The Annals of Statistics
2000, Vol. 28, No. 2, 461–482

GENERAL NOTIONS OF STATISTICAL DEPTH FUNCTION

By Yijun Zuo and Robert Serfling[1]

Arizona State University and University of Texas

Statistical depth functions are being formulated *ad hoc* with increasing popularity in nonparametric inference for multivariate data. Here we introduce several general structures for depth functions, classify many existing examples as special cases, and establish results on the possession, or lack thereof, of four key properties desirable for depth functions in general. Roughly speaking, these properties may be described as: affine invariance, maximality at center, monotonicity relative to deepest point, and vanishing at infinity. This provides a more systematic basis for selection of a depth function. In particular, from these and other considerations it is found that the *halfspace* depth behaves very well overall in comparison with various competitors.

1. Introduction. Statistical depth functions have become increasingly pursued as a useful tool in nonparametric inference for multivariate data. Roughly speaking, for a distribution P in $\mathbf{R}^d$, a corresponding depth function is any function $D(x; P)$ which provides a P-based center-outward ordering of points $x \in \mathbf{R}^d$. Tukey (1975) proposed a "halfspace" depth and suggested its role in defining multivariate analogues of univariate rank and order statistics via depth-induced "contours." The *halfspace depth* (HD) of a point x in $\mathbf{R}^d$ with respect to a probability measure P on $\mathbf{R}^d$ is defined as the minimum probability mass carried by any closed halfspace containing x, that is,

$$HD(x; P) = \inf\{P(H) : H \text{ a closed halfspace}, x \in H\}, \quad x \in \mathbf{R}^d.$$

Based on this depth, Donoho and Gasko (1992) studied multivariate location estimators and Yeh and Singh (1997) developed confidence regions. Properties of the corresponding contours have been studied by various authors including Eddy (1985), Nolan (1992), Donoho and Gasko (1992) and Massé and Theodorescu (1994). See Carrizosa (1996) for a characterization of halfspace depth relating to problems of facility location analysis in the operations research literature.

The "center-outward ordering" interpretation of a depth function suggests that (i) a relevant notion of "center" is available, and (ii) points near the center should have higher depth. From this standpoint, the "center" consists of the set of points globally maximizing depth, in which case a depth function should tend to ignore multimodality features of the underlying distribution P. If, on the other hand, sensitivity to multimodality is desirable, then the

Received December 1998; revised December 1999.

[1]Supported by NSF Grant DMS-97-05209.

AMS 1991 *subject classifications*. Primary 62H05; secondary 62G20.

Key words and phrases. Statistical depth functions, halfspace depth, simplicial depth, multivariate symmetry.

"center" should include local maxima as well, in which case the notion of center-outward ordering becomes compromised and "inner" points can have low depth. It is thus important, in considering depth functions, to make a choice on this issue. In the present paper we opt for the center to be given by global maxima, with low depth corresponding to large distance from the center. For further discussion, see Remark A.1 in Appendix A.

Liu (1990) introduced a notion of "simplicial" depth and corresponding multivariate location estimators. Namely, the *simplicial depth* (SD) of a point x in $\mathbf{R}^d$ with respect to a probability measure P on $\mathbf{R}^d$ is defined to be the probability that x belongs to a random simplex in $\mathbf{R}^d$, that is,

$$SD(x; P) = P(x \in S[X_1, \ldots, X_{d+1}]), \quad x \in \mathbf{R}^d,$$

where $X_1, \ldots, X_{d+1}$ is a random sample from P and $S[x_1, \ldots, x_{d+1}]$ denotes the d-dimensional simplex with vertices $x_1, \ldots, x_{d+1}$, that is, the set of all points in $\mathbf{R}^d$ that are convex combinations of $x_1, \ldots, x_{d+1}$.

Liu and Singh (1993) considered the above two depth functions and two more, "Mahalanobis" depth and "majority" depth, which they applied in formulating a "quality index" for use in connection with manufacturing processes. Rousseeuw and Hubert (1999) introduced "regression depth" and Rousseeuw and Ruts (1996), Ruts and Rousseeuw (1996) and Rousseeuw and Struyf (1998) studied computing issues concerning depth functions and contours. Liu, Parelius and Singh (1999) considered seven examples of depth function, including a "convex hull peeling" version and a "likelihood" type, and developed methodology for their practical use in exploratory statistical analysis. Likelihood-based depth functions have also been considered by Fraiman and Meloche (1996) and Fraiman, Liu and Meloche (1997). Koshevoy and Mosler (1997) introduced a "zonoid" depth function based on "zonoid trimming." Bartoszyński, Pearl and Lawrence (1997) introduced a depth function based on interpoint distances in the context of a multivariate goodness-of-fit test. Depth functions also arise in the theory of social choice [see Caplin and Nalebuff (1988, 1991a, b)]. Nonparametric notions of multivariate "scatter measure" and "more scattered" based on general depth functions have been formulated and studied by Zuo and Serfling (2000a). Mizera (1998) has introduced a differential calculus for depth functions. Finally, Vardi and Zhang (1999) have introduced a method for constructing depth functions from notions of multivariate median.

Depth functions thus have been introduced *ad hoc* in great variety, without regard to whether they meet any particular set of criteria that ought to be satisfied. Consequently, there is no systematic basis for preferring one such function over another. In the present paper, we address this issue by asking:

(i) What desirable properties should a statistical depth function possess?
(ii) What constructive approaches lead to attractive depth functions?
(iii) Do existing depth functions possess all desired properties?

In Section 2 we list several desirable properties first introduced by Liu (1990), on the basis of which we formulate a general definition of "statistical depth function". Roughly speaking, these properties may be described as:

affine invariance, maximality at center, monotonicity relative to deepest point, and vanishing at infinity. Also, several distinct structures for construction of depth functions are introduced and investigated with respect to possession of these properties, and a number of presently popular depth functions are classified with respect to these different structural types.

In Section 3 we evaluate and critically compare, from the above perspectives as well as from robustness considerations, a number of existing depth functions and some new ones introduced via the above-mentioned constructions. It is found that the half-space depth and a closely related "projection depth," both of which reflect projection pursuit methodology, are distinctly more attractive than popular competitors.

Various supplementary notes are provided in Appendix A, including discussion of almost sure uniform convergence of sample depth functions to their population counterparts. Finally, proofs of the results in Section 2 are provided in Appendix B.

2. General notions of statistical depth. Here we consider general notions of depth function on $\mathbf{R}^d$, defined with respect to arbitrary distributions which may be either continuous or discrete. In the spirit of Liu (1990), Section 2.1 presents four *desirable properties* that an ideal depth function should possess. In Section 2.2 the halfspace and simplicial depth functions are examined with respect to these criteria, and it is found that the halfspace depth possesses all four properties (see Theorem 2.1), whereas the simplicial depth lacks certain properties in some cases (see Remark 2.1). In Section 2.3, several *general structures* for depth functions are introduced and investigated with respect to the four properties (see Theorems 2.2–2.11). Also, familiar existing versions of depth function as well as some new ones are reviewed in the context of these structures.

2.1. *Desirable properties and a general definition.* We confine attention to depth functions that are *nonnegative* and *bounded*. In order that a depth function serve most effectively as a tool providing a center-outward ordering of points in $\mathbf{R}^d$, it should ideally satisfy the following further properties, which we state informally first and then more precisely in Definition 2.1.

P1. *Affine invariance.* The depth of a point $x \in \mathbf{R}^d$ should not depend on the underlying coordinate system or, in particular, on the scales of the underlying measurements.

P2. *Maximality at center.* For a distribution having a uniquely defined "center" (e.g., the point of symmetry with respect to some notion of symmetry), the depth function should attain maximum value at this center.

P3. *Monotonicity relative to deepest point.* As a point $x \in \mathbf{R}^d$ moves away from the "deepest point" (the point at which the depth function attains maximum value; in particular, for a symmetric distribution, the center) along any fixed ray through the center, the depth at x should decrease monotonically.

P4. *Vanishing at infinity.* The depth of a point x should approach zero as $\|x\|$ approaches infinity.

We note that P1–P4 are introduced and investigated for the simplicial depth in Liu (1990).

