surely) continuously differentiable, $\psi$ has a zero at $x = 0$ and $\chi$ has a minimum at $x = 0$, and $\psi' > 0$ and $\chi'/\psi'$ is strictly monotone, then $\Lambda_{u}$ is nonsingular [specific examples of such $\psi$ and $\chi$ include $\psi(\cdot) = c^{2}\arctan(\cdot)$ for any constant $c \neq 0$ and $\chi(\cdot) = \psi(\cdot)^{2} + \beta$; strict monotonicity of $\psi$ can be slightly relaxed; see pages 137–139 of Huber (1981) for the argument and other examples of $\psi$ and $\chi$]. The part (ii) of the theorem thus holds in the light of continuity and compactness as long as $E(\psi^{2}(Y_{u}))$ and $E(\chi^{2}(Y_{u}))$ exist. The part (iii) of the theorem holds for the given $\psi$ and $\chi$ [see Pollard (1984), Examples II.26 and VII.18 and Problem II.18]. Condition (4) [i.e., (C5)] holds for the given $\psi$ and $\chi$; see Lemma 3.2 of Zuo, Cui and He (2001) for further discussions related to (C5). The uniformity in $u$ of the remainder term is needed to handle stochastic processes (and $\sup_{u}$) involved. (ii) The limiting distribution in the theorem is not convenient for use in practice. However, bootstrapping techniques can be used to approximate the distribution of $\sqrt{n}(PM(\hat{F}_{n}) - PM(F))$ and to construct confidence regions for $PM(F)$.

For the special case $(\mu, \sigma) = (\text{Med, MAD})$, we have the following theorem.

Theorem 3.3. Let $(\mu, \sigma) = (\text{Med, MAD})$. Assume that $F$ is $\mu$-symmetric about $\theta \in \mathbb{R}^{d}$ with density $f$, $F_{u}$ is twice differentiable at $\mu(F_{u})$ with the first-
order derivative $f_u$, $\inf_{\|u\| = 1} f_u(\mu(F_u)) > 0$, and $\inf_{\|u\| = 1} (f_u(\mu(F_u) + \sigma(F_u)) + f_u(\mu(F_u) - \sigma(F_u))) > 0$. Then

$$
\sqrt{n}(PM(\hat{F}_n) - PM(F)) \xrightarrow{d} \arg \inf_t \sup_{\|u\| = 1} \left| \frac{u't - Z(u)}{\sigma(F_u)} \right|,
$$

where $Z(u)$ is a Gaussian process on the unit sphere with mean zero and covariance structure $\text{Cov}(Z(u_1), Z(u_2)) = (1/4 - \arccos(u_1^t u_2)/(2\pi))/(f_\sigma(0))^2$ with $f_\sigma$ being any marginal density. (ii) Tyler (1994) stated, based on a heuristic argument, the limiting distribution of $PM(\hat{F}_n)$ for spherically symmetric $F$. The above theorem includes his result as a special case.

**Remark 3.3.** (i) The conditions in the theorem are satisfied if $F$ is a smooth elliptically symmetric distribution. When $F$ is spherically symmetric about the origin, the covariance structure becomes $\text{Cov}(Z(u_1), Z(u_2)) = (1/4 - \arccos(u_1^t u_2)/(2\pi))/(f_\sigma(0))^2$ with $f_\sigma$ being any marginal density. (ii) Tyler (1994) stated, based on a heuristic argument, the limiting distribution of $PM(\hat{F}_n)$ for spherically symmetric $F$. The above theorem includes his result as a special case.

**Remark 3.4.** By Theorems 3.2 and 3.3, it is seen that the limiting distribution of $PM_n$ does not depend on the choice of $\sigma$ as long as $\sigma(\cdot)$ is Fisher consistent (up to a fixed scalar) w.r.t. a given scale parameter of $F_u$ for each $u$. For example, consider the $\mu$-symmetric $F$ in (i) of Remark 3.2. Suppose $\eta$ is the scale parameter of $Y$. Then any $\sigma$ with $\sigma(F_u) = \kappa_\sigma a(u)\eta$ for a constant $\kappa_\sigma$ leads to the same limiting distribution of $PM_n$.

### 3.2. Finite sample breakdown point

For an appropriate choice of $(\mu, \sigma)$, $PM_n$ is a robust location estimator in the sense that it possesses a very high breakdown point (in fact, its breakdown point can be higher than that of any existing affine equivariant location estimators). The notion of a finite sample breakdown point was introduced in Donoho and Huber (1983) and has become the most prevalent quantitative assessment of global robustness of estimators. Let $X^n = \{X_1, \ldots, X_n\}$ be a sample of size $n$ in $\mathbb{R}^d$, $d \geq 1$. The replacement breakdown point (RBP) of an estimator $T$ at $X^n$ is defined as

$$
RBP(T, X^n) = \min \left\{ \frac{m}{n} : \sup_{X_m^n} \| T(X_m^n) - T(X^n) \| = \infty \right\},
$$

where $X_m^n$ denotes a contaminated sample from $X^n$ by replacing $m$ points of $X^n$ with arbitrary values. In other words, the RBP of an estimator is the minimum replacement fraction that could drive the estimator beyond any bound.

In the following discussion, $(\mu, \sigma) = (\text{Med}, \text{MAD}_k)$, where $\text{MAD}_k$ is a modified version of MAD. $\text{MAD}_k(x^n) = \text{Med}_k(\{|x_1 - \text{Med}(x^n)|, \ldots, |x_n - \text{Med}(x^n)|\})$, where

$$
\text{Med}_k(x^n) = (x_{[(n+k)/2]} + x_{[(n+1+k)/2]})/2, \quad 1 \leq k \leq n.
$$
$x^n = \{x_1, \ldots, x_n\}$, $x(1) \leq \cdots \leq x(n)$, are ordered values of $x_1, \ldots, x_n$ in $\mathbb{R}^1$, and $\lfloor x \rfloor$ is the largest integer less than or equal to $x$. Denote the corresponding projection median by $PM_n^k$. The Med–MAD combination corresponds to $PM_n^k$ with $k = 1$. A random sample $X^n$ in $\mathbb{R}^d$, $d \geq 1$, is said to be in general position if there are no more than $d$ data points of $X^n$ contained in any $(d - 1)$-dimensional subspace.

**THEOREM 3.4.** Let $(\mu, \sigma) = (\text{Med, MAD}_k)$ and $X^n$ be in general position with $n \geq 2(d - 1)^2 + k + 1$, where $k \leq (d - 1)$ when $d \geq 2$. Then

$$RBP(\text{PM}_n^k, X^n) = \begin{cases} 
\lfloor (n + 1)/2 \rfloor, & \text{for } d = 1, \\
\frac{n}{\lfloor (n - 2d + k + 3)/2 \rfloor}, & \text{for } d \geq 2.
\end{cases}$$

**REMARK 3.5.** (i) For $X^n$ in general position, $d = 1$, and $1 \leq k \leq n$, the affine equivariant, hence necessarily translation equivariant, location estimator $PM_n^k$ achieves the best possible RBP of any translation equivariant location estimators [see Lopuhaä and Rousseeuw (1991)]. (ii) For $X^n$ in general position and $d \geq 2$, when $k = d - 1$ and $n \geq 2(d - 1)^2 + d$, $PM_n^k$ achieves its maximum RBP, $\lfloor (n - d + 2)/2 \rfloor/n$, which is the highest among (and can be higher than) the RBPs of any existing affine equivariant location estimators in the literature (the best RBP in the literature is $\lfloor (n - d + 1)/2 \rfloor/n$). (iii) The idea of modifying the Med or the MAD to achieve a higher breakdown point for the related estimators appeared in a personal communication of Siegel and Rousseeuw; see Rousseeuw (1984). Tyler (1994) employed the same idea and modified the MAD in the S–D location and scatter estimators. The above modification of MAD is similar to that of Gather and Hilker (1997) and related to (but different from) Tyler’s. (iv) Tyler (1994) stated the breakdown point of the projection median. The RBP result above is general and does not follow from his. (v) The approach to the breakdown point here is somewhat different from some existing approaches in the literature in the sense that we define $|u' x - \text{Med}(u'X^n)|/\text{MAD}(u'X^n) = 0$ when $|u' x - \text{Med}(u'X^n)| = \text{MAD}(u'X^n) = 0$, because we think $u' x$ is at the center in this case (and hence has an outlyingness 0), whereas some other authors think the estimator breaks down whenever $\text{MAD}(u'X^n) = 0$.

