*The Annals of Statistics*
1991, Vol. 19, No. 3, 1639–1650

# GENERALIZATIONS OF JAMES–STEIN ESTIMATORS UNDER SPHERICAL SYMMETRY

By Ann Cohen Brandwein and William E. Strawderman[1]

*Baruch College and Rutgers University*

This paper is primarily concerned with extending the results of Stein to spherically symmetric distributions. Specifically, when $X \sim f(\|X - \theta\|^2)$, we investigate conditions under which estimators of the form $X + ag(X)$ dominate $X$ for loss functions $\|\delta - \theta\|^2$ and loss functions which are concave in $\|\delta - \theta\|^2$. Additionally, if the scale is unknown we investigate estimators of the location parameter of the form $X + aVg(X)$ in two different settings. In the first, an estimator $V$ of the scale is independent of $X$. In the second, $V$ is the sum of squared residuals in the usual canonical setting of a generalized linear model when sampling from a spherically symmetric distribution. These results are also generalized to concave loss.

The conditions for domination of $X + ag(X)$ are typically (a) $\|g\|^2 + 2\nabla \circ g \le 0$, (b) $\nabla \circ g$ is superharmonic and (c) $0 < a < 1/pE_0(1/\|X\|^2)$, plus technical conditions.

**1. Introduction.** This paper is concerned with estimating a location vector perhaps in the presence of an unknown scale parameter. The underlying distributions for the location estimator are assumed to be spherically (or elliptically) symmetric. The loss functions are either quadratic, or a concave function of quadratic loss. Information about the scale parameter is either independent of the location information, or in the form of the squared norm of a residual vector. The estimators considered are of the form $X + ag(X)$, where $g$ satisfies the inequality $\|g\|^2 + 2\nabla \circ g \le 0$. In the unknown scale case where $V$ is the estimator of scale, the estimators are of the form $X + aVg(X)$.

The paper is concerned with extending the results of Stein (1981) to the spherically symmetric case and the case of concave loss.

Specifically, we show that the estimator $X + ag(X)$ dominates $X$ for quadratic loss $\|\delta - \theta\|^2$, where $X$ has a spherically symmetric density $f(\|X - \theta\|^2)$, when satisfying (a) $\|g\|^2 + 2\nabla \circ g \le 0$, (b) $\nabla \circ g$ is superharmonic and (c) $0 < a < 1/pE_0(1/\|X\|^2)$, plus technical conditions. Stein's (1981) result for a normal distribution with identity covariance is equivalent to (a) and (c) $0 < a < 1$. Since, in this case $E_0(1/\|X\|^2) = 1/(p - 2)$, our result gives $0 < a < (p - 2)/p$.

The present results then, for the additional price of superharmonicity of $\nabla \circ g$ and a slight reduction of the range of shrinkage values, extend the class

Received August 1989; revised July 1990.

[1]Research supported by NSF Grant DMS-88-22622.

*AMS* 1980 *subject classifications*. 62C99, 62F10, 62H99.

*Key words and phrases*. Spherical symmetry, minimaxity, squared error loss, concave loss, James–Stein estimation, superharmonic, location parameters.

of distributions for which such estimators dominate $X$. Section 2 is concerned with these results.

We also consider extensions in two other directions. First, we consider loss functions which are monotone-concave functions of squared error and show general dominance results for estimators of the form $X + ag(X)$. These results then are extensions of both Stein (1981) and Brandwein and Strawderman (1980). Results of this type are developed in Section 3, where we also consider elliptically symmetric distributions and general quadratic loss.

Second, we consider the case of unknown scale in two different settings. In Section 4 we consider the case where an estimator of scale is available which is independent of $X$. In Section 5 we consider the canonical form of a general linear model in which case the estimator of scale is the squared norm of the residual vector and hence is not generally independent of the estimator of the mean. In both these cases we consider estimators of the form $X + aVg(X)$, where $V$ is the estimate of scale. The loss functions considered include concave functions of squared error loss.

Hence, the results are quite general from the perspective of underlying distribution, estimation procedure and loss function.