We now formally define "statistical depth function". Denote by $\mathscr{F}$ the class of distributions on the Borel sets of $\mathbf{R}^d$ and by F_ξ the distribution of a given random vector ξ.

DEFINITION 2.1. Let the mapping $D(\cdot\,;\cdot) : \mathbf{R}^d \times \mathscr{F} \to \mathbf{R}^1$ be bounded, nonnegative, and satisfy P1–P4. That is, assume:

(i) $D(Ax + b; F_{AX+b}) = D(x; F_X)$ holds for any random vector X in $\mathbf{R}^d$, any $d \times d$ nonsingular matrix A, and any d-vector b;
(ii) $D(\theta; F) = \sup_{x\in\mathbf{R}^d} D(x; F)$ holds for any $F \in \mathscr{F}$ having center θ;
(iii) for any $F \in \mathscr{F}$ having deepest point θ, $D(x; F) \le D(\theta + \alpha(x - \theta); F)$ holds for $\alpha \in [0, 1]$; and
(iv) $D(x; F) \to 0$ as $\|x\| \to \infty$, for each $F \in \mathscr{F}$.
Then $D(\cdot\,; F)$ is called a *statistical depth function*.

A sample version of $D(x; P)$, denoted by $D_n(x) \equiv D(x; \widehat{P}_n)$, may be defined by replacing P by a suitable empirical measure $\widehat{P}_n$.

In the above we have used the term "center" to denote a point of symmetry. Various notions of multivariate symmetry are possible. In particular, a standard notion widely used in the literature is that a random vector X in $\mathbf{R}^d$ is *centrally symmetric* about θ if $X - \theta \overset{d}{=} \theta - X$, where "$\overset{d}{=}$" denotes "equal in distribution." A broader notion due to Liu (1990) defines X to be *angularly symmetric* about θ if $(X - \theta)/\|X - \theta\|$ is centrally symmetric about the origin. A still broader notion, which we here introduce, defines X to be *halfspace symmetric* about θ if $P(X \in H) \ge 1/2$ for every closed halfspace H containing θ. In an obvious terminology, it is easily established that C-symmetry $\to$ A-symmetry $\to$ H-symmetry. For characterizations of H-symmetry motivating its relevance in nonparametric multivariate location inference, see Zuo and Serfling (2000c). Thus the most favorable manifestation of property P2 for a depth function $D(\cdot;\cdot)$ is that maximality at center should hold for $D(\cdot; F)$ as generally as possible, that is, for every H-symmetric F. A similar remark holds with respect to property P3. [For further comparison of angular and halfspace symmetry, and of these with notions in Beran and Millar (1997), see Remark A.2 in Appendix A.]

One might view property P4 as rather too strict and thus instead consider some weaker variant. If, for example, the depth function has a lower limit $L > 0$, one might normalize the depth function by subtracting L. But when L depends on F (as for the majority depth when $d \ge 2$), this is computationally and technically very burdensome.

Or one might require merely that $R(x; F) \to 0$ as $\|x\| \to \infty$, where $R(x; F) = P_F(\{y : D(y; F) \le D(x; F)\})$, the proportion of the distribution F having depth $\leq$ the depth of x. [This quantity is used by Liu and Singh (1993) in defining their "quality index."] Under P2 and P3, however, convergence of $R(x; F)$ to 0 is seen to hold already and thus does not offer anything productive in addition to P2 and P3.

In Zuo and Serfling [(2000b), Theorem 3.1(iv)], the present form of P4 is useful in establishing compactness of depth-trimmed regions. Further, it plays a role in using truncation arguments to establish almost sure uniform convergence of sample depth functions to population versions.

2.2. *A further look at the halfspace and simplicial depth functions.* We now investigate whether the halfspace depth function $HD(x; P)$ and the simplicial depth function $SD(x; P)$ are "statistical depth functions" in the sense of Definition 2.1. These are treated, respectively, in the following theorem and remark.

THEOREM 2.1. *The halfspace depth function $HD(x; P)$ is a statistical depth function in the sense of Definition* 2.1.

REMARK 2.1. For *continuous* angularly symmetric distributions, it follows from results of Liu (1990) that the simplicial depth function $SD(\cdot\,; P)$ is a statistical depth function in the sense of Definition 2.1. For *discrete* distributions, however, $SD(x; P)$ can for H-symmetric distributions fail to satisfy the "maximality" property P2 and even for C-symmetric distributions fail to satisfy the "monotonicity" property P3. This is seen from the following counterexamples.

COUNTEREXAMPLE 1. Let $d = 1$ and $P(X = 0) = 1/5$, $P(X = \pm 1) = 1/5$, and $P(X = \pm 2) = 1/5$. Then clearly X is centrally symmetric about 0. It is not difficult to show that $SD(1/2;\ P) = 12/25$ and $SD(1;\ P) = 15/25$, violating P3.

COUNTEREXAMPLE 2. Let $d = 2$ and $P(X = (\pm 1, 0)) = P(X = (\pm 2, 0)) = P(X = (0, \pm 1)) = 1/6$. Then X is centrally symmetric about $(0, 0)$ and

$$SD((1,0); P) - SD((1/2,0); P) = 3! \cdot 2 \cdot (1/6)^3 = 1/18 > 0,$$

again violating P3.

COUNTEREXAMPLE 3. Let $d = 2$ and $P(X = \theta = (0, 0)) = 19/40$, $P(X = A = (-1, 1)) = 3/40$, and $P(X = B = (-1, -1)) = P(X = C = (1, 0)) = 1/40$. Let $B\,\theta$ intersect AC at D, x be a point inside the triangle $\triangle\, A\,\theta D$, and $P(X = x) = 16/40$. Then it is not difficult to verify, based on results established in Zuo and Serfling (2000c), that X is H-symmetric about θ, which is thus the center of the distribution. However, we have

$$SD(x; P) - SD(\theta; P) = \frac{3!}{40^3}\,(2 \times 16 \times 1 \times 3 - (3 \times 1 \times 19 + 1 \times 1 \times 19)) > 0,$$

that is, the "maximality" property P2 fails to hold.

For the above two well-known notions of depth function, we thus have found that one behaves well overall, while in some discrete cases the other is not completely satisfactory. This leads one to investigate whether other attractive statistical depth functions can be defined, indeed to explore general structures for such functions and to seek to identify the more favorable types.

2.3. *General structures for statistical depth functions.* Four general structures for construction of statistical depth functions are introduced and investigated with respect to properties P1–P4. Various existing depth functions are classified according to these types.

2.3.1. *Type A depth functions.* Let $h(x; x_1, \ldots, x_r)$ be any bounded nonnegative function which in some sense measures the *closeness* of x to the points $x_1, \ldots, x_r$. A corresponding *Type A depth function* is then defined by the average closeness of x to a random sample of size r:

$$D(x; P) = Eh(x; X_1, \ldots, X_r), \tag{1}$$

where $X_1, \ldots, X_r$ is a random sample from P. For such depth functions the corresponding sample versions $D(x; \widehat{P}_n)$ turn out to be *U-statistics* or *V-statistics*.

Taking $r = d+1$ and $h(x; x_1, \ldots, x_{d+1}) = \mathbf{I}\{x \in S[x_1, \ldots, x_{d+1}]\}$, we obtain the *simplicial depth*, whose properties have been covered in Section 2.2. Another example is the following.

EXAMPLE 2.1. [Majority depth (Singh, 1991)] For given points $x_1, \ldots, x_d$ in $\mathbf{R}^d$ which determine a unique hyperplane containing themselves, there correspond two closed halfspaces with this hyperplane as boundary. Denote by $H^P_{x_1,\ldots,x_d}$ the one which carries probability mass $\geq 1/2$ under the distribution P on $\mathbf{R}^d$. Then the *majority depth function* is defined by

$$MJD(x; P) = P(x \in H^P_{X_1,\ldots,X_d}), \quad x \in \mathbf{R}^d, \tag{2}$$

where $X_1, \ldots, X_d$ is a random sample from P. Clearly, the majority depth function is of Type A with $r = d$ and $h(x; x_1, \ldots, x_d) \equiv \mathbf{I}\{x \in H^P_{x_1,\ldots,x_d}\}$.

