The Annals of Statistics
1991, Vol. 19, No. 4, 2102–2119

REDESCENDING *M*-ESTIMATES OF MULTIVARIATE LOCATION AND SCATTER

BY JOHN T. KENT AND DAVID E. TYLER[1]

University of Leeds and Rutgers University

A class of redescending M-estimates of multivariate location and scatter are investigated. Sufficient conditions are given to ensure the existence and uniqueness of the estimates. These results are applied to the multivariate t-distribution with degrees of freedom $\nu \geq 1$.

1. Introduction. M-estimates are usually defined as the solution of a set of implicit equations. As such, it is possible for an M-estimate not to exist for a given sample, and if it exists it is possible for it not to be unique. It is therefore of interest to know under what conditions on the sample does an M-estimate exist and under what conditions is it unique. This paper studies the existence and uniqueness of the simultaneous M-estimates of multivariate location and scatter.

One way to pose the problem of M-estimation of location and scatter is as follows. Let $\{\mathbf{y}_i,\ i = 1, \dots, n\}$ be a data set in $\mathbb{R}^p$. Let $\rho(s)$, $s \geq 0$, be a given function of s and let $\mathscr{P}_p$ denote the set of $p \times p$ symmetric positive definite matrices. A location estimate $\hat{\boldsymbol{\mu}} \in \mathbb{R}^p$ and an estimate of scatter $\hat{\boldsymbol{\Sigma}} \in \mathscr{P}_p$ are defined to minimize the objective function

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^{n} \rho\{(\mathbf{y}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})\} + \tfrac{1}{2}n \log|\boldsymbol{\Sigma}|. \tag{1.1}$$

When $\exp\{-\rho(\mathbf{y}'\mathbf{y})\}$ is integrable over $\mathbb{R}^p$, (1.1) can be regarded as the negative log-likelihood from an elliptically symmetric distribution. Throughout the paper, it is assumed without further mention that ρ is continuous.

If ρ is differentiable, then setting the derivative of (1.1) with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ to 0 yields the estimating equations

$$\hat{\boldsymbol{\mu}} = \operatorname{ave}\{w_i \mathbf{y}_i\}/\operatorname{ave}\{w_i\}, \tag{1.2}$$

$$\hat{\boldsymbol{\Sigma}} = \operatorname{ave}\{w_i(\mathbf{y}_i - \hat{\boldsymbol{\mu}})(\mathbf{y}_i - \hat{\boldsymbol{\mu}})'\}, \tag{1.3}$$

where $w_i = u(s_i)$, $u(s) = 2\rho'(s)$ and $s_i = (\mathbf{y}_i - \hat{\boldsymbol{\mu}})'\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\mu}})$. Here "ave" stands for the arithmetic average over $i = 1, \dots, n$.

If $s^{1/2}u(s)$ is an increasing function of $s > 0$, then $\exp\{-\rho(\mathbf{y}'\mathbf{y})\}$ is a strongly unimodal function of $\mathbf{y} \in \mathbb{R}^p$; that is, $\rho(\mathbf{y}'\mathbf{y})$ is convex in $\mathbf{y}$. The

---

Received April 1989; revised August 1990.
[1]Research supported by NSF Grant DMS-87-01198.
AMS 1980 *subject classifications*. Primary 62H12, 62H10.
Key words and phrases. Redescending M-estimate, robustness, multivariate t-distribution.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*.
www.jstor.org

existence and uniqueness of solutions to (1.2) and (1.3) follow from straightforward arguments provided the hyperplane generated by the data set $\{\mathbf{y}_i;\ 1 \le i \le n\}$ corresponds to $\mathbb{R}^p$; see, for example, Barndorff-Nielsen and Blaesild (1980). The emphasis in this paper is on redescending M-estimates, that is, estimates for which $s^{1/2}u(s)$ is increasing near 0 and decreasing for s near ∞. Curiously, the case for which $s^{1/2}u(s)$ is not redescending is of limited interest. Although the influence function of the location statistic is bounded if $s^{1/2}u(s)$ is bounded, the influence function of the scatter statistic is bounded only if $su(s)$ is bounded. If $su(s)$ is bounded, then $s^{1/2}u(s)$ must redescend to 0 as $s \to \infty$. The redescending case is considerably more delicate than the strongly unimodal case. In particular, when the scatter matrix $\hat{\Sigma}$ is fixed, then it is well known that (1.2) is susceptible to multiple solutions in $\hat{\mu}$; see, for example, Reeds (1985) who studied in detail the likelihood equation for the univariate location family of Cauchy distributions.

