The Annals of Statistics
1984, Vol. 12, No. 1, 127–141

ON PROJECTION PURSUIT MEASURES OF MULTIVARIATE LOCATION AND DISPERSION

BY JAMES ALLEN FILL[1] AND IAIN JOHNSTONE[2]

Stanford University

Huber (1981b) and Li and Chen (1981) have proposed the use of projection pursuit techniques to construct and estimate measures of location and dispersion for multivariate distributions. In this paper we show that such measures do not ordinarily commute with affine coordinate reexpressions unless the distribution is centrosymmetric (location case) or elliptically contoured (dispersion case).

0. Introduction. Projection pursuit (PP) techniques for the analysis of high dimensional data were introduced by Kruskal (1972) and Switzer (1970; Switzer and Wright, 1971) and Friedman and Tukey (1974) and later applied by Friedman and Stuetzle to a variety of problems, including regression (1980a), classification (1980b), and density estimation (1981). The idea underlying PP is to select "interesting" low dimensional projections by iteratively maximizing an appropriate projection index, usually with the aid of a computer.

Huber (1981b) and Li and Chen (1981) have proposed the use of PP to construct and estimate measures of location and dispersion for multivariate distributions. When is such a measure equivariant under affine tranformations? In essence we show that a PP location (dispersion) measure is equivariant if and only if the univariate measure from which it is built is linear (quadratic) in an appropriate sense (Lemmas 1, 4). This equivariance holds for all distributions only when the univariate measure is expectation (standard deviation) (Theorems 2, 5). For a fixed multivariate distribution, *all* PP location (dispersion) measures will be equivariant only when the distribution is symmetric through its center (elliptically contoured) (Theorems 3, 6). We emphasize that these theorems pertain to functionals on probability distributions. Corresponding estimators are constructed in the usual way by evaluating these functionals at the empirical distribution.

Thus PP measures of location and dispersion possess limited equivariance properties. Nevertheless, for data generated from elliptically contoured distributions the PP dispersion estimators of Li and Chen (1981) perform promisingly.

Our results are developed in Section 1 (location) and 2 (dispersion) and proved in Sections 4 and 5, respectively. Section 3 contains further comments and discussion of some implications for data analysis.

Received October 1982; revised July 1983.

[1] Research partially supported by a grant to Stanford University from SIAM Institute for Mathematics and Society.

[2] Research partially supported by Project Orion, Stanford University, and Mathematical Sciences Research Institute, Berkeley.

AMS 1980 *subject classifications.* Primary 62H05, secondary 62A05, 62F35.

Key words and phrases. Projection pursuit, multivariate location, multivariate dispersion, equivariance, centrosymmetry, elliptically contoured distributions.

1. PP location. Let Q be a real-valued functional on a given class of distribution functions F on $\mathbb{R}^1$. It will be convenient to abuse notation by writing QX for $Q(F)$ when X is a random variable with cdf F. To define a PP measure of location, we require that Q be a *location functional*:

$$Q(bX + c) = bQX + c$$

for all choices of b, c, X. Examples of location functionals include expectation, median, and M- and L-measures of location. Given a random vector $\mathbf{X} \in \mathbb{R}^p$, $p \geq 2$ fixed, define $Q_0 = Q_0(\mathbf{X})$ by

$$Q_0 = \sup\{Q(\mathbf{a}'\mathbf{X}): \|\mathbf{a}\| = 1\}$$

and let $\mathbf{a}_0 = \mathbf{a}_0(\mathbf{X})$ be a maximizing direction. Huber (1981b) defines the PP measure of location for $\mathscr{L}(\mathbf{X})$ based on Q as

$$\mathbf{TX} = \mathbf{T}(\mathscr{L}(\mathbf{X})) = Q_0\mathbf{a}_0.$$

In the most familiar example, Q is expectation and $\mathbf{TX} = \mathbf{EX}$ for $\mathbf{X}$ with finite expectation.

When $\mathscr{L}(\mathbf{X})$ has an unambiguous center $\boldsymbol{\mu}$, $\mathbf{TX} = \boldsymbol{\mu}$ for *any* choice of Q. More precisely, suppose that $\mathscr{L}(\mathbf{X})$ is *centrosymmetric* through $\boldsymbol{\mu}$:

$$\mathbf{X} - \boldsymbol{\mu} =_d -(\mathbf{X} - \boldsymbol{\mu}).$$

From characteristic functions an equivalent condition is that $\mathscr{L}(\mathbf{a}'(\mathbf{X} - \boldsymbol{\mu}))$ be symmetric about zero for each $\mathbf{a} \in \mathbb{R}^p$. For *any* location functional Q we find $Q(\mathbf{a}'(\mathbf{X} - \boldsymbol{\mu})) = 0$, so that $Q(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\boldsymbol{\mu}$ and consequently $\mathbf{TX} = \boldsymbol{\mu}$.

The PP location measure $\mathbf{T}$ is *translation equivariant* on the translation family

$$\mathrm{Tr}(\mathbf{X}) = \{\mathscr{L}(\mathbf{X} + \mathbf{b}): \mathbf{b} \in \mathbb{R}^p\}$$

generated by $\mathscr{L}(\mathbf{X})$ if

$$\mathbf{T}(\mathbf{X} + \mathbf{b}) = \mathbf{TX} + \mathbf{b}, \quad \mathbf{b} \in \mathbb{R}^p.$$

It is easy to see that $\mathbf{T} = \mathbf{E}$ is translation equivariant on each $\mathrm{Tr}(\mathbf{X})$, and that *every* PP location measure $\mathbf{T}$ is translation equivariant on $\mathrm{Tr}(\mathbf{X})$ if the given $\mathscr{L}(\mathbf{X})$ is centrosymmetric. In fact, these properties characterize expectation and centrosymmetry, respectively (Theorems 2, 3).

Huber (1981b) shows that translation equivariance forces Q to be linear:

LEMMA 1. $\mathbf{T}$ *is translation equivariant on* $\mathrm{Tr}(\mathbf{X})$ *if and only if there exists a vector* $\boldsymbol{\mu}$ *such that*

$$Q(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\boldsymbol{\mu}, \quad \mathbf{a} \in \mathbb{R}^p,$$

and then

$$\mathbf{TX} = \boldsymbol{\mu} = (Q(X_1), \cdots, Q(X_p))'.$$

A data analyst desiring a measure of location not biased toward any placement or scaling of coordinate axes might require *affine equivariance* of $\mathbf{T}$:

$$\mathbf{T}(A\mathbf{X} + \mathbf{b}) = A(\mathbf{TX}) + \mathbf{b}, \quad A \in \mathbb{R}^{p\times p}, \quad \mathbf{b} \in \mathbb{R}^p.$$

It is, however, an immediate consequence of Lemma 1 that a PP measure $\mathbf{T}$ is affinely equivariant for $\mathbf{X}$ if and only if it is translation equivariant on $\mathrm{Tr}(\mathbf{X})$.

Choice of the projection index Q will depend on the family of possible distributions of the input $\mathbf{X}$ to the PP location algorithm. At one extreme, if in each instance $\mathbf{X}$ has a multivariate normal distribution, then $Q = E$ seems a sensible choice. On the other hand, if $\mathscr{L}(\mathbf{X})$ can have heavy "tails," then a more robust location functional Q might be desired.