Let us explore the majority depth function with respect to properties P1–P4. Clearly P1 is satisfied. Also, as remarked by Liu and Singh (1993), for any A-symmetric distribution P, $MJD(x; P)$ decreases monotonically as x moves away from the center along any fixed ray originating from the center, that is, P2 and P3 hold. Indeed, the following result establishes this more generally.

THEOREM 2.2. *For H-symmetric distributions P, $MJD(x; P)$ satisfies* P2 *and* P3.

The majority depth fails to satisfy property P4, however. As a counterexample, take $d = 2$ and define P by $P(X = (\pm 1, 0)) = 1/3$ and $P(X = (0, 1)) = 1/3$. Then it is easy to see that $\lim_{\|x\| \to \infty} MJD(x; P) = 2/3$. As another counter example, for $d = 1$ one can show for *any* P that $MJD(x; P) = 1/2 + \min\{P(x), 1 - P(x)\} \to 1/2$ as $x \to \infty$.

2.3.2. *Type B depth functions.* Let $h(x; x_1, \ldots, x_r)$ be an *unbounded* nonnegative function which measures in some sense the *distance* of x from the points $x_1, \ldots, x_r$. A corresponding *Type B depth function* is then defined by

$$D(x; F) \equiv (1 + Eh(x; X_1, \ldots, X_r))^{-1}, \tag{3}$$

for $X_1, \ldots, X_r$ a random sample from F. Closely related to (3), but not equivalent, is the structure $E[1 + h(x; X_1, \ldots, X_r)]^{-1}$, which is a further example of the Type A structure. For the sake of tractability, we prefer the form (3).

As a measure of dispersion of a point cloud $\{x; x_1, \ldots, x_r\}$, the function $h(x; x_1, \ldots, x_r)$ possibly may not possess the affine invariance property P1, but in many such cases it satisfies at least *rigid-body invariance*, that is, $h(Ax + b; Ax_1 + b, \ldots, Ax_r + b) = h(x; x_1, \ldots, x_r)$, for any $d \times d$ orthogonal matrix A and any vector $b \in \mathbf{R}^d$. For example, see the L^p depth treated below. Or, a suitable modification of the function h sometimes yields an affine invariant version, as in the case of the "simplicial volume depth" as well as the L^2 depth treated below. Regarding properties P2–P4, Type B depth functions are rather well behaved, as shown by the following examples and theorems.

EXAMPLE 2.2 (Simplicial volume depth). Take

$$h(x; x_1, \ldots, x_d) = \Delta^\alpha(S[x, x_1, \ldots, x_d]),$$

where $\Delta(S[x; x_1, \ldots, x_d])$ denotes the volume of the d-dimensional simplex $S[x, x_1, \ldots, x_d]$ and $\alpha > 0$. This is a measure of the dispersion of the point cloud $\{x, x_1, \ldots, x_d\}$ and accordingly

$$(1 + E[\Delta^\alpha(S[x, X_1, \ldots, X_d])])^{-1} \tag{4}$$

defines a Type B depth function. This depth function usually is not affine invariant, however, since

$$\Delta^\alpha(S[Ax + b, Ax_1 + b, \ldots, Ax_d + b]) = |\det(A)|^\alpha \Delta^\alpha(S[x, x_1, \ldots, x_d]),$$

where b is any vector in $\mathbf{R}^d$, and the determinant $\det(A)$ of the nonsingular matrix A is not always equal to 1. This problem can be rectified by a modification. Rather than (4), we define the *simplicial volume depth function* by

$$SVD^\alpha(x; F) \equiv \left(1 + E\left[\left(\frac{\Delta(S[x, X_1, \ldots, X_d])}{\sqrt{\det(\Sigma)}}\right)^\alpha\right]\right)^{-1}, \tag{5}$$

where Σ is the covariance matrix of F. This version is affine invariant.

REMARK 2.2. Oja (1983) introduced for C-symmetric distributions a family of *location measures* utilizing simplicial volume, as follows. For each $\alpha > 0$, a location measure μ_α: $\mathscr{F} \to \mathbf{R}^d$ is defined by

$$E[\Delta^\alpha(S[\mu_\alpha(F), X_1, \ldots, X_d])] = \inf_{\mu \in \mathbf{R}^d} E[\Delta^\alpha(S[\mu, X_1, \ldots, X_d])].$$

However, he did not develop it into a depth function, nor did he consider the affine invariant version (5).

EXAMPLE 2.3. [L^p depth ($p > 0$.)] Another way to measure distance is via the L^p norm $\|\cdot\|_p$. Taking $h(x; x_1) = \|x - x_1\|_p$, a corresponding Type B depth function is given by

$$L^p D(x; F) \equiv \left(1 + E\|x - X\|_p\right)^{-1}. \tag{6}$$

Note that $L^p D(x; F)$ generally does not possess the affine invariance property, however, since

$$E\|Ax + b - (AX + b)\|_p = E\|A(x - X)\|_p,$$

which is not equal to $E\|x - X\|_p$ for every nonsingular matrix A. On the other hand, taking $p = 2$, it is easy to see that $L^2 D(x; F)$ is rigid-body invariant. Moreover, a modification of the L^2 norm yields an affine invariant version. Following Rao (1988), for a positive definite $d \times d$ matrix M, define a norm $\|\cdot\|_M$ as

$$\|x\|_M \equiv \sqrt{x' M x} \quad \forall x \in \mathbf{R}^d. \tag{7}$$

Then, for $p = 2$, the depth function defined in (6) may be modified to an affine invariant version,

$$\widetilde{L}^2 D(x; F) \equiv (1 + E\,[\,\|x - X\|_{\Sigma^{-1}}\,])^{-1}, \tag{8}$$

where Σ is the covariance matrix of F.

Under some conditions on $h(x; x_1, \dots, x_r)$, Type B depth functions necessarily satisfy P2 and P3, as shown in the following two results.

THEOREM 2.3. *Suppose θ is the point of symmetry of a distribution F with respect to a given notion of symmetry. Then Type B depth functions $D(x; F)$ possess the "maximality at center" property P2 if*:

(i) $h(x + b; x_1 + b, \dots, x_r + b) = h(x; x_1, \dots, x_r)$;
(ii) $h(-x; -x_1, \dots, -x_r) = h(x; x_1, \dots, x_r)$;
(iii) $h(x; x_1, \dots, x_r)$ *is convex in the argument* x; *and*
(iv) *for x, b and $x_1, \dots, x_r$ arbitrary vectors in $\mathbf{R}^d$ and $X_1, \dots, X_r$ a random sample from F, the set*

$$\left(\arg\inf_{x \in \mathbf{R}^d} Eh(x; X_1 - \theta, \dots, X_r - \theta)\right) \cap \left(\arg\inf_{x \in \mathbf{R}^d} Eh(x; \theta - X_1, \dots, \theta - X_r)\right)$$

is nonempty.

REMARK 2.3. For any distribution C-symmetric about a point θ in $\mathbf{R}^d$, there is always a point $y \in \mathbf{R}^d$ satisfying condition (iv) above.

THEOREM 2.4. *If $h(x; x_1, \dots, x_r)$ is convex in x, then the corresponding Type B depth function $D(x; F)$ decreases monotonically as x moves outward along any ray starting at a deepest point of F.*

Equipped with the above two results, we now take a further look at $SVD^{\alpha}(x;F)$ and $L^pD(x;F)$.

COROLLARY 2.1. *For* $\alpha \geq 1$, $SVD^{\alpha}(x;F)$ *satisfies* P3 *and* P4.

Since $\Delta^{\alpha}(S[x, x_1, \ldots, x_d])$ is convex and rigid-body invariant, according to Theorem 2.3 we obtain

COROLLARY 2.2. *For C-symmetric distributions and* $\alpha \geq 1$, $SVD^{\alpha}(x;F)$ *satisfies* P2.

The affine invariance and Corollaries 2.1 and 2.2 thus yield:

THEOREM 2.5. *For C-symmetric distributions and* $\alpha \geq 1$, $SVD^{\alpha}(x;F)$ *is a statistical depth function in the sense of Definition* 2.1.

The next three results treat P2–P4 for $L^pD(x;F)$, $p \geq 1$ and $\widetilde{L}^2D(x;F)$.

