Science in China Series A: Mathematics
Jun., 2009, Vol. 52, No. 6, 1–21
www.scichina.com math.scichina.com
www.springerlink.com

On a robust and efficient maximum depth estimator

ZUO YiJun[1,2]† & LAI ShaoYong[1]

[1] School of Economic Mathematics, Southwestern University of Finance and Economics, Chengdu 610074, China
[2] Department of Statistics and Probability, Michigan State University, East Lansing, MI 48823, USA
(email: zuo@msu.edu, laishaoy@swufe.edu.cn)

Abstract The best breakdown point robustness is one of the most outstanding features of the univariate median. For this robustness property, the median, however, has to pay the price of a low efficiency at normal and other light-tailed models. Affine equivariant multivariate analogues of the univariate median with high breakdown points were constructed in the past two decades. For the high breakdown robustness, most of them also have to sacrifice their efficiency at normal and other models, nevertheless. The affine equivariant maximum depth estimator proposed and studied in this paper turns out to be an exception. Like the univariate median, it also possesses a highest breakdown point among all its multivariate competitors. Unlike the univariate median, it is also highly efficient relative to the sample mean at normal and various other distributions, overcoming the vital low-efficiency shortcoming of the univariate and other multivariate generalized medians. The paper also studies the asymptotics of the estimator and establishes its limit distribution without symmetry and other strong assumptions that are typically imposed on the underlying distribution.

Keywords: data depth, maximum depth estimator, median, location estimator, breakdown point, asymptotic distribution, robustness, efficiency

MSC(2000):

1 Introduction

The univariate median is well known for its robustness. Indeed, it is a valid location (center) estimator even if up to half of data points are "bad" (contaminated). It is said to have the highest breakdown point. The notion of breakdown point, introduced by Donoho and Huber[1], has become the most prevailing quantitative assessment of robustness of estimators. Roughly speaking, the breakdown point of an estimator is the minimum fraction of bad points in a data set that could render the estimator useless. Since one bad point in a data set of size n can force the sample mean to be unbounded (hence useless), its breakdown point then is $1/n$, the lowest possible value. On the other hand, to make the univariate median useless, 50% of original data points need to be contaminated. That is, the breakdown point of the univariate median is $\lfloor (n+1)/2 \rfloor /n$, which turns out to be the best possible value for any reasonable estimators of

Received January 8, 2009; accepted January 20, 2009
DOI: 10.1007/s11425-009-0064-0
† Corresponding author
This work was supported by Natural Science Foundation of USA (Grant Nos. DMS-0071976, DMS-0234078) and by the Southwestern University of Finance and Economics Third Period Construction Item Funds of the 211 Project (Grant No. 211D3T06)

Citation: Zuo Y J, Lai S Y. On a robust and efficient maximum depth estimator. Sci China Ser A, 2009, 52(6): 1–21, DOI: 10.1007/s11425-009-0064-0

location parameters. Here $\lfloor x \rfloor$ denotes the largest integer no larger than x.

Multivariate analogues of the univariate median are desirable and the L_1 (or spatial) and the coordinate-wise medians are two that have the same breakdown point as that of the univariate counterpart (see [2]). Unlike the univariate median, the two, however, lack the affine equivariance property, which, just like robustness, is also highly desirable. Roughly speaking, affine equivariance of an estimator implies that the estimator is coordinate system free and measurement scale free. Constructing affine equivariant multivariate location estimators with a high breakdown point has been one of the primary goals of researches in robust statistics for two decades. Important results were obtained by Stahel[3], Donoho[4], Davies[5], Tyler[6], and Hettmansperger and Randles[7], for example. Related estimators can have breakdown points as high as (but no higher than) $\lfloor (n-d+1)/2 \rfloor / n$ in $\mathbb{R}^d$.

Recently "data depth" has been utilized to serve the above goal. The primary purpose of data depth is to provide a center-outward order for multi-dimensional data based on their depth; see [8–10] for example. Points deep inside a data cloud get high depth and those on the outskirts get lower depth. The point with maximum depth is then defined as a multi-dimensional location estimator (or median). Maximum depth estimators (or medians) are usually affine equivariant and can have high breakdown points. Indeed, the half-space depth median (see [11]) can have a breakdown point about 1/3 in $\mathbb{R}^d$. The projection depth median[12], on the other hand, can have an unprecedentedly high breakdown point $\lfloor (n-d+2)/2 \rfloor / n$ ($d \geqslant 1$) (the best result before this one is $\lfloor (n-d+1)/2 \rfloor / n$ in $\mathbb{R}^d$).

Unfortunately, most affine equivariant multivariate estimators (medians), just like the univariate median, have to pay the price of a low efficiency at normal models for their high breakdown robustness. This is also true for above depth medians. A natural question raised is: Is there any affine equivariant multivariate median-type estimator that has the highest breakdown point $\lfloor (n-d+2)/2 \rfloor / n$ while enjoying a very high efficiency at normal and other models?

This paper provides a positive answer to the question by conducting a thorough study of a maximum depth estimator defined in Section 2. It investigates the large sample properties of the estimator, establishing its $\sqrt{n}$- (and strong) consistency and limiting distribution; it examines the finite sample properties of the estimator, studying its finite sample breakdown point robustness and finite (as well as large) sample efficiency. Developing the asymptotic theory for the maximum depth estimator is rather challenging. Empirical process theory turns out to be vital. In pioneer studies such as He and Portnoy[13], Nolan[14], Bai and He[15] and Zuo[12], a symmetry as well as other strong (e.g., everywhere differentiable density and finite moments) assumptions are imposed on the underlying distribution. Without these assumptions, the task turns out to be much more challenging and technically demanding and is fulfilled in this current paper. Unlike the univariate median and multivariate competitors, the maximum depth estimator can enjoy the highest breakdown point robustness without having to pay the price of a low efficiency. Indeed, it is highly efficient at a variety of light- and heavy-tailed distributions in one and higher dimensions.

The rest of the paper is organized as follows: Section 2 defines a maximum depth estimator and discusses its primary properties. Asymptotic properties of the estimator are thoroughly investigated in Section 3. Section 4 is devoted to the study of finite sample properties of the estimator where its breakdown point robustness and relative efficiency are examined. Real data examples are presented in Section 5. Selected proofs are reserved for the Appendix.

2 A maximum depth estimator and its primary properties

2.1 Outlyingness, projection depth, and maximum depth estimators

In $\mathbb{R}^1$, the outlyingness of a point x with respect to (w.r.t.) a univariate data set (sample) $X^n = \{X_1, \ldots, X_n\}$ is simply $|x - \mu(X^n)|/\sigma(X^n)$, the deviation of x to the center of X^n standardized by the scale of X^n. Here μ and σ are univariate location and scale estimators with typical choices including (mean, standard deviation), (median, median absolute deviation) and, more generally, (M-estimator of location, M-estimator of scale). Mosteller and Tukey see [16, p. 205] introduced and discussed an outlyingness weighted mean in the univariate setting. Stahel[17] and Donoho[4] considered a multivariate analog and defined the outlyingness of a point x w.r.t. X^n in $\mathbb{R}^d$ ($d \geqslant 1$) as

$$O(x, X^n) = \sup_{\{u:\ u \in S^{d-1}\}} \frac{u'x - \mu(u'X^n)}{\sigma(u'X^n)}, \tag{1}$$

where $S^{d-1} = \{u : \|u\| = 1\}$ and $u'X^n = \{u'X_1, \ldots, u'X_n\}$. If $u'x - \mu(u'X^n) = \sigma(u'X^n) = 0$, then we define $(u'x - \mu(u'X^n))/\sigma(u'X^n) = 0$. Along with other notions of data depth, Liu[18], Zuo and Serfling[10] and Zuo[12] defined and discussed "projection depth" (PD) of a point x w.r.t. X^n in $\mathbb{R}^d$

$$\mathrm{PD}(x, X^n) = 1/(1 + O(x, X^n)), \tag{2}$$

and treated the maximum depth (deepest) point (w.r.t. a general depth notion) as a multi-dimensional location estimator (median). With μ and σ being general m-estimators of location and scale, Zuo[12] defined and thoroughly studied

$$T_n := T(X^n) = \arg\sup_{x \in \mathbb{R}^d} \mathrm{PD}(x, X^n), \tag{3}$$

the maximum projection depth estimator (taking average if there is more than one maximizer). For μ and σ being the median (Med) and the median absolute deviation (MAD), T_n is very robust with a breakdown point highest among all competitors. It is also quite efficient with this μ and σ. Indeed, T_n is about 78% efficient relative to the mean at bivariate normal models. This relative efficiency, albeit higher than that of Med (64%) and comparable with those of multivariate competitors, is still low. The low efficiency of Med seems to be the main source. A remedial measure is to replace Med with an efficient location estimator μ. To keep the high breakdown point robustness of T_n, μ should be as robust as Med. We introduce and study such a μ next.

2.2 A maximum projection depth estimator and its primary properties

Though properties in this subsection hold for many general μ and σ, hereafter μ is taken to be the special univariate "projection depth weighted mean" (PWM) for X^n in $\mathbb{R}^1$ (see [19] for a multi-dimensional PWM)

$$\mu(X^n) := \mathrm{PWM}\,(X^n) = \frac{\sum_{i=1}^n w_i X_i}{\sum_{i=1}^n w_i}, \tag{4}$$

where $w_i = w(\mathrm{PD}(X_i, X^n))$, w is a weight function on $[0,1]$, and $\mathrm{PD}(X_i, X^n) = 1/(1 + |X_i - \mathrm{Med}(X^n)|/\mathrm{MAD}(X^n))$; σ is taken to be MAD (though any other robust scale estimator can serve the same purpose). We now confine attention to the corresponding maximum projection depth estimator with this μ and σ. Note that μ is not an M-estimator but an outlyingness (depth) weighted mean. Hence many results in [2] are not applicable to T_n in this paper.

It is readily seen that the μ and the σ are affine equivariant, that is, $\mu(sX^n + b) = s\mu(X^n) + b$ and $\sigma(sX^n + b) = |s|\sigma(X^n)$ for any X^n and scalars s and b in $\mathbb{R}^1$, where $sX^n + b = \{sX_1 + b, \ldots, sX_n + b\}$. With (any) affine equivariant μ and σ, we see that $O(x, X^n)$ is affine invariant, that is, $O(Ax + b, AX^n + b) = O(x, X^n)$ for any nonsingular $d \times d$ matrix A and vector $b \in \mathbb{R}^d$. Consequently, T_n is affine equivariant, that is, $T(AX^n + b) = AT(X^n) + b$ since PD is also affine invariant.