If equivariance guides the choice of a location measure, one can ask for which choices of Q the corresponding PP location measure $\mathbf{T}$ will be translation equivariant on $\mathrm{Tr}(\mathbf{X})$ for each $\mathscr{L}(\mathbf{X})$ in a given class. The answer is in principle provided by Lemma 1. However, Theorem 2 provides a more explicit solution when $\mathscr{L}(\mathbf{X})$ is completely unspecified, and Theorem 3 specifies when the choice of Q is totally unrestricted by equivariance.

Huber asked if there are any PP location measures other then expectation which are translation equivariant for all vectors $\mathbf{X}$. The answer is no if Q is nonnegative ($X \geq 0$ implies $QX \geq 0$).

THEOREM 2. *If the* PP *location measure* $\mathbf{T}$ *corresponding to a nonnegative location functional* Q *is translation equivariant on* $\mathrm{Tr}(\mathbf{X})$ *for each* $\mathbf{X}$ *with finite expectation, then* $Q = E$.

Changing the point of view, we now regard the distribution of $\mathbf{X}$ as fixed and let Q vary. The next theorem shows that *every* PP measure $\mathbf{T}_Q$ is equivariant for a genuinely multivariate $\mathbf{X}$ only when $\mathbf{X}$ has a centrosymmetric distribution.

We call $\mathbf{X}$ *one-dimensional* if $\mathbf{X} =_d \boldsymbol{\mu} + Z\mathbf{b}$ for some $\boldsymbol{\mu} \in \mathbb{R}^p$, $\mathbf{b} \in \mathbb{R}^p$, and scalar random variable Z. Since

$$Q(\mathbf{a}'\mathbf{X}) = \mathbf{a}'(\boldsymbol{\mu} + (QZ)\mathbf{b}), \quad \mathbf{a} \in \mathbb{R}^p,$$

it follows from Lemma 1 that each $\mathbf{T}_Q$ is translation equivariant on $\mathrm{Tr}(\mathbf{X})$ with $\mathbf{T}_Q\mathbf{X} = \boldsymbol{\mu} + (QZ)\mathbf{b}$. If $\mathscr{L}(Z)$ is not symmetric, then $\mathbf{T}_Q\mathbf{X}$ varies with Q.

THEOREM 3. *Suppose that* $\mathbf{T}_Q$ *is translation equivariant on* $\mathrm{Tr}(\mathbf{X})$ *for each nonnegative location functional* Q *defined on* $\{\mathscr{L}(\mathbf{a}'\mathbf{X}): \mathbf{a} \in \mathbb{R}^p\}$. *Then* $\mathscr{L}(\mathbf{X})$ *is either centrosymmetric or one-dimensional.*

2. PP dispersion. Analogous results apply to multivariate PP measures of dispersion. We now require the projection index Q to be a *scale functional* on cdf's:

$$Q(bX + c) = |b|\, QX$$

for all b, c, X. Let $\mathbf{X} \in \mathbb{R}^p$ be a given random vector. Following Huber (1980, 1981b) and Li and Chen (1981), we construct a PP pseudocovariance matrix $S(\mathbf{X})$

for $\mathscr{L}(\mathbf{X})$ using the principal components algorithm. Let

$$\lambda_1 = \sup\{Q^2(\mathbf{a}'\mathbf{X}): \|\mathbf{a}\| = 1\}$$

and let α_1 be a maximizing vector. Subsequent pseudo-eigenvalues λ_i and pseudo-eigenvectors α_i are found successively by restricting $\mathbf{a}$ to the orthogonal complement of the space spanned by the eigenvectors found previously. We then define

$$S(\mathbf{X}) = \textstyle\sum_{i=1}^{p} \lambda_i \alpha_i \alpha_i'.$$

In particular, if Q is standard deviation SD and $\mathbf{X}$ has covariance matrix Σ, then $S(\mathbf{X}) = \Sigma$ and the PP algorithm yields the principal components of variation in $\mathscr{L}(X)$.

Let $\mathscr{L}(\mathbf{X})$ be spherically symmetric about zero, i.e., suppose that $\Gamma\mathbf{X} =_d \mathbf{X}$ for each orthogonal matrix $\Gamma \in \mathbb{R}^{p\times p}$. An equivalent condition is that $\mathbf{a}'\mathbf{X} =_d X_1$ for all $\|\mathbf{a}\| = 1$, from which follows

$$Q^2(\mathbf{a}'\mathbf{X}) = k\,\|\mathbf{a}\|^2, \quad \mathbf{a} \in \mathbb{R}^p,$$

where $k = k_Q = Q^2(X_1)$. Thus $S(\mathbf{X}) = kI$; the lack of uniqueness of the sequence $(\alpha_1, \cdots, \alpha_p)$ poses no difficulty.

More generally, $\mathbf{X}$ is said to have an *elliptically contoured* distribution with center μ and nonnegative definite *shape matrix* Σ if there exists a (necessarily symmetric) scalar random variable W such that

$$\mathbf{a}'(\mathbf{X} - \mu) =_d (\mathbf{a}'\,\Sigma\mathbf{a})^{1/2} W, \quad \mathbf{a} \in \mathbb{R}^p$$

(cf. Anderson and Fang, 1982). Through an appropriate change of coordinates, it is again easy to check (cf. Li and Chen, 1981) that $S(\mathbf{X})$ is a scalar multiple (depending on Q) of Σ.

In the one-dimensional case $\mathbf{X} =_d \mu + Z\mathbf{b}$ we find likewise that $S(\mathbf{X})$ is a scalar multiple of $\mathbf{bb}'$.

As remarked by Li and Chen, the PP matrix $S(\mathbf{X})$ is in general symmetric, nonnegative definite, location invariant, and, by virtue of its construction, orthogonally equivariant:

$$S(\Gamma\mathbf{X}) = \Gamma S(\mathbf{X})\Gamma' \quad \text{for all orthogonal} \quad \Gamma.$$

It is natural to posit also that S be equivariant with respect to changes of coordinate scale in $\mathbb{R}^p$: $S(D\mathbf{X}) = DS(\mathbf{X})D$ for all diagonal D. Since any $p \times p$ matrix can be expressed as a product $\Psi D\Gamma$ of orthogonal matrices Ψ and Γ with a diagonal matrix $D \in \mathbb{R}^{p\times p}$, we are led to require *affine equivariance*:

$$S(A\mathbf{X} + \mathbf{b}) = AS(\mathbf{X})A', \quad A \in \mathbb{R}^{p\times p}, \quad \mathbf{b} \in \mathbb{R}^p.$$

Clearly covariance has this property, as does *any* PP operator S when $\mathscr{L}(\mathbf{X})$ is elliptically contoured.

The following analogue of Lemma 1 shows that affine equivariance forces $Q^2(\mathbf{a}'\mathbf{X})$ to be a quadratic form in $\mathbf{a}$.