**REMARK 3.6.** Theorem 3.4 focuses on the choice $(\text{Med, MAD}_k)$. The result in the theorem, however, can be extended for general $(\mu, \sigma)$. Call the RBP of $\mu(F_nu)$ or $\sigma(F_nu)$ over all directions $u$ the uniform RBP of $\mu$ or $\sigma$ [Tyler (1994)]. Then the RBP of $PM_n$ based on general $(\mu, \sigma)$ will be no less than the minimum of the uniform RBPs of $\mu$ and $\sigma$ [Tyler (1994)]. The RBP of the projection medians thus depend on the uniform RBPs of $\mu$ and $\sigma$. 
Both halfspace and projection medians can resist six contaminating points in a data set of 20 points.

The halfspace median $HM_n$, induced from the Tukey halfspace depth (HD), is one of the most popular depth-based medians. $HM_n$, however, has a relatively lower breakdown point since “rank-based” HD focuses mainly on relative positions (not “distances”) of points to the center of data. In fact, the large sample breakdown point of $HM_n$ was shown no higher than $1/3$ for continuous and angularly symmetric $F$ [Donoho and Gasko (1992)]. PD, on the other hand, appreciates the information of relative positions as well as distances of points to the center of data. Consequently, the induced median (with robust choice of $\mu$ and $\sigma$) is expected to have a higher breakdown point. The difference between the breakdown points of $HM_n$ and $PM_n$ ($k = 1$) is illustrated in Figures 3 and 4. Figure 3 shows that both $HM_n$ and $PM_n$ can resist six contaminating points in a data set of 20 standard bivariate normal points without breakdown. Figure 4 shows halfspace median can break down with seven contaminating points in a data set of 20 points, while to break down projection median, 50% points need to be contaminated.
that contaminating 1/3 of data points can break down \( HM_n \) while to break down \( PM_n \), 50% of original points need to be contaminated. So from the breakdown point of view, \( PM_n \) is a better alternative as an affine equivariant location estimator than \( HM_n \). The questions now are: What is the relative efficiency of \( PM_n \)? Is it less or more efficient than \( HM_n \) or is it at least comparable to \( HM_n \) in efficiency?

3.3. Finite sample relative efficiency. We generate 1000 samples from the bivariate standard normal distribution for several sample sizes. A slightly modified version of HALFMED of Rousseeuw and Ruts (1998) is used for computing \( HM_n \). The time complexity of the algorithm is \( O(n^2 \log^2 n) \) for fixed \( d = 2 \). An approximate algorithm with time complexity \( O(n^3 + Nn^2) \) for fixed \( d = 2 \) is utilized for computing \( PM_n \). Here \( N \) is the iteration number in the downhill simplex algorithm [Press, Teukolsky, Vetterling and Flannery (1997)] employed in the computation. First we consider \((\mu, \sigma) = (\text{Med}, \text{MAD})\). We calculate for estimator \( T \) the empirical mean squared error (EMSE): 
\[
\frac{1}{m} \sum_{i=1}^{m} \| T_i - \theta \|^2,
\]
where \( m = 1000\), \( \theta = (0,0)' \) and \( T_i \) is the estimate for the \( i \)th sample. The relative efficiency (RE) of \( T \) is then obtained by dividing the EMSE of the sample mean by that of \( T \). See Table 1 for the results. It turns out that the RE of \( HM_n \) is around 77%, which is consistent with what obtained in Rousseeuw and Ruts (1998), where they studied the finite sample relative efficiency of the halfspace and coordinatewise medians. (The coordinatewise median is not affine equivariant and has RE about 66%.) The RE of the projection median for the choice of \((\text{Med, MAD})\) is around 78% and is slightly higher than that of the halfspace median. The latter is also true for bivariate Cauchy distributions.

The RE of \( PM_n \) with \((\text{Med, MAD})\), albeit slightly higher than that of \( HM_n \), is not very high. This, in fact, is common for many high breakdown affine equivariant

<table>
<thead>
<tr>
<th>( n )</th>
<th>( HM_n )</th>
<th>( PM_n ) (Med, MAD)</th>
<th>( PM_n ) (PWM, MAD)</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>EMSE 0.2390</td>
<td>0.2413</td>
<td>0.2077</td>
<td>0.1885</td>
</tr>
<tr>
<td>RE 0.79</td>
<td>0.78</td>
<td>0.91</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>EMSE 0.1290</td>
<td>0.1290</td>
<td>0.1085</td>
<td>0.0987</td>
</tr>
<tr>
<td>RE 0.77</td>
<td>0.77</td>
<td>0.91</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>EMSE 0.0947</td>
<td>0.0935</td>
<td>0.0790</td>
<td>0.0717</td>
</tr>
<tr>
<td>RE 0.76</td>
<td>0.77</td>
<td>0.91</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>EMSE 0.0698</td>
<td>0.0682</td>
<td>0.0585</td>
<td>0.0530</td>
</tr>
<tr>
<td>RE 0.76</td>
<td>0.78</td>
<td>0.91</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>EMSE 0.0579</td>
<td>0.0570</td>
<td>0.0477</td>
<td>0.0433</td>
</tr>
<tr>
<td>RE 0.75</td>
<td>0.76</td>
<td>0.91</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>EMSE 0.0457</td>
<td>0.0445</td>
<td>0.0386</td>
<td>0.0355</td>
</tr>
<tr>
<td>RE 0.78</td>
<td>0.80</td>
<td>0.92</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>
location estimators. Note that the choice $\mu = \text{mean}$ can lead to the best possible RE of $PM_n$, 100%, but $PM_n$ in this case is no longer robust (see Remark 3.6). A natural question raised here is: Can we improve the efficiency of $PM_n$ without sacrificing its robustness? The answer is positive. With a robust $M$-functional $\mu$, the RE of $PM_n$ can lie between 76% to 100%. Take $\mu$ to be the projection depth weighted mean (PWM) in $\mathbb{R}^1$,

$$PWM(x^n) = \frac{\sum_{i=1}^{n} w(PD_n(x_i))(x_i)}{\sum_{i=1}^{n} w(PD_n(x_i))},$$

where $w(r) = I(r < c)(\exp(-k(1-r/c)^2)-\exp(-k))/(1-\exp(-k))+I(r \geq c)$, $PD_n(x_i) = 1/(1+|x_i - \text{Med}(x^n)|/\text{MAD}(x^n))$ and $x^n = \{x_1, \ldots, x_n\}$ with $x_i \in \mathbb{R}^1$, the RE of $PM_n$ can be higher than 90% in two dimensions (and yet higher in higher dimensions) while keeping its RBP in Theorem 3.4. For discussions of $w$ and parameters $k$ and $c$, see Zuo, Cui and He (2001). The RE of $PM_n$, with $k = 2$ and $c = \text{Med}(PD_n(x_1), \ldots, PD_n(x_n))$, is listed in Table 1.

**Remark 3.7.** The RE of $PM_n$ depends on the choice of $\mu$. It, however, does not depend on the choice of $\sigma$. This latter assertion is also confirmed by our simulation study. For example, the RE of $PM_n$ is almost the same with $\sigma = \text{MAD}$ or standard deviation (SD) and $\mu = \text{Med}$. The same is true with $\sigma = \text{MAD}$ or SD and $\mu = \text{mean}$. The phenomenon here is not surprising because the limiting distribution of $PM_n$ depends on the choice of $\mu$ but not $\sigma$ (see Remark 3.4).

**4. Concluding remarks.** This paper introduces and studies a class of projection-based depth functions and their associated medians. The depth functions enjoy desirable properties and their sample versions possess strong and $\sqrt{n}$ uniform consistency. Depth regions and contours induced from these functions have nice structural properties. Multivariate medians associated with these functions share many desirable properties. For example, they are affine equivariant and can identify the center of any $\mu$-symmetric distribution. Sample projection medians are unbiased for the center of centrally symmetric distributions and strongly and $\sqrt{n}$ consistent and possess a very high breakdown point (which can be higher than that of any existing affine equivariant location estimators) with robust choices of univariate location and scale estimators. Furthermore, under mild conditions limiting distributions of the sample projection medians exist. The limiting distributions are nonnormal in general, which makes them difficult to be used in practical inference; nevertheless, bootstrapping techniques can be employed for this end. The complex and non-Gaussian limiting distributions also make it difficult to obtain clear insight into the asymptotic relative efficiency of projection medians. Instead, finite sample relative efficiency of the medians are investigated. Compared with the leading depth-based median, the Tukey halfspace median, the projection medians are favored in the sense that with appropriate univariate location and scale
estimators they have a much higher finite sample breakdown point as well as relative efficiency. (In view of the high breakdown point and relative efficiency, the projection medians remain highly competitive among leading affine equivariant location estimators.)

Like other high breakdown estimators, projection medians (and depth) are computationally intensive. Algorithms for projection medians for the moment take a longer time than those for the halfspace median; faster ones for projection medians are expected to be developed, though. Computing issues of projection medians will be addressed elsewhere. For the computing of halfspace median, see Rousseeuw and Ruts (1998) and Struyf and Rousseeuw (2000).