If the distribution of X_i is symmetric about a point $\theta \in \mathbb{R}^d$, that is, $\pm(X_i - \theta)$ have the same distribution, then it is seen that T_n is also symmetric about θ. Furthermore, if $E(X_i)$ exists, then T_n is unbiased for θ, that is, $E(T_n) = \theta$.

Let F_n (F_{nu}) be the empirical distribution based on X^n ($u'X^n$) which places mass $1/n$ at points X_i ($u'X_i$). We sometimes write F_n (F_{nu}) for X^n ($u'X^n$) for convenience. Let F (F_u) be the distribution of X_i ($u'X_i$). Replacing X^n (or F_n) and $u'X^n$ (or F_{nu}) with F and F_u in the above definitions, we obtain the population versions. For example, the popular version of PWM for $F \in \mathbb{R}^1$ is

$$\mu(F) := \mathrm{PWM}\,(F) = \frac{\int w(\mathrm{PD}(x, F))x dF(x)}{\int w(\mathrm{PD}(x, F)) dF(x)}. \tag{5}$$

It can be seen that the above affine invariance or equivariance properties in the sample case hold true in the population case. For example, $T(F)$ is affine equivariant, that is, $T(F_{AX+b}) = AT(F_X) + b$ for any nonsingular $d \times d$ matrix A and any $b \in \mathbb{R}^d$, where F_Z denotes the distribution of Z. Furthermore, if F is symmetric about θ, then $T(F) = \theta$. That is, $T(\cdot)$ is Fisher consistent.

3 Asymptotics for the maximum depth estimator

3.1 Preliminary lemmas

To ensure that μ (PWM) is well defined, we assume that $w(r) > 0$ on $(0, 1]$ and $w(0) = 0$ (and the derivative $w^{(1)}(r)$ is continuous on $[0, 1]$ for the technical sake). Such a w is exemplified in Subsection 4.2. To ensure the scale σ (MAD) is > 0 at F_u, we assume $P(u'X = a) < 1/2$ for any $u \in S^{d-1}$ and $a \in \mathbb{R}^1$. Further to ensure that Med and MAD are unique at F_u for any $u \in S^{d-1}$, we assume

Assumption A1. $F_{u'X}$ and $F_{|u'X - \mathrm{Med}(F_u)|}$ are non-flat in a small right neighborhood of $\mathrm{Med}(F_u)$ and $\mathrm{MAD}(F_u)$, respectively, for any $u \in S^{d-1}$.

Assumption A1 guarantees the continuity of $\mathrm{Med}(F_u)$ and $\mathrm{MAD}(F_u)$ in $u \in S^{d-1}$ and hence $0 < \inf_{u \in S^{d-1}} \mathrm{MAD}(F_u)$. Clearly, $\sup_{u \in S^{d-1}} \mathrm{MAD}(F_u) < \infty$. This boundedness property turns out to be true for μ as well. We have

Lemma 1. *Under Assumption* A1, $0 < \inf_{u \in S^{d-1}} \sigma(F_u) \leqslant \sup_{u \in S^{d-1}} \sigma(F_u) < \infty$, *and* $-\infty < \inf_{u \in S^{d-1}} \mu(F_u) \leqslant \sup_{u \in S^{d-1}} \mu(F_u) < \infty$.

Assumption A1 also ensures the uniform strong consistency of μ and σ.

Lemma 2. *Under Assumption* A1 *and assume F is continuous at* $\mathrm{Med}(F_u) \pm \sigma(F_u)$, $\sup_{u\in S^{d-1}} |\mu(F_{nu}) - \mu(F_u)| = o(1)$ *and* $\sup_{u\in S^{d-1}} |\sigma(F_{nu}) - \sigma(F_u)| = o(1)$, *a.s.*

This lemma implies that Lemma 1 holds a.s. in the sample case for large n. The lemmas also ensure the (Lipschitz) continuity of $\mathrm{PD}(\cdot, G)$, which in turn implies the existence of $T(G)$ since $\mathrm{PD}(x, G) \to 0$ as $\|x\| \to \infty$, for $G = F, F_n$ and for large n. Further, the set of maximum depth points for $G = F$ has an empty interior (see Theorem 2.3 of [12]) and is a singleton in many cases. For example, if F is symmetric about θ, or even just μ-symmetric about θ (i.e. $\mu(F_u) = u'\theta$, $\forall\, u \in S^{d-1}$; see [12]), then $T(F) = \theta$ is unique. Throughout we assume that $T(F)$ is unique. The lemmas also lead to the SLL for $T(F_n)$.

Lemma 3. *Under the conditions of Lemma* 2, $T(F_n) - T(F) = o(1)$, *a.s.*

To discuss the limit distribution of T_n, we need uniform Bahadur type representations of Med and MAD and especially of μ at F_{nu}. To this end, assume

Assumption A2. $F_u^{(2)}$ exists at $\mu(F_u)$ with $F_u^{(1)} = f_u$, $\inf_{u\in S^{d-1}} f_u(\mathrm{Med}(F_u)) > 0$, and $\inf_{u\in S^{d-1}}(f_u(\mathrm{Med}(F_u) + \mathrm{MAD}(F_u)) + f_u(\mathrm{Med}(F_u) - \mathrm{MAD}(F_u))) > 0$.

Remark 1. Assumption A2 is standard for the asymptotic representation of the sample Med and MAD for a fixed u (see, for example, [20, 21]). The differentiability requirement guarantees the expansion of the functionals near $\mu(F_u)$ and $\sigma(F_u)$, and the positivity requirement guarantees the existence of the variances of the functionals (with a non-zero denominators) and the uniqueness of the functionals at $\mu(F_u)$ and $\sigma(F_u)$.

Note that Assumptions A1 and A2 are nested Assumption A2 implies A1. Indeed we have

$$\mathrm{Med}(F_{nu}) - \mathrm{Med}(F_u) = \frac{1}{n}\sum f_1(X_i, u) + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{6}$$

$$\mathrm{MAD}(F_{nu}) - \mathrm{MAD}(F_u) = \frac{1}{n}\sum f_2(X_i, u) + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{7}$$

uniformly in $u \in S^{d-1}$. Let $a_u = \mathrm{Med}(F_u)$, $b_u = \mathrm{MAD}(F_u)$. If F is symmetric (or if $F(a_u+b_u) = 1 - F(a_u - b_u)$, $f_u(a_u + b_u) = f_u(a_u - b_u)$ only), then

$$f_1(x, u) = \frac{1/2 - I(u'x \leqslant a_u)}{f_u(a_u)}, \quad f_2(x, u) = \frac{1/2 - I(|u'x - a_u| \leqslant b_u)}{2f_u(a_u + b_u)}. \tag{8}$$

These representations, proved in Appendix, lead to the following result.

Lemma 4. *Under Assumption* A2, *we have uniformly in* $u \in S^{d-1}$,

$$\mu(F_{nu}) - \mu(F_u) = \int g(x, u) d(F_n - F)(x) + o_p(n^{-1/2}), \tag{9}$$

where

$$\begin{aligned} g(x, u) &= \bigg(\int (y - \mu(F_u)) w^{(1)}(\mathrm{PD}(y, F_u)) h(y, u, x) dF_u(y) \\ &\quad + \frac{(u'x - \mu(F_u)) w(\mathrm{PD}(u'x, F_u))}{\int w(\mathrm{PD}(y, F_u)) dF_u(y)} \bigg), \\ h(y, u, x) &= \frac{\mathrm{PD}^2(y, F_u)}{\mathrm{MAD}(F_u)} \left(\mathrm{sign}(y - a_u) f_1(x, u) + \frac{1 - \mathrm{PD}(y, F_u)}{\mathrm{PD}(y, F_u)} f_2(x, u) \right), \end{aligned}$$

Remark 2. (i) When F is symmetric about a point, say 0, then g, with $\int |y|w^{(1)}(\mathrm{PD}(y, F_u))(\mathrm{PD}^2(y, F_u)f_1(x, u)/b_u)dF_u(y) + u'xw(\mathrm{PD}(u'x, F_u))$ as its numerator, is greatly simplified. (ii) The lemma provides a Bahadur representation of $T(F_n)$ in $\mathbb{R}^1$ since $T(F_n) = \mu(F_n)$ and $T(F) = \mu(F)$ $(u = 1)$ in this case. Hence $\sqrt{n}\,(T(F_n) - T(F)) \xrightarrow{d} N(0, v)$ with $v = E(g^2(x, 1))$ in $\mathbb{R}^1$.

Equipped with results above, we now discuss the $\sqrt{n}$-consistency and asymptotic distribution of the sample maximum projection depth estimator T_n in $\mathbb{R}^d$.

3.2 $\sqrt{\mathrm{n}}$-consistency and limiting distribution

The strong consistency of T_n above raises a natural question: how fast does T_n converge to $T(F)$? We answer this question with a $\sqrt{n}$-consistency result below. Under Assumption A1, Med$(\cdot)$ and MAD$(\cdot)$ and consequently $\mu(\cdot)$ at F_u are continuous in $u \in S^{d-1}$. Hence we have a non-empty set defined as follows:

$$V(T) = \{v \in S^{d-1}:\ O(T, F) = (v'T - \mu(F_v))/\sigma(F_v)\}. \tag{10}$$

It is seen that F is μ-symmetric if and only if $V(T) = S^{d-1}$, which holds if and only if $\{v, -v\} \in V(T)$ for some $v \in S^{d-1}$. Instead of imposing a μ-symmetry constraint on F (i.e. requiring $V(T) = S^{d-1}$) in the discussion below, we assume

Assumption A3. $\sup_{v\in V(T)} u'v \geqslant c$, for some fixed $c \in (0, 1]$ and any $u \in S^{d-1}$.

Clearly, 1 is a choice for c if F is μ-symmetric. Assumption A3 means that any cap of S^{d-1} with hight c always contains at least one member of $V(T)$. We have

Theorem 1. *Under Assumptions* A2 *and* A3, $n^{1/2}(T(F_n) - T(F)) = O_p(1)$.

The theorem extends a corresponding result of Zuo[12] where the $\sqrt{n}$-consistency of general maximum projection depth estimators is established under a (μ-) symmetry assumption, which is vital in the proof there.