LEMMA 4. *S is affinely equivariant on*

$$\mathrm{Aff}(\mathbf{X}) = \{\mathscr{L}(A\mathbf{X} + \mathbf{b}): A \in \mathbb{R}^{p\times p}, \mathbf{b} \in \mathbb{R}^p\}$$

if and only if there exists a symmetric matrix $\Sigma \in \mathbb{R}^{p\times p}$ *such that*

$$Q^2(\mathbf{a}'\mathbf{X}) = \mathbf{a}'\Sigma\mathbf{a}, \quad \mathbf{a} \in \mathbb{R}^p,$$

and then

$$S(\mathbf{X}) = \Sigma = (\sigma_{ij}) = (1/2)[Q^2(X_i + X_j) - Q^2(X_i) - Q^2(X_j)]).$$

To obtain a characterization of covariance analogous to Theorem 2, we shall require the scale functional Q to satisfy two monotonicity conditions, the first designed for symmetric distributions and the second for arbitrary variables. Let X and Y be random variables symmetric about μ and ν, respectively. X is said to be *less dispersed* than Y ($X <_d Y$) if

$$|X - \mu| \le |Y - \nu| \text{ stochastically.}$$

Now drop the symmetry requirement and let X and Y have distribution functions F and G, respectively. X is said to be *less spread out* than Y($X <_s Y$) if

$$F^{-1}(v) - F^{-1}(u) \le G^{-1}(v) - G^{-1}(u), \quad 0 < u < v < 1.$$

For definiteness we use left-continuous inverses, but any fixed convex combination of left and right continuous inverses would be an equivalent choice. These definitions are discussed further by Bickel and Lehmann (1976, 1978).

A scale functional is *dispersion* (resp., *spread*) *monotone* if it is monotone relative to the ordering $<_d$ (resp., $<_s$). Both orderings are needed because in the symmetric case $<_s$ compares too few distributions to be useful alone; for example, $(1/2)(\delta_{(-1)} + \delta_{(+1)})$ is not spread-comparable to the uniform density on $(-1, 1)$.

Let $L_2(L_2^s)$ denote the class of all distribution functions (symmetric about zero) with second moment. Since kQ is a scale functional if Q is, we suppose henceforth that Q has been normalized so that

$$Q(1/2\ \delta_{(-1)} + 1/2\ \delta_{(+1)}) = 1.$$

THEOREM 5. *Suppose that S is affinely equivariant for all square integrable random vectors of a fixed dimension $p \ge 2$.*

(i) *If Q is dispersion monotone, then* $Q^2 = \mathrm{Var}$ *on* L_2^s.

(ii) *If, in addition, Q is spread monotone, then* $Q^2 = \mathrm{Var}$ *on all of* L_2.

Consider again the situation of $\mathcal{L}(\mathbf{X})$ fixed and Q varying. Theorem 6 says that if *every* PP dispersion measure S_Q is affinely equivariant for a genuinely multivariate $\mathbf{X}$, then $\mathcal{L}(\mathbf{X})$ is elliptically contoured.

THEOREM 6. *Suppose that* $\mathbf{X}$ *has finite expectation, and that S_Q is affinely equivariant on* $\mathrm{Aff}(\mathbf{X})$ *for each scale functional Q defined on* $\{\mathcal{L}(\mathbf{a}'\mathbf{X}): \mathbf{a} \in \mathbb{R}^p\}$. *Then* $\mathcal{L}(\mathbf{X})$ *is either elliptically contoured or one-dimensional.*

3. Further comments and discussion.

A. *Higher dimensional starts.* The question naturally arises whether wider classes of affinely equivariant PP measures can be constructed if one begins with projections onto subspaces of dimension two (or higher). We examine this issue for the more interesting dispersion problem; analogous comments apply to the location case.

Let $\mathbb{R}_+^{2\times 2}$ denote the class of nonnegative definite 2×2 matrices. One approach is to replace the real-valued pseudovariance functional Q^2 on univariate distributions with an affinely equivariant $\mathbb{R}_+^{2\times 2}$-valued functional Q^2 on bivariate distributions. Choosing a total ordering on $\mathbb{R}_+^{2\times 2}$, we then search for the $p \times 2$ matrix A with orthonormal columns which optimizes $Q^2(A'\mathbf{X})$; call this matrix A_1. By successively restricting the columns of A to the orthogonal complement of those of the preceding A_i's, we obtain $A_1, A_2, \cdots, A_{p/2}$ (assume p is even) and define

$$S(\mathbf{X}) = \sum_{i=1}^{p/2} A_i Q^2(A_i' \mathbf{X}) A_i'.$$

Suppose, however, that the ordering on $\mathbb{R}_+^{2\times 2}$ is consistent with the partial ordering of unit rank matrices $(a\ b)'\ (a\ b)$ induced by the trace $a^2 + b^2$. Then one can show that no new affinely equivariant PP dispersion measures S are generated from this recipe: $S(\mathbf{X})$ can also be built up from a univariate Q satisfying Lemma 4.

B. *Directional data.* We have seen that the term PP "location measure" is in fact a misnomer since there exist distributions for which $\mathbf{T}$ is not translation equivariant. As Donoho (1982, Section 4.3) notes, however, the corresponding location measures for directional data *do* possess directional equivariance. For example, let Q = median and let $\mathscr{L}(\mathbf{X})$ be concentrated on the unit sphere in $\mathbb{R}^p$. Then $\mathbf{TX}$ is the central ray of the (infinite) cone of least angle containing at least half the data. $\tilde{\mathbf{T}}\mathbf{X} = \mathbf{TX}/\|\mathbf{TX}\|$ is a measure of directional location: it is equivariant with respect to rotations about the origin. We note in passing that the empirically derived $\tilde{\mathbf{T}}(F_n)$ converges slowly to $\tilde{\mathbf{T}}(F)$; in the case $p = 2$, for example, the "shorth" argument of Andrews, *et al.* (1971, pages 50–52) implies a rate of order $n^{-1/3}$.

C. *Implications for data analysis.* The location and dispersion measures have been considered as functions on theoretical distributions. Point configurations (empirical measures) of samples from a continuous distribution are, with probability one, neither centrosymmetric nor ellipically contoured, even when the parent distribution is precisely symmetric.

In fact, not even the *distribution* of a PP estimator is equivariant. We illustrate the dispersion case. Let $S_n(\mathbf{X}) = S(F_n)$ denote the PP estimator of $S(\mathbf{X})$, where F_n is the empirical measure based on $X_1, \cdots, X_n$. Then in general $\mathscr{L}(S_n(A\mathbf{X})) \neq \mathscr{L}(AS_n(\mathbf{X})A')$, even if $\mathscr{L}(\mathbf{X})$ is spherically symmetric. For a simple example, take three observations from the uniform distribution on the unit sphere in $\mathbb{R}^2$, and let Q = range and $A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$. While the condition number of $S_3(A\mathbf{X})$ is at least $4/3$, there is positive probability that the condition number of $AS_3(\mathbf{X})A'$ is strictly smaller than $4/3$.

At this point one asks—how important is equivariance? The answer of course depends on the statistician and the data analytic situation. Equivariance may be less compelling, for example, when the original coordinates have valuable interpretations. On the other hand, robust affinely equivariant location estimators in $\mathbb{R}^P$ *are* available (Donoho, 1982, Section 2.3; Stahel, 1981), but are expensive to compute—currently, the best algorithms require about $n\binom{n}{p}$ operations.

How much equivariance is lost by using the relatively inexpensive PP "location estimators" in nonsymmetric situations? Unfortunately, the effect of coordinate shifts can be arbitrarily large even for a standard normal distribution with a fixed proportion ε of contamination at the point $(x, 0)'$. Using $Q =$ median and measuring location from $(x, y)'$, where $y > x^2/(2\theta)$ with $\theta = \Phi^{-1}(1/(2(1 - \varepsilon)))$, one finds $\mathbf{TX} = (x, 0)'$. In contrast, $\mathbf{TX}$ measured from the origin is $(\theta, 0)'$. Note that the unsatisfactory behavior of $\mathbf{T}$ is achieved only by moving the contamination far (in the sense of Euclidean distance) from the symmetric center.