APPENDIX:
SELECTED PROOFS AND AUXILIARY LEMMAS

PROOF OF REMARK 2.4. The proof for (ii) is straightforward and thus skipped. We now show (iii). By the continuity of $\mu(F_u)$ and $\sigma(F_u)$ in $u$ (see Lemma A.1) and the compactness of $\{u: \|u\| = 1\}$, (C0) and (C1) follow immediately. The given conditions permit the uniform asymptotic representations

$$\mu(F_{nu}) - \mu(F_u) = \int f_1(x, u)(F_n - F)(dx) + R_{1n},$$

$$\sigma(F_{nu}) - \sigma(F_u) = \int f_2(x, u)(F_n - F)(dx) + R_{2n},$$

with $\sup_u |R_{in}| = O(n^{-3/4} \log n)$, $i = 1, 2$, and

$$f_1(\cdot, u) \in F_1 = \left\{ \frac{I_{u \cdot \leq \mu(F_u)}}{f_u(\mu(F_u))} : \|u\| = 1 \right\},$$

$$f_2(\cdot, u) \in F_2 = \left\{ \frac{I_{|u \cdot - \mu(F_u)| \leq \sigma(F_u)}}{f_u(\mu(F_u) + \sigma(F_u)) + f_u(\mu(F_u) - \sigma(F_u))} : \|u\| = 1 \right\}.$$

Note that the graphs of both $F_1$ and $F_2$ form a polynomial class of sets [see Example II.26 of Pollard (1984)]. (C2)–(C4) follow immediately from Theorem II.24 and Lemma VII.15 and Theorem VII.21 of Pollard (1984). □

LEMMA A.1. Let $f(u) = \text{Med}(F_u)$ and $g(u) = \text{MAD}(F_u)$ for any unit vector $u$. If $F_u$ and $F_{u'X - f(u)}$ are not flat in a right-neighborhood of $f(u)$ and $g(u)$, respectively, then $f(u)$ and $g(u)$ are continuous.

Invoking Slutsky’s theorem and Lemma A.2, we obtain the desired result.

LEMMA A.2. If $F_n \rightarrow_d F$ as $n \rightarrow \infty$ and $F(F^{-1}(p) + \varepsilon) > p$ for any $\varepsilon > 0$ and $0 < p < 1$, then $F_n^{-1}(p) \rightarrow F^{-1}(p)$ as $n \rightarrow \infty$. 
By mimicking the proof of Theorem 2.3.1 of Serfling (1980), the desired result follows.

**REMARK A.1.** The nonflatness condition in Lemmas A.1 and A.2 cannot be dropped. For example, consider \( X \in \mathbb{R}^2 \) such that (1) the mass on point \((1, 0)\) is \(1/2\), (2) the mass on \(\{(0, y) : y \geq a\}\) and \(\{(0, y) : y \leq -a\}\) is positive for any \(a > 0\) and (3) \(\text{Med}((1, 0)X) = 0\). Then as unit vectors \(u_n(\neq u_0) \rightarrow u_0 = (1, 0)\), \(\text{Med}(u_n'X) - \text{Med}((1, 0)X) > 1/2\) as \(n \rightarrow \infty\).

**PROOF OF THEOREM 2.2.** Since for any \(x, y \in \mathbb{R}^d\),
\[
|\text{PD}(x, F) - \text{PD}(y, F)| = \frac{|O(x, F) - O(y, F)|}{(1 + O(x, F))(1 + O(y, F))} \leq |O(x, F) - O(y, F)|,
\]
most parts of the following proof thus are focused on the outlyingness functions. The corresponding results for projection depth functions follow immediately from the above inequality.

For any given \(\varepsilon > 0\), let \(x, y \in \mathbb{R}^d\) such that \(\|x - y\| \leq \inf \|u\| = 1 \sigma(F_u)\varepsilon\). We have
\[
|O(x, F) - O(y, F)| \leq \sup_{\|u\| = 1} \frac{\|u'x - \mu(F_u)\| - |u'y - \mu(F_u)|}{\sigma(F_u)} \leq \sup_{\|u\| = 1} \frac{\|x - y\|}{\sigma(F_u)}.
\]
The uniform (and Lipschitz) continuity of \(O(x, F)\) in \(x\) follows. This gives part (i).

Now we show part (iii); the proof for part (ii) is similar and omitted. Let \(L_n(u) = |\mu(F_{nu}) - \mu(F_u)|\) and \(S_n(u) = |\sigma(F_{nu}) - \sigma(F_u)|\) for fixed \(F\). Then they approach 0 uniformly w.r.t. \(u\) as \(n \rightarrow \infty\). Note that
\[
|O(x, F_n) - O(x, F)| \leq \sup_{\|u\| = 1} \frac{|u'x|S_n(u) + |\mu(F_u)|S_n(u) + \sigma(F_u)L_n(u)}{\sigma(F_{nu})\sigma(F_u)}.
\]
By (C0) and (C1), \(\mu(F_u)\) and \(\sigma(F_u)\) are uniformly bounded above and \(\sigma(F_u)\) is uniformly bounded below from 0 w.r.t. \(u\). Thus if we can show that \(\sigma(F_{nu})\) is uniformly bounded below from 0 w.r.t. \(u\) for sufficiently large \(n\), then \(O(x, F_n) \rightarrow O(x, F)\) as \(n \rightarrow \infty\) for a fixed \(x\). Since
\[
\inf_{\|u\| = 1} \sigma(F_{nu}) - \inf_{\|u\| = 1} \sigma(F_u) \leq \sup_{\|u\| = 1} |\sigma(F_{nu}) - \sigma(F_u)|,
\]
thus \(\inf_{\|u\| = 1} \sigma(F_{nu}) \rightarrow \inf_{\|u\| = 1} \sigma(F_u)\) as \(n \rightarrow \infty\) and consequently \(\sigma(F_{nu})\) is uniformly bounded below from 0 w.r.t. \(u\) for sufficiently large \(n\). It follows that for any fixed \(M > 0\) and bounded \(x\) with \(\|x\| \leq M\),
\[
\sup_{\|x\| \leq M} |\text{PD}(x, F_n) - \text{PD}(x, F)| \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.
\]
Part (iii) follows if we can show that the above is also true for \( \| x \| > M \).

By Theorem 2.1, \( PD(x, F) \to 0 \) as \( \| x \| \to \infty \). So we need only show that \( PD(x, F_n) \to 0 \) as \( \| x \| \) and \( n \to \infty \). However, this follows from the fact that

\[
O(x, F_n) \geq \frac{\| x \| - \sup_{\| u \|=1} \mu(F_{nu})}{\sup_{\| u \|=1} \sigma(F_{nu})}
\]

and the conditions given. Part (iii) now follows.

From the proof for part (iii), we see that

\[
\sqrt{n} |O(x, F_n) - O(x, F)|
\]

\[
\leq \sqrt{n} \sup_{\| u \|=1} \frac{|u'x| S_n(u) + |\mu(F_u)| S_n(u) + \sigma(F_u) L_n(u)}{\sigma(F_{nu}) \sigma(F_u)}
\]

\[
\leq \| x \| Q_n + R_n,
\]

where

\[
Q_n = \frac{\sqrt{n} \sup_{\| u \|=1} S_n(u)}{\inf_{\| u \|=1} (\sigma(F_{nu}) \sigma(F_u))}
\]

and

\[
R_n = \frac{\sup_{\| u \|=1} |\mu(F_u)| \sqrt{n} \sup_{\| u \|=1} S_n(u) + \sup_{\| u \|=1} \sigma(F_u) \sqrt{n} \sup_{\| u \|=1} L_n(u)}{\inf_{\| u \|=1} (\sigma(F_{nu}) \sigma(F_u))}.
\]

By the given conditions and the proof above, it is readily seen that \( Q_n \) and \( R_n \) are bounded in probability. Thus for any fixed \( M > 0 \),

\[
\sqrt{n} \sup_{\| x \| \leq M} |PD(x, F_n) - PD(x, F)| = O_P(1).
\]

For any \( \| x \| > M \) (\( M \) sufficiently large), we see that for sufficiently large \( n \),

\[
\sqrt{n} |PD(x, F_n) - PD(x, F)|
\]

\[
\leq \sup_{\| u \|=1} \sigma(F_{nu}) \sup_{\| u \|=1} \sigma(F_u) (\| x \| Q_n + R_n) / (\| x \| - \sup_{\| u \|=1} \mu(F_{nu})) (\| x \| - \sup_{\| u \|=1} \mu(F_u)).
\]

Part (iv) now follows immediately. \( \square \)

**Proof of Remark 2.5.** We show that (iv)* does not hold for the halfspace depth in general. Consider a spherically symmetric Cauchy distribution with marginal p.d.f. \( f(x) = \pi^{-1}(1 + x^2)^{-1} \). In light of Massé (1999), it is readily seen that \( H_n(x) \equiv \sqrt{n}(HD(x, F_n) - HD(x, F)) = \int h(x, y)v_n(dy) + o_P(1) \) uniformly for \( x \) over any closed set \( S_n \) with \( 0 \notin S_n \), where \( h(x, y) = I(y \in H(x)) \) and \( H(x) \) is the unique closed hyperplane with \( x \) on its boundary such that \( HD(x, F) = P(H(x)) \).