The risk function of an estimator $X + ag(X)$ is given by

$$E\|X + ag(X) - \theta\|^2 = E_\theta\|X - \theta\|^2 + a^2\|g(X)\|^2 + 2aE_\theta(X - \theta)'g(X). \tag{1.1}$$

The main difficulty in proving dominance results is handling the cross-product term $E_\theta(X - \theta)'g(X)$. Stein's (1981) beautiful solution in the normal case is to use integration by parts to show $E_\theta(X - \theta)'g(X) = E_\theta \nabla \circ g(X)$. The main idea in the present paper is to evaluate (1.1) conditionally on $\|X - \theta\| = R$ and to apply the divergence theorem to the cross-product term. This gives an expression of the form

$$\begin{aligned} E[(X - \theta)'g(X) \mid \|X - \theta\| = R] &= \frac{R}{A(S)} \oint_S [g(V + \theta)]' \left(\frac{V}{R}\right) dA(V) \\ &= \frac{R}{A(S)} \int_B \nabla \circ g(\theta + V)\, dM(V), \end{aligned}$$

where $S$ and $B$ are the sphere and ball of radius $R$ respectively and $dA$ and $dM$ are differential "surface area" and "volume" elements. This device gives the conditional expected value of the cross-product in terms of $\nabla \circ g$ as in Stein; however, the expected value is now over the ball ("volume") instead of the sphere ("surface area"). Superharmonicity of $\nabla \circ g$ (or $\|g\|^2$ or some function related to $g$) is then used to obtain an inequality relating the conditional expected value on the ball to the conditional expected value on the sphere. The basic fact here is that if $h(X)$ is superharmonic, its average over the ball ("volume") is greater than its average over the sphere ("surface area"). These ideas are more fully developed in succeeding sections.

It is worth noting that the results using the techniques here typically require at least four dimensions instead of three. This is related to the fact that if $X \sim \mathscr{U}\{\|X - \theta\| = R\}$, the James–Stein type estimators will not improve $X$ unless the dimension is at least four [see Brandwein (1979)].

**2. Minimax estimators with respect to quadratic loss.** Consider a $p \times 1$ random vector $X = [X_1, X_2, \dots, X_p]'$ having a $p$-dimensional spherically symmetric distribution about $\theta$ [$X \sim$ s.s.($\theta$)]. The problem of estimating the mean vector $\theta$ with respect to quadratic loss

$$L(\delta, \theta) = \|\delta - \theta\|^2 = \sum_{i=1}^{p} (\delta_i - \theta_i)^2, \tag{2.1}$$

where $\delta = [\delta_1, \delta_2, \dots, \delta_p]'$ and $\theta = [\theta_1, \theta_2, \dots, \theta_p]'$ has been studied by many [see Brandwein and Strawderman (1990) for a review and references].

In this section we will look at estimators of the form $\delta_a(X) = X + ag(X)$, when $X$ is one observation from a s.s.($\theta$) distribution, and the loss is (2.1).

The following lemma, which is necessary for the main theorems of this section, is well known and follows immediately from du Plessis (1970), page 54.

Lemma 2.1. *If $h(X)$ is a superharmonic function* [$\nabla^2 h(X) = \Sigma(d/dX_i^2)h(X) \le 0$] *and $X$ has a uniform distribution on the sphere centered at $\theta$ with radius $R$, denoted $X \sim \mathscr{U}\{\|X - \theta\|^2 = R^2\}$, then*

$$E_\theta h(X) \le E_\theta h(Y),$$

*where $Y \sim \mathscr{U}\{\|Y - \theta\|^2 \le R^2\}$. That is,*

$$\frac{1}{A(S)} \oint_S h(X)\, dA(X) \le \frac{1}{M(B)} \int_B h(Y)\, dM(Y),$$

*where $A(S)$ and $M(B)$ represent the areas of the sphere and ball, respectively.*

The importance of this lemma in the present study stems from the fact that if $X \sim$ s.s.($\theta$), then $X|R \sim \mathscr{U}\{\|X - \theta\|^2 = R^2\}$.

The following theorem presents minimax estimators which are better than $X$ with respect to quadratic loss.