With a $\sqrt{n}$-consistency result, an immediate question raised is: does T_n possess any limit distribution? The answer is positive. Lemma 4, in conjunction with results in empirical process theory, implies the weak convergence of $\sqrt{n}\,(\mu(F_{nu}) - \mu(F_u))$, as a process indexed in $u \in S^{d-1}$, to a centered Gaussian process $Z(u)$, $u \in S^{d-1}$, with $Z(u) = -Z(-u)$ and a covariance structure:

$$\mathrm{Cov}(Z(u_1), Z(u_2)) = E(Z(u_1)Z(u_2)) = E(g(X, u_1)g(X, u_2)). \tag{11}$$

This, in conjunction with Assumptions A2 and A3, leads to the following result:

Theorem 2. *Under Assumptions* A2 *and* A3, *we have*

$$\sqrt{n}\,(T(F_n) - T(F)) \xrightarrow{d} \arg\inf_{x\in\mathbb{R}^d} \sup_{u\in V(T)} (u'x - Z(u))/\sigma(F_u).$$

The theorem is proved via an Argmax continuous mapping theorem (see, e.g., [22]). The main idea and the key steps of the proof are as follows. First, in virtue of the uniform representation in Lemma 4 and empirical process theory, we show that $\sqrt{n}\,(\mu(F_{nu}) - \mu(F_u))$ converges weakly to the Gaussian process $Z(u)$. Second, we show that the arg inf in the theorem is unique. Third, we show that $\sup_{u\in S^{d-1}}(u't - \sqrt{n}\,(\mu(F_{nu}) - \mu(F_u)))/\sigma(F_{nu})$ converges weakly to $\sup_{u\in V(T)}(u't - Z(u))/\sigma(F_u)$ in $\ell^\infty(K)$ for any compact $K \subset \mathbb{R}^d$. Finally, the uniform tightness in Theorem 1 completes the proof.

Remark 3. (i) The theorem indicates that the (standardized) sample maximum depth estimator converges weakly to a random variable. It is $Z(1)$ (the same one as $N(0, v)$ discussed in Remark 1 in $\mathbb{R}^1$ and a minimizer of a random function involving a Gaussian process in $\mathbb{R}^d$ ($d \geqslant 2$). (ii) The limiting distribution depends largely on the choice of μ (via $Z(u)$) less on σ. Indeed, any Fisher consistent scale σ at F_u leads to the same result. (iii) We note that Zuo[12] also studied the limiting distribution of maximum depth estimators. However, (μ, σ) there are simultaneous M-estimators of location and scale, which do not cover more involved cases such as (PWM, MAD) here. Furthermore μ-symmetry as well as many other very restrictive conditions are imposed on F in [12]. The theorem here is established without these strong assumptions on F.

4 Robustness and efficiency of the maximum depth estimator

4.1 Finite sample breakdown point robustness

Following Donoho and Huber[1], we define the finite sample replacement breakdown point (BP) of a location estimator L at $X^n = \{X_1, \dots, X_n\}$ in $\mathbb{R}^d$ as

$$\mathrm{BP}(L, X^n) = \min\left\{\frac{m}{n} : \sup_{X^n(m)} \|L(X^n(m)) - L(X)\| = \infty\right\}, \tag{12}$$

where $X^n(m)$ denotes a contaminated sample resulting from replacing m original points of X^n with arbitrary m points in $\mathbb{R}^d$. For a scale estimator S, the same definition can be used with L replaced by $\log(S)$ on the right hand side.

Clearly, the breakdown point of the maximum depth estimator depends on those of μ and σ. More precisely, it depends mainly on the BP of μ and the implosion breakdown point of σ (and less on the explosion breakdown point of σ). As a scale estimator, σ breaks down if it becomes 0 or ∞, corresponding to implosion or explosion. It can be shown that μ (PWM), like the univariate median, also has the best possible BP, $\lfloor (n+1)/2 \rfloor / n$, of any affine equivariant location estimator. The BP of T_n thus depends essentially on that of σ (MAD). Since it is easier to implode MAD with the projected data $u'X^n$ of X^n in high dimension along a special direction u, a modified version, MAD_k ($1 \leqslant k \leqslant n$), has been utilized in the literature to enhance the implosion BP of MAD; see, e.g., [6, 12, 23]. It is the average of the $\lfloor (n+k)/2 \rfloor$th and $\lfloor (n+k+1)/2 \rfloor$th absolute deviations among the n absolute deviations $|x_1 - \mathrm{Med}(x^n)|, \dots, |x_n - \mathrm{Med}(x^n)|$ for data x^n in $\mathbb{R}^1$. Note that MAD_1 is just the standard MAD. A data set X^n is in general position if no more than d data points lie in any $(d-1)$-dimensional hyperplane. If F of X_i is absolute continuous, then X^n is in general position almost surely.

Theorem 3. *Let $w(r)$ be continuous and $0 \leqslant w(r) \leqslant Cr$ for $r \in [0, 1]$ and some $C > 0$ and $w(r) > 0$ for $r \in [1/2, 1]$. Let $(\mu, \sigma) = (\mathrm{PWM}, \mathrm{MAD}_k)$. Then*

$$\mathrm{BP}\,(T_n, X^n) = \begin{cases} \lfloor (n+1)/2 \rfloor / n, & d = 1, \\ \lfloor (n - 2d + k + 3)/2 \rfloor / n, & d \geqslant 2, \end{cases}$$

for X^n in general position and $n \geqslant 2(d-1)^2 + k + 1$, where $k \leqslant (d-1)$ if $d \geqslant 2$.

Remark 4. The theorem indicates that (i) in $\mathbb{R}^1$, T_n has the best possible breakdown point of any affine equivariant location estimator; and (ii) in higher dimension $\mathbb{R}^d$, it has a BP

$\lfloor (n-d+2)/2 \rfloor /n$ for $\sigma = \mathrm{MAD}_{d-1}$, which again is the best for any affine equivariant location estimators constructed in the past two decades. Thus T_n has the best breakdown robustness among existing estimators.

4.2 Finite and large sample relative efficiency

It is well known that the univariate median pays the price of a low efficiency at normal and light-tailed models for its best breakdown robustness. Now T_n also possesses the best breakdown point among any existing affine equivariant competitors in one and higher dimensions. Does it have to pay the same price?

The asymptotic normality of T_n in $\mathbb{R}^1$ established in Section 3 allows one to calculate the asymptotic relative efficiency and absolute efficiency directly (the latter notion is the relative efficiency with respect to Cramér-Rao lower bound (crlb); see, e.g., [24, p. 363] for definition and related discussions). Indeed, we have these efficiency results listed in Table 1. In this subsection, the weight function w is taken to be the continuously differentiable function

$$W(r,c,k) = I(r \geqslant c) + (e^{-k(1-r^2/c^2)^2} - e^{-k})/(1-e^{-k}) I(r < c), \tag{13}$$

where $0 \leqslant c \leqslant 1$ and $k > 0$ are two parameters. The main idea behind this weight function w is as follows. For deep data points, we simply average them to gain high efficiency. For less depth points or points on the outskirts, we exponentially down-weight them to alleviate their influence and consequently to gain robustness. In general, a small c is favorable at light-tailer F's and a large c at heavy-tailer ones. The same is true for k. A weight function similar to W above was first used and discussed in [19]. We consider here normal, logistic, double exponential and cauchy distributions, representing very light- to very heavy-tailed distributions. For simplicity, we take $k = 3$ and c to be 0.2 and 0.8 for light- and heavy-tailed distributions, respectively.

Table 1 Asymptotic and absolute efficiency of T_n with $w(r) = W(r,c,3)$

F	$N(0,1)$	LG(0,1)	DE(0,1)	CAU(0,1)
	$c = 0.2$	$c = 0.2$	$c = 0.8$	$c = 0.8$
$\sigma^2_{\bar{X}_n}/\sigma^2_{T_n}$	0.9595	1.0340	1.9563	∞
$\mathrm{crlb}/\sigma^2_{T_n}$	0.9595	0.9409	0.9782	0.8960

In $\mathbb{R}^1$, T_n has astonishingly high efficiencies, uniformly across a number of light- and heavy-tailed distributions. Indeed, it is about 90% (or higher) efficient relative to the most efficient estimator at each model. It is about 96% efficient relative $\bar{X}_n$ at normal models and is more efficient than $\bar{X}_n$ at other models. By appropriately tuning c and k, one can get even better results in practice.

In $\mathbb{R}^d$ ($d > 1$), the complex limiting distribution of T_n makes it difficult to calculate the asymptotic relative efficiency directly. In the following, we consider the finite sample efficiency of T_n relative to $\bar{X}_n$ at normal and other t models for $d = 2$, 3, and 4. We assume that both T_n and $\bar{X}_n$ aim at estimating a location parameter θ. We generate m samples, each with size n, from each model and calculate the "empirical mean squared error" (EMSE) of T_n: $\sum_{i=1}^m \|T_n^i - \theta\|^2/m$, where T_n^i is the estimate obtained based on the i-th sample. The relative efficiency of T_n with respect to $\bar{X}_n$ is obtained by dividing the EMSE of $\bar{X}_n$ by that of T_n. In our simulation, $m = 1000$, $n = 50$, and θ is the origin.

The efficiency results of T_n relative to $\bar{X}_n$ are given in Table 2, where the EMSE's of $\bar{X}_n$ and T_n are given in the parentheses with those of $\bar{X}_n$ on the numerators. For simplicity, we take $k = 3$ and let $c = 0.8$ for $t(3)$ and Cauchy distribution cases. At normal models, we take $c = 1/(1 + 2\sqrt{2d-3})$ which decrease at a rate $O(1/\sqrt{d})$ so that the relative efficiency of T_n increases.

Table 2 Finite sample efficiency of T_n relative to $\bar{X}_n$ with $w(r) = W(r, c, 3)$

F	$N_d(0,1)$	$t_d(3)$	$\mathrm{CAU}_d(0,1)$
	$c = \frac{1}{1+2\sqrt{2d-3}}$	$c = 0.8$	$c = 0.8$
$d = 2$	0.959 $(\frac{0.03950}{0.04118})$	1.962 $(\frac{0.11637}{0.05931})$	18157.730 $(\frac{1552.02}{0.08547})$
$d = 3$	0.964 $(\frac{0.06309}{0.06547})$	2.076 $(\frac{0.17487}{0.08422})$	845324.01 $(\frac{99147.37}{0.11729})$
$d = 4$	0.970 $(\frac{0.07971}{0.07465})$	2.070 $(\frac{0.23417}{0.11312})$	5694367.8 $(\frac{889577.3}{0.15622})$

Inspecting the table entries reveals that (i) like in the univariate case, the maximum depth estimator is highly efficient relative to the mean at normal and other heavy-tailed distributions in the high dimensional cases; (ii) the relative efficiency increases as the dimension d increases; and (iii) the maximum depth estimator is highly robust against heavy-tailed t distributions. Indeed T_n, unlike the sample mean $\bar{X}_n$, possesses a sample covariance matrix that has finite entries at multi-dimensional cauchy distributions and that is approximately $0.04 * I_d$.