Now for any fixed \( M > 0 \),

\[
P \left( \sup_{x \in \mathbb{R}^d} \| x \| | H_n(x) | > M \right) \geq P \left( \int \frac{2M}{\| x_0 \|} |h(x_0, y)v_n(dy) > \| x_0 \| \right) \geq 1/4.
\]
for large $\|x_0\|$ and $n$, since $\text{var}(h(x_0, X)) = P(H(x_0))(1 - P(H(x_0)))$ with $P(H(x_0)) = (1/2 - \pi^{-1}\arctan(\|x_0\|))$ and $r^2(1/2 - \pi^{-1}\arctan(r)) \to \infty$ as $r \to \infty$. The proof is complete. $\square$

PROOF OF THEOREM 2.3. Parts (i)–(iii) are trivial. We now show the rest.

Part (iv). We show first that $\{x : PD(x, F) = \alpha\} \subseteq \delta PD^\alpha(F)$ under (C0). Let $PD(x, F) = \alpha$. Assume that $x \notin \delta PD^\alpha(F)$, then there is a ball (centered at $x$ with radius $r$) contained in the interior of $PD^\alpha(F)$. Since $O(x, F) = \sup_{\|u\|=1} |u'x - \mu(F_u)|/\sigma(F_u)$, then there exits a unit vector $a$ such that

$$g(x, a; F) = \frac{|a'x - \mu(F_u)|}{\sigma(F_u)} > O(x, F) - \frac{r}{\sup_{\|u\|=1} \sigma(F_u)}.$$ 

On the other hand, it can be seen that there exists a point $y' \in PD^\alpha(F)$ such that

$$g(y', a; F) \geq g(x, a; F) + \frac{r}{\sup_{\|u\|=1} \sigma(F_u)} > O(x, F).$$

This leads to a contradiction that $PD(y', F) < PD(x, F)$.

Now we show that $\delta PD^\alpha(F) \subseteq \{x : PD(x, F) = \alpha\}$ under (C0) and (C1). Let $x \in \delta PD^\alpha(F)$. By the closedness of $PD^\alpha(F)$, $PD(x, F) \geq \alpha$. If $PD(x, F) > \alpha$, then by Theorem 2.2 there exists a neighborhood of $x$ such that $PD(y, F) > \alpha$ for any $y$ in that neighborhood, which contradicts the assumption that $x \in \delta PD^\alpha(F)$. Thus part (iv) follows.

Part (v). We first show that $PD^\alpha \neq \emptyset$. It is readily seen that for any $0 < \alpha < \alpha^* = \sup_x PD(x, F)$, $PD^\alpha(F)$ is nonempty. By part (ii), $PD^\alpha(F)$ is also closed and bounded. Since $\forall 0 \leq \beta < \alpha$, $PD^\alpha(F) \subseteq PD^\beta(F)$, thus $PD^{\alpha^*} = \bigcap_{\alpha < \alpha^*} PD^\alpha(F) \neq \emptyset$ [see Theorem 2.6 on page 37 of Rudin (1987)].

Now we show the emptiness of the interior of $PD^{\alpha^*}$. Assume that there is a point $y$ and a ball (centered at $y$ with radius $r$) contained in the interior of $PD^{\alpha^*}$. Following the same argument in the proof of part (iv), there is a point $y' \in PD^{\alpha^*}$ such that $PD(y', F) < PD(y, F) = \alpha$. This is a contradiction. Part (v) follows. $\square$

PROOF OF THEOREM 2.4. Define for any $0 \leq \alpha < \alpha^* PD^\alpha = \bigcup_{\alpha_0 > \alpha} PD^{\alpha_0}$ and $PD^\alpha = \bigcap_{\alpha_0 < \alpha} PD^{\alpha_0}$, then it is readily seen that $PD^- \subseteq PD^\alpha \subseteq PD^\alpha^+$. Furthermore, it is easy to see that $PD^\alpha = PD^{\alpha^*}$. Now we show that

$$PD^\alpha \subseteq \liminf_{n \to \infty} PD^{\alpha_n} \subseteq \limsup_{n \to \infty} PD^{\alpha_n} \subseteq PD^\alpha^+,$$

if $\alpha_n \to \alpha$ as $n \to \infty$. For any $\alpha_0 < \alpha$, it can be seen that there exists some $m$ such that when $n \geq m$, $\alpha_n > \alpha_0$ and consequently $PD^{\alpha_n} \subseteq PD^{\alpha_0}$. Hence

$$\limsup_{n \to \infty} PD^{\alpha_n} = \bigcap_{m=1}^\infty \bigcup_{n=m}^\infty PD^{\alpha_n} \subseteq PD^{\alpha_0}.$$
Therefore, \( \limsup_{n \to \infty} PD^{a_n} \subseteq \bigcap_{\alpha_0 < \alpha} PD^{a_0} = PD^{a_+} \). Likewise, we can show that \( \bigcup_{\alpha_0 < \alpha} PD^{a_0} = PD^{a_-} \subseteq \liminf_{n \to \infty} PD^{a_n} \). This, in conjunction with the fact that \( D^{a_+} = \{ x : PD(x, F) \geq \alpha \} \) and \( D^{a_-} = \{ x : PD(x, F) > \alpha \} \), gives part (i).

For part (ii), following the proof of Theorem 4.1 of Zuo and Serfling (2000b) and utilizing the results in our Theorems 2.1 and 2.2, the desired result can be obtained. \( \square \)

**Proof of Theorem 2.5.** For any \( \varepsilon > 0 \), by Theorem 2.2, there exists a \( \delta > 0 \) such that when \( \| x_1 - x_2 \| < \delta \), \( |PD(x_1) - PD(x_2)| < \varepsilon \). Assume that \( \rho(PD^a, PD^{a_n}) \to 0 \) as \( n \to \infty \). Then there exists an \( N \) such that when \( n \geq N \), \( \rho(PD^a, PD^{a_n}) < \delta / 2 \). We now show that \( |\alpha - \alpha_n| < \varepsilon \) when \( n \geq N \). Assume there is a point belonging to the set \( (\partial PD^a - \partial PD^{a_n}) \cup (\partial PD^{a_n} - \partial PD^a) \) (if no such point exists then, by Theorem 2.3, \( \alpha = \alpha_n \)) and, without loss of generality, assume that it belongs to \( \partial PD^a - \partial PD^{a_n} \). Denote it by \( x_{\alpha_n} \). Since \( \rho(PD^a, PD^{a_n}) < \delta / 2 \) when \( n \geq N \), thus \( d(x_{\alpha_n}, PD^{a_n}) < \delta / 2 \) and there is a point \( x_{\alpha_n} \in PD^{a_n} \) such that \( \| x_{\alpha_n} - x_{\alpha_n} \| < \delta / 2 \) when \( n \geq N \). By Theorem 2.3, \( PD(x_{\alpha_n}) = \alpha \) and \( PD(x_{\alpha_n}) = \alpha_n \).

Since now \( \| x_{\alpha_n} - x_{\alpha_n} \| < \delta / 2 \), then \( |PD(x_{\alpha_n}) - PD(x_{\alpha_n})| = |\alpha - \alpha_n| < \varepsilon \) when \( n \geq N \). Part (i) follows.

Now we prove part (ii). By the given conditions and Theorem 2.2, \( PD(x, F) \) has an inverse function along any ray stemming from \( \theta \) and the inverse function is continuous along the ray. For any given \( \varepsilon > 0 \), consider the inverse function of \( PD(x, F) \) along the ray \( u \). Since it is continuous at \( \alpha \), then there exists a \( \delta \) such that when \( |\alpha - \alpha_0| < \delta \), \( \| x_{\alpha_n}^u - x_{\alpha_n}^m \| < \varepsilon \), where \( PD(x_{\alpha_n}^u, F) = \alpha \) and \( PD(x_{\alpha_n}^m, F) = \alpha_0 \).