Theorem 2.1. *If $X \sim$ s.s.($\theta$) and $\delta_a(X) = X + ag(X)$, then with respect to quadratic loss* (2.1), *$\delta_a(X)$ has smaller risk than $X$ provided*:

(i) $\|g\|^2/2 \le -h \le -\nabla \circ g$,

(ii) $-h$ *is superharmonic and $E_\theta[R^2 h(W)]$ is a nonincreasing function in $R$, where $W \sim \mathscr{U}\{\|W - \theta\|^2 \le R^2\}$ and*

(iii) $0 < a \le 1/pE_0(1/\|X\|^2)$.

PROOF.

$$\begin{aligned} R(\delta_a, \theta) &= E_\theta \|\delta_a - \theta\|^2 \\ &= E_\theta \|(X - \theta) + ag(X)\|^2 \\ &= E_\theta\left[R^2 + a^2\|g(X)\|^2 + 2a(X-\theta)'g(X)\right] \\ &= E\left[R^2 + a^2 E_0\left[\|g(V+\theta)\|^2 \middle| R\right] + 2aE_0\left[V'g(V+\theta)\middle|R\right]\right], \end{aligned}$$

where $R = \|X - \theta\|$, $V = (X - \theta) \sim$ s.s.(0) and thus $V|R \sim \mathscr{U}\{\|V\|^2 = R^2\}$.

By the divergence theorem and assumption (i),

$$\begin{aligned} E_0[V'g(V+\theta)|R] &= \frac{R}{A(S)} \oint_S g'(V+\theta) \circ \frac{V}{R}\, dA(V) \\ &= \frac{M(B)}{A(S)} R \int_B \frac{1}{M(B)} \nabla \circ g(V+\theta)\, dM(V) \\ &= \frac{R^2}{p} \int_B \frac{1}{M(B)} \nabla \circ g(V+\theta)\, dM(V) \\ &\le \frac{R^2}{p} \int_B \frac{1}{M(B)} h(V+\theta)\, dM(V), \end{aligned}$$

where $B$ is the ball of radius $R$ centered at the origin.

Moreover,

$$\begin{aligned} E_0\left[\|g(V+\theta)\|^2|R\right] &\le -2E_0[h(V+\theta)|R] \\ &= -2\oint_S \frac{1}{A(S)} h(V+\theta)\, dA(V) \\ &\le -2\int_B \frac{1}{M(B)} h(V+\theta)\, dM(V). \end{aligned}$$

The first inequality follows from assumption (i); the second inequality is true by assumption (ii) and Lemma 2.1. Therefore,

$$\begin{aligned} R(\delta_a, \theta) &\le E\left[R^2 + \left(-2a^2 + \frac{2aR^2}{p}\right) E_0[h(V+\theta)|R]\right] \\ &= E\left[R^2 + \left(\frac{-2a^2}{R^2} + \frac{2a}{p}\right) E_\theta[R^2 h(W)|R]\right], \end{aligned}$$

where $W|R \sim \mathscr{U}[\|W - \theta\|^2 \le R^2\}$. Clearly, by assumption (ii), $E_\theta[R^2 h(W)|R]$

is a nonincreasing function of $R$, and since $(-2a^2/R^2 + 2a/p)$ is nondecreasing in $R$, we have

$$R(\delta_a, \theta) - R(X, \theta) = R(\delta_a, \theta) - ER^2$$
$$\le 2a\left(\frac{1}{p} - aE\left(\frac{1}{R^2}\right)\right)E\left[E_\theta[R^2 h(W)|R]\right].$$

Since $h \le 0$, $R(\delta_a, \theta) - R(X, \theta) \le 0$ provided

$$0 < a \le \frac{1}{pE(1/R^2)} = \frac{1}{pE_0(1/\|X\|^2)}. \qquad \square$$

COMMENT 2.1. If $g(\cdot)$ is homogeneous of degree $-1$, $h$ can typically be chosen to be homogeneous of degree $-2$. The monotonicity part of condition (ii) may then be replaced by $E_\theta h(W)$ is nondecreasing in $\theta$ for $R = 1$. If $-h$ is unimodal (either globally or one variable at a time), this monotonicity will follow from Anderson's theorem.

We now indicate some applications.