The efficiency results in Table 2 do not depend much on the sample size n, Indeed, our extensive simulation studies confirm this. For example, the relative efficiency of T_n at bivariate normal models for a variety of sample sizes including $100, 200, \ldots, 1000$ ranges from 95% to 97%. The high efficiency of T_n relative to $\bar{X}_n$ at bivariate normal models is illustrated in the scatter plots of 400 sample means and sample maximum depth estimators in Figure 1 below.

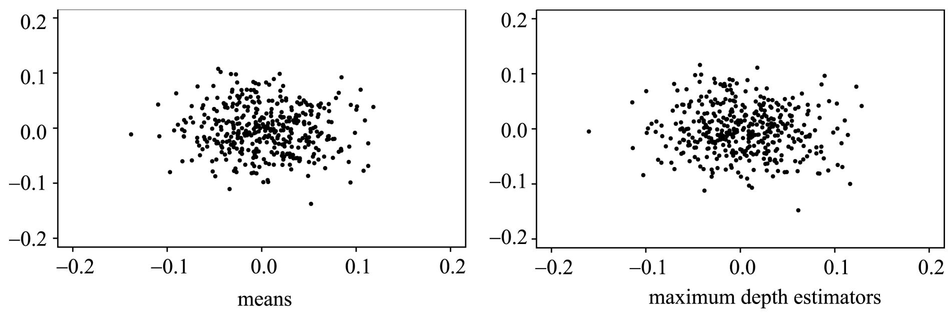

Figure 1 400 bivariate means and 400 maximum projection depth estimators based on $N_2(0, I_2)$ samples of size 500 are scattered roughly in the same pattern.

The results in Table 2 depend on the choices of c and k. Appropriately tuning these parameters, one can get better results. For example, with $c = 0$ at normal models, the efficiencies of T_n relative to $\bar{X}_n$ increase to 100%. But T_n $(= \bar{X}_n)$ is no longer robust and conditions on w in Theorem 3 are violated.

There are many competitors of T_n in the literature. The breakdown points of these estimators, however, are not as high as that of T_n. Furthermore, most leading competitors are less efficient. For example, at bivariate normal models, the efficiency relative to $\bar{X}_n$ is 72% for transformation-retransformation median (see [25]), 76% for Tukey halfspace median (see [26]), 78.5% for Hettmansperger-Randles median (see [7]), 78% for the projection median (see [12]), but 96% for T_n.

We also studied the efficiency of T_n at contaminated (or mixed) normal and other distributions. Simulation results indicate, as we expect, that the robust T_n is overwhelmingly more efficient than the sample mean in all those cases.

5 Real data examples

We apply the maximum projection depth estimator to the real data sets to illustrate its (i) high robustness, (ii) high efficiency, and (iii) computability.

5.1 Gilgai survey data

This data set is generated from a line transect survey in gilgai territory in New South Wales, Australia. Gilgais are natural gentle depressions in otherwise flat land, and sometimes seem to be regularly distributed. The data collection was stimulated by the question: are these patterns reflected in soil properties?

At each of 365 sampling locations on a linear grid of 4 meters spacing, samples were taken at depths 0–10 cm, 30–40 cm and 80–90 cm below the surface. pH, electrical conductivity (EC) and chloride content (CC) were measured on a 1:5 soil:water extract from each sample. Further details may be found in [27]. This data set was also analyzed in [28].

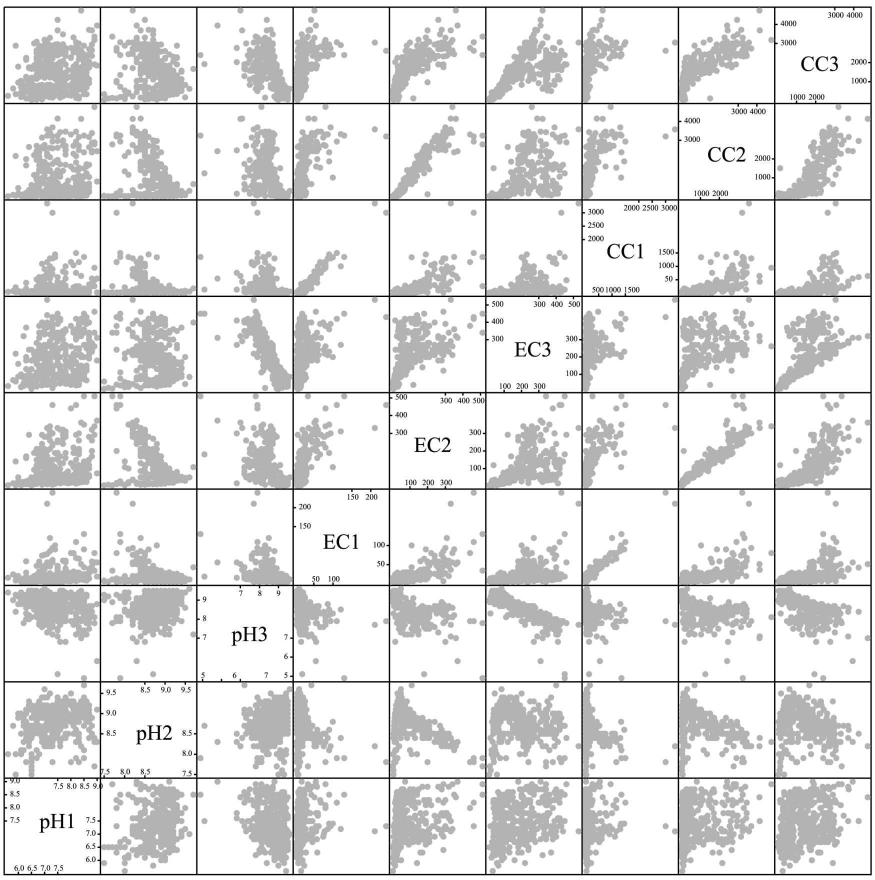

Figure 2 Pair-wise scatter plot of gilgai soil data[27, 28].

A pair-wise scatter plot of this data set (365×9) is given in Figure 2. Inspecting the plot reveals that there are quite a lot variation and unusual points in the data. Histograms of the individual variables confirm this observation. It seems that the measurements of the electrical conductivity and chloride content at $0-10$ cm and $30-40$ cm depths have extremely high variability whereas those at $80-90$ cm depth are quite stable except some extreme measurements. The measurements of pH have much less variability except some extreme measurements at the deepest level. The coordinate-wise minimum, the sample mean, the coordinate-wise median, T_n, and the coordinate-wise maximum, of this data set are obtained (in less than 3 minutes) with a program in Fortran and listed below respectively

(5.6	7.5	4.9	5.0	6.0	15.0	20.0	20.0	50.0),
(7.406	8.787	8.585	20.56	95.77	184.3	171.7	894.3	1450.0),
(7.3	8.9	8.7	13.0	54.0	170.0	38.0	355.0	1400.0),
(8.578	8.128	9.225	18.564	76.14	164.3	98.91	698.4	1327.0),
(9.0	9.7	9.6	240.0	510.0	530.0	3350.0	4770.0	4725.0).

The three location estimates in the middle are quite different from each other. This is due to the high variability of the data as well as the robustness and efficiency properties of the estimators. The 7th and 8th components of the sample mean are especially different from those of the other two. A scatter plot of 7th and 8th components (CC1 and CC2) is given in Figure 3(a) which displays the strong abnormality of these two variables. A histogram of CC1 in Figure 3(b) indicates that the majority of CC1 values are less than 100. Indeed there are 264 out of 365 (72.3%) of CC1 values no greater than 100. The 7th component of both T_n and the coordinate-wise median is less than 100 while that of the sample mean is 171.7425. On the other hand, there are only about 21.6% of CC1 values that are greater than 171.7425. All the evidence manifests the strong robustness of T_n (and the coordinate-wise median) and the vulnerability of the sample mean with respect to unusual observations. Geometrically, T_n is closer to the sample mean than the coordinate-wise median (and the coordinate-wise minimum and maximum). This reflects that T_n can be very efficient while enjoying the high breakdown robustness.

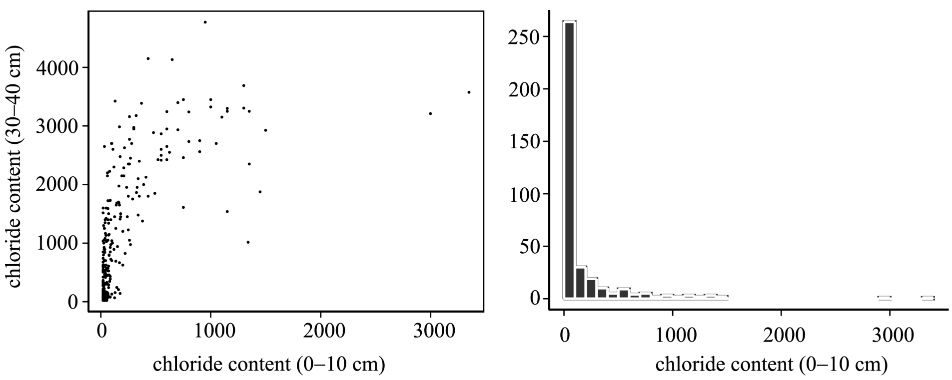

Figure 3 (a) Scatter plot of CC1 and CC2. (b) Histogram of CC1.

6 Concluding remarks

Univariate median possesses the highest breakdown point robustness among all affine equivariant location estimators. It has long served as a very robust measure of the center of univariate data. Constructing its affine equivariant analogues in $\mathbb{R}^d$ $(d > 1)$ is very desirable but turns out to be non-trivial. Among all constructed, except the projection median in [12], none possesses a breakdown point higher than $\lfloor (n-d+1)/2 \rfloor/n$. Furthermore, like the univariate median, most of these estimators (including the projection median) have to sacrifice their efficiency for achieving their high breakdown point robustness.

The maximum depth estimator proposed and studied in this paper not only breaks the breakdown-point barrier $\lfloor (n-d+1)/2 \rfloor/n$ that lasted two decades in the literature but also possesses simultaneously very high efficiencies at a variety of light- and heavy-tailed distributions in one and higher dimensions.

The limiting distribution of the maximum depth estimator in $\mathbb{R}^d$ $(d > 1)$ is somewhat uncommon. It is the minimizer of a random function that involves a Gaussian process. Establishing such a limiting distribution without the usual symmetry and other strong assumptions on the underlying distribution is rather challenging and is one of major objectives and contributions of this paper. In practical inference, bootstrapping techniques can be invoked to estimate the variance-covariance structure of the estimator. Details will be pursed elsewhere.