Since \( \alpha_n \to \alpha \), then there exists an \( N \) such that when \( n \geq N \), \( |\alpha - \alpha_n| < \delta \). Now we have that for any given \( \varepsilon > 0 \), there exists an \( N \) such that when \( n \geq N \), \( \| x_{\alpha_n}^u - x_{\alpha_n}^m \| < \varepsilon \), where \( PD(x_{\alpha_n}^u, F) = \alpha_n \). For \( x_{\alpha_n}^u \) and \( x_{\alpha_n}^m \) on any other ray \( v \) stemming from \( \theta \) with \( PD(x_{\alpha_n}^v, F) = \alpha_n \) and \( PD(x_{\alpha_n}^v, F) = \alpha_n \), by the condition given, \( \| x_{\alpha_n}^u - x_{\alpha_n}^m \| < \varepsilon \). From the definition of \( \rho \), it is seen that \( \rho(PD^a, PD^{a_n}) < \varepsilon \). Part (ii) follows. \( \square \)

**Lemma A.3.** Let \( D(x, \cdot) \) be any given general depth function satisfying:

(i) \( D(x, \cdot) \) is upper semicontinuous in \( x \) and \( \to 0 \) as \( \| x \| \to \infty \),
(ii) \( \theta = \arg \sup_{x \in \mathbb{R}^d} D(x, F) \) is unique and \( D(\theta, F) > 0 \),
(iii) \( \sup_{x \in \mathbb{R}^d} |D(x, F_n) - D(x, F)| \to 0 \) a.s. as \( n \to \infty \).

Define \( M(F_n) = \arg \sup_{x \in \mathbb{R}^d} D(x, F_n) \). Then \( M(F_n) - M(F) \to 0 \) a.s.

**Proof.** For any \( \varepsilon > 0 \), by Lemma A.4 it is seen that \( D(\theta, F) > \sup_{x \in N_\varepsilon} D(x, F) \) for \( N_\varepsilon = \{ y \in \mathbb{R}^d : \| y - \theta \| \geq \varepsilon \} \). Set \( \alpha_0 = D(\theta, F) \), \( \alpha_1 = \sup_{x \in N_\varepsilon} D(x, F) \) and \( \alpha_0 = (\alpha_0 + \alpha_1) / 2 \). Then \( \alpha_0 > \alpha_0 \). By the conditions given, there exists an \( N \) such that for any \( n \geq N \),

\[
\sup_{x \in \mathbb{R}^d} |D(x, F) - D(x, F_n)| \leq \sup_{x \in \mathbb{R}^d} |D(x, F) - D(x, F_n)| \leq \alpha_0 - \alpha_0 \quad \forall n > N.
\]
Hence
\[
D(\theta, F) - D(\theta, F_n) \leq \sup_{x \in \mathbb{R}^d} |D(x, F) - D(x, F_n)| \overset{a.s.}{\prec} \alpha_0 - \alpha \quad \forall n > N,
\]
which implies \(D(\theta, F_n) \overset{a.s.}{\succ} \alpha \quad \forall n > N\). On the other hand, we have that
\[
\alpha - \sup_{x \in N_\varepsilon} D(x, F) = \alpha_0 - \frac{\alpha_1}{2} + \sup_{x \in N_\varepsilon} D(x, F) - \sup_{x \in N_\varepsilon} D(x, F_n) \overset{a.s.}{\succ} (\alpha_0 - \alpha) + (\alpha - \alpha_0) \quad \forall n > N.
\]
That is, almost surely \(D(\theta, F_n) > \alpha \varepsilon > \sup_{x \in N_\varepsilon} D(x, F_n) \forall n > N\). This implies that \(|M(F_n) - \theta| \overset{a.s.}{\prec} \varepsilon \forall n > N\). The desired result now follows. □

**Remark A.2.** Applying Lemma A.3 to simplicial and halfspace depth functions, we obtain, as special cases, Theorem 5(b) of Liu (1990), Theorem 6.9 of Arcones and Giné (1993) and Lemma 1 of Nolan (1999).

**Lemma A.4.** Let depth function \(D(x, F)\) be upper semicontinuous in \(x\) and \(D(x, F) \to 0\) as \(\|x\| \to \infty\). Let \(\theta\) be a unique point in \(\mathbb{R}^d\) such that \(\theta = \arg \sup_{x \in \mathbb{R}^d} D(x, F)\) and \(D(\theta, F) > 0\). Then for any \(\varepsilon > 0\), \(D(\theta, F) > \sup_{x \in N_\varepsilon} D(x, F)\) with \(N_\varepsilon = \{y \in \mathbb{R}^d : \|y - \theta\| \geq \varepsilon\}\).

**Proof.** If \(\sup_{x \in N_\varepsilon} D(x, F) = D(\theta, F)\), then there exists a sequence \(\{x_n\}\) with \(\|x_n - \theta\| \geq \varepsilon\) such that \(D(x_n, F) \to D(\theta, F)\) as \(n \to \infty\). Since \(D(x, F) \to 0\) as \(\|x\| \to \infty\), then \(\{x_n\}\) is bounded. Consequently there is a subsequence \(\{x_{n_k}\}\) of \(\{x_n\}\) such that \(x_{n_k} \to x_0\) as \(k \to \infty\) for some \(x_0 \in \mathbb{R}^d\) with \(\|x_0 - \theta\| \geq \varepsilon\). Since \(D(\theta, F) > D(x_0, F)\), the upper semicontinuity of \(D(x, F)\) implies that
\[
D(x_{n_k}, F) < D(x_0, F) + (D(\theta, F) - D(x_0, F))/2
\]
\[
= (D(\theta, F) + D(x_0, F))/2
\]
\[
< D(\theta, F).
\]
This, however, contradicts the fact that \(D(x_{n_k}, F) \to D(\theta, F)\). The proof is complete. □

**Proof of Theorem 3.1.** Part (i) follows in a straightforward fashion from Lemma A.3 and Theorem 2.2. We now show part (ii). Without loss of generality, assume that \(\theta = 0\). Write \(T_n\) for \(PM(F_n)\). Observe that
\[
O(T_n, F_n) = \sup_{\|u\| = 1} |u^T T_n - \mu(F_nu)| / \sigma(F_nu) \leq O(0, F_n) \leq \frac{\sup_{\|u\| = 1} |\mu(F_nu)|}{\inf_{\|u\| = 1} \sigma(F_nu)}
\]
and
\[ O(T_n, F_n) \geq \sup_{\|u\|=1} \frac{|u'T_n|}{\sigma(F_{nu})} - \frac{\sup_{\|u\|=1} |\mu(F_{nu})|}{\inf_{\|u\|=1} \sigma(F_{nu})}, \]
which implies that
\[ \frac{\sup_{\|u\|=1} |u'\sqrt{n}T_n|}{\sup_{\|u\|=1} \sigma(F_{nu})} \leq \sup_{\|u\|=1} \frac{|u'\sqrt{n}T_n|}{\sigma(F_{nu})} \leq 2 \frac{\sup_{\|u\|=1} |\sqrt{n}\mu(F_{nu})|}{\inf_{\|u\|=1} \sigma(F_{nu})}. \]
From the given conditions and results in the proof of Theorem 2.2, the desired result follows. □

**Lemma A.5.** Let \( \{\sqrt{n}(\mu(\hat{F}_{nu}) - \mu(F_u)) : \|u\| = 1\} \to_d \{Z(u) : \|u\| = 1\} \), where \( Z(u) \) is a Gaussian process with uniformly continuous sample paths in \( L^\infty(U) \) (the space of real bounded functions on the unit sphere that is equipped with the supremum norm) and \( E(Z(u)) = 0 \) and \( \text{cov}(Z(u_1), Z(u_2)) = E(Z(u_1)Z(u_2)) \). Let (C0) and (C1) hold and \( \sigma_n(\hat{F}_{nu}) - \sigma(F_u) = o_p(1) \) uniformly in \( u \). Then for \( F \) \( \mu \)-symmetric about \( \theta \in \mathbb{R}^d \),
\[ \sqrt{n}(PM(\hat{F}_n) - PM(F)) \Rightarrow \arg \inf_t \sup_{\|u\|=1} \left| \frac{u't - Z(u)}{\sigma(F_u)} \right|, \]
provided that the \( \arg \inf \) is unique almost surely.