EXAMPLE 2.1 (James–Stein estimators). Theorem 2.1 and Comment 2.1 give a brief and elegant proof of Brandwein's (1979) result on minimaxity of James–Stein estimators in the spherically symmetric case. Here

$$\delta_a(X) = X + a\left(\frac{-b}{X'X}\right)X \tag{2.2}$$

so that $g(X) = -b/X'X$, where $b$ must be chosen to satisfy $0 < b \le p - 2$ in order for condition (i) of Theorem 2.1 to hold. It is easily seen that $g(X)$ is homogeneous of degree $-1$, and that $-\nabla \circ g$ is unimodal and superharmonic if $p \ge 4$. Hence choosing $h = \nabla \circ g$, it follows that $\delta_a(X)$ dominates $X$ for

$$0 < ab < \frac{2(p-2)}{p}\frac{1}{E_0(1/\|X\|^2)}.$$

In addition, using Theorem 2.1 when $0 < r(\cdot) \le 1$ and $r(\|X\|^2)$ is nondecreasing and concave, the estimator

$$\delta(X) = \left(1 - \frac{abr(\|X\|^2)}{\|X\|^2}\right)X \tag{2.3}$$

is minimax for $0 < ab \le 2(p-2)/p(1/E_0(1/\|X\|^2))$.

EXAMPLE 2.2 (Nonspherical shrinkage estimators).

$$\delta_a(X) = X + a\left(\frac{-bA}{X'BX}\right)X, \tag{2.4}$$

where $A$ and $B$ are positive definite matrices.

Here we use Theorem 2.1 and Comment 2.1 with $g(X) = (-bA/X'BX)X$. Note that

$$-\|g\|^2 = \frac{-b^2X'A'AX}{(X'BX)^2} \geq \frac{-b^2a_L^2}{b_M}\frac{1}{X'BX} = 2h, \tag{2.5}$$

where $b_M$ is the minimum eigenvalue of $B$ and $a_L$ is the maximum eigenvalue of $A$. Clearly, $g$ is homogeneous of degree $-1$, and it is easily shown that $-h$ is superharmonic if $\operatorname{tr} B \geq 4b_L$, where $b_L$ is the maximum eigenvalue of $B$. An easy calculation shows that it is possible to choose $b$ to satisfy condition (i) of Theorem 2.1. It follows that $\delta_a$ is minimax provided

$$0 < ab \leq 2\frac{(\operatorname{tr} A - 2a_L)b_M}{pE_0(1/\|X\|^2)a_L^2} \tag{2.6}$$

and $\operatorname{tr} A > 2a_L$.

It is interesting to note that we may obtain a different result using

$$2h = \frac{-b^2a_L^2}{b_M^2}\frac{1}{X'X}. \tag{2.7}$$

Here the condition for minimaxity becomes

$$0 < ab \leq 2\frac{(\operatorname{tr} A - 2a_L)}{pE_0(1/\|X\|^2)}\frac{b_M^2}{b_La_L^2}.$$

Although this is a smaller bound than (2.6), it holds for $p \geq 4$ regardless of whether $\operatorname{tr} B - 4b_L \geq 0$.

To our knowledge, this is the first general minimaxity result for estimators of the form (2.4) with $B$ not equal to a multiple of the identity matrix. In any event, the ease of proving minimaxity in this example illustrates the utility of this approach.

EXAMPLE 2.3 ("Limited translation rule" for spherically symmetric distributions). Suppose we consider the "limited translation" rule based on order statistics given by Stein (1981), for $X \sim \text{s.s.}(\theta)$ with respect to quadratic loss (2.1).