Although the maximum depth estimator can possess both the affine equivariance and the high breakdown point robustness without sacrificing its extremely high efficiency, for these gains it does pay a high price in the computing. Just like all other affine equivariant location estimators that have really high breakdown points, the maximum depth estimator, with no exception, is computationally intensive. The key difficulty here lies in the calculation of the outlyingness that is defined based on projections to all unit directions. The related computing is very involved, if not impossible. Recent studies of this author indicate, however, that the outlyingness function can actually be computed exactly in two and higher dimensions. An implementable algorithm in $\mathbb{R}^d$ $(d > 2)$, however, is yet to be developed. In the simulation studies of this paper, we employed an approximate algorithm that can compute the depth of a point in $\mathbb{R}^d$ $(d \geqslant 2)$ quite fast. For the global minimum (maximum), a downhill simplex algorithm is also utilized in our calculation. (For the related program, please see http://www.stt.msu.edu/˜zuo/pmsdmadrdRE.f.htm). A detailed account of the computing issue of the depth estimator will be pursued elsewhere.

Finally, we remark that the general definition of the location estimator in (3) is not new and was suggested in [6] under the outlyingness framework. Zuo[12] defined and studies the estimator (called projection median) under the data depth framework. In those studied μ, however, is either Med[6] or a general M-estimator[12]. The choice of $\mu = \mathrm{PWM}$, with a primary version suggested first in [12] but defined with a different weight w and with the parameter tuning here, is first studied in this current paper.

Appendix

Proof of Lemma 1. The assertion on σ is trivial. This assertion and the continuity of $w^{(1)}$ on $[0,1]$ imply that uniformly in $x \in \mathbb{R}^1$ and in $u \in S^{d-1}$

$$|x| w(\mathrm{PD}(x, F_u)) = |x|\, w^{(1)}(\eta(x, F_u))/(1 + |x - \mathrm{Med}(F_u)|/\mathrm{MAD}(F_u)) < \infty,$$

where $\eta(x, F_u)$ is between 0 and $\mathrm{PD}(x, F_u)$. Hence

$$\sup_{u \in S^{d-1}} \sup_{x \in \mathbb{R}^1} (|x| + 1) w(\mathrm{PD}(x, F_u)) < \infty. \tag{14}$$

Clearly $\int w(PD(x, F_u))dF_u > 0$ uniformly in $u \in S^{d-1}$ since $w(PD(x, F_u)) > 0$ uniformly in $u \in S^{d-1}$ for any $\|x\| \leqslant M < \infty$. This, in conjunction with the last display and the fact that $-\mu(F_{-u}) = \mu(F_u)$, gives the desired result.

Proof of Lemma 2. The continuity of F_u and $\mathrm{Med}(F_u)$ in $u \in S^{d-1}$, the compactness of S^{d-1}, and Assumption A1 implies that for any $\epsilon > 0$

$$\begin{aligned}\delta_\epsilon &= \min\Big\{ \inf_{u \in S^{d-1}} (F_u(\mathrm{Med}(F_u) + \epsilon) - 1/2),\ \inf_{u \in S^{d-1}} (1/2 - F_u(\mathrm{Med}(F_u) - \epsilon)) \Big\} \\ &= \min\{F_{u1}(\mathrm{Med}(F_{u1}) + \epsilon) - 1/2,\ 1/2 - F_{u2}(\mathrm{Med}(F_{u2}) - \epsilon)\} > 0,\end{aligned}$$

for some fixed $u1, u2 \in S^{d-1}$. By Theorem 2.3.2 of [3],

$$P(|\mathrm{Med}(F_{nu}) - \mathrm{Med}(F_u)| > \epsilon) \leqslant 2\, e^{-2n\delta_\epsilon^2},\ \forall\, n,$$

which, together with Borel-Cantelli lemma, ensures the strong consistency of $\mathrm{Med}(F_{nu})$ with $\mathrm{Med}(F_u)$ uniformly in $u \in S^{d-1}$. The same argument leads to the same conclusion for $\mathrm{MAD}(F_{nu})$ (with $\mathrm{MAD}(F_u)$). That is, almost surely

$$\sup_{u \in S^{d-1}} |\mathrm{Med}(F_{nu}) - \mathrm{Med}(F_u)| = o(1), \quad \sup_{u \in S^{d-1}} |\mathrm{MAD}(F_{nu}) - \mathrm{MAD}(F_u)| = o(1). \tag{15}$$

We now show the uniform strong consistency of $\mu(F_{nu})$ with $\mu(F_u)$. Write

$$\mu(F_{nu}) - \mu(F_u) = \frac{\int (x - \mu(F_u)) w(\mathrm{PD}(x, F_{nu})) dF_{nu}(x)}{\int w(\mathrm{PD}(x, F_{nu})) dF_{nu}(x)}. \tag{16}$$

The numerator can be decomposed into

$$\begin{aligned} I_{1n} + I_{2n} := &\int (x - \mu(F_u))(w(\mathrm{PD}(x, F_{nu})) - w(\mathrm{PD}(x, F_u))) dF_{nu}(x) \\ &+ \int (x - \mu(F_u)) w(\mathrm{PD}(x, F_u)) d(F_{nu} - F_u)(x).\end{aligned}$$

By Lemma 1 and its proof, we see that the integrand of I_{2n} is bounded uniformly in $u \in S^{d-1}$. It is readily seen that $\mu(F_u)$ is continuous in $u \in S^{d-1}$ and so is I_{2n} consequently. Employing this continuity and the compactness of S^{d-1}, and invoking Hoeffding's inequality (see Proposition 2.3.2 of [3], for example) and Borel-Cantelli lemma, we see that I_{2n} converges to 0 a.s. and uniformly in $u \in S^{d-1}$. If we can now show the same is true for I_{1n}, then with the same (slightly less involved) argument we can show that the denominator converges a.s. to $\int w(\mathrm{PD}(x, F_u)) dF_u(x)$ and uniformly in u. The result follows.

To treat I_{1n}, we observe that the continuity of $w^{(1)}$ on $[0, 1]$ implies

$$\begin{aligned} &\Big| \int (x - \mu(F_u))(w(\mathrm{PD}(x, F_{nu})) - w(\mathrm{PD}(x, F_u))) dF_{nu}(x) \Big| \\ &\leqslant \int |x - \mu(F_u)||w^{(1)}(\eta_n(x, u))||\mathrm{PD}(x, F_{nu}) - \mathrm{PD}(x, F_u)|\, dF_{nu}(x) \\ &\leqslant C \sup_{x \in \mathbb{R}^1} |x - \mu(F_u)||\mathrm{PD}(x, F_{nu}) - \mathrm{PD}(x, F_u)|,\end{aligned}$$

for some $C > 0$, where $\eta_n(x, u)$ is a point between $\mathrm{PD}(x, F_{nu})$ and $\mathrm{PD}(x, F_u)$. In virtue of Lemma 1, the uniform boundedness of $\mathrm{Med}(F_u)$ and $\mathrm{MAD}(F_u)$ in u, Display (15) and Theorem 2.2 and Remark 2.5 of [12], we have that

$$\sup_{x \in \mathbb{R}^1} (|x| + |\mu(F_u)|) |\mathrm{PD}(x, F_{nu}) - \mathrm{PD}(x, F_u)| = o(1), \text{ a.s.} \tag{17}$$

uniformly in $u \in S^{d-1}$. We thus complete the proof of the lemma.

Proof of Lemma 3. Lemmas 1 and 2 imply that there is a $C > 0$ such that

$$|O(x, F) - O(y, F)| \leqslant C \|x - y\|, \quad \sup_{x \in S} |O(x, F_n) - O(x, F)| = o(1), \text{ a.s.} \tag{18}$$

for any bounded subset S of $\mathbb{R}^d$. Further, for $\theta = T(F)$, $G = F$, F_n, we have

$$O(\theta, F) < \infty, \qquad O(x, G) \to \infty, \quad \text{a.s.} \quad \text{as} \quad \|x\| \to \infty. \tag{19}$$

This, in conjunction with the continuity of $O(x, F)$ in x, implies that

$$\alpha_\epsilon := \inf_{\|x - \theta\| \geqslant \epsilon} O(x, F) > O(\theta, F), \quad \text{for any } \epsilon > 0, \tag{20}$$

since otherwise there exist x_n $(n \geqslant 1)$ and x_0 in $\mathbb{R}^d$ such that $\|x_n - \theta\| \geqslant \epsilon$, $\|x_0 - \theta\| \geqslant \epsilon$, and $O(\theta, F) = \lim_{n \to \infty} O(x_n, F) = O(x_0, F)$, which, however, is a contradiction. Now in the light of Displays (18)–(20), there exists an $M > 0$ such that $\|T_n - \theta\| \leqslant M$ a.s. and for n sufficiently large

$$\begin{aligned} O(\theta, F_n) &< O(\theta, F) + \frac{\alpha_\epsilon - O(\theta, F)}{2} = \alpha_\epsilon + \frac{O(\theta, F) - \alpha_\epsilon}{2} \\ &= \inf_{\epsilon \leqslant \|x - \theta\| \leqslant M} O(x, F) + \frac{O(\theta, F) - \alpha_\epsilon}{2} \\ &< \inf_{\epsilon \leqslant \|x - \theta\| \leqslant M} O(x, F_n) \ = \ \inf_{\|x - \theta\| \geqslant \epsilon} O(x, F_n), \quad \text{a.s.,} \end{aligned}$$

which implies the strong consistency of T_n.

Proof of Displays (6) *and* (7). Assumption A2 permits an asymptotic representation of the sample Med for any $u \in S^{d-1}$; see Theorem 2.5.1 of [3], for example. A careful examination of the proof of that theorem reveals that such representation actually can be uniform in $u \in S^{d-1}$ by virtue of Assumption A2. Hence

$$\mathrm{Med}(F_{nu}) - \mathrm{Med}(F_u) = \frac{1}{n} \sum_{i=1}^{n} \frac{1/2 - I(u' X_i \leqslant \mathrm{Med}(F_u))}{f_u(\mathrm{Med}(F_u))} + R_n^1. \tag{21}$$

where $R_n^1 = O(n^{-3/4} (\log n)^{3/4})$ uniformly in $u \in S^{d-1}$. A similar but more involved argument leads to the asymptotic representation of the sample MAD

$$\mathrm{MAD}(F_{nu}) - \mathrm{MAD}(F_u) = \frac{1}{n} \sum_{i=1}^{n} \frac{1/2 - I(|u' X_i - a_u| \leqslant b_u)}{f_u(a_u + b_u) + f_u(a_u - b_u)} + R_n^2, \tag{22}$$

where $R_n^2 = O(n^{-3/4} (\log n)^{3/4})$ uniformly in $u \in S^{d-1}$ if F is symmetric. The proof for non-symmetric F is the same with more complicated $f_2(x, u)$.