One possible remedy for nonequivariance is to choose the coordinate origin via a true location measure $\mathbf{T}_0$ (e.g., coordinatewise median) and then apply a PP operator $\mathbf{T}$. The resulting measure $\mathbf{T}^*\mathbf{X} = \mathbf{T}_0\mathbf{X} + \mathbf{T}(\mathbf{X} - \mathbf{T}_0\mathbf{X})$ is then translation equivariant, and even rigid motion equivariant if $\mathbf{T}_0$ is. Such a measure *is* typically continuous with respect to ε-contamination from centrosymmetry. To be specific, let d denote the Kolmogorov metric $d(F, G) = \sup\{|F(t) - G(t)|: t \in \mathbb{R}\}$ on univariate df's and define the PP multivariate extension $d(\mathbf{W}, \mathbf{X}) = \sup\{d(\mathbf{a}'\mathbf{W}, \mathbf{a}'\mathbf{X}): \|\mathbf{a}\| = 1\}$. Let $b(\varepsilon, F) = b(\varepsilon, F, Q) = \sup\{|Q(G) - Q(F)|: d(F, G) \le \varepsilon\}$ be the bias contamination function (e.g., Huber, 1981a), and define $b(\varepsilon, \mathbf{W}) = \sup\{b(\varepsilon, \mathbf{a}'\mathbf{W}): \|\mathbf{a}\| = 1\}$. If $\mathbf{W}$ is centrosymmetric about some point and $d(\mathbf{W}, \mathbf{X}) = \varepsilon$, then we show

$$\|\mathbf{T}^*\mathbf{X} - \mathbf{T}^*\mathbf{W}\| \le (p + 1)^{1/2} b(\varepsilon, \mathbf{W})$$

provided $\mathbf{T}_0\mathbf{X}$ has ith coordinate $Q(\mathbf{e}_i'\mathbf{X})$, $i = 1, \cdots, p$. By the translation equivariance of $\mathbf{T}^*$ and invariance of the multivariate d and b, we may suppose $\mathbf{T}_0\mathbf{X} = \mathbf{0}$, so that $\mathbf{T}$ and $\mathbf{T}^*$ agree at both $\mathbf{W}$ and $\mathbf{X}$. But the projection of $\mathbf{TX} - \mathbf{TW}$ onto $\mathbf{TX}$ has length $\le b(\varepsilon, \mathbf{W})$, while the length of the residual does not exceed $\|\mathbf{TW}\| = (\sum_{i=1}^p Q^2(\mathbf{e}_i'\mathbf{W}))^{1/2} \le p^{1/2} b(\varepsilon, \mathbf{W})$.

Other translation equivariant possibilities—computationally more expensive—are to minimize $J(\mathbf{x}) = \sup\{Q_1(\mathbf{a}'(\mathbf{X} - \mathbf{x})): \|\mathbf{a}\| = 1\}$ or $K(\mathbf{x}) = \sup\{Q_1(\mathbf{a}'(\mathbf{X} - \mathbf{x}))/Q_2(\mathbf{a}'(\mathbf{X} - \mathbf{x})): \|\mathbf{a}\| = 1\}$, where Q_1 and Q_2 are respectively location and scale functionals. The latter solution is even affinely equivariant and, as with other PP techniques, will inherit resistance properties from the projection indices Q_1 and Q_2.

4. Proofs for PP location.

A. PROOF OF LEMMA 1. In Section 1 we tacitly assumed that the supremum of $Q(\mathbf{a}'\mathbf{X})$ over unit vectors $\mathbf{a}$ in $\mathbb{R}^p$ is achieved by some $\mathbf{a}_0$. More generally we now suppose only that the supremum Q_0 is finite, let $\mathbf{a}_0$ be any limit point of a maximizing sequence, and denote by $\{\mathbf{TX}\}$ the set of all vectors $Q_0\mathbf{a}_0$ so obtained.

For an example demonstrating lack of uniqueness of $\mathbf{TX}$, let Q be median and $\mathscr{L}(\mathbf{X}) = (1/3)(\delta_{\mathbf{e}_1} + \delta_{\mathbf{e}_2} + \delta_{-\mathbf{e}_2})$, where $\mathbf{e}_i$ are the coordinate vectors in $\mathbb{R}^2$. In this case $\{\mathbf{TX}\} = \{(½, ½), (½, -½)\}$. Nonuniqueness also occurs if $\mathscr{L}(\mathbf{X})$ has uniform density on the interior of the triangle with vertices $\mathbf{e}_1$, $\mathbf{e}_2$, $-\mathbf{e}_2$.

In this more general setting translation equivariance of $\mathbf{T}$ requires equality of the sets $\{\mathbf{T}(\mathbf{X} + \mathbf{b})\}$ and $\{\mathbf{TX}\} + \mathbf{b}$ for each $\mathbf{b} \in \mathbb{R}^p$. For an alternative proof of Lemma 1, see Fill and Johnstone (1982).

B. Proof of Theorem 2. A proof of Theorem 2 may be based on the usual Riesz representation theorem for L^2 (Critchlow, 1981; described in Fill and Johnstone, 1982, pages 17–18). We present here an elementary proof which extends more conveniently to the dispersion problem of Section 5. Our argument has four parts, the first three developing a characterization of expectation on any atomless probability space in which the usual continuity requirements are replaced by the assumptions of nonnegativity of Q. We begin by establishing the result for two point distributions.

1°. Let $(\Omega, \mathscr{A}, P)$ be an atomless probability space and $Q(\cdot) = f(P(\cdot))$ a finitely additive set function on $\mathscr{A}$ with $Q(\Omega) = 1$. If Q is nonnegative valued, then $Q = P$.

To see this, choose p_1 and p_2 from $[0, 1]$ with $p_1 + p_2 \le 1$. P is atomless, so there exist disjoint sets $A_i \in \mathscr{A}$ with $P(A_i) = p_i$ for $i = 1, 2$. Additivity of Q entails $f(p_1 + p_2) = f(p_1) + f(p_2)$. Combined with $f(1) = 1$, this easily yields $f(1/n) = 1/n$ for $n = 1, 2, \cdots$, and then $f(p) = p$ for rational $p \in [0, 1]$. Nonnegativity of Q forces f to be monotone, and the conclusion follows.

2°. If μ is a finite Borel measure on $[0, \infty)$, then there exists an unbounded increasing nonnegative function on $[0, \infty)$ which is μ-integrable. Consequently, if $X \ge 0$ is an integrable random variable on a probability space $(\Omega, \mathscr{A}, P)$, then there exists an unbounded increasing function g with $E(Xg(X)) < \infty$.

To prove this, it suffices to consider a cdf F on $[0, \infty)$ having unbounded support, so that $t_n \equiv F^{-1}(1 - n^{-3}) \uparrow \infty$ as $n \uparrow \infty$, where F^{-1} denotes the left continuous inverse of F. The monotone unbounded function $g(t) = \sum_{n=1}^{\infty} nI_{(t_n, t_{n+1}]}(t)$ satisfies

$$\int g\,d\mu \le \sum_{n=1}^{\infty} n[1 - F(t_n)] \le \sum_{n=1}^{\infty} n^{-2} < \infty.$$

3°. Let $(\Omega, \mathscr{A}, P)$ be an atomless probability space and suppose that $Q(X) = Q(\mathscr{L}(X))$ is defined for $X \in L_1(\Omega)$ and satisfies

(i) $Q1 = 1$,

(ii) $Q(X + Y) = QX + QY$ regardless of the joint $\mathscr{L}(X, Y)$, and

(iii) $X \ge 0$ implies $QX \ge 0$.