**Proof.** For simplicity, assume that \( \theta = 0 \). Write for \( t \in \mathbb{R}^d \),
\[ Q_n(t) = \sup_{\|u\|=1} \frac{u't - \sqrt{n}\mu(\hat{F}_{nu})}{\sigma(\hat{F}_{nu})} \quad \text{and} \quad Q(t) = \sup_{\|u\|=1} \frac{u't - Z(u)}{\sigma(F_u)}. \]
Then for each finite subset \( S \) of \( \mathbb{R}^d \), \( \{Q_n(t) : t \in S\} \to_d \{Q(t) : t \in S\} \) by virtue of the given conditions and the continuous mapping theorem. On the other hand, it is straightforward to verify that (ii) of Theorem 2.3 of Kim and Pollard (1990) holds. In light of Theorem 2.3 of Kim and Pollard (1990), we conclude that (i) of Theorem 2.7 of Kim and Pollard (1990) holds (with \( Z_n \) there equals \( Q_n \) here). Write \( T_n \) for \( PM(\hat{F}_n) \). By Remark 3.1 and Theorem 3.1, \( t_n = \sqrt{n}T_n = O_p(1) \). Note that
\[ Q_n(t_n) = \sqrt{n} \sup_{\|u\|=1} \frac{u'T_n - \mu(\hat{F}_{nu})}{\sigma(\hat{F}_{nu})} \leq \inf_t Q_n(t). \]
Invoking Theorem 2.7 of Kim and Pollard (1990), we obtain the desired result. □

**Proof of Theorem 3.2.** By condition (iv), we have
\[ \sqrt{n}(\mu_{nu} - \mu_u) = -\frac{1}{\sqrt{n}} \sum_{i=1}^n (d_u\psi(Y_{iu}) - b_uX(Y_{iu})) \frac{a_u}{a_u - b_u} + o_p(1), \]
almost surely and uniformly in \( u \), where \( Y_{i,u} = (u'X_i - \mu_u)/\sigma_u \). By the conditions in (ii), \( \mathcal{F} \) has an envelope

\[
G = \frac{\sup_u |d_u| \sup_u |\psi(h_u)| + \sup_u |b_u| \sup_u |\chi(h_u)|}{\inf_u \det \Lambda_u}
\]

and \( E(G^2) < \infty \). By condition (iii), the equicontinuity lemma of VII.15 and Theorem VII.21 of Pollard (1984), we deduce that \( \{\sqrt{n}(\mu_{nu} - \mu_u) : \|u\| = 1\} \) converges in distribution to \( \{Z(u) : \|u\| = 1\} \), a \( F \)-Brownian bridge [see page 149 of Pollard (1984) or page 82 of Van der Vaart and Wellner (1996)], a zero-mean Gaussian process satisfying conditions in Lemma A.5. By condition (iv) and the strong law of large numbers, we have that \( \sigma_n(\hat{F}_{nu}) - \sigma(F_u) \to 0 \) almost surely and uniformly in \( u \). Note that \( Z(u) = -Z(-u) \).

By Lemma A.5, the desired result follows if we can show that the \( \arg \inf \) in the theorem is unique. Without loss of generality, assume that \( F \) is \( \mu \)-symmetric about \( \theta = 0 \) in \( \mathbb{R}^d \). We now employ Lemma 3.3 of Bai and He (1999), a powerful result for checking the uniqueness of this sort of projection-based estimators. Note that \((u't - Z(u)) \) in the theorem is odd in \( u \). So we can drop the absolute value symbols in the theorem. Corresponding to \( \mu(u) \) in the Lemma 3.3 of Bai and He (1999) for unit vector \( u \), we have here \(-u/\sigma_u \). Condition (W1) in Bai and He (1999) holds trivially and (W3) also holds since otherwise \( Z(u) \) would equal constant 0 for a unit vector \( u \). So we need only verify (W2). Following the proof of Theorem 4.1 of Bai and He (1999), we now find \( D_u \) by calculating the derivative of \( \mu(u) \) with respect to \( u \). First, consider directional derivatives of \( u \) and \( \sigma_u \) with respect to a direction \( l \). The product rule is utilized. Write \( u_t = (u + tl)/\|u + tl\| \) for any \( t \in \mathbb{R}^l \). Consider the derivative of \( u_t \) with respect to \( t \) and let \( t \) approach 0, we obtain the derivative \((l - uu')l \), which then contributes \(-I(u - uu')/\sigma_u \) to \( D_u \). The contribution from \( \sigma_u \) to \( D_u \) will be \(-ub'_u \). We now derive \( b'_u \) from the two equations defining \( \mu_u \) and \( \sigma_u \).

Denote by \( f_u \) the density of \( F_u \). First we express \( f_{u,t} \) in terms of \( f \), the density of \( F \). Write

\[
P(u'_tX \leq a) = \int_{u'_tx + tl'x \leq a\|u + tl\|} f(x) \, dx,
\]

where \( x \in \mathbb{R}^d \). Take an orthogonal transformation to the underlying coordinate system with the orthogonal matrix \( B = (u, C) \) and \( x = (x_1, \ldots, x_d)' = B y \) with \( y = (v, z)' \) where \( y \in \mathbb{R}^d \) and \( z \in \mathbb{R}^{d-1} \). Thus, \( x = uv + Cz \). It follows that

\[
P(u'_tX \leq a) = \int \left[ \int_{v \leq (a\|u + tl\| - tl'Cz)/(1 + tl'u)} f(uv + Cz) \, dv \right] \, dz.
\]

Taking the derivative with respect to \( a \) in both sides yields

\[
f_{u,t}(a) = \int f \left( \frac{(u(a\|u + tl\| - tl'Cz))/(1 + tl'u) + Cz}{1 + tl'u} \right) \|u + tl\| \, dz.
\]
Now taking the derivative with respect to $t$ in the two equations defining $\mu_{ut}$ and $\sigma_{ut}$, using the above relation between $f_{ut}$ and $f$, and letting $t \to 0$, we have

$$\int \left[ \int \psi \left( \frac{v - \mu_u}{\sigma_u} \right) \frac{-\sigma_u (\partial \mu_u / \partial l) - (v - \mu_u) (\partial \sigma_u / \partial l)}{\sigma_u^2} f(uv + Cz) \, dz \right] \, dv$$

$$= \int \left[ \int \psi \left( \frac{v - \mu_u}{\sigma_u} \right) u' \hat{f}(uv + Cz)l' \, dz \right] \, dv,$$

$$\int \left[ \int \chi \left( \frac{v - \mu_u}{\sigma_u} \right) -\sigma_u (\partial \mu_u / \partial l) - (v - \mu_u) (\partial \sigma_u / \partial l) \right] \frac{\sigma_u^2 f(uv + Cz) \, dz}{\sigma_u^2} \, dv$$

$$= \int \left[ \int \chi \left( \frac{v - \mu_u}{\sigma_u} \right) u' \hat{f}(uv + Cz)l' \, dz \right] \, dv.$$

Note that $\int f(uv + Cz) \, dz = f_u(v)$, hence

$$- \left( \begin{array}{c} a_u \\ b_u \\ c_u \\ d_u \end{array} \right) \left( \begin{array}{c} \partial \mu_u / \partial l \\ \partial \sigma_u / \partial l \end{array} \right) = \left( \begin{array}{c} \int \left[ \int \psi((v - \mu_u)/\sigma_u)u' \hat{f}(uv + Cz)l' \, dz \right] \, dv \\ \int \left[ \int \chi((v - \mu_u)/\sigma_u)u' \hat{f}(uv + Cz)l' \, dz \right] \, dv \end{array} \right).$$

Therefore,

$$\frac{\partial \sigma_u}{\partial l} = - \frac{a_u \int \left[ \int \psi((v - \mu_u)/\sigma_u)u' \hat{f}(uv + Cz)l' \, dz \right] \, dv}{a_u d_u - b_u c_u} - \frac{c_u \int \left[ \int \chi((v - \mu_u)/\sigma_u)u' \hat{f}(uv + Cz)l' \, dz \right] \, dv}{a_u d_u - b_u c_u}.$$

That is,

$$b_u = - \frac{a_u \int \left[ \int \chi((v - \mu_u)/\sigma_u)u' \hat{f}(uv + Cz)Cz \, dz \right] \, dv}{a_u d_u - b_u c_u} - \frac{c_u \int \left[ \int \psi((v - \mu_u)/\sigma_u)u' \hat{f}(uv + Cz)Cz \, dz \right] \, dv}{a_u d_u - b_u c_u}.$$

Now we have shown that $D_u = -(I - uu')/\sigma_u - ub'_u$. It is readily seen that $b'_u u = 0$ and consequently $D_u u = 0$ for any unit vector $u$. Thus, $\{ D_u a : \|a\| = 1 \} = \{ D_u a : a' u = 0 \}$, which has the same dimension as $\{ a' D_u : a' u = 0 \} = \{ -a' / \sigma_u : a' u = 0 \}$. This implies that the rank of $D_u$ is $d - 1$. Thus $D_u$ is well defined with rank $d - 1$ and $D_u u = 0$ for all $u$, implying that (W2) holds with no exceptional direction $\alpha$. The proof is complete. \hfill $\Box$