Convexity of $h(x;x_1) = \|x - x_1\|_p$ in the argument x follows in straightforward fashion from Minkowski's inequality. Thus Theorem 2.4 yields P3 for $L^pD(x;F)$, while P4 is obvious. Thus we have

COROLLARY 2.3. *For* $p \geq 1$, $L^pD(x;F)$ *satisfies* P3 *and* P4.

Since $h(x;x_1)$ is *location invariant* and *even*, that is, $h(x+b, x_1+b) = h(x, x_1)$ for any vector $b \in \mathbf{R}^d$ and $h(-x, -x_1) = h(x, x_1)$, by the convexity just established and Theorem 2.3 we obtain:

COROLLARY 2.4. *For C-symmetric distributions and for* $p \geq 1$, $L^pD(x;F)$ *satisfies* P2.

For $\widetilde{L}^2\mathrm{D}(x;F)$ we have:

THEOREM 2.6. *For any distribution F A-symmetric about a unique point* $\theta \in \mathbf{R}^d$, $\widetilde{L}^2D(x;F)$ *defined in* (8) *is a statistical depth function in the sense of Definition* 2.1.

REMARK 2.4. In the foregoing proof, condition (iv) of Theorem 2.3 was established for $\widetilde{L}^2(x;F)$ for all A-symmetric F. For the depth function $L^2(x;F)$, it follows from results established in Zuo and Serfling (2000c) that this condition holds for all H-symmetric F.

2.3.3. *Type C depth functions.* Let $O(x;F)$ be a measure of the *outlyingness* of the point x in $\mathbf{R}^d$ with respect to the center or the deepest point of the distribution F. Usually $O(x;F)$ is *unbounded*, but a corresponding *bounded* depth function is defined by

$$D(x;F) \equiv (1 + O(x;F))^{-1}. \tag{9}$$

We call these *Type C depth functions*.

REMARK 2.5. Although Type B and Type C depth functions are clearly similar in form, it is convenient to treat them separately, as they arise from somewhat different conceptual points of view.

EXAMPLE 2.4. Projection depth. Define the outlyingness of a point x to be the worst case outlyingness of x with respect to the one-dimensional median in any one-dimensional projection, that is,

$$O(x; F) \equiv \sup_{\|u\|=1} \frac{| u'x - \mathrm{Med}(u'X) |}{\mathrm{MAD}(u'X)}, \tag{10}$$

where X has distribution F, Med denotes the univariate median, MAD denotes the univariate median absolute deviation defined for univariate Y as $\mathrm{MAD}(Y) = \mathrm{Med}(|Y - \mathrm{Med}(Y)|)$, and $\| \cdot \|$ is the Euclidean norm. We call the corresponding Type C depth function *projection depth* and denote it by $PD(x; F)$, $x \in \mathbf{R}^d$.

REMARK 2.6. For one-dimensional datasets $X = \{X_1, \dots, X_n\}$,

$$O_n(x) \equiv |x - \mathrm{Med}_{1\le i\le n}\{X_i\}| / \mathrm{MAD}_{1\le i\le n}\{X_i\}$$

has long been used as a robust measure of outlyingness of $x \in \mathbf{R}$ with respect to the center (median) of the dataset. See Mosteller and Tukey [(1977), pages 205–208]. Here

$$\mathrm{Med}_{1\le i\le n}\{X_i\} = \frac{1}{2}\left(X_{(\lfloor \frac{n+1}{2} \rfloor)} + X_{(\lfloor \frac{n+2}{2} \rfloor)}\right),$$

$$\mathrm{MAD}_{1\le i\le n}\{X_i\} = \mathrm{Med}_{1\le i\le n}\{|X_i - \mathrm{Med}_{1\le j\le n}\{X_j\}|\},$$

and $X_{(1)} \le \cdots \le X_{(n)}$ are the ordered $X_1, \dots, X_n$. Donoho and Gasko (1992) generalized this to arbitrary dimension d, defining $O_n(x)$ to be the worst case outlyingness of $x \in \mathbf{R}^d$ in any one-dimensional projection of x and the dataset X. A sample version of the projection depth function $PD(x; F)$ is thus given by

$$PD_n(x) = (1 + O_n(x))^{-1}. \tag{11}$$

Liu (1992) suggested the use of (11) as a data depth function, but did not provide any treatment of it.

EXAMPLE 2.5 (Mahalanobis depth). Mahalanobis (1936) introduced a distance between two points x and y in $\mathbf{R}^d$, with respect to a positive definite $d \times d$ matrix M, as

$$d_M^2(x, y) = (x - y)' M^{-1} (x - y).$$

Based on this *Mahalanobis distance*, one can define a *Mahalanobis depth* as the corresponding Type C depth function,

$$MHD(x; F) = \left(1 + d^2_{\Sigma(F)}(x, \mu(F))\right)^{-1}, \tag{12}$$

where F is a given distribution and $\mu(F)$ and $\Sigma(F)$ are any corresponding location and covariance measures, respectively. The case that $\mu(F)$ and $\Sigma(F)$ are the mean and covariance matrix of F was suggested by Liu (1992). For these choices, however, $MHD(\cdot\,; F)$ is not "robust" [since $\mu(F)$ = mean is not robust, as noted by Liu and Singh (1993)], and it can fail to achieve maximum value at the center of A-symmetric distributions.

For Type C depth functions, the following analogues of Theorems 2.3 and 2.4 hold and can be proved similarly. It is convenient to write $O(x; X)$ for $O(x; F_X)$.

THEOREM 2.7. *Suppose θ in $\mathbf{R}^d$ is the point of symmetry of a distribution F with respect to a given notion of symmetry. The Type C depth functions $D(x; F)$ possess the "maximality at center" property P2 if for arbitrary vectors x, b in $\mathbf{R}^d$:*

(i) $O(x + b; X + b) = O(x; X)$;
(ii) $O(-x; -X) = O(x; X)$;
(iii) $O(x; X)$ *is convex in the argument x; and*
(iv) *the set*

$$y \in \left(\arg\inf_{x\in\mathbf{R}^d} O(x; X - \theta)\right) \cap \left(\arg\inf_{x\in\mathbf{R}^d} O(x; \theta - X)\right)$$

is nonempty.

THEOREM 2.8. *If $O(x; F)$ is convex in the argument x, then the corresponding Type C depth function $D(x; F)$ decreases monotonically as x moves outward along any ray starting at a deepest point of F.*

The following two theorems establish that $PD(x; F)$ and $MHD(x; F)$ are proper statistical depth functions.

THEOREM 2.9. *The projection depth function $PD(x; F)$ is a statistical depth function in the sense of Definition* 2.1.

A *location measure* μ is *affine equivariant* if $\mu(AX + b) = A\mu(X) + b$ for any affine transformation $AX + b$ of X. A *covariance measure* Σ is *affine equivariant* if $\Sigma(AX + b) = A\Sigma(X)A'$ for any affine transformation $AX + b$ of X.

THEOREM 2.10. *Let F be symmetric. Then the Mahalanobis depth function $MHD(x; F)$ is a statistical depth function in the sense of Definition* 2.1 *if μ and Σ are affine equivariant and $\mu(F)$ agrees with the point of symmetry of F.*

The proof is straightforward.

2.3.4. *Type D depth functions.* One can interpret the "tailedness" of a point with respect to a given distribution as an index related to its relative depth with respect to the center or deepest point of the distribution. Let $\mathscr{C}$ be a class of *closed* subsets of $\mathbf{R}^d$ and P a probability measure on $\mathbf{R}^d$. A corresponding *Type D depth function* is defined by

$$D(x; P, \mathscr{C}) \equiv \inf_{C} \{ P(C) \mid x \in C \in \mathscr{C} \}. \tag{13}$$

Thus the $\mathscr{C}$-depth of a point x with respect to a probability measure P on $\mathbf{R}^d$ is defined to be the minimum probability mass carried by a set C in $\mathscr{C}$ that contains x. In essence, this form of depth function is equivalent, via $D = 1 - I$, to the "index function" $I[x, P, \mathscr{C}]$ introduced by Small (1987) for measuring the "tailedness" of points x in some space. Such functions have antecedents in game theoretical work of Hotelling (1929) and Chamberlin (1937).

We confine attention to classes $\mathscr{C}$ satisfying the following conditions:

C1. If $C \in \mathscr{C}$, then $\overline{C^c} \in \mathscr{C}$.