Then $Q = E$.

First we note that conditions (i)–(iii) insure that 1° applies here; hence $Q(I_A) = P(A)$ for each $A \in \mathscr{A}$. Together with repeated use of (ii), this shows that $Q(aI_A) = aP(A) = E(aI_A)$ for rational $a \ge 0$. Monotonicity of Q yields the same relation for all $a \in [0, \infty)$. Additivity then establishes $QX = EX$ for simple random variables $X \geq 0$, and an approximation argument based on the monotonicity of Q extends the equality to all bounded $X \geq 0$.

For unbounded $0 \leq X \in L_1$, observe that for any $c \in [0, \infty)$

$$QX = E(XI_{\{X \leq c\}}) + Q(XI_{\{X>c\}}) = EX + RX,$$

where

$$0 \leq RX \equiv \lim_{c \uparrow \infty} \downarrow Q(XI_{\{X>c\}}) < \infty.$$

Let g be an unbounded increasing function with $E(Xg(X)) < \infty$, as provided by 2°. Choose $M \geq 1$ and c sufficiently large that $g(c) \geq M$. It follows that

$$Q(Xg(X)I_{\{Xg(X)>c\}}) \geq Q(Xg(X)I_{\{X>c\}}) \leq MQ(XI_{\{X>c\}}),$$

whence

$$\infty > R(Xg(X)) \geq M\, RX.$$

Therefore $RX = 0$ and $QX = EX$ for all $0 \leq X \in L_1$.

Now for general $X \in L_1$, write $X = X^+ - X^-$ and use (ii) to extend the identity to all of L_1.

4°. To finish the proof of the theorem, take $(\Omega, \mathscr{A}, P)$ in 3° to be ([0, 1], Borel field, uniform probability). Since Q is a location functional, $Q1 = 1$. Translation equivariance of $\mathbf{T}$ implies that Q is finitely additive. Since the random variable $F^{-1}(U)$ has cdf F when U is uniformly distributed on [0, 1], it follows that $QX = \int_0^1 F^{-1}(u)\, du = EX$ whenever $E\,|X| < \infty$.

C. Proof of Theorem 3. We again fix dimension $p \geq 2$.

Let $\mathbf{X}$ be a given random vector, and $\mathscr{Q} = \mathscr{Q}_{\mathbf{X}}$ the class of nonnegative location functionals with domain at least $\{\mathscr{L}(\mathbf{a}'\mathbf{X}): \mathbf{a} \in \mathbb{R}^p\}$. For each $Q \in \mathscr{Q}$ let $\mathbf{T}_Q\mathbf{X}$ be the PP location measure for $\mathbf{X}$.

For an arbitrary scalar random variable Z with distribution function F, define $G_\alpha Z$ to be the α-quantile $(1/2)(F^{-L}(\alpha) + F^{-R}(\alpha))$, where F^{-L} and F^{-R} are, respectively, the left and right continuous inverses of F. Define the "α-mid"

$$M_\alpha Z = (1/2)(G_\alpha Z + G_{1-\alpha} Z).$$

With our definition of G_α, M_α is easily seen to be a monotone location functional.

Suppose now that $\mathscr{L}(\mathbf{X})$ is not centrosymmetric. Rotating and translating $\mathscr{L}(\mathbf{X})$ if necessary, we may suppose (Lemma 2.1) that $\text{med}(\mathbf{a}'\mathbf{X}) = 0$ for $\mathbf{a} \in \mathbb{R}^p$ and that $\mathscr{L}(X_1)$ is not symmetric (about zero); here med $\equiv M_{1/2}$. In this case there exists $\alpha \in (0, 1/2)$ for which $M_\alpha X_1 \neq 0$.

We define the relation of *affine equivalence* on distribution functions by decreeing for scalar random variables Y, Z that $Y \sim Z$ if there exist real constants $b \neq 0$ and c such that $Z =_d bY + c$. Now construct a nonnegative location functional Q_α in $\mathscr{Q}$ by setting

$$Q_\alpha Z = \begin{cases} bM_\alpha X_1 + c, & \text{if} \quad Z =_d bX_1 + c; \\ \text{med}\, Z, & \text{if} \quad Z \not\sim X_1. \end{cases}$$

We show that location equivariance of T_{Q_α} forces all projections $\mathbf{a}'\mathbf{X}$ to be equivalent. For suppose to the contrary that $Y = \mathbf{a}_1'\mathbf{X} \not\sim X_1$. Then $Q_\alpha Y = \text{med}(\mathbf{a}_1'\mathbf{X}) = 0$ and hence, by Lemma 1, $Q_\alpha(cX_1 + Y) = cM_\alpha X_1 \neq 0 = \text{med}(cX_1 + Y)$ for $c \neq 0$. Consequently $cX_1 + Y \sim X_1$ for all $c \neq 0$. But $cX_1 + Y \to_d Y$ as $c \to 0$, so by the convergence of types theorem (e.g. Breiman, 1968) $Y \sim X_1$. This contradiction forces the conclusion that for each $\mathbf{a} \in \mathbb{R}^p$ there are scalars $b(\mathbf{a})$, $c(\mathbf{a})$ with $\mathbf{a}'\mathbf{X} =_d b(\mathbf{a})X_1 + c(\mathbf{a})$.

Equating medians we find that $c(\mathbf{a}) \equiv 0$. Lemma 1 shows that

$$b(\mathbf{a})QX_1 = Q(\mathbf{a}'\mathbf{X}) = \mathbf{a}'(QX_1, \cdots, QX_p)'$$

for all $\mathbf{a} \in \mathbb{R}^p$ and $Q \in \mathscr{Q}$. Let $b_j = b(\mathbf{e}_j)$ and $\mathbf{b} = (b_1, \cdots, b_p)'$. It follows that $QX_j = b_j QX_1$ and hence that $b(\mathbf{a})QX_1 = \mathbf{a}'\mathbf{b}QX_1$. Taking for Q the Q_α constructed above allows us to divide by $Q_\alpha X_1 \neq 0$ and conclude that $b(\mathbf{a}) = \mathbf{a}'\mathbf{b}$. Therefore $\mathbf{a}'\mathbf{X} =_d \mathbf{a}'\mathbf{b}X_1$ for all $\mathbf{a} \in \mathbb{R}^p$, so that $\mathbf{X} =_d \mathbf{b}X_1$ is one-dimensional.

REMARK 4.1. The location functional Q_α employed in the proof of Theorem 3 is indeed nonnegative, but quite irregular. If the hypotheses of Theorem 3 are strengthened by the requirement that $T_Q(\mathbf{X}) = \mu$ be independent of $Q \in \mathscr{Q}$, then the class $\mathscr{Q}$ may be reduced to the collection $\{M_\alpha: 0 < \alpha \leq ½\}$ of α-mids. See Fill and Johnstone (1982, page 20).

5. Proofs for PP dispersion.

A. PROOF OF LEMMA 4. In proving Lemma 4 we do not require that $S(\mathbf{X})$ be uniquely defined, the general requirement for affine (linear) equivariance being that the sets $\{S(A\mathbf{X})\}$ and $A\{S(\mathbf{X})\}A'$ agree for each $A \in \mathbb{R}^{p\times p}$; the notation here is as in Section 4A.