Proof of Theorem 3.3. Condition (C0) automatically holds. In virtue of Remark 2.4, Lemma A.1, and the given conditions, (C1) holds. By the given
conditions and Theorem 1 of Hall and Welsh (1985), we can show that \( \sigma_n(\hat{F}_{nu}) - \sigma(F_u) = o_P(1) \) uniformly in \( u \). In light of the Bahadur type representation of the sample median [Serfling (1980) or Jurečková and Sen (1996)], we can show that

\[
\sqrt{n}(\text{Med}_n(\hat{F}_{nu}) - \text{Med}(F_u)) = \sqrt{n}\left(\frac{1}{n} \sum_{i=1}^{n} \frac{1/2 - I(u'X_i \leq u'\theta)}{f_u(u'\theta)}\right) + R_n(u)
\]

with \( \sup_{\|u\|=1} |R_n(u)| = O(n^{-3/4} \log n) \) almost surely; see also Cui (1994). Write

\[
F = \left\{ \frac{-I(u' \cdot \leq u'\theta)}{f_u(u'\theta)} : \|u\| = 1 \right\}
\]

Then it is seen that \( F \) is a permissible class with envelope \( 1/\inf_{\|u\|=1} f_u(u'\theta) \) and the graphs of functions in \( F \) form a polynomial class of sets. [Assume (w.l.o.g.) that \( \theta = 0 \), the points \( \{(x, t)\} \) contained by the graphs of functions in \( F \) can be written as \( (t = 0) \cap \{u'x > 0\} \cup (-1/f_u(0) \leq t < 0) \cap \{u'x \leq 0\} \), now invoking Lemmas II.15 and 18 and Examples II.26 and VII.18 of Pollard (1984) gives the desired result.] By Lemma VII.15 and Theorem VII.21 of Pollard (1984), \( \{\sqrt{n}(\text{Med}_n(\hat{F}_{nu}) - \text{Med}(F_u)) : \|u\| = 1\} \) converges in distribution to \( \{Z(u) : \|u\| = 1\} \), a Gaussian process with bounded and uniformly continuous sample paths in \( L^\infty(F) \) with mean and covariance matrix specified in the theorem.

Note that \( Z(u) = -Z(-u) \).

Following the proof of Theorem 3.2, we now calculate \( b_u \). Write

\[
1/2 = P(-\sigma_{u,\|u+tl\|} \leq (u+tl)'X \leq \sigma_{u,\|u+tl\|}).
\]

With the orthogonal transformation in Theorem 3.2, we have that

\[
1/2 = \int \int (f(Cz + \sigma_uu) - f(Cz - \sigma_uu)) f(\mu v + Cz) d\mu d\nu
\]

d\nu.

Taking derivative with respect to \( t \) in both sides above and consider \( t \to 0 \), we have

\[
0 = \int \left( f(Cz + \sigma_uu) \left( -l'Cz + \frac{\partial \sigma_u}{\partial l} \right) + f(Cz - \sigma_uu) \left( l'Cz + \frac{\partial \sigma_u}{\partial l} \right) \right) dz,
\]

where \( \frac{\partial \sigma_u}{\partial l} \) is the directional derivative of \( \sigma_u \) with respect to \( l \). Thus

\[
\frac{\partial \sigma_u}{\partial l} \int (f(Cz + \sigma_uu) + f(Cz - \sigma_uu)) dz
\]

\[
= \int (f(Cz + \sigma_uu) - f(Cz - \sigma_uu)) l'Cz dz.
\]
Note that \( Cz \pm \sigma_u u = By \) with \( y_1 = v = \pm \sigma_u \). By the relationship between \( f \) and \( f_u \) established in the proof of Theorem 3.2, we have
\[
\int (f(Cz + \sigma_u u) + f(Cz - \sigma_u u)) \, dz = f_u(\sigma_u) + f_u(-\sigma_u) > 0.
\]
Therefore
\[
b_u = \frac{\int (f(Cz + \sigma_u u) - f(Cz - \sigma_u u)) \, Cz \, dz}{f_u(\sigma_u) + f_u(-\sigma_u)}.
\]
Since \( b'_u u = 0 \), the same argument in the proof of Theorem 3.2 now gives the desired result. \( \square \)

**Proof of Theorem 3.4.** Denote by \( Z = \{Z_1, \ldots, Z_n\} \) the contaminated data set \( X^n_m \) (with \( m \) points in \( X^n \) contaminated) and \( u'Z = \{u'Z_1, \ldots, u'Z_n\} \) for unit vector \( u \). Let \( \mu(u'Z) \) and \( \sigma(u'Z) \) be, respectively, the univariate location and scale estimators based on the sample \( u'Z \). Recall the convention that \( |u'x - \text{Med}(u'Z)|/\text{MAD}(u'Z) = 0 \) if \( |u'x - \text{Med}(u'Z)| = \text{MAD}_k(u'Z) = 0 \).

(a) \( d = 1 \). (i) \( m = \lfloor (n + 1)/2 \rfloor \) points are sufficient to break down \( PM^k_n \). This follows from the upper bound of RBP of translation equivariant location estimators [Lopuhäär and Rousseeuw (1991)].

(ii) \( m = \lfloor (n + 1)/2 \rfloor - 1 \) points are insufficient to break down \( PM^k_n \). It is readily seen that in this case \( \mu(u'Z) \) is uniformly bounded w.r.t. \( u \) and \( Z \). Hence, \( |u'x - \mu(u'Z)| \) is uniformly bounded w.r.t. \( u \) and \( Z \) for any \( x \) in the range \( R \) formed by \( X^n \). If more than \( (n + k - 1)/2 \) points in \( Z \) lie at (or approach to) the same point, say \( y \), then \( \sigma(u'Z) \) is (or approaches) zero. (For simplicity, we skip “approaching” case in the following discussion unless otherwise stated.) Note that for \( Z_i \)’s at \( y \), \( O_{n_b}(Z_i, Z) = 0 \) by the convention. Thus there always exists some \( Z_i \) of \( Z \) in \( R \) such that \( O(Z_i, Z) \) is bounded above for any \( Z \) and \( k \geq 1 \). When \( k = 1 \), it is not difficult to see that \( O(x, Z) \rightarrow \infty \) as \( \|x\| \rightarrow \infty \) [since \( \sigma(u'Z) \) is bounded] for any \( Z \). This is also true when \( k > 1 \) but \( \sigma(u'Z) \) is bounded. When \( k > 1 \), \( \sigma(u'Z) \) may approach infinity, but in such a case \( O(Z_i, Z) \) approaches zero for any \( Z_i \) inside the range \( R \). In any case, \( \inf_{x \in R} O(x, Z) \) is less than \( O(x, Z) \) for \( x \) outside the range \( R \). Therefore, \( m \leq \lfloor (n + 1)/2 \rfloor - 1 \) points never break down \( PM^k_n \).

(b) \( d = 2 \). (i) \( m = \lfloor (n-2d+k+3)/2 \rfloor \) points are sufficient to break down \( PM^k_n \). Let \( l(X^n) \) be a line determined by two points in \( X^n \) and \( X_j \) be a point in \( X^n \) and not on line \( l \). Move \( m \) points from \( X^n \) (not from \( X_j \)) to the same site \( y \) far away from the origin and outside the original convex hull of \( X^n \) and on the line \( l \), leaving at least one untouched original point on \( l \). Choose \( u_0 \) and \( u_1 \) to be the two unit vectors perpendicular to the line \( l \) and the line connecting \( y \) and \( X_j \), respectively. Since \( m + d - 1 \) is greater than \( (n + k - 1)/2 \), then \( \sigma(u_0'Z) \) and \( \sigma(u_1'Z) \) equal zero.
Thus $O(x, Z) = \infty$ for any $x \in \mathbb{R}^d$ except $y$. Hence $\|PM^k_n(Z)\| = \|y\| \to \infty$ as $\|y\| \to \infty$.