C2. For $C \in \mathscr{C}$ and $x \in C^\circ$, there exists $C_1 \in \mathscr{C}$ with $x \in \partial C_1$, $C_1 \subset C^\circ$,

where ∂C, C^c, C° and $\overline{C}$ denote, respectively, the *boundary*, *complement, interior* and *closure* of C.

The class of all closed halfspaces $\mathscr{H}$ on $\mathbf{R}^d$ satisfies C1 and C2 and thus the *halfspace depth* is a typical example of Type D depth function. As shown in Theorem 2.1, $HD(x; P)$ is a statistical depth function. Useful further properties of $HD(x; P)$ that in fact hold more generally are given in the following result.

THEOREM 2.11. *Let $\mathscr{C}$ be a class of closed Borel sets satisfying C1 and C2. Further, for a given probability measure P on $\mathbf{R}^d$, assume that if $x \in C \in \mathscr{C}$ and $P(C) < \alpha$, then there is a $C_1 \in \mathscr{C}$ such that $x \in C_1^\circ$ and $P(C_1) < \alpha$. Then*:

(i) *$D(x; P, \mathscr{C})$ is upper semicontinuous*;

(ii) *$D^\alpha \equiv \{x \in \mathbf{R}^d \mid D(x; P, \mathscr{C}) \geq \alpha\}$, $\alpha \in (0, 1]$, are compact and nested (i.e., $D_{\alpha_1} \subset D_{\alpha_2}$ if $\alpha_1 > \alpha_2$); and*

(iii) *D^α is convex if every $C \in \mathscr{C}$ is convex.*

REMARK 2.7. If C2 is replaced by

C2′. $P(\partial C) = 0, \ \forall\, C \in \mathscr{C}$,

the above theorem remains true.

3. Concluding remarks. Here we examine and compare a number of depth functions with respect to the criteria given by properties P1–P4.

We begin with four cases having central importance because the corresponding versions of multidimensional median generated by their points of maximal depth are among the most popular competitors for nonparametric and robust estimation of multidimensional location. These are the *halfspace depth* (Type D, Example 2.7), the *simplicial depth* (Type A, Example 2.1), the *simplicial volume depth* (Type B, Example 2.3), and the L^2 *depth* (Type B, Example 2.4), which generate, respectively, the so-called Tukey/Donoho halfspace median (H), the Liu simplicial depth median (S), the Oja median (O) and the spatial or L^2 median. [See Small (1990) for an overview of these and other multidimensional medians.] With respect to affine invariance P1, all but the L^2 version are fully satisfactory, the L^2 depth function being invariant only under rotational and rigid-body transformations. The "maximality at center" property P2 is satisfied by the halfspace depth function for H-symmetric distributions (see the proof of Theorem 2.1) and can be shown to be satisfied by the L^2 depth function for all H-symmetric distributions (see Remark 2.4) and the simplicial volume depth function for C-symmetric distributions (see Corollary 2.2). Also, P2 is satisfied by the simplicial depth function for continuous A-symmetric distributions but not necessarily for discrete H-symmetric distributions (see Remark 2.1). The "monotonicity relative to deepest point" P3 is satisfied arbitrarily by the halfspace, simplicial volume, and L^2 depth functions, and also by the simplicial depth function except in some discrete cases (see Theorem 2.1, Remark 2.1, and Corollaries 2.1 and 2.3). Finally, "vanishing at infinity" P4 is satisfied by all four of these depth functions (see Theorem 2.1 and Corollaries 2.1 and 2.3). Thus, from consideration of P1–P4, the halfspace and simplicial volume depth functions appear to be the most comprehensively attractive among these four competitors. If, however, we in addition consider breakdown points of the corresponding location estimators [for details, see Small (1990), Niinimaa, Oja and Tableman (1990), Donoho and Gasko (1992) and Chen (1995)], we find that the estimator based on the simplicial volume depth, unlike the others, has breakdown point 0, while that based on the halfspace depth has breakdown point 1/3 for typical data sets, leading us to prefer the halfspace depth function more exclusively.

Let us now consider the *projection depth* and the *Mahalanobis depth*. By Theorems 2.9 and 2.10, these both satisfy properties P1–P4. Regarding robustness, however, the multidimensional median corresponding to sample projection depth has large-sample breakdown point 1/2 [see Tyler (1994), page 1033, and Zuo (1999)] as does the closely related Donoho-Stahel estimator [Stahel (1981), Donoho (1982) and Donoho and Gasko (1992)], whereas the robustness of the median generated by the Mahalanobis depth depends critically on the choice of location and covariance measures in defining this depth. We anticipate that suitable choices exist which yield high breakdown point. Therefore, we consider both of these depth functions to be competitive.

Another approach toward construction of depth functions consists of "peeling" methods, such as *convex hull peeling*. This latter approach, however, not only lacks a population analogue but also exhibits very unfavorable robustness properties. See discussion of Donoho and Gasko (1992), Nolan (1992) and Liu, Parelius and Singh (1999).

Likelihood-based depth functions have also been considered. See Fraiman and Meloche (1996), Fraiman, Liu and Meloche (1997) and Liu, Parelius and

Singh (1999). These, however, fail to satisfy in general any of P1–P4, and their effectiveness appears to be confined primarily to models with ellipsoidal densities, or to situations where sensitivity to multimodality is paramount. For further discussion, see Remark A.1 in Appendix A.

The *zonoid depth function* of Koshevoy and Mosler (1997) has some nice properties but can fail to satisfy "maximality at center" P2 for A- or H-symmetric distributions, because it attains maximum value always at the expectation $E(X)$ for any random variable X in $\mathbf{R}^d$. Also, the sample zonoid depth function is not robust, as a single corrupted data point can move the "center point of zonoid data depth" to infinity.

In conclusion, the *halfspace* and *projection* depth functions appear to represent very favorable choices. Both are implementations of the "projection pursuit" method, which utilizes all of the one-dimensional views of a dataset as a foundation for data analysis, thus producing the advantage of great power at extraction of information, although at the expense of a substantial computational burden. Also, competitively, the $\widetilde{L}^2$ and *Mahahalanobis* depth functions appear to have strong potential for development.

APPENDIX A: SUPPLEMENTARY NOTES

REMARK A.1. As pointed out and pictorially illustrated in Baggerly and Scott (1999), the near convexity of the simplicial depth contours limits their interpretability for multimodal data, whereas the likelihood depth contours follow the multimodality structure. In the usual sense of "center-outward ordering," and from the common standpoint of desiring connectedness of depth-trimmed regions, the likelihood "depth" has less of a role as a depth function than as simply what it is by definition: a density function, which keeps the information on multimodality structure when present.

REMARK A.2. As *broadenings* of central symmetry, angular and halfspace symmetry are opposite in character and purpose to several notions of nonparametric multivariate symmetry introduced by Beran and Millar (1997) which in fact are *narrowings* — see their formula (17). Also, their use of halfspaces is essentially for the purpose of indexing the empirical measure, rather than as a fundamental element in defining symmetry.

As shown in Zuo and Serfling (2000c), halfspace symmetry of P about θ reduces to angular symmetry about θ except when P is discrete with positive mass at θ. These exceptions are of practical relevance, since underlying distributions for actually observed phenomena are invariably discrete (and asymmetric), and it is reasonable to permit an approximating symmetric distribution to have mass at the center of symmetry.

REMARK A.3. An important aspect of any depth function is whether its sample version converges to the population counterpart. In particular, we desire that almost surely $[P]$

$$\text{(A.1)} \qquad \sup_x |D_n(x) - D(x; P)| \to 0, \quad n \to \infty.$$

Besides carrying intrinsic interest, (A.1) plays a supporting role for other purposes. For example, it underlies the convergence of sample depth *contours* to their population counterparts, as in He and Wang (1997) especially for elliptical models and in Zuo and Serfling (2000b) for more general models. In Liu and Singh (1993), it is basic to the convergence of a certain "quality index", while in Liu, Parelius and Singh (1999) it supports various practical methods such as "DD-plots."