Suppose first that S is equivariant. Let $A \in \mathbb{R}^{p\times p}$ be the matrix with first row $\mathbf{a}'$ and all other rows $\mathbf{0}'$, so that $AS(\mathbf{X})A' = \text{diag}\{\mathbf{a}'S(\mathbf{X})\mathbf{a}, 0, \cdots, 0\}$ and $A\mathbf{X} = (\mathbf{a}'\mathbf{X})\mathbf{e}_1$. It is easy to see that $S(A\mathbf{X}) = \text{diag}\{Q^2(\mathbf{a}'\mathbf{X}), 0, \cdots, 0\}$ uniquely. Thus affine equivariance implies that $Q^2(\mathbf{a}'\mathbf{X}) = \mathbf{a}'S(\mathbf{X})\mathbf{a}$ for all $\mathbf{a} \in \mathbb{R}^p$. Hence $\{S(\mathbf{X})\}$ is a singleton whose element we denote Σ

Conversely, if the *quadratic form equation*

$$Q^2(\mathbf{a}'\mathbf{X}) \equiv \mathbf{a}'\Sigma\mathbf{a}$$

holds, then the relations $Q^2(\mathbf{a}'A\mathbf{X}) = Q^2((A'\mathbf{a})'\mathbf{X}) = \mathbf{a}'A\Sigma A'\mathbf{a}$ show that $S(A\mathbf{X}) = A\Sigma A'$ uniquely, with $\Sigma = S(\mathbf{X})$.

We now identify the entries σ_{ij} of $\Sigma = S(\mathbf{X})$. Putting $\mathbf{a} = \mathbf{e}_i$ and then $\mathbf{a} = \mathbf{e}_i + \mathbf{e}_j$ in the quadratic form equation shows that

$$\sigma_{ii} = Q^2(X_i), \quad \sigma_{ij} = C(X_i, X_j), \quad i = 1, \cdots, p, \quad j = 1, \cdots, p,$$

where the *pseudocovariance functional* C is defined by

$$C(Y, Z) \equiv (1/2)[Q^2(Y+Z) - Q^2(Y) - Q^2(Z)]$$

and depends only on the joint distribution of Y and Z. In these terms the quadratic form equation takes the form

$$Q^2(\sum_{i=1}^{p} a_i X_i) = \sum_{i=1}^{p} \sum_{j=1}^{p} a_i a_j C(X_i, X_j).$$

For future reference we note first that if Y and Z have the same joint distribution as Y and $(-Z)$, then $C(Y, Z) = 0$. Secondly, if Q^2 is of quadratic form on a linear space $\mathscr{M}$ of random variables on a given probability space, then C and Q define a semidefinite inner product and seminorm, respectively, on $\mathscr{M}$. In particular, the parallelogram law and triangle inequality then hold for Q.

B. PROOF OF THEOREM 5. We prove a weakened version of Theorem 5 assuming affine equivariance of S for *every* dimension $p \geq 2$; see Fill and Johnstone (1982, page 26) for the extension to fixed dimension.

1°. We first establish that $Q^2 = \text{Var}$ for symmetric three-point distributions: with $F_p \equiv p\delta_{(-1)} + (1-2p)\delta_0 + p\delta_{(+1)}$ we show

$$f(p) \equiv Q^2(F_p) = 2p, \quad 0 \leq p \leq 1/2.$$

Since by convention $f(1/2) = 1$, it is enough to show that $f(p_1 + p_2) = f(p_1) + f(p_2)$ whenever p_1, p_2, and $p_1 + p_2$ all lie in $[0, 1/2]$. For then the argument of 1° of Theorem 2 establishes $f(p) \equiv 2p$ via monotonicity of f, which here is a consequence of additivity of f and the fact that $Q^2 \geq 0$.

To show additivity, pick four mutually exclusive events A_1, A_1', A_2, A_2' from an atomless probability space with $P(A_i) = P(A_i') = p_i$, $i = 1, 2$. Then

$$\begin{aligned} f(p_1 + p_2) &= Q^2([I_{A_1'} - I_{A_1}] + [I_{A_2'} - I_{A_2}]) \\ &= Q^2(I_{A_1'} - I_{A_1}) + Q^2(I_{A_2'} - I_{A_2}) = f(p_1) + f(p_2); \end{aligned}$$

$C(I_{A_1'} - I_{A_1}, I_{A_2'} - I_{A_2}) = 0$ since $\mathscr{L}(I_{A_1'} - I_{A_1}, I_{A_2'} - I_{A_2}) = \mathscr{L}(I_{A_1'} - I_{A_1}, -(I_{A_2'} - I_{A_2}))$.

2°. Here we show that $Q^2 = \text{Var}$ on L_2^s. Represent a simple symmetric random variable X as $X = \sum_{j=1}^{r} a_j(I_{A_j'} - I_{A_j})$, where all $2r$ events A_j, A_j' are mutually exclusive and $P(A_j) = P(A_j')$ for each j. As in 1°, $C(I_{A_j'} - I_{A_j}, I_{A_k'} - I_{A_k}) = 0$ for $j \neq k$; thus from the quadratic form equation and 1°

$$Q^2(X) = 2 \sum_{j=1}^{r} a_j^2 P(A_j) = \text{Var } X.$$

An approximation argument using the dispersion monotonicity of Q extends the result to bounded symmetric X. For arbitrary $X \in L_2^s$, approximation "from within", that is, approximation to $|X|$ from below, shows that $Q^2(X) \geq \text{Var } X$ on L_2^s.

For the reverse inequality we argue as in 3° of Theorem 2. By the triangle inequality for Q

$$QX \le SD(XI_{\{|X|\le c\}}) + Q(XI_{\{|X|>c\}}) = SD(X) + RX,$$

where

$$0 \le R^2(X) \equiv \lim_{c\uparrow\infty} \downarrow Q^2(XI_{\{|X|>c\}}) < \infty.$$

Using 2° of Theorem 2, we let g be an increasing unbounded nonnegative function on $[0, \infty)$ such that the symmetric variable $Xg^{1/2}(X^2)$ has finite variance. As in 3° of Theorem 2, the finiteness of $R^2(Xg^{1/2}(X^2))$ implies that $R^2(X) = 0$, whence $Q^2 = \text{Var}$ on L_2^s.

3°. We show that $Q^2 = \text{Var}$ on all of L_2 if Q *preserves the law of large numbers* (LLN), in the sense that if $X, X_1, X_2, \cdots$ is an iid sequence in L_2 and $\bar{X}_n \equiv (1/n) \sum_{i=1}^n X_i$, then $Q\bar{X}_n \to 0$. Since $X_1 - X_2$ is symmetric, we have from the quadratic form equation and 2° that $2 \text{ Var } X = \text{Var}(X_1 - X_2) = Q^2(X_1 - X_2) = 2[Q^2(X) - C(X_1, X_2)]$, i.e., that $Q^2(X) = \text{Var } X + C(X_1, X_2)$. On the other hand, $Q^2(\bar{X}_n) = (1/n)Q^2(X) + (1 - 1/n)C(X_1, X_2)$ and hence $Q^2(\bar{X}_n) = (1/n)\text{Var } X + C(X_1, X_2)$ and

$$0 \le C(X_1, X_2) = \lim_{n\uparrow\infty} \downarrow Q^2(\bar{X}_n) < \infty.$$

Therefore $Q^2(X) \ge \text{Var } X$ in general, with equality if Q preserves the LLN.