(ii) $m = \left[ (n - 2d + k + 3)/2 \right] - 1$ points are insufficient to break down $PM^k_n$.
Since $k \leq d - 1$, then $n - m > (n + k - 1)/2$. Thus $\mu(u'Z)$ and $\sigma(u'Z)$ are uniformly bounded w.r.t. $u$ and $Z$. Consequently, $O(x, Z) \to \infty$ as $\|x\| \to \infty$ uniformly w.r.t. $Z$. The desired result follows if we can show that for fixed large $M > 0$ and $\|x\| \leq M$, $\inf_{\|x\|\leq M} O(x, Z) < M_0$ uniformly w.r.t. $Z$ for some $M_0 > 0$. Let $M > 0$ be a fixed value large enough so that $\|y\| < M$ for any $y$ that is a intersecting point of two lines determined by points of $X^n$. Suppose there is no $M_0$ for this fixed $M$. Then there is a sequence of contaminated data sets $\{Z_t\}$ and $u_t$ such that $\sigma(u_t'Z_t) \to 0$ and $\inf_{\|x\|\leq M} O(x, Z_t) \to \infty$ as $t \to \infty$.
Since $m + d - 1 < (n + k)/2$, there must be a sequence of lines $l_t$ $(l_t \perp u_t)$ such that there are $d$ original points from $X^n X_{t1}, \ldots, X_{td}$ on $l_t$ and $m$ contaminating points on (or approaching) $l_t$. Since $l_t$ contains $X_{t1}, \ldots, X_{td}, O(X_{t1}, Z_t) \to \infty$ for $1 \leq j \leq d$ only if there exists another sequence of unit vectors $u_t^*$ perpendicular to lines $l_t^*$ which contain no $X_{tj}$, $1 \leq j \leq d$ and there are more than $(n + k - 1)/2$ points from $Z_t$ approaching or on $l_t^*$. Since $m + d - 1 < (n + k)/2$, $l_t^*$ must contain $d$ original points of $X^n$ and intersect with $l_t$ with all $m$ contaminating points at (or approaching) the intersecting point. Let the intersecting point be $y_t$, then $\|y_t\| \leq M$. Since $\{u_t\}$ is a sequence of unit vectors on the unit sphere, then there is a subsequence $\{u_{t_k}\}$ of $\{u_t\}$ which converges to a unit vector $v$.
Then we have the corresponding subsequences $\{Z_{t_k}\}, \{l_{t_k}\}, \{u_{t_k}^*\}, \{l_{t_k}^*\}$ and $\{y_{t_k}\}$ such that $l_{t_k} \to l$ with $l \perp v, u_{t_k}^* \to v^*$ with $l^* \perp v^*$, $\sigma(u_{t_k}^*Z_{t_k}) \to 0$, $\sigma(u_{t_k}^*Z_{t_k}) \to 0$, and $\inf_{\|x\|\leq M} O(x, Z_{t_k}) \to \infty$ as $t_k \to \infty$. Note that there are only finite many different $y_{t_k}$’s [indeed no more than $\binom{d}{2}$], then we may assume w.l.o.g. that for sufficiently large $t_k$, $y_{t_k} = y$, $l_{t_k} = l$, $u_{t_k} = v$, $l_{t_k}^* = l^*$, $u_{t_k}^* = v^*$ and $y$ is the intersecting point of $l$ and $l^*$ (this can always be achieved by taking subsequences of subsequences if necessary). We now show that $O(y, Z_{t_k})$ is uniformly bounded w.r.t. $Z_{t_k}$ for sufficiently large $t_k$, which contradicts the assertion that $\inf_{\|x\|\leq M} O(x, Z_{t_k}) \to \infty$ as $t_k \to \infty$.

For simplicity, assume that $n$ is odd. Since $\mu(u'Z)$ and $\sigma(u'Z)$ are continuous in unit vector $u$ for any fixed $Z$, assume, without loss of generality, that

$$O(y, Z_{t_k}) = \frac{|a_{t_k}y - \mu(a_{t_k}'Z_{t_k})|}{\sigma(a_{t_k}'Z_{t_k})}.$$ 

Since $a_{t_k}$ is a unit vector on the unit sphere, there is a subsequence of $\{a_{t_k}\}$ which converges to a unit vector $a$. For simplicity, just assume that $a_{t_k} \to a$ as $t_k \to \infty$.
If $a$ is $v$, $v^*$ or any other unit vector $v^*$ which is perpendicular to a line through $y$ determined by $d$ points of $X^n$ which also belong to infinitely many $Z_{t_k}$’s, then for sufficiently large $t_k$, $|a'y - \mu(a'Z_{t_k})|/\sigma(a'Z_{t_k}) \leq 1$ for any $Z_{t_k}$ or infinitely many $Z_{t_k}$’s. Since $|a'y - \mu(a'Z)|/\sigma(a'Z)$ is continuous in $u$, then the unit vector $a$ cannot be $v$, $v^*$ or the $v^*$ in order to have $\inf_{\|x\|\leq M} O(x, Z_{t_k}) \to \infty$ as $t_k \to \infty$. 

Consider the axis through $y$ with the same direction of $a_t$. Denote by $R_{t_s}$ the smallest closed interval on the axis containing all the projections of $m$ contaminating points and the projection of $y$ on to the axis. For simplicity, also use $R_{t_s}$ to denote the length of $R_{t_s}$. Then $R_{t_s} \to 0$ as $t_s \to \infty$.

Assume that $X_{1,1}, \ldots, X_{i_m}$ are original points from $X^n$ but also belonging to $Z_{t_s}$ such that $|\mu(a_t' Z_{t_s}) - a_t' X_{1,1}| \leq |\mu(a_t' Z_{t_s}) - a_t' X_{i_m}| \leq \cdots \leq |\mu(a_t' Z_{t_s}) - a_t' X_{i_m} - \cdots - m|$. Assume that $X_{i_1}, \ldots, X_{i_d}$ are on a line $L_{t_s}$, then $L_{t_s}$ cannot be a line through $y$ and perpendicular to $a$ for sufficiently large $t_s$. Assume that $Z_{t_s}, \ldots, Z_{t_s,i_n}$ are points of $Z_{t_s}$ such that $|\mu(a_t' Z_{t_s}) - a_t' Z_{t_s})| \leq \cdots \leq |\mu(a_t' Z_{t_s}) - a_t' Z_{t_s})|$. Now we consider two cases.

(1) Assume there is at least one point $Z_{t_s,j} \notin X^n$ such that $|\mu(a_t' Z_{t_s}) - a_t' Z_{t_s,j}| \leq \sigma(a_t' Z_{t_s})$. Write $D(L_{t_s}) = \max_{1 \leq q \leq d} |a_t' X_{i_q} - a_t'y|/8$. Then $D(L_{t_s}) > 0$. For sufficiently large $t_s$ and all possible $Z_{t_s}$, there are only finite many $L_{t_s}$'s. Thus $D = \min L_{t_s} D(L_{t_s}) > 0$. For sufficiently large $t_s$,

$$
\sigma(a_t' Z_{t_s}) \geq \max_{1 \leq q \leq d} |a_t' X_{i_q} - a_t' Z_{t_s,j}|/2 \\
\geq \max_{1 \leq q \leq d} |a_t' X_{i_q} - a_t'y|/2 - |a_t'y - a_t' Z_{t_s,j}|/2 \\
\geq \max_{1 \leq q \leq d} |a_t' X_{i_q} - a_t'y|/4 - D \\
\geq D.
$$

Therefore $\sigma(a_t' Z_{t_s}) \geq D > 0$ for any sufficiently large $t_s$.

(2) Assume there is no point $Z_{t_s,j} \notin X^n$ such that $|\mu(a_t' Z_{t_s}) - a_t' Z_{t_s,j}| \leq \sigma(a_t' Z_{t_s})$. Write $D = \inf_{1 \leq i_1, \ldots, i_{d+1}} \max_{1 \leq j_1, j_2 \leq d+1} |a_t' X_{i_{j_1}} - a_t' X_{i_{j_2}}|/4$, where $i_1, \ldots, i_{d+1}$ are distinct numbers from $1, \ldots, n$. Since $X^n$ is in the general position, $D > 0$. For sufficiently large $t_s$, we have

$$
\sigma(a_t' Z_{t_s}) \geq \max_{1 \leq j_1, j_2 \leq d+1} |a_t' X_{i_{j_1}} - a_t' X_{i_{j_2}}|/2 \\
\geq \max_{1 \leq j_1, j_2 \leq d+1} |a_t' X_{i_{j_1}} - a_t' X_{i_{j_2}}|/4 \\
\geq \inf_{i_1, \ldots, i_{d+1}} \max_{1 \leq j_1, j_2 \leq d+1} |a_t' X_{i_{j_1}} - a_t' X_{i_{j_2}}|/4.
$$

Therefore $\sigma(a_t' Z_{t_s}) \geq D > 0$.

In both cases, $O(y, Z_{t_s})$ is uniformly bounded w.r.t. $Z_{t_s}$ for sufficiently large $t_s$, contradicting the assertion that $\inf_{\|x\| \leq M} O(x, Z_{t_s}) \to \infty$ as $t_s \to \infty$.

(c) $d > 2$. The proof for $d = 2$ can be extended in a straightforward fashion for this case. \(\square\)
Acknowledgments. The author greatly appreciates the insightful and constructive remarks of an Associate Editor and two referees, which led to distinct improvements in the paper. He thanks Miguel Arcones, Xuming He and Peter Rousseeuw for insightful comments and helpful suggestions and is very grateful to Robert Serfling for stimulating discussions, thoughtful and constructive remarks and constant encouragement.

REFERENCES


Department of Statistics and Probability
Michigan State University
East Lansing, Michigan 48824
E-mail: zuo@msu.edu