Results on (A.1) are now available for several cases of depth function. Donoho and Gasko (1992) proved it for the sample halfspace depth,

$$HD_n(x) = \inf\{\widehat{P}_n(H) : H \text{ a closed halfspace}, \ x \in H\}, \quad x \in \mathbf{R}^d,$$

where $\widehat{P}_n$ denotes the usual empirical measure, and Liu (1990), Dümbgen (1990), and Arcones and Giné (1993) for the sample simplicial depth

$$SD_n(x) = \binom{n}{d+1}^{-1} \sum_{1 \le i_1 < \cdots < i_{d+1} \le n} \mathbf{I}\{x \in S[X_{i_1}, \ldots, X_{i_{d+1}}]\}, \quad x \in \mathbf{R}^d.$$

For the sample majority and Mahalanobis depths, under suitable conditions on F, (A.1) is established by Liu and Singh (1993). For sample versions of the "projection" depth function and the "Type D" depth functions introduced above, (A.1) is established in Appendix B of Zuo and Serfling (2000b).

APPENDIX B: PROOFS

PROOF OF THEOREM 2.1. Clearly, $HD(x; P)$ is bounded and nonnegative. We need only check P1–P4.

(a) *Affine invariance*. Straightforward.

(b) *Maximality at center*. Suppose that P is H-symmetric about a unique point $\theta \in \mathbf{R}^d$. By the definition of H-symmetry, we have $P(H_\theta) \ge 1/2$, for any closed halfspace H with $\theta \in \partial H$. It follows that $HD(\theta; P) \ge 1/2$. Now suppose that there is a point $x_0 \in \mathbf{R}^d$, $x_0 \ne \theta$, such that $HD(x_0; P) > 1/2$. Then $P(H) > 1/2$ for any closed halfspace H with $x_0 \in \partial H$, which implies that P is also H-symmetric about x_0, contradicting the assumption that P is H-symmetric about a unique point $\theta \in \mathbf{R}^d$. Therefore, $HD(\theta;\ P) = \sup_{x \in \mathbf{R}^d} HD(x; P)$.

(c) *Monotonicity relative to deepest point*. Suppose θ is a deepest point with respect to the underlying distribution. To compare $HD(x; P)$ and $HD(\theta + \alpha(x - \theta); P)$, we need only consider the infimum in the definition of HD over all closed halfspaces which do not contain θ. For any $H_{\theta+\alpha(x-\theta)}$ [closed halfspace with $(\theta+\alpha(x-\theta)) \in \partial H$], by the separating hyperplane theorem there always exists a closed halfspace H_x such that $H_x \subset H_{\theta+\alpha(x-\theta)}$. It follows that $HD(x; P) \le HD(\theta + \alpha(x - \theta); P)$, $\forall \alpha \in (0, 1)$.

(d) *Vanishing at infinity*. It is easy to see that $P(\|X\| \geq \|x\|) \to 0$ as $\|x\| \to \infty$ and that for each x and X there exists a closed halfspace H_x such that $H_x \subset \{\|X\| \geq \|x\|\}$. Thus $HD(x; P) \to 0$ as $\|x\| \to \infty$, completing the proof. □

Proof of Theorem 2.2. (a) Let θ be the center of an H-symmetric distribution P and x an arbitrary point in $\mathbf{R}^d$. Then, by the definition of H-symmetry, for any random sample $X_1, \ldots, X_d$ from P we have $x \in H^P_{X_1,\ldots,X_d} \Rightarrow \theta \in H^P_{X_1,\ldots,X_d}$ and thus $MJD(\theta; P) = \sup_{x\in\mathbf{R}^d} MJD(x; P)$.

(b) Let $\lambda \in (0, 1)$ and $x_0 \equiv \lambda\theta + (1-\lambda)x$. Then

$$\begin{aligned} MJD(x_0; P) - MJD(x; P) &= P\left(x_0 \in H^P_{X_1,\ldots,X_d}\right) - P\left(x \in H^P_{X_1,\ldots,X_d}\right) \\ &= P\left(x_0 \in H^P_{X_1,\ldots,X_d} \text{ and } x \notin H^P_{X_1,\ldots,X_d}\right) \\ &\geq 0. \end{aligned}$$

□

Proof of Theorem 2.3. By (i) and (ii) we have

$$\begin{aligned} Eh(x; X_1 - \theta, \ldots, X_r - \theta) &= Eh(\theta + x; X_1, \ldots, X_r), \\ Eh(x; \theta - X_1, \ldots, \theta - X_r) &= Eh(\theta - x; X_1, \ldots, X_r). \end{aligned}$$

Let y be a point in the set in (iv). It follows that

$$y \in \left(\arg\inf_{x\in\mathbf{R}^d} Eh(\theta + x; X_1, \ldots, X_r)\right) \cap \left(\arg\inf_{x\in\mathbf{R}^d} Eh(\theta - x; X_1, \ldots, X_r)\right).$$

The convexity of $h(x; x_1, \ldots, x_r)$ in x now yields

$$h(\theta; X_1, \ldots, X_r) \leq \tfrac{1}{2}h(\theta + y; X_1, \ldots, X_r) + \tfrac{1}{2}h(\theta - y; X_1, \ldots, X_r).$$

It follows that

$$\begin{aligned} Eh(\theta; X_1, \ldots, X_r) &\leq \tfrac{1}{2}Eh(\theta + y; X_1, \ldots, X_r)] + \tfrac{1}{2}Eh(\theta - y; X_1, \ldots, X_r) \\ &= \inf_{x\in\mathbf{R}^d} Eh(\theta + x; X_1, \ldots, X_r) \\ &= \inf_{x\in\mathbf{R}^d} Eh(x; X_1, \ldots, X_r). \end{aligned}$$

Hence $D(\theta; F) = \sup_{x\in\mathbf{R}^d} D(x; F)$, completing the proof. □

Proof of Theorem 2.4. Let θ in $\mathbf{R}^d$ be a deepest point with respect to the underlying distribution F, that is, $D(\theta; F) = \sup_{x\in\mathbf{R}^d} D(x; F)$. Let $x \neq \theta$ be an arbitrary point in $\mathbf{R}^d$, let $\lambda \in (0, 1)$ and set $x_0 \equiv \theta + \lambda(x - \theta)$. Then $D(x; F) \leq D(\theta; F)$. The convexity of $h(x; x_1, \ldots, x_r)$ in x yields

$$h(x_0; X_1, \ldots, X_r) \leq \lambda h(x; X_1, \ldots, X_r) + (1-\lambda)h(\theta; X_1, \ldots, X_r).$$

Thus

$$Eh(x_0; X_1, \ldots, X_r) \le \max\{Eh(x; X_1, \ldots, X_r), Eh(\theta; X_1, \ldots, X_r)\}$$
$$= Eh(x; X_1, \ldots, X_r),$$

and hence $D(x_0; F) \ge D(x; F)$, completing the proof. □

PROOF OF COROLLARY 2.1. (a) By Theorem 2.4, to show P3 we check convexity of $\Delta^\alpha(S[x, x_1, \ldots, x_d])$ in the argument x for $\alpha \in [1, \infty)$. Let x, y be two points in $\mathbf{R}^d$, take $\lambda \in (0, 1)$, and put $x_0 \equiv \lambda x + (1-\lambda)y$. Then

$$\Delta(S[x_0, x_1, \ldots, x_d]) = \left| \frac{1}{d!} \det \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_{01} & x_{11} & \cdots & x_{d1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{0d} & x_{1d} & \cdots & x_{dd} \end{vmatrix} \right|$$

$$= \left| \frac{1}{d!} \det \begin{vmatrix} \lambda + (1-\lambda) & 1 & \cdots & 1 \\ \lambda \tilde{x}_1 + (1-\lambda)\tilde{y}_1 & x_{11} & \cdots & x_{d1} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda \tilde{x}_d + (1-\lambda)\tilde{y}_d & x_{1d} & \cdots & x_{dd} \end{vmatrix} \right|$$

$$\le \lambda \Delta(S[x, x_1, \ldots, x_d]) + (1-\lambda)\Delta(S[y, x_1, \ldots, x_d]),$$

where $x = (\tilde{x}_1, \ldots, \tilde{x}_d)'$, $y = (\tilde{y}_1, \ldots, \tilde{y}_d)'$ and $x_i = (x_{i1}, \ldots, x_{id})'$ for $0 \le i \le d$. Now the convexity of the function x^α for $0 < x < \infty$ and $\alpha \ge 1$ yields

$$\Delta^\alpha(S[x_0, x_1, \ldots, x_d]) \le \lambda \Delta^\alpha(S[x, x_1, \ldots, x_d]) + (1-\lambda)\Delta^\alpha(S[y, x_1, \ldots, x_d]).$$

(b) It is obvious that $\Delta^\alpha(S[x; x_1, \ldots, x_d]) \to \infty$ as $\|x\| \to \infty$. Thus $SVD^\alpha(x; F) \to 0$ as $\|x\| \to \infty$, completing the proof. □

PROOF OF THEOREM 2.6. Since $\widetilde{L}^2 D(x; F)$ defined in (8) is affine invariant, and P4 is evident, we check P2 and P3.