When the M-estimates of multivariate location and scatter are defined via implicit equations, the form is usually more general than that given by (1.2) and (1.3). Maronna (1976) defines a more general class of M-estimates by replacing w_i in (1.2) by $w_{1,i} = u_1(s_i)$ and replacing w_i in (1.3) by $w_{2,i} = u_2(s_i)$. The functions u_1 and u_2 need not be the same function and hence the more general M-estimates need not be related to a minimization problem of the form (1.1). Viewing M-estimates as a minimization problem though has some advantages. The function ρ need not be differentiable, and so M-estimates of the form (1.1) include estimates which cannot be represented as solutions to implicit equations. Also, when multiple solutions exist to M-estimating equations, not all solutions may be "desired" solutions. Minimizing (1.1) gives a natural definition of a "right" solution for equations of the form (1.2) and (1.3).

Sufficient conditions on the data and weight functions for the existence of the M-estimates of location and scatter are known; see, for example, Maronna (1976), Huber (1981), Chapter 8, or Hampel, Ronchetti, Rousseeuw and Stahel (1986). Application of these conditions to (1.2) and (1.3) requires $u(s)$ to be continuous and nonincreasing and for $su(s)$ to be nondecreasing. The results on existence given in this paper essentially require only that $\rho(s)$ be continuous. Some conditions are also imposed on the observed data set $\{\mathbf{y}_i;\ 1 \le i \le n\}$. These sample conditions are dependent on the "tail" behavior of ρ. They are milder than previously known sample conditions and are perhaps the best possible. These conditions, however, exclude strong redescenders, that is, whenever $u(s) = 0$ for large s. This may be unavoidable since for strong redescenders, (1.1) goes to $-\infty$ when Σ approaches a singular matrix; see Section 2. The nonexistence of a solution to minimizing (1.1) does not necessarily imply the nonexistence of a solution to (1.2) and (1.3), although the desirability of such a solution, if it exists, is questionable.

Results on the uniqueness of the M-estimates of location and scatter are more scarce, at least in the multivariate setting ($p \ge 2$). Huber (1981) notes on page 223 that "Uniqueness of the (estimators) so far has been proved only under the assumption that the distribution... has a center of symmetry; in the sample distribution this is of course unrealistic." This problem is still largely unresolved; see, for example, Hampel, Ronchetti, Rousseeuw and Stahel (1986), Chapter 5. We believe this paper is the first to give sufficient conditions for the uniqueness of the solution of the M-estimating equations for multivariate M-estimates of location and scatter.

Results on uniqueness in the univariate case do exist. Within the "robustness" literature, the results on uniqueness for $p = 1$ assume $s^{1/2}u(s)$ is nondecreasing; see Maronna (1976) or Huber (1981), Chapter 6. Thus these earlier uniqueness results do not apply to (1.2) and (1.3) when $su(s)$ is bounded, that is, when the scale estimate has bounded influence. In particular, they do not apply to bounded influence maximum likelihood estimates in symmetric location-scale families of distributions.

The paper is set out as follows. Section 3 gives the main results on existence and uniqueness. The key observation in our approach is that the location-scatter problem in p dimensions can be viewed as a scatter-only problem in $(p + 1)$ dimensions. Results on the existence and uniqueness of the solution $\hat{\boldsymbol{\Sigma}}$ to (1.3) with $\hat{\boldsymbol{\mu}}$ set equal to 0, which is a well-studied problem, can then be used to obtain results on the location-scatter problem. Some background results on the M-estimates of scatter are presented in Section 2. The results on existence of the M-estimates of scatter both extend and simplify earlier results and the approach used to prove existence is novel. For clarity of exposition, the technical proofs are relegated to Section 5.

To help illustrate the main results of the paper, the class of multivariate t M-estimates are discussed in Section 4. The multivariate t M-estimates refer to the M-estimates which correspond to the solutions to the likelihood equations for the location-scatter families of elliptically symmetric t-distributions on $\nu > 0$ degrees of freedom. The density of the t-distribution is given by

$$(1.4)\quad f(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = c_{\nu,p}|\boldsymbol{\Sigma}|^{-1/2}\{1 + (\mathbf{y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})/\nu\}^{-(\nu+p)/2}, \quad \mathbf{y} \in \mathbb{R}^p,$$

for some suitable normalizing constant $c_{\nu,p}$ which is not dependent on $\boldsymbol{\mu} \in \mathbb{R}^p$ or $\boldsymbol{\Sigma} \in \mathscr{P}_p$; see, for example, Mardia, Kent and Bibby (1979), page 57. The log-likelihood function up to an additive constant is the negative of (1.1) with $\rho(s)$ taken as