4°. It remains to establish that a spread monotone Q preserves the LLN. For a given $X \in L_2$, we construct a symmetric variable $Z >_s X$ for which $\text{Var } Z \le 4 \text{ Var } X$. Performing this construction for each $\bar{X}_n$ we obtain Z_n with

$$Q^2(\bar{X}_n) \le Q^2(Z_n) = \text{Var } Z_n \le 4\text{Var } \bar{X}_n = \frac{4}{n} \text{Var } X \to 0.$$

For the construction, let F^{-1} denote the left continuous inverse cdf for X and define

$$Z = F^{-1}(1 - U) - F^{-1}(U),$$

where U is uniformly distributed on $(0, 1)$. Symmetry and the variance bound are clear. Write G^{-1} for the inverse cdf for Z; then $G^{-1}(u) = F^{-1}(u) - F^{-1}(1 - u)$ off of the at most countable discontinuity set of F^{-1}. From this and left continuity of the inverses follows

$$\begin{aligned} G^{-1}(v) - G^{-1}(u) &= F^{-1}(v) - F^{-1}(1 - v) - F^{-1}(u) + F^{-1}(1 - u) \\ &\ge F^{-1}(v) - F^{-1}(u), \quad 0 < u < v < 1, \end{aligned}$$

which shows that $X <_s Z$ and completes the proof of Theorem 5.

REMARK 5.1 (a). We note that a scale functional can possess one of the properties of dispersion and spread monotonicity and lack the other. Bickel and Lehmann (1978, Example 11) give an example of a functional which is spread monotone but not dispersion monotone. On the other hand, $QZ = \text{med}(|Z -$

$EZ|$) is dispersion monotone but not spread monotone. For example, let $\mathscr{L}(X) = (1/3)(\delta_{(-4)} + \delta_0 + \delta_{(+4)})$ and

$$\mathscr{L}(Y) = \frac{1}{12}(4\delta_{(-8)} + 4\delta_{(-3)} + 3\delta_{(+3)} + \delta_{(+35)});$$

then $X <_s Y$ but $QX = 4 > 3 = QY$.

(b) Theorem 5 is also valid (with a simpler, more standard proof) if the monotonicity conditions are replaced by the requirement that Q be L_2-continuous, or more generally that $QX_n \to Q0 = 0$ whenever $EX_n = 0$ for all n and $\operatorname{Var} X_n \downarrow 0$. See Fill and Johnstone (1982, page 28), who also discuss the difficulties encountered in attempting to prove Theorem 5 via Riesz representation methods.

C. PROOF OF THEOREM 6. Let $\mathbf{X}$ be a given random vector of dimension $p \geq 2$, and $\mathscr{Q} = \mathscr{Q}_{\mathbf{X}}$ the class of scale functionals with domain at least $\{\mathscr{L}(\mathbf{a}'\mathbf{X})$: $\mathbf{a} \in \mathbb{R}^p\}$. For each $Q \in \mathscr{Q}$ let $S_Q(\mathbf{X})$ be the PP dispersion measure for $\mathbf{X}$.

For the proof of Theorem 6 we shall require a scale functional $R \in \mathscr{Q}$ which is *nondegenerate*, in the sense that Z of the form $\mathbf{a}'\mathbf{X}$ is degenerate if $RZ = 0$. If the expected squared length of $\mathbf{X}$ is finite, then R = standard deviation will do. To discuss the general case, define the "α-spread"

$$Q_\alpha Z = (1/2)(G_{1-\alpha}Z - G_\alpha Z)$$

for $\alpha \in (0, 1/2)$; the α-quantile function G_α was defined precisely in Section 4C. Let $\mathscr{M}$ be the linear space of vectors $\mathbf{a}$ for which $\mathscr{L}(\mathbf{a}'\mathbf{X})$ is degenerate. Now $Q_\alpha^2(\mathbf{a}'\mathbf{X})$ decreases as α increases, and had a quadratic form representation $\mathbf{a}'\Sigma_\alpha\mathbf{a}$ in view of the affine equivariance of S_{Q_α}. It follows that $\mathscr{M}_\alpha = \{\mathbf{a}: Q_\alpha(\mathbf{a}'\mathbf{X}) = 0\}$ form a monotone family of linear spaces with intersection equal to $\mathscr{M}$; the linearity of $\mathscr{M}_\alpha$ is a consequence of the triangle inequality for Q_α. Hence for $\alpha_0 > 0$ sufficiently small $\mathscr{M}_{\alpha_0} = \mathscr{M}$, and $\mathbf{a}'\mathbf{X}$ is degenerate if $Q_{\alpha_0}(\mathbf{a}'\mathbf{X}) = 0$. Thus $R \equiv Q_{\alpha_0} \in \mathscr{Q}$ is nondegenerate.

One could more simply choose

$$RZ = \begin{cases} 0, & \text{if } \mathscr{L}(Z) \text{ is degenerate} \\ \operatorname{med}(|Z - \operatorname{med} Z| \mid \{|Z - \operatorname{med} Z| > 0\}), & \text{otherwise.} \end{cases}$$

Note, however, that each Q_α is a very regular functional, whereas R is not even dispersion monotone. Theorem 6 is clearly strengthened if we require the functionals $Q \in \mathscr{Q}$ to be as regular as possible.

In the proof below we drop the unnatural assumption that $\mathbf{X}$ have finite expectation and show instead of elliptical contouring of $\mathbf{X}$ that

$$\mathbf{a}'\mathbf{X} =_d (\mathbf{a}'\Sigma\mathbf{a})^{1/2}W + c(\mathbf{a}), \quad \mathbf{a} \in \mathbb{R}^p$$

with $c(\mathbf{a}) \in \mathbb{R}$, W symmetric about zero, and Σ nonnegative definite.

Of course, if $\boldsymbol{\mu} = E\mathbf{X} \in \mathbb{R}^p$ exists, then $c(\mathbf{a}) \equiv \mathbf{a}'\boldsymbol{\mu}$ and $\mathbf{X}$ has an elliptically contoured distribution. More generally, if $\mathscr{L}(W)$ satisfies the weak law of large numbers (see Feller, 1971, XVII.2a for equivalent conditions), or if $\mathbf{X}$ is centrosymmetric about some $\boldsymbol{\mu} \in \mathbb{R}^p$ (in particular, if the hypotheses of Remark 4.1 are satisfied), then $\mathbf{X}$ is elliptically contoured about μ. Indeed it is natural to conjecture that the (apparently) weaker conclusion above forces $\mathbf{X}$ to be elliptically contoured without any further assumptions.

We begin the proof of Theorem 6 with a reduction to the full rank, spherically symmetric case. Let $R \in \mathscr{Q}$ be the nondegenerate scale functional obtained above. Consider $\mathbf{Y} = V^{-1}\mathbf{X}$, where the nonsingular matrix $V \in \mathbb{R}^{p\times p}$ satisfies $VPV' = S_R(\mathbf{X})$, with $P = \operatorname{diag}\{1, \cdots, 1, 0, \cdots, 0\}$ having rank r equal to the rank of $S_R(\mathbf{X})$. Partitioning $\mathbf{Y}' = (\mathbf{Y}_1', \mathbf{Y}_2')$ and $\mathbf{a}' = (\mathbf{a}_1', \mathbf{a}_2')$ correspondingly gives $R^2(\mathbf{a}_1'\mathbf{Y}_1) = \mathbf{a}_1'\mathbf{a}_1$ and $R^2(\mathbf{a}_2'\mathbf{Y}_2) = 0$ for all $\mathbf{a} \in \mathbb{R}^p$. Since R is nondegenerate we conclude that $\mathbf{Y}_2$ is degenerate. In particular, if $r \le 1$ then it is clear that $\mathbf{X}$ is one-dimensional.