For $k$ a positive integer, let $\delta(X) = X + ag(X)$, where

$$g_i(X) = \begin{cases} \dfrac{-bX_i}{\Sigma\left(X_j^2 \wedge Z_{(k)}^2\right)}, & \text{if } |X_i| \leq Z_{(k)}, \\[2ex] \dfrac{-b}{\Sigma\left(X_j^2 \wedge Z_{(k)}^2\right)} Z_{(k)} \operatorname{sgn} X_i, & \text{if } |X_i| > Z_{(k)}, \end{cases} \tag{2.8}$$

where $Z_i = |X_i|$ and $Z_{(1)} < Z_{(2)} < \cdots < Z_{(p)}$ are the order statistics and $c \wedge d = \min(c, d)$. It can be checked that this example satisfies the conditions of

Theorem 2.1 and so, $\delta(X)$ is minimax if

$$0 < ab \le \frac{2(k-2)}{p} \frac{1}{E_0(1/\|X\|^2)} \quad \text{provided } k \ge 4.$$

**3. Minimax estimators with respect to nonquadratic loss functions.** When considering estimators of the James–Stein type for spherically symmetric distributions, there are two major lines of development relating to generalizations of quadratic loss.

The first is to consider general quadratic loss given by

$$L(\delta, \theta) = (\delta - \theta)' D(\delta - \theta), \tag{3.1}$$

where $D$ is a given $p \times p$ positive-definite matrix. This problem was considered in Brandwein (1979).

The second relates to nonquadratic loss of the form

$$L(\delta, \theta) = f(\|\delta - \theta\|^2), \tag{3.2}$$

where $f(\cdot)$ is a nondecreasing concave function. Brandwein and Strawderman (1980) and Bock (1985) have results for losses of this form.

For the general estimator $\delta_a(X) = X + ag(X)$, Chou and Strawderman (1990) have studied the case when $X$ has a distribution which is a mixture of normals and the loss is general quadratic loss (3.1).

Here is our result for loss (3.1).

THEOREM 3.1. *If $X \sim$ s.s.$(\theta)$ and the loss is* (3.1), *then $\delta_a(X) = X + ag(X)$ has smaller risk than $X$ provided the conditions of Theorem* 2.1 *hold with* (i) *replaced by*

(i′) $g'Dg/2 \le -h \le \nabla \circ Dg$.

EXAMPLE 3.1 (An extension of Example 2.2 to general quadratic loss). Consider $\delta_a(X)$ defined by (2.4) and $2h = (-a_L^2 d_L/b_M)(1/X'BX)$ analogous to (2.5). If $\operatorname{tr} B > 4b_L$ and $\operatorname{tr} AD > 2a_L d_L$, then $\delta_a$ is minimax for $0 < ab \le (\operatorname{tr} AD - 2a_L d_L)b_M/pE_0(1/\|X\|^2)a_L^2$. Moreover, if we use $2h = (-a_L^2 d_L/b_M^2)(1/X'X)$, analogous to (2.7), we have minimaxity for $p \ge 4$ when

$$0 < ab \le \frac{(\operatorname{tr} AD - 2a_L d_L)b_M^2}{pE_0(1/\|X\|^2)a_L^2 b_L}.$$

COMMENT 3.1. Note that when $X$ has an elliptical distribution about $\theta$ there is a natural extension of Theorem 3.1 which can be proved by standard invariance arguments.

We now look at generalizations of the results of Section 2 when the loss is a nondecreasing concave function of quadratic loss given by (3.2).

THEOREM 3.2. *Suppose for the* $p \times 1$ *random vector* $X \sim$ s.s.$(\theta)$ *conditions* (i) *and* (ii) *of Theorem* 2.1 *are satisfied, where* $\delta_a(X) = X + ag(X)$ *and the loss is given by* (3.2). *Then* $\delta_a(X)$ *is better than* $X$ *for* $0 < a \leq 1/pE_H(1/R^2)$, *where* $H(R) = \int_0^R f'(S^2)\,dG(S)/\int_0^\infty f'(S^2)\,dG(S)$, $G$ *is the cumulative distribution function of* $R$, $E_H$ *denotes the expected value when* $R$ *has cdf* $H$ *and* $0 < E_G f'(R^2) < \infty$.

PROOF. If $\Delta(X) = \|\delta_a(X) - \theta\|^2 - \|X - \theta\|^2$ then, as shown by Brandwein and Strawderman (1980), $E_\theta f(\|\delta_a(X) - \theta\|^2) - E_\theta f(\|X - \theta\|^2) \leq E_\theta f'(R^2)\Delta(X)$.