Proof of Lemma 4. By Display (16) and the proof of Lemma 2, the denominator of $\mu(F_{nu}) - \mu(F_u) = \int w(\mathrm{PD}(x, F_u))dF_u(x) + o(1)$, a.s. and uniformly in $u \in S^{d-1}$ and its numerator can be decomposed into $I_{1n} + I_{2n}$ with

$$I_{1n} = \int (x - \mu(F_u))w^{(1)}(\eta_n(x,u))(\mathrm{PD}(x, F_{nu}) - \mathrm{PD}(x, F_u))dF_{nu}(x),$$

where $\eta_n(x,u)$ is a point between $\mathrm{PD}(x, F_{nu})$ and $\mathrm{PD}(x, F_u)$. By Displays (21) and (22), Lemma 1, and Theorem 2.2 and Remark 2.5 of [12], we have

$$\sup_{u\in S^{d-1}} \sup_{x\in\mathbb{R}^1} (|x| + |\mu(F_u)|)|\mathrm{PD}(x, F_{nu}) - \mathrm{PD}(x, F_u)| = O_p(n^{-1/2}). \tag{23}$$

This, together with the continuity of $w^{(1)}$ and Display (17), implies that

$$I_{1n} = \int (x - \mu(F_u))w^{(1)}(\mathrm{PD}(x, F_u))(\mathrm{PD}(x, F_{nu}) - \mathrm{PD}(x, F_u))dF_{nu}(x) + o_p\left(\frac{1}{\sqrt{n}}\right),$$

uniformly in $u \in S^{d-1}$. Now we appeal empirical process theory to show that

$$I_{1n} = \int (x - \mu(F_u))w^{(1)}(\mathrm{PD}(x, F_u))(\mathrm{PD}(x, F_{nu}) - \mathrm{PD}(x, F_u))dF_u(x) + o_p(n^{-1/2}), \tag{24}$$

uniformly in $u \in S^{d-1}$. Adopt Pollard[29] notation system and define

$$\mathcal{F} = \left\{\frac{(u'\cdot - \mu(F_u))w^{(1)}(\mathrm{PD}(u'\cdot, F_u))}{1 + |u'\cdot - \alpha|/\beta} : u \in S^{d-1}, \alpha \in I_a, \beta \in I_b\right\},$$

where $I_a = (a_1 - \delta, a_2 + \delta)$, $I_b = (b_1 - \delta, b_2 + \delta)$, $a_1 = \inf_{u\in S^{d-1}} a_u$ and $a_2 = \sup_{u\in S^{d-1}} a_u$ are bounded, $b_1 = \inf_{u\in S^{d-1}} b_u > 0$ and $b_2 = \sup_{u\in S^{d-1}} b_u < \infty$, and $0 < \delta < b_1$ is a fixed number. Let $\gamma = (u, \alpha, \beta)$. Then the space Γ formed by all γ's is a bounded subspace of $\mathbb{R}^{d+1}$. Define

$$f_n(x) = (u'x - \mu(F_u))w^{(1)}(\mathrm{PD}(u'x, F_u))\mathrm{PD}(u'x, F_{nu}),$$
$$f(x) = (u'x - \mu(F_u))w^{(1)}(\mathrm{PD}(u'x, F_u))\mathrm{PD}(u'x, F_u),$$

for $x \in \mathbb{R}^d$. Then $f \in \mathcal{F}$ and $f_n \in \mathcal{F}$ a.s. for large n. A straightforward but tedious analysis reveals that for any $f_i \in \mathcal{F}$ corresponding to γ_i, $i = 1, 2$,

$$|f_1(x) - f_2(x)| \leqslant M\,\|\gamma_1 - \gamma_2\|, \quad \forall \gamma_1, \gamma_2 \in \Gamma,$$

where $\gamma_i = (u_i, \alpha_i, \beta_i)$, $i = 1, 2$, and the constant M does not depend on γ_i, $i = 1, 2$. In the light of empirical process theory, $\mathcal{F}$ is a Donsker class; see Theorem 2.7.11 of [22] or Example 19.7 of [30], for example. By Display (23) and the boundedness of $w^{(1)}(\mathrm{PD}(x, F_u))$ we have

$$\int ((x - \mu(F_u))w^{(1)}(\mathrm{PD}(x, F_u))(\mathrm{PD}(x, F_{nu}) - \mathrm{PD}(x, F_u)))^2 dF_u(x) = o_p(1),$$

uniformly in $u \in S^{d-1}$. Now invoking Equicontinuity lemma VII.4.15 of [29] or Proposition 29.24 of [30], we obtain Display (24).

A tedious (technically not very challenging) derivation yields that for $x \neq a_u$

$$\mathrm{PD}(x, F_{nu}) - \mathrm{PD}(x, F_u) = \frac{\mathrm{PD}^2(x, F_u)}{\mathrm{MAD}(F_u)}\left[\mathrm{sign}(x - a_u)\, l_{nu} + \frac{1 - \mathrm{PD}(x, F_u)}{\mathrm{PD}(x, F_u)}\, s_{nu}\right]$$
$$+ o_p\left(\frac{1}{\sqrt{n}}\right), \quad \text{uniformly in } x \in \mathbb{R}^1 \text{ and in } u \in S^{d-1},$$

where $l_{nu} := \mathrm{Med}(F_{nu}) - \mathrm{Med}(F_u)$, and $s_{nu} := \mathrm{MAD}(F_{nu}) - \mathrm{MAD}(F_u)$. Thus,

$$\mathrm{PD}(x, F_{nu}) - \mathrm{PD}(x, F_u) = \int h(x, u, y)\, d(F_n - F)(y) + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{25}$$

uniformly in $u \in S^{d-1}$ for $x \neq a_u$. By Display (24) and the boundedness of $(x - \mu(F_u))w^{(1)}$ $(\mathrm{PD}(x, F_u))h(x, u, y)$, and invoking Fubini's theorem, we have

$$I_{1n} = \int \left[\int (y - \mu(F_u))w^{(1)}(\mathrm{PD}(y, F_u))h(y, u, x)dF_u(y)\right] d(F_n - F)(x) + o_p\left(\frac{1}{\sqrt{n}}\right),$$

uniformly in $u \in S^{d-1}$. This and I_{2n} lead to the desired result.

Proof of Theorem 1. Assume, w.l.o.g., that $T(F) = 0$. It is seen that

$$O(T_n, F_n) \leqslant O(0, F_n) = O(0, F) + (O(0, F_n) - O(0, F)) = O(0, F) + O_p(n^{-1/2}).$$

where the last equality follows from Assumption A2, Lemma 4, and the proof of Theorem 2.2 of [12]. On the other hand, for any $v \in V(0)$,

$$O(T_n, F_n) \geqslant \frac{v'T_n - \mu(F_{nv})}{\sigma(F_{nv})} = \frac{v'T_n}{\sigma(F_{nv})} + O(0, F) + \frac{\mu(F_v)}{\sigma(F_v)} - \frac{\mu(F_{nv})}{\sigma(F_{nv})},$$

which, combining with Lemma 4 and the last display, gives

$$\sup_{v \in V(0)} \frac{v'(n^{1/2}T_n)}{\sigma(F_{nv})} \leqslant O_p(1). \tag{26}$$

Assumption A3 implies there are $v_n \in V(0)$ such that $v_n'(n^{1/2}T_n) \geqslant c\|n^{1/2}T_n\|$, which, the last display, and Lemmas 1 and 2, yield the desired result.

Proof of Theorem 2. We invoke an Argmax continuous mapping theorem (Theorem 3.2.2 of [22]) to fulfill our task here. Assume, w.l.o.g., that $T(F) = 0$. Define $M(t, v) := (v't - Z(v))/\sigma(F_v)$, $M(t) := \sup_{v \in V(0)} M(t, v)$, $M_n(t) := \sup_{u \in S^{d-1}} (u't - \sqrt{n}\,(\mu(F_{nu}) - \mu(F_u))/\sigma(F_{nu})$ with Z specified before the theorem. It takes several steps to complete the proof.

Firstly, we show that $\sqrt{n}\,(\mu(F_{nu}) - \mu(F_u))$, a process indexed by $u \in S^{d-1}$, converges weakly to the gaussian process $Z(u)$. This is done if we can show the functions of $g(x, u)$ with $u \in S^{d-1}$ form a Donsker class (see [22] or [30] for definition and discussions). Define

$$\mathcal{G} = \{g(\cdot, u) : u \in S^{d-1}\}, \quad \mathcal{G}_1 = \left\{\frac{(u' \cdot - \alpha_1)w(\mathrm{PD}(u'\cdot, F_u))}{\zeta_u} : u \in S^{d-1},\ |\alpha_1| \leqslant \eta\right\},$$
$$\mathcal{G}_2 = \left\{f_1(\cdot, u)\beta_1 + f_2(\cdot, u)\beta_2 : u \in S^{d-1},\ |\beta_i| \leqslant \sup_{u \in S^{d-1}} |g_i(u)|,\ i = 1, 2\right\}$$

where $\eta = \sup_{u \in S^{d-1}} \mu(F_u)$, $\zeta_u = \int w(\mathrm{PD}(y, F_u))dF_u(y)$, and

$$g_1(u) = \int \frac{(y - \mu(F_u))\,\mathrm{sign}(y - \mathrm{Med}(F_u))\, w^{(1)}(\mathrm{PD}(y, F_u))\,\mathrm{PD}^{\,2}(y, F_u)}{\mathrm{MAD}(F_u)\zeta_u}\, dF_u(y),$$
$$g_2(u) = \int \frac{(y - \mu(F_u))\, w^{(1)}(\mathrm{PD}(y, F_u))\,(\mathrm{PD}(y, F_u) - \mathrm{PD}^2(y, F_u))}{\mathrm{MAD}(F_u)\zeta_u}\, dF_u(y).$$

Following the argument given in the proofs of Theorem 3.3 and Remark 2.4 of [12], we see that all functions $f_i(\cdot, u)$ for $u \in S^{d-1}$ form a Donsker class for $i = 1, 2$, respectively. Note that

$g_i(u)$ are bounded for $u \in S^{d-1}$ and so are the functions in $\mathcal{G}_2$. Now invoking Theorem 2.10.6 of [22], we conclude that $\mathcal{G}_2$ is Donsker. Employing the same argument for $\mathcal{F}$ in the proof of Lemma 4 and invoking Theorem 2.7.11 of [22], we see that $\mathcal{G}_1$ is also Donsker. Obviously $\mathcal{G} \subset \mathcal{G}_1 + \mathcal{G}_2$. The latter is the pairwise sums of functions from the two classes. Now in virtue of Theorems 2.10.1 and 2.10.6 and Example 2.10.7 of [22], we conclude that $\mathcal{G}$ is Donsker. By Lemma 4, we conclude (in virtue of Theorem VII. 21 of [29] or Section 2.1 of [22]) that $\{\sqrt{n}\,(\mu(F_{nu}) - \mu(F_u)) : u \in S^{d-1}\}$ converges weakly to $\{Z(u) : u \in S^{d-1}\}$, where the Brownian bridge Z is a tight Borel measurable element in $l^\infty(\mathcal{G})$ and $Z(-u) = -Z(u)$ with the covariance specified before the theorem.