Replacing $\mathbf{X}$ by $\mathbf{Y}_1$ we may now suppose that $p \ge 2$ and $R^2(\mathbf{a}'\mathbf{X}) = \|\mathbf{a}\|^2$ for all $\mathbf{a} \in \mathbb{R}^p$ and must show that

$$\mathbf{a}'\mathbf{X} =_d X_1 + c(\mathbf{a}), \quad \mathbf{a} \in \mathbb{S}^{p-1}$$

where $\mathbb{S}^{p-1}$ is the unit sphere in $\mathbb{R}^p$. The main step is to prove that affine equivariance forces all projections $\mathbf{a}'\mathbf{X}$ to be affinely equivalent; to wit, that $\mathbf{a}'\mathbf{X} =_d b(\mathbf{a})X_1 + c(\mathbf{a})$ for $\mathbf{a} \in \mathbb{S}^{p-1}$. Assuming this for the present, it then follows by equating R-values that $b(\mathbf{a}) = \pm 1$. The mapping $a \mapsto \mathscr{L}(\mathbf{a}'\mathbf{X})$ on the connected set $\mathbb{S}^{p-1}$ is continuous in the topology of weak convergence and takes values in both of the closed sets $\{\mathscr{L}(X_1 + c)\colon c \in \mathbb{R}\}$ and $\{\mathscr{L}(-X_1 + c)\colon c \in \mathbb{R}\}$. The intersection of these two sets is therefore nonempty, so that $\mathscr{L}(X_1)$ is symmetric and we may take $b(\mathbf{a}) \equiv 1$.

To show the affine equivalence of all $\mathbf{a}'\mathbf{X}$, we suppose that some $\mathbf{a}'\mathbf{X} \not\sim X_1$ and deduce from the convergence of types theorem the existence of $\gamma > 0$ such that $X_1 \not\sim \mathbf{a}(\beta)'\mathbf{X}$ for $\beta = 0, \frac{1}{2}, \gamma, \gamma$; here

$$\mathbf{a}(\beta) \equiv (1 - \beta)\mathbf{a} + \beta\mathbf{e}_1.$$

Affine equivariance requires that

$$\begin{aligned} Q^2(X_1) &= Q^2(\gamma^{-1}\mathbf{a}(\gamma)'\mathbf{X} - \gamma^{-1}(1-\gamma)\mathbf{a}'\mathbf{X}) \\ &= \gamma^{-2}Q^2(\mathbf{a}(\gamma)'\mathbf{X}) + \gamma^{-2}(1-\gamma)^2Q^2(\mathbf{a}'\mathbf{X}) \\ &\quad -\gamma^{-2}(1-\gamma)[4Q^2(\mathbf{a}(\gamma/2)'\mathbf{X}) - Q^2(\mathbf{a}(\gamma)'\mathbf{X}) - Q^2(\mathbf{a}'\mathbf{X})], \end{aligned}$$

but it is an easy matter to construct a scale functional Q violating this condition. This contradiction shows that all projections $\mathbf{a}'\mathbf{X}$ must be equivalent, and completes the proof.

REMARK 5.2. It is not clear how to construct the violating Q so as to be regular, say, dispersion and spread monotone. However, if the hypotheses of Theorem 6 are strengthened by the requirement that $S_Q(\mathbf{X}) = \Sigma_Q$ depend on Q only through a multiplicative scalar, there is a direct proof of Theorem 6 (Fill and Johnstone, 1982, page 32) which uses only the class $\mathscr{Q}_{ds}$ of dispersion and spread monotone functionals Q and avoids the contradiction argument.

Acknowledgments. We thank David Donoho for insightful comments which make up much of Section 3C, Doug Critchlow for providing us with his (prior) proof of Theorem 2, and Persi Diaconis, Jerry Friedman, and a referee for helpful discussions and suggestions.

REFERENCES

ANDREWS, D. F., *et al.* (1972). *Robust estimates of location: survey and advances.* Princeton University Press, Princeton, N.J.

ANDERSON, T. W. and FANG, K. T. (1982). On the theory of multivariate elliptically contoured distributions and their applications. Techical Report No. 54, Department of Statistics, Stanford University.

BICKEL, P. and LEHMANN, E. L. (1976). Descriptive statistics for nonparametric models, III Dispersion. *Ann. Statist.* **4** 1139–1158.

BICKEL, P. and LEHMANN, E. L. (1979). Descriptive statistics for nonparametric models, IV Spread. In *Contributions to Statistics: Jaroslav Hájek Memorial Volume* 33–40. Editor: J. Jurecková. D. Reidel, Dordrecht, Holland.

BREIMAN, L. (1968). *Probability.* Addison, Reading, Mass.

CHEN, Z. and LI, G. (1981). Robust principal components and dispersion matrices via projection pursuit. Research Report PJH-8, Department of Statistics, Harvard University.

CRITCHLOW, D. (1981). Unpublished manuscript.

DONOHO, D. (1982). Breakdown properties of multivariate location estimators. Technical Report, Department of Statistics, Harvard University.

FELLER, W. (1971). *An Introduction to Probability Theory and its Applications, II. 2nd ed.* Wiley, New York.

FILL, J. and JOHNSTONE, I. (1982). On projection pursuit measures of multivariate location and dispersion. Technical Report No. 12, Orion Project, Stanford University.

FRIEDMAN, J. H. and STUETZLE, W. (1980a). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76** 817–823.

FRIEDMAN, J. H. and STUETZLE, W. (1980b). Projection pursuit classification. Unpublished manuscript.

FRIEDMAN, J. H. and STUETZLE, W. (1981). Projection pursuit density estimation. Technical Report No. Orion 002, Department of Statistics, Stanford University.

FRIEDMAN, J. H. and TUKEY, J. W. (1974). A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. Computers* **C-23** 881–889.

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

HUBER, P. J. (1981b). Projection pursuit. Research Report PJH-6, Department of Statistics, Harvard University.

KRUSKAL, J. B. (1972). Linear transformation of multivariate data to reveal clustering. In *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences, Vol. 1* Theory 181–191. Editors: R. N. Shepard, A. K. Romney, and S. B. Nerlov. Seminar Press, New York.

LI, G. and CHEN, Z. (1981). Robust projection pursuit estimator for dispersion matrices and principal components. Research Report PJH-11, Department of Statistics, Harvard University.

STAHEL, WERNER A. (1981). Breakdown of covariance estimators. Research Report No. 31, Fachgruppe für Statistik, Eidgenössische Technische Hochschule, Zürich.

SWITZER, P. (1970). Numerical classification. In *Geostatistics* 31–43. Editor: D. F. Merriam. Plenum Press, New York.

SWITZER, P. and WRIGHT, R. (1971). Numerical classification applied to certain Jamaican eocene nummulitids. *Mathematical Geology* **3** No. 3, 297–311.

DEPARTMENT OF STATISTICS
SEQUOIA HALL
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305