The result then follows from the proof of Theorem 2.1 with respect to the new distribution on $R$. □

COMMENT 3.2. Note that for loss (3.2), Examples 2.1, 2.2 and 2.3 will work. As an indication of the possible amount of shrinkage in the case when the loss is $L(\delta, \theta) = \|\delta - \theta\|$ and $X \sim MVN(\theta, I)$, $E_H(1/R^2) = 1/(p-3)$. In all of these examples, the ratio of the maximum shrinkage factors is $(p-3)/(p-2)$.

## 4. The unknown scale case with an independent estimate of scale.

Suppose the $p \times 1$ random vector $X$ has a density $(1/\sigma^p) f_x(\|x - \theta\|^2/\sigma^2)$, where $\sigma$ is unknown, and consider the random variable $V$, with density $(1/\sigma^2) f_v(v/\sigma^2)$, independent of $X$. In this section we will find minimax estimators of $\theta$ which are better than $X$ with respect to three types of loss functions: a scaled quadratic loss

$$L(\delta, \theta) = \|\delta - \theta\|^2/\sigma^2, \tag{4.1}$$

nondecreasing concave functions of this scaled quadratic loss and a general scaled quadratic loss

$$L_D(\delta, \theta) = (\delta - \theta)' D (\delta - \theta)/\sigma^2. \tag{4.2}$$

Bravo and MacGibbon (1988) have given results for this problem in the "variance mixture of normals" case for the loss $\|\delta - \theta\|^2/\sigma^2$.

For the general spherically symmetric problem, we will consider estimators similar in form to those in Sections 2 and 3. Specifically, estimators

$$\delta_{a,V}(X) = X + aVg(X).$$

These estimators will dominate $X$ not only for the scaled losses just given, but will also dominate $X$ for quadratic loss $\|\delta - \theta\|^2$ since $E_\theta(\|\delta - \theta\|^2) = \sigma^2 E_\theta(\|\delta - \theta\|^2/\sigma^2)$.

We now present without proof the conditions under which $\delta_{a,V}(X)$ dominates $X$ and is minimax.

THEOREM 4.1. *Suppose $X$, a $p \times 1$ random vector, has a density $(1/\sigma^p) f_x(\|x - \theta\|^2/\sigma^2)$, where $\sigma$ is unknown. Moreover, suppose the random vector $V$ has a density $(1/\sigma^2) f_v(v/\sigma^2)$ and $X$ and $V$ are independent. If $\delta_{a,V}(X) = X + aVg(X)$, then with respect to scaled quadratic loss* (4.1), *$\delta_{a,V}(X)$ dominates $X$ provided $g$ satisfies conditions* (i) *and* (ii) *of Theorem* 2.1 *and*

(iii)
$$0 < a \leq \frac{1}{pE_{0,\sigma=1}(1/\|X\|^2)} \left[ \frac{E_{\sigma=1}V}{E_{\sigma=1}V^2} \right].$$

A similar result holds for nondecreasing concave functions of scaled quadratic loss (4.1) and general scaled quadratic loss (4.2).

**5. The unknown scale case with scale estimated from the residual vector.** Consider the problem of estimating the mean vector $\theta = [\theta_1, \theta_2, \dots, \theta_p]'$ when $X$, a $p \times 1$ random vector, and $U$, an $m \times 1$ random vector, are distributed such that $X^* = [X_1, X_2, \dots, X_p, U_1, U_2, \dots, U_m]'/\sigma$ has a spherically symmetric distribution about $\theta^* = [\theta_1, \theta_2, \dots, \theta_p, 0, 0, \dots, 0]'$. We say $X^* \sim \text{s.s.}(\theta^*, \sigma^2 I)$.

The assumptions on the distribution of $X^*$ coincide with the canonical form of the general linear model [see Scheffé (1959)].

The improved estimators will be of the form $\delta_a(X^*) = X + aU'Ug(X)$, but unlike the estimators considered in Sections 2–4, the bounds of $\delta_a(X^*)$ will not depend on the distribution of $X^*$. This type of robustness phenomenon has been observed by Cellier, Fourdrinier and Robert (1988) for the James-Stein estimator.