(a) We first show that $\|\cdot\|_M$ is convex for any positive definite $d \times d$ matrix M. Since M is positive definite, there is a nonsingular matrix S such that $M = S'S$. Let x, y be two points in $\mathbf{R}^d$ and $\lambda \in (0, 1)$. Then

$$\begin{aligned} \|\lambda x + (1-\lambda)y\|_M^2 &= (\lambda x + (1-\lambda)y)' M (\lambda x + (1-\lambda)y) \\ &= \lambda^2 x'Mx + 2\lambda(1-\lambda)x'My + (1-\lambda)^2 y'My \\ &= \lambda^2 x'Mx + 2\lambda(1-\lambda)(Sx)'(Sy) + (1-\lambda)^2 y'My. \end{aligned}$$

The Schwarz inequality implies that

$$\begin{aligned} \|\lambda x + (1-\lambda)y\|_M^2 &\le \lambda^2 x'Mx + 2\lambda(1-\lambda)\|Sx\|\|Sy\| + (1-\lambda)^2 y'My \\ &= \lambda^2 \|x\|_M^2 + 2\lambda(1-\lambda)\|x\|_M \|y\|_M + (1-\lambda)^2 \|y\|_M^2 \\ &= (\lambda \|x\|_M + (1-\lambda)\|y\|_M)^2. \end{aligned}$$

It follows that

$$\|\lambda x + (1-\lambda)y\|_M \le \lambda\|x\|_M + (1-\lambda)\|y\|_M.$$

(b) Now we show that there is a point $y \in \mathbf{R}^d$ satisfying condition (4) of Theorem 2.3. Equivalently, we need to show that

$$\theta \in \arg\inf_{x\in\mathbf{R}^d} E\,[\,\|x - X\|_{\Sigma^{-1}}\,], \tag{B.1}$$

where Σ is the covariance matrix of F.

We first show that

$$E\left[\frac{\theta - X}{\|X-\theta\|_{\Sigma^{-1}}}\right] = 0. \tag{$*$}$$

Since F is angularly symmetric about θ, it can be shown [see Zuo and Serfling (2000c)] that $P(X \in H_\theta) = P(X \in -H_\theta)$, for any closed halfspace H_θ with θ on the boundary, where $-H_\theta$ is the reflection of H_θ about θ. Since Σ^{-1} is positive definite, there is a nonsingular matrix R such that $\Sigma^{-1} = R'R$. Thus

$$P(RX \in RH_\theta) = P(RX \in -RH_\theta),$$

for any closed halfspace H_θ with θ on the boundary. By nonsingularity and results established in Zuo and Serfling (2000c), we conclude that RX is angularly symmetric about $R\theta$. Hence

$$\frac{R\,(X-\theta)}{\|R\,(X-\theta)\|} \overset{\mathrm{d}}{=} \frac{R\,(\theta - X)}{\|R\,(\theta-X)\|},$$

which is equivalent to

$$\frac{R\,(X-\theta)}{\|(X-\theta)\|_{\Sigma^{-1}}} \overset{\mathrm{d}}{=} \frac{R\,(\theta - X)}{\|(\theta-X)\|_{\Sigma^{-1}}}.$$

This implies $(*)$.

Now we show that (B.1) holds true. Consider the derivative of $E[\,\|\mu - X\|_{\Sigma^{-1}}\,]$ with respect to $\mu \in \mathbf{R}^d$. By vector differentiation, we have

$$\begin{aligned}
\frac{d\,(E\,[\,\|\mu - X\|_{\Sigma^{-1}}\,])}{d\mu} &= \frac{d\left(\int_{\mathbf{R}^d}\|\mu - x\|_{\Sigma^{-1}}\,dF(x)\right)}{d\mu}\\
&= \int_{\mathbf{R}^d}\frac{d\,(\|\mu - x\|_{\Sigma^{-1}})}{d\mu}\,dF(x)\\
&= \int_{\mathbf{R}^d}\frac{\Sigma^{-1}(\mu - x)}{\|\mu - x\|_{\Sigma^{-1}}}\,dF(x)\\
&= \Sigma^{-1}E\left[\frac{\mu - X}{\|\mu - X\|_{\Sigma^{-1}}}\right].
\end{aligned}$$

Then by convexity and $(*)$ we conclude that (B.1) holds.

The result now follows from Theorems 2.3 and 2.4. □

PROOF OF THEOREM 2.9. Since $PD(x; F)$ is nonnegative and bounded, we need only check P1–P4.

(a) *Affine invariance*. Straightforward.

(b) *Maximality at center*. Suppose that F is H-symmetric about a unique point $\theta \in \mathbf{R}^d$. Then [see Zuo and Serfling (2000c)] we have $\text{Med}(u'X) = u'\theta$, for any unit vector $u \in \mathbf{R}^d$ and it follows that $PD(\theta; F) = \sup_{x \in \mathbf{R}^d} PD(x; F)$.

(c) *Monotonicity relative to deepest point*. We show that $O(x; X)$ is convex in its first argument. Let θ and x be two arbitrary points in $\mathbf{R}^d$, $0 < \alpha < 1$, and put $x_0 \equiv (1-\alpha)\theta + \alpha x$. Then we have

$$\begin{aligned} |\, u'x_0 - \text{Med}(u'X)\,| &= |\, u'((1-\alpha)\theta + \alpha x) - \text{Med}(u'X)\,| \\ &= |\,(1-\alpha)(u'\theta - \text{Med}(u'X)) + \alpha(u'x - \text{Med}(u'X))\,| \\ &\le (1-\alpha)\,|\,(u'\theta - \text{Med}(u'X))\,| + \alpha\,|\,(u'x - \text{Med}(u'X))\,|. \end{aligned}$$

It follows that

$$\begin{aligned} O(x_0; X) &= \sup_{\|u\|=1} \frac{|\, u'x_0 - \text{Med}(u'X)\,|}{\text{MAD}(u'X)} \\ &\le \sup_{\|u\|=1} \frac{(1-\alpha)\,|\,(u'\theta - \text{Med}(u'X))\,| + \alpha\,|\,(u'x - \text{Med}(u'X))\,|}{\text{MAD}(u'X)} \\ &\le (1-\alpha) O(\theta; F) + \alpha O(x; F). \end{aligned}$$

"Monotonicity" now follows from Theorem 2.8.

(d) *Vanishing at infinity*. Straightforward. □

PROOF OF THEOREM 2.11. (i) We first show that

$$(*) \qquad \{x \in \mathbf{R}^d \mid D(x; P, \mathscr{C}) \ge \alpha\} = \cap\{C \mid P(C) > 1-\alpha, C \in \mathscr{C}\}.$$

(a) If $x \in \{x \in \mathbf{R}^d \mid D(x; P, \mathscr{C}) \ge \alpha\}$ and there exists a $C \in \mathscr{C}$ such that $P(C) > 1-\alpha$, $x \notin C$, then $x \in C^c$, $P(C^c) < \alpha$. By C1 and C2, there is a $C_1 \in \mathscr{C}$ such that $x \in \partial C_1$, $C_1 \subset C^c$. It follows that $P(C_1) < \alpha$, and hence $D(x; P, \mathscr{C}) < \alpha$, which is a contradiction to the assumption that $x \in \{x \in \mathbf{R}^d \mid D(x; P, \mathscr{C}) \ge \alpha\}$. This implies

$$\{x \in \mathbf{R}^d \mid D(x; P, \mathscr{C}) \ge \alpha\} \subset \cap\{C \mid P(C) > 1-\alpha, C \in \mathscr{C}\}.$$

(b) If $x \in \cap\{C \mid P(C) > 1-\alpha, C \in \mathscr{C}\}$, and there is a $C \in \mathscr{C}$ such that $x \in C$, $P(C) < \alpha$, then by the condition given, there exists a $C_1 \in \mathscr{C}$ such that $x \in C_1^\circ$, $P(C_1) < \alpha$, and thus $x \notin \overline{C_1^c}$, $P(\overline{C_1^c}) > 1-\alpha$, which contradicts the assumption that $x \in \cap\{C \mid P(C) > 1-\alpha, C \in \mathscr{C}\}$. This implies

$$\{x \in \mathbf{R}^d \mid D(x; P, \mathscr{C}) \ge \alpha\} \supset \cap\{C \mid P(C) > 1-\alpha, C \in \mathscr{C}\}.$$

Now (a) and (b) yield $(*)$, which implies that D^α is closed, and thus $D(x; P, \mathscr{C})$ is upper semicontinuous.