Secondly, we show that $M(t)$ has continuous sample paths and unique minimum points $\hat{t}$ almost surely and $\hat{t}$ is tight. The continuity is trivial. The tightness is equivalent to the measurability which is straightforward (see [29], for example) if there exists a minimum. We now focus on the existence and the uniqueness of $\hat{t}$. Clearly, $M(t) \to \infty$ as $\|t\| \to \infty$ and it is readily seen that $M(\cdot)$ is convex. The existence of $\hat{t}$ thus follows. To show the uniqueness, we follow the line of Massé[31]. Let $d > 1$. Define $\mathcal{M}(\hat{t}) := \{v \in V(0) :\ M(\hat{t}) = M(\hat{t}, v)\}$. Clearly, it is non-empty. Further, if $\hat{s}$ is another minimizer, then so is $\alpha\hat{s} + (1-\alpha)\hat{t}$ for any $\alpha \in [0, 1]$ by the convexity of $M(t)$. Following [14], we can show in a straightforward fashion that

$$\sup_{v \in \mathcal{M}(\hat{t})} v'u \geqslant 0, \quad \forall\, u \in S^{d-1};\ \mathcal{M}(\alpha\hat{s} + (1-\alpha)\hat{t}) = \mathcal{M}(\hat{t}) \cap \mathcal{M}(\hat{s}), \quad \forall\, \alpha \in (0, 1). \tag{27}$$

The linear space spanned by $\mathcal{M}(\alpha\hat{s} + (1-\alpha)\hat{t})$ for any $\alpha \in (0, 1)$ has dimension $r > 1$. Otherwise, by the last display $\mathcal{M}(\alpha\hat{s} + (1-\alpha)\hat{t}) = \{v, -v\} \subseteq V(0)$ for some v, which implies that $V(0) = S^{d-1}$. Since $M(t, u)$ is odd in u,

$$M(\hat{t}) = M(\hat{t}, v) = \sup_{u \in S^{d-1}} M(\hat{t}, u) \geqslant \inf_{u \in S^{d-1}} M(\hat{t}, u) = M(\hat{t}, -v) = M(\hat{t}),$$

which implies $r = d$ since $0 = M(\hat{t}) = M(\hat{t}, u) = M(\hat{s}, u)$ for any $u \in V(0)$. Let $v_1, \ldots, v_r$ be linearly independent members of $\mathcal{M}(\alpha\hat{s} + (1-\alpha)\hat{t})$ and G be any r-dimensional linear space containing $\hat{t}$ and $\hat{s}$. Then Display (27) implies that both points $\hat{t}$ and $\hat{s}$ meet the linear system $v_i's = Z(v_i) - M(\hat{t})\sigma(F_{v_i})$, $s \in G$, $i = 1, \ldots, r$, which implies that $\hat{s} = \hat{t}$.

Thirdly, we show that M_n converges weakly to M in $\ell^\infty(K)$ for every compact $K \in \mathbb{R}^d$. First, employing an argument similar to that in the proof of Theorem 3.5 of [31], we can show that asymptotically for weak convergence in $\ell^\infty(K)$, $M_n(t)$ is equivalent to $M_n^*(t) := \sup_{u \in V(0)}(u't - \sqrt{n}\,(\mu(F_{nu}) - \mu(F_u)))/\sigma(F_{nu})$. Now invoking continuous mapping theorem and following the proof of lemma A.5 of [12], we can show that M_n^* converges weakly to M in $\ell^\infty(K)$ and hence so does M_n.

Finally, the uniform tightness of $\sqrt{n}\,(T(F_n) - T(F))$ based on Theorem 1 and the Argmax theorem complete the proof.

Proof of Theorem 3. Recall (see Definition (1)) that if $|x - \mu(x^n)| = \sigma(x^n) = 0$, then the outlyingness $O(x, x^n) = 0$ for x^n in $\mathbb{R}^1$ since x is at the "center".

First, considering the case $d = 1$, we note that $T(X^n) = \mu(X^n)$. Since $\lfloor (n+1)/2 \rfloor / n$ is the upper bound of BP of any affine equivariant estimators[4], we need only show that if

$m = \lfloor (n+1)/2 \rfloor - 1$, then m contaminating points can not break down T_n. Clearly for this m

$$\sup_{X^n(m)} |\mathrm{Med}(X^n(m))| < \infty, \quad \sup_{X^n(m)} \mathrm{MAD}(X^n(m) < \infty.$$

Let $X_{(1)}(m) \leqslant \cdots \leqslant X_{(n)}(m)$ be the ordered value of $\{X_1(m), \ldots, X_n(m)\} = X^n(m)$. Then it is not difficult to see that $O(X_{(\lfloor (n+1)/2 \rfloor)}(m), X^n(m)) = 0$ or 1 for n odd or even respectively. Hence the denominator of T_n (PWM) is bounded away from 0 uniformly in $X^n(m)$. On the other hand, we observe

$$\begin{aligned} & w(\mathrm{PD}(X_i(m), X^n(m)))|X_i(m)| \\ & \leqslant w(\mathrm{PD}(X_i(m), X^n(m)))(\mathrm{MAD}(X^n(m))O(X_i(m), X^n(m)) + |\mathrm{Med}(X^n(m))|) \\ & \leqslant C(\mathrm{MAD}(X^n(m)) + |\mathrm{Med}(X^n(m))|) < \infty, \end{aligned}$$

in virtue of $0 \leqslant w(r) \leqslant Cr$ for $r \in [0,1]$. This implies that the numerator of T_n is bounded uniformly in $X^n(m)$. The proof for the case $d = 1$ is completed.

Second, consider the case $d = 2$. We first show that if $m = \lfloor (n - 2d + k + 3)/2 \rfloor$, then m contaminating points are enough to break down T_n. Let l_{12} be a line determined by $X_1, X_2 \in X^n$. Move m points from $\{X_4, \ldots, X_n\}$ to the same site y on l_{12} far away from the original convex hull formed by X^n. Denote the resulting data set $X^n(m)$. Let l_{3y} be the line connecting y and X_3. Let $u_{12}(\perp l_{12})$ and $u_{3y}(\perp l_{3y})$ be two unit vectors. Since $m+d-1 = \lfloor (n+k+1)/2 \rfloor$, then $\sigma(u_{12}' X^n(m)) = \sigma(u_{3y}' X^n(m)) = 0$. Thus $O(x, X^n(m)) = \infty$ for $x \neq y$ and $O(y, X^n(m)) < \infty$. Hence $T_n(X^n(m)) = y$, which breaks down as $\|y\| \to \infty$.

Now we show that if $m = \lfloor (n - 2d + k + 3)/2 \rfloor - 1$, then m contaminating points are not enough to break down T_n. Since $m < \lfloor (n+1)/2 \rfloor$ and $n - m \geqslant \lfloor (n+k+1)/2 \rfloor$, this, combining with the result above for the case $d = 1$, yields

$$\sup_{X^n(m)} \sup_{u \in S^{d-1}} |\mu(u' X^n(m))| < \infty, \quad \sup_{X^n(m)} \sup_{u \in S^{d-1}} \sigma(u' X^n(m)) < \infty. \tag{28}$$

Consequently, $O(x, X^n(m)) \to \infty$ as $\|x\| \to \infty$. Hence it suffices to show that

$$\sup_{X^n(m)} \inf_{\|x\| \leqslant M} O(x, X^n(m)) < \infty, \tag{29}$$

for some large $M > 0$. Suppose that this is not true, then for any large $M > 0$, there is a sequence of contaminated data set $X^{n1}(m), X^{n2}(m), \ldots,$ such that

$$\inf_{\|x\| \leqslant M} O(x, X^{ns}(m)) \to \infty, \quad \text{as } s \to \infty. \tag{30}$$

Let M be a fixed large number so that any intersecting points of two lines determined by the original data points of X^n are in the ball with a radius M. Displays (28) and (30) imply that there is a sequence $u_s \in S^{d-1}$ such that

$$\sigma(u_s' X^{ns}(m)) \to 0, \quad \text{as } s \to \infty. \tag{31}$$

Since $m + d - 1 = \lfloor (n + k - 1)/2 \rfloor$, thus there must be lines $l_s \perp u_s$ containing d points from X^n, say, $X_{s1}, \ldots, X_{sd}$, and all other m contaminating points of $X^{ns}(m)$ are approaching (as

$s \to \infty$) or on l_s. Now since X_{si}, $i = 1, \ldots, d$, are on l_s, to have $O(X_{si}, X^{ns}(m)) \to \infty$ as $s \to \infty$ there must be another sequence of $u_s^* \in S^{d-1}$ that are perpendicular to lines l_s^* that do not contain any X_{si}, $i = 1, \ldots, d$. Furthermore, l_s^* must contains some other d points of X^n and all the m contaminating points of $X^{ns}(m)$ must be approaching (as $s \to \infty$) or at the intersecting point y_s of l_s and l_s^*.