As we discussed in the previous section, these improved estimators will dominate $X$ with respect to quadratic loss $\|\delta - \theta\|^2$ and general quadratic loss $(\delta - \theta)'D(\delta - \theta)$.

The following theorem makes similar assumptions to those of Theorem 2.1 about the improved estimators.

THEOREM 5.1. *Suppose $X$ is a $p \times 1$ random vector and $U$ is an $m \times 1$ random vector and $X^* = \begin{bmatrix} X \\ U \end{bmatrix} \sim \text{s.s.}(\theta^*, \sigma^2 I)$, where*

$$\theta^* = [\theta_1, \theta_2, \dots, \theta_p, 0, 0, \dots, 0]'$$

*and $\sigma^2$ is an unknown scale. If $\delta_a(X^*) = X + aU'Ug(X)$, then with respect to scaled quadratic loss $L(\delta, \theta) = \|\delta - \theta\|^2/\sigma^2$, $\delta_a(X)$ dominates $X$ provided conditions* (i) *and* (ii) *of Theorem* 2.1 *hold and*

(iii)
$$0 < a \leq \frac{1}{p} \frac{(p-2)}{m+2}.$$

PROOF.

$$\begin{aligned}
\Delta &= R(\delta_a, \theta) - R(X, \theta) \\
&= \frac{1}{\sigma^2} E_\theta\left[a^2 (U'U)^2 \|g(X)\|^2 + 2aU'U(X-\theta)' g(X)\right] \\
&= \frac{1}{\sigma^2} E_0\left[a^2 (U'U)^2 \|g(V+\theta)\|^2 + 2aU'UV' g(V+\theta)\right] \\
&\qquad\qquad\qquad\qquad (\text{where } V = X - \theta) \\
&= \frac{1}{\sigma^2} E\Big[ E\left[a^2 (U'U)^2 \|g(V+\theta)\|^2 \,\big|\, \|V\| = R, \|U\| = S\right] \\
&\qquad + 2aE\left[U'UV'g(V+\theta) \,\big|\, \|V\| = R, \|U\| = S\right]\Big].
\end{aligned}$$

By the divergence theorem,

$$\begin{aligned}
&E\left[U'Ug'(V+\theta)V \,\big|\, \|V\| = R\ h, \|U\ 2\| = S\right] \\
&= S^2 \frac{R^2}{A(S)} \oint_S g'(V+\theta) \frac{V}{R}\, dA(V) \\
&= S^2 \frac{R^2}{p} \frac{1}{M(B)} \int_B \nabla \circ g(V+\theta)\, dM(V) \\
&\le S^2 \frac{R^2}{p} \frac{1}{M(B)} \int_B h(V+\theta)\, dM(V),
\end{aligned}$$

where $B$ is the ball of radius $R$ centered at the origin.

Moreover, by assumption (i),

$$\begin{aligned}
&E\left[(U'U)^2 \|g(V+\theta)\|^2 \,\big|\, \|V\| = R, \|U\| = S\right] \\
&\le -2S^4 E\left[h(V+\theta) \,\big|\, \|V\| = R, \|U\| = S\right] \\
&= -2S^4 \oint_S \frac{1}{A(S)} h(V+\theta)\, dA(V) \\
&\le -2S^4 \int_B \frac{1}{M(B)} h(V+\theta)\, dM(V)
\end{aligned}$$

by Lemma 2.1, since $-h$ is superharmonic. Therefore,

$$\begin{aligned}
\Delta &\le \frac{2}{\sigma^2} E\left[\left(\frac{a^2 S^4}{R^2} - \frac{aS^2}{p}\right) G(R^2)\right] \\
&= \frac{2a}{\sigma^2} E\left[\left(\frac{a(T^2 - R^2)^2}{R^2} - \frac{(T^2 - R^2)}{p}\right) G(R^2)\right],
\end{aligned} \tag{5.1}$$

where $G(R) = E_\theta[R^2 h(W)]$ is a nondecreasing function of $R^2$ by assumption (ii) and $T^2 = R^2 + S^2$.