(ii) The nestedness of D^α is trival. The boundedness of D^α follows from the fact that $D(x; P, \mathscr{C}) \to 0$ as $\|x\| \to \infty$. The compactness of D^α now follows from its being bounded and closed.

(iii) The convexity follows from $(*)$, since the intersection of convex sets is convex. □

Acknowledgments. The authors greatly appreciate the thoughtful and constructive remarks of an Associate Editor and two referees, which led to distinctive improvements in the paper.

REFERENCES

ARCONES, M. A. and GINÉ, E. (1993). Limit theorems for U-processes. *Ann. Probab.* **21** 1494–1542.

BAGGERLY, K. A. and SCOTT, D. W. (1999). Comment on "Multivariate analysis by data depth: Descriptive statistics, graphics and inference," by R. Y. Liu, J. M. Parelius and K. Singh. *Ann. Statist.* **27** 843–844.

BARTOSZYŃSKI, R., PEARL, D. K. and LAWRENCE, J. (1997). A multidimensional goodness-of-fit test based on interpoint distances. *J. Amer. Statist. Assoc.* **92** 577–586.

BERAN, R. J. and MILLAR, P. W. (1997). Multivariate symmetry models. In *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics* (D. Pollard, E. Torgerson and G. L. Yang, eds.) 13–42. Springer, Berlin.

CAPLIN, A. and NALEBUFF, B. (1988). On 64%-majority rule. *Econometrica* **56** 787–814.

CAPLIN, A. and NALEBUFF, B. (1991a). Aggregation and social choice: A mean voter theorem. *Econometrica* **59** 1–23.

CAPLIN, A. and NALEBUFF, B. (1991b). Aggregation and imperfect competition: On the existence of equilibrium. *Econometrica* **59** 25–59.

CARRIZOSA, E. (1996). A characterization of halfspace depth. *J. Multivariate Anal.* **58** 21–26.

CHAMBERLIN, E. (1937). *The Theory of Monopolistic Competition*. Harvard Univ. Press.

CHEN, Z. (1995). Bounds for the breakdown point of the simplicial median. *J. Multivariate Anal.* **55** 1–13.

DONOHO, D. L. (1982). Breakdown properties of multivariate location estimators. Ph. D. qualifying paper, Dept. Statistics, Harvard Univ.

DONOHO, D. L. and GASKO, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827.

DÜMBGEN, L. (1990). Limit theorems for the empirical simplicial depth. *Statist. Probab. Lett.* **14** 119–128.

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

FRAIMAN, R. and MELOCHE, J. (1996). Multivariate L-estimation. Preprint.

FRAIMAN, R., LIU, R. Y. and MELOCHE, J. (1997). Multivariate density estimation by probing depth. In *L_1-Statistical Procedures and Related Topics* (Y. Dodge, ed.) 415–430. IMS, Hayward, CA.

HE, X. and WANG, G. (1997). Convergence of depth contours for multivariate datasets. *Ann. Statist.* **25** 495–504.

HOTELLING, H. (1929). Stability in competition. *Econom. J.* **39** 41–57.

KOSHEVOY, G. and MOSLER, K. (1997). Zonoid trimming for multivariate distributions. *Ann. Statist.* **25** 1998–2017.

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

LIU, R. Y. (1992). Data depth and multivariate rank tests. In *L_1-Statistics and Related Methods* (Y. Dodge, ed.) 279–294. North-Holland, Amsterdam.

LIU, R. Y., PARELIUS, J. M. and SINGH, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference (with discussion). *Ann. Statist.* **27** 783–858.

LIU, R. Y. and SINGH, K. (1993). A quality index based on data depth and multivariate rank tests. *J. Amer. Statist. Assoc.* **88** 252–260.

MAHALANOBIS, P. C. (1936). On the generalized distance in statistics. *Proc. Nat. Acad. Sci. India* **12** 49–55.

MASSÉ, J. C. and THEODORESCU, R. (1994). Halfplane trimming for bivariate distributions. *J. Multivariate Anal.* **48** 188–202.

MIZERA, I. (1998). On depth and deep points: a calculus. Preprint.

MOSTELLER, C. F. and TUKEY, J. W. (1977). *Data Analysis and Regression.* Addison-Wesley, Reading, MA.

NIINIMAA, A., OJA, H. and TABLEMAN, M. (1990). On the finite sample breakdown point of the Oja bivariate median and of the corresponding half-samples version. *Statist. Probab. Lett.* **10** 325–328.

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

OJA, H. (1983). Descriptive statistics for multivariate distributions. *Statist. Probab. Lett.* **1** 327–333.

RAO, C. R. (1988). Methodology based on the L_1 norm in statistical inference. *Sankhyā Ser. A* **50** 289–313.

ROUSSEEUW, P. J. and HUBERT, M. (1999). Regression depth (with discussion). *J. Amer. Statist. Assoc.* **94** 388–433.

ROUSSEEUW, P. J. and RUTS, I. (1996). Bivariate location depth. *J. Roy. Statist. Soc. Ser. C* **45** 516–526.

ROUSSEEUW, P. J. and STRUYF, A. (1998). Computing location depth and regression depth in higher dimensions. *Statist. Comput.* **8** 193–203.

RUTS, I. and ROUSSEEUW, P. J. (1996). Computing depth contours of bivariate point clouds. *Comput. Statist. Data Anal.* **23** 153–168.

SERFLING, R. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

SINGH, K. (1991). A notion of majority depth. Preprint.

SMALL, C. G. (1987). Measures of centrality for multivariate and directional distributions. *Canad. J. Statist.* **15** 31–39.

SMALL, C. G. (1990). A survey of multidimensional medians. *Internat. Statist. Inst. Rev.* **58** 263–277.

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

TUKEY, J. W. (1975). Mathematics and picturing data. In *Proceedings of the International Congress on Mathematics* (R. D. James, ed.) **2** 523–531 Canadian Math. Congress.

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

VARDI, Y. and ZHANG, C.-H. (1999). The multivariate L_1-median and associated data depth. Preprint.

YEH, A. B. and SINGH, K. (1997). Balanced confidence regions based on Tukey's depth and the bootstrap. *J. Roy. Statist. Soc. Ser. B* **59** 639–652.

ZUO, Y. (1999). Affine equivariant multivariate location estimates with best possible breakdown points. Preprint.

ZUO, Y. and SERFLING, R. (2000a). Nonparametric notions of multivariate "scatter measure" and "more scattered" based on statistical depth functions. *J. Multivariate Anal.* To appear.

ZUO, Y. and SERFLING, R. (2000b). Structural properties and convergence results for contours of sample statistical depth functions. *Ann. Statist.* **28** 483–499.

ZUO, Y. and SERFLING, R. (2000c). On the performance of some robust nonparametric location measures relative to a general notion of multivariate symmetry. *J. Statist. Plann. Inference* **84** 55–79.

DEPARTMENT OF MATHEMATICS
ARIZONA STATE UNIVERSITY
TEMPE, ARIZONA 85287-1804
E-MAIL: zuoyijun@math.la.asu.edu

DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF TEXAS AT DALLAS
RICHARDSON, TEXAS 75083-0688
E-MAIL: serfling@utdallas.edu