By the compactness of S^{d-1}, there must be subsequences, $\{u_{s_t}\}$ and $\{u_{s_t}^*\}$, of u_s and u_s^* respectively such that $u_{s_t} \to u$ and $u_{s_t}^* \to u^*$ as $t \to \infty$. Let $l_{s_t} \to l$ and $l_{s_t}^* \to l^*$. Then $u \perp l$ and $u^* \perp l^*$. Since there are only finitely many y_{s_t}'s, we can assume without loss of generality that for sufficiently large t, $y_{s_t} = y$, $l_{s_t} = l$, $l_{s_t}^* = l^*$, $u_{s_t} = u$, and $u_{s_t}^* = u^*$ and y is the intersecting point of l and l^* (we can take subsequences if necessarily to achieve this). Note that $\sigma(u'X^n(m))$ is continuous in u for any given $X^n(m)$. So is $\mu(u'X^n(m))$ by the continuity of w. This and Display (30) imply that there is a sequence $v_{s_t} \in S^{d-1}$ such that

$$O(y, X^{ns_t}(m)) = \frac{v_{s_t}'y - \mu(v_{s_t}'X^{ns_t}(m))}{\sigma(v_{s_t}'X^{ns_t}(m))} \to \infty, \quad \text{as } t \to \infty. \tag{32}$$

We now seek a contradiction. Since there is a subsequence of v_{s_t} that converges to v, assume for simplicity that $v_{s_t} \to v$ as $t \to \infty$. We first show that

Claim 1. *v cannot be u, u^* or any unit vector perpendicular to a line through y that contains d points of X^n which also belong to infinitely many $X^{ns_t}(m)$'s.*

It suffices to show that $v \neq u$. Suppose that $v = u$. Assume for convenience that $X_i^{ns_t}(m)$, $i = 1, \ldots, m + d$, are the points on the line l or approaching y and X_{s_tj}, $j = 1, \ldots, n - m - d$, are the point of $X^n \cap X^{ns_t}(m)$. Then it is seen that

$$|v'(X_i^{ns_t}(m) - y)| \leqslant 2\sigma(v'X^{ns_t}(m)), \quad i = 1, 2, \ldots, m + d, \quad \text{for } t \text{ large}.$$

By the proof above for the case $d = 1$ we see that uniformly in $X^{ns_t}(m)$

$$\frac{1}{\sum_{i=1}^n w(\mathrm{PD}(v'X_i^{ns_t}(m), v'X^{ns_t}(m)))} \leqslant \frac{1}{\min\{w(1/2), w(1)\}} := N.$$

The last two displays and the conditions w imply that for t sufficiently large

$$\begin{aligned}
&\frac{|v'y - \mu(v'X^{ns_t}(m))|}{\sigma(v'X^{ns_t}(m))} \\
&\leqslant \frac{|\sum_{i=1}^{m+d} v'(X_i^{ns_t}(m) - y)w(\mathrm{PD}(v'X_i^{ns_t}(m), v'X^{ns_t}(m)))|}{\sigma(v'X^{ns_t}(m))\sum_{i=1}^n w(\mathrm{PD}(v'X_i^{ns_t}(m), v'X^{ns_t}(m))} \\
&\quad + \frac{|\sum_{j=1}^{n-m-d} v'(X_{s_yj} - y)w(\mathrm{PD}(v'X_{s_tj}, v'X^{ns_t}(m)))|}{\sigma(v'X^{ns_t}(m))\sum_{i=1}^n w(\mathrm{PD}(v'X_i^{ns_t}(m), v'X^{ns_t}(m)))} \\
&\leqslant 2 + \sum_{j=1}^{n-m-d} \frac{(|v'X_{s_tj} - \mathrm{Med}(v'X^{ns_t}(m))| + |v'y - \mathrm{Med}(v'X^{ns_t}(m))|)CN}{\sigma(v'X^{ns_t}(m)) + |v'X_{s_tj} - \mathrm{Med}(v'X^{ns_t}(m))|} \\
&< 2(1 + (n - m - d)CN),
\end{aligned}$$

which implies that $O(y, X^{ns_t}(m))$ is bounded for large t, a contradiction to (30).

In the above discussion, we have assumed that $\sigma(v'X^{ns_t}(m)) > 0$. If it is 0, then $|v'(X_i^{ns_t}(m) - y)| = 0$. Hence $\mathrm{Med}(v'X^{ns_t}(m)) = v'y$ and $\mathrm{MAD}(v'X^{ns_t}(m)) = 0$. We have that $\mu(v'X^{ns_t}(m))$

$= v'y$ and consequently $O(y, X^{ns_t}(m)) = 0$, again a contradiction to (30). Hence $v \neq u$ and we have proved the claim.

Let $|v'_{s_t} X^{ns_t}_{i1} - \mathrm{Med}(v'_{s_t} X^{ns_t}(m))| \leqslant \cdots \leqslant |v'_{s_t} X^{ns_t}_{in} - \mathrm{Med}(v'_{s_t} X^{ns_t}(m))|$. Since $m + d = \lfloor (n+k+1)/2 \rfloor$, among $X^{ns_t}_{ij}$, $j = 1, \ldots, \lfloor (n+k+1)/2 \rfloor$, there are at least d points of X^n, say, $X_{s_t 1}, \ldots, X_{s_t d}$. Let l_{s_t} be the line determined by these d points. Then by the claim above it can not be the line through y and perpendicular to v for t sufficiently large. Let $D_{s_t} = \max_{1 \leqslant j \leqslant d} |v'(X_{s_t j} - y)|$. Then $D_{s_t} > 0$. Since there are at most finitely many such l_{s_t}, thus

$$D = \min_{s_t} D_{s_t} > 0, \quad \text{for } t \text{ sufficiently large.} \tag{33}$$

Let $X^{ns_t}_{i0}(m)$ be any one of m contaminating points in $X^{ns_t}(m)$. Then

$$|v'_{s_t}(y - X^{ns_t}_{i0}(m))| < \frac{D}{4}, \quad \text{for } t \text{ sufficiently large.} \tag{34}$$

If there are exactly d points of X^n among $X^{ns_t}_{ij}$, $j = 1, \ldots, m+d$, then, for a contaminating point $X^{ns_t}_{i0}(m)$ of $X^{ns_t}(m)$ that is among the $m+d$ points, by (33) and (34) it is seen that for t sufficiently large

$$\begin{aligned}
\sigma(v'_{s_t} X^{ns_t}(m)) &\geqslant \max_{1 \leqslant j \leqslant d} \frac{|v'_{s_t}(X_{s_t j} - X^{ns_t}_{i0}(m))|}{2} \\
&\geqslant \max_{1 \leqslant j \leqslant d} \frac{|v'_{s_t}(X_{s_t j} - y)|}{2} - \frac{|v'_{s_t}(y - X^{ns_t}_{i0}(m))|}{2} \\
&\geqslant \max_{1 \leqslant j \leqslant d} \frac{|v'(X_{s_t j} - y)|}{4} - \frac{D}{8} \geqslant \frac{D}{8} > 0.
\end{aligned}$$

If there are at least $d+1$ points, $X_{s_t 1}, \ldots, X_{s_t (d+1)}$, among $X^{ns_t}_{ij}$, $j = 1, \ldots, m+d$, then $\inf_{s_{t1}, \ldots, s_{t(d+1)}} \max_{1 \leqslant j,\, k \leqslant (d+1)} |v'(X_{s_t j} - X_{s_t k})| > 0$ since X^n is in general position. Hence for t sufficiently large

$$\begin{aligned}
\sigma(v'_{s_t} X^{ns_t}(m)) &\geqslant \max_{1 \leqslant j,\, k \leqslant (d+1)} \frac{|v'_{s_t}(X_{s_t j} - X_{s_t k})|}{2} \\
&\geqslant \max_{1 \leqslant j,\, k \leqslant (d+1)} \frac{|v'(X_{s_t j} - X_{s_t k})|}{4} \\
&\geqslant \inf_{s_{t1}, \ldots, s_{t(d+1)}} \max_{1 \leqslant j,\, k \leqslant (d+1)} \frac{|v'(X_{s_t j} - X_{s_t k})|}{4} := \frac{E}{8} > 0.
\end{aligned}$$

Both cases above lead to a bounded $O(y, X^{ns_t}(m))$ for t sufficiently large, a contradiction with Display (32). This completes the proof for the case $d = 2$.

Third, consider the case $d > 2$. This can be done by mimicking the proof above for $d = 2$, replacing lines with hyper-planes and 2 with d.

References

1 Donoho D L, Huber P J. The notion of breakdown point. In: A Festschrift for Erich L. Lehmann. Bickel P J, eds. CA: Wadsworth, 1993, 157–184

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

3 Serfling R. Approximation Theorems of Mathematical Statistics. New York: John Wiley & Sons, 1980

4 Donoho D L. Breakdown properties of multivariate location estimators. PhD Thesis Dept. Statistics, Harvard University, 1982

5 Davies P L. Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *Ann Statist*, **15**: 1269–1292 (1987)

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

7 Hettmansperger T P, Randles R H. A practical affine equivariant multivariate median. *Biometrika*, **89**: 851–860 (2002)

8 Tukey J W. Mathematics and picturing data. In: Proceeding of the International Congress of Mathematicians. Vol 2. Vancouver: Canadian Mathematical Congress, 1975, 523–531

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

10 Zuo Y, Serfling R. General notions of statistical depth function. *Ann Statist*, **28**(2): 461–482 (2000)

11 Donono D L, Gasko M. Breakdown properties of multivariate location parameters and dispersion matrices. *Ann Statist*, **20**: 1803–1827 (1992)

12 Zuo Y. Projection based depth functions and associated medians. *Ann Statist*, **31**(5): 1460–1490 (2003)

13 He X, Portnoy S. Asymptotics of the deepest line. In: Appl Statist Sci, III, Commack, New York: Nova Sci Publ, 1998, 71–81

14 Nolan D. On min-max majority and deepest points. *Statist Prob Lett*, **43**: 325–333 (1999)

15 Bai Z D, He X. Asymptotic distributions of the maximum depth estimators for regression and multivariate location. *Ann Statist*, **27**: 1616–1637 (1999)

16 Mosteller F, Tukey J W. Data Analysis and Regression. MA: Addison, 1977

17 Stahel W A. Breakdown of covariance estimators. Research Report 31, Fachgruppe für Statistik, ETH, Zürich, 1981

18 Liu R Y. Data Depth and Multivariate Rank Tests. In: L_1-Statistical Analysis and Related Methods, Y Dodge, ed. New York: Springer, 1992, 279–294

19 Zuo Y, Cui H, He X. On the Stahel-Donoho estimators and depth weighted means for multi variate data. *Ann Statist*, **32**(1): 167–188 (2004)

20 Hall P, Welsh A H. Adaptive estimates of parameters of regular Variation. *Ann Statist*, **13**: 331–341 (1985)

21 Falk M. Asymptotic independence of median and MAD. *Statist Prob Lett*, **34**: 341–345 (1997)

22 Van der Vaart A W, Wellner J A. Weak convergence and Empirical Processes: With Applications to Statistics. New York: Springer, 1986

23 Gather U, Hilker T. A note on Tyler's modification of the MAD for the Stahel-Donoho estimator. *Ann Statist*, **25**: 2024–2026 (1997)

24 Lehmann E L. Theory of Point Estimation. New York: Wiley, 1983

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

26 Rousseeuw P J, Ruts I. Constructing the bivariate Tukey median. *Statist Sinica*, **8**: 828–839 (1998)

27 Webster R. Spectral analysis of gilgai soil. *Australian J Soil Research*, **15**: 191–204 (1977)

28 Laslett G M. Kriging and splines: an empirical comparison of their predictive performance in some applications (with discussion). *J Amer Statist Assoc*, **89**: 391–409 (1994)

29 Pollard D. Convergence of Stochastic Processes. New York: Springer-Verlag Nolan, 1984

30 Van der Vaart A W. Asymptotic Statistics. Cambridge: Cambridge University Press, 1998

31 Massé J C. Asymptotics for the Tukey median. *J Multivariate Anal*, **81**: 286–300 (2002)