Let

$$H(R,T) = \frac{a(T^2 - R^2)^2}{R^2} - \frac{(T^2 - R^2)}{p}$$
$$= T^2\left(\frac{T^2}{R^2} - 1\right)\left[a - \frac{R^2}{T^2}\left(a + \frac{1}{p}\right)\right].$$

Now $G(R^2)$ is nondecreasing in $R$ and $H(R,T) > 0$ if $R^2 \le b(T^2)$ and $H(R,T) < 0$ for $R^2 \ge b(T^2)$ and crosses 0 when $R^2 = b(T^2)$. Thus, $H(R,T)G(R^2) \le H(R,T)G(b(T^2))$. Therefore, (5.1) becomes

$$\text{(5.2)}\quad \begin{aligned}\Delta &\le \frac{2}{\sigma^2}E\left[E\left[a\left(\frac{T^2}{R^2} - 1\right)\left(a - \frac{R^2}{T^2}\left(a + \frac{1}{p}\right)\right)T^2G(b(T^2))\middle|T\right]\right]\\ &= \frac{2a}{\sigma^2}E\left[\left(\frac{T^2}{R^2} - 1\right)\left(a - \frac{R^2}{T^2}\left(a + \frac{1}{p}\right)\right)\right]E\left[T^2G(b(T^2))\right],\end{aligned}$$

since $T^2$ and $R^2/T^2$ are independent.

So, since $E(T^2G(b(T^2))) \ge 0$, $\Delta \le 0$ if

$$E\left[\left(\frac{T^2}{R^2} - 1\right)\left(a - \frac{R^2}{T^2}\left(a + \frac{1}{p}\right)\right)\right] \le 0.$$

Returning to the original notation, put $T^2 = R^2 + S^2$. Then using the fact that $R^2/T^2 \sim \text{Beta}[p/2, m/2]$, it follows that

$$\Delta \le 0 \quad \text{if } aE\left[\left[\frac{S^2}{R^2}\right]\left[\frac{S^2}{R^2 + S^2}\right]\right] \le \frac{1}{p}E\left[\frac{S^2}{R^2 + S^2}\right]$$

$$\equiv 0 < a \lesssim \frac{\frac{1}{p}E\left[\frac{S^2}{R^2 + S^2}\right]}{E\left[\left(\frac{S^2}{R^2}\right)\left(\frac{S^2}{R^2 + S^2}\right)\right]} = \frac{1}{p}\frac{(p-2)}{m+2}. \qquad \square$$

## REFERENCES

BOCK, M. E. (1985). Minimax estimators that shift towards a hypersphere for location vectors of spherically symmetric distributions. *J. Multivariate Anal.* **17** 127–147.

BRANDWEIN, A. C. (1979). Minimax estimation of the mean of spherically symmetric distributions under general quadratic loss. *J. Multivariate Anal.* **9** 579–588.

BRANDWEIN, A. C. and STRAWDERMAN, W. E. (1980). Minimax estimation of location parameters for spherically symmetric distributions with concave loss. *Ann. Statist.* **8** 279–284.

BRANDWEIN, A. C. and STRAWDERMAN, W. E. (1990). Stein estimation: The spherically symmetric case. *Statist. Sci.* **5** 356–369.

BRAVO, G. and MACGIBBON, B. (1988). Improved shrinkage estimators for the mean vector of a scale mixture of normals with unknown variance. *Canad. J. Statist.* **16** 237–245.

CELLIER, D., FOURDRINIER, D. and ROBERT, C. (1988). Robust shrinkage estimators of the location parameter for elliptically symmetric distributions. *J. Multivariate Anal.* **29** 39–52.

CHOU, J. P. and STRAWDERMAN, W. E. (1990). Minimax estimation of means of multivariate normal mixtures. *J. Multivariate Anal.* **35** 141–150.

DU PLESSIS, N. (1970). *An Introduction to Potential Theory*. Hafner, Darien, Conn.

SCHEFFÉ, H. (1959). *The Analysis of Variance*. Wiley, New York.

STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.

DEPARTMENT OF STATISTICS AND
COMPUTER INFORMATION SYSTEMS
BOX 513
BARUCH COLLEGE
17 LEXINGTON AVE.
NEW YORK, NEW YORK 10010

DEPARTMENT OF STATISTICS
HILL CENTER
BUSCH CAMPUS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903