$$(1.5)\qquad \rho_\nu(s) = \tfrac{1}{2}(\nu + p)\log\{(\nu + s)/\nu\}$$

and the likelihood equations correspond to (1.2) and (1.3) with $u(s)$ taken as

$$(1.6)\qquad u_\nu(s) = (\nu + p)/(\nu + s).$$

Uniqueness of the multivariate t M-estimates can be shown for $\nu \geq 1$. Uniqueness in the univariate setting was established in the Cauchy case ($\nu = 1$) by Copas (1975) and for $\nu > 1$ by Märkeläinen, Schmidt and Styan (1981). Our approach for proving uniqueness differs substantially from the approach taken in these earlier papers and is as easy to apply in the multivariate setting as it is in the univariate setting.

2. Scatter-only problem. In this section, the M-estimates of scatter are considered when the location parameter $\boldsymbol{\mu}$ is known or fixed. Without loss of generality, assume $\boldsymbol{\mu} = 0$ and so interest lies in minimizing over $A \in \mathscr{P}_p$ the function

$$L(A) = \sum_{i=1}^{n} \rho\{\mathbf{y}_i' A^{-1} \mathbf{y}_i\} + \tfrac{1}{2} n \log|A|. \tag{2.1}$$

Before discussing conditions for the existence and uniqueness of a minimum of (2.1), the following definition is introduced.

DEFINITION 2.1. Given a continuous function $\rho(s)$, $s \geq 0$, define the "sill" a_0, $-\infty \leq a_0 \leq \infty$, of $\rho(s)$ by

$$a_0 = \sup\{a | s^{a/2} \exp[-\rho(s)] \to 0 \text{ as } s \to \infty\}. \tag{2.2}$$

If $\rho(s)$ is differentiable with $\rho'(s) = 2u(s)$ and if $\psi(s) = su(s)$ converges to a limit as $s \to \infty$, then the sill can also be characterized by

$$\lim_{s \to \infty} \psi(s) = a_0. \tag{2.3}$$

A proof of (2.3) is given in Section 5. For statistical purposes, only the case $a_0 \geq 0$ is of interest.

The following condition specifies that the data should not be "too concentrated" in low-dimensional subspaces, where "too concentrated" depends on the sill of $\rho(s)$. Fix $0 \leq q \leq p - 1$ and let $P_n(\cdot)$ denote the empirical distribution of the $\{\mathbf{y}_i;\ 1 \leq i \leq n\}$. Here "D" stands for "data."

CONDITION D_q. Let $\rho(s)$ have sill $a_0 > 0$ and suppose that for all linear subspaces $V \subset R^p$ with $q \leq \dim(V) \leq p - 1$,

$$P_n(V) < 1 - \{p - \dim(V)\}/a_0. \tag{2.4}$$

Condition D_0 will be important in this section; Condition D_1 will be important in the next section. Note that Condition D_q gets less stringent as q increases and as a_0 increases. In particular, for any choice of q, if $a_0 = \infty$, then Condition D_q merely implies that the $\{\mathbf{y}_i\}$ span R^p. Some further simple properties are listed below.

LEMMA 2.1. (i) *If Condition* D_q *holds, then* $a_0 > p - q$ *and* $n \geq p$.
(ii) *If* $a_0 > p$, $n \geq p$ *and if every subset of* p *or fewer data points is linearly independent, then Condition* D_0 *holds.*

PROOF. See Section 5.

The next lemma is needed for the existence theorem. Let $\{A_k;\ k = 1, 2, \dots\}$ be a sequence in $\mathscr{P}_p$ and $\lambda_{1,k} \geq \cdots \geq \lambda_{p,k} > 0$ represent the eigenvalues of A_k. Note that A_k goes to the boundary of $\mathscr{P}_p$ only if $\lambda_{1,k} \to \infty$ or $\lambda_{p,k} \to 0$.

LEMMA 2.2. *Let $\rho(\cdot)$ have sill $a_0 > 0$. Then $L(A_k) \to \infty$ if either*:

(i) $\lambda_{1,k} \to \infty$ *and* $\{\lambda_{p,k};\ k = 1, 2, \dots\}$ *is bounded away from* 0,
(ii) *Condition* D_0 *holds and* $\lambda_{p,k} \to 0$, *or*
(iii) *for some* $q \geq 1$, *Condition* D_q *holds*, $\{\lambda_{q,k};\ k = 1, 2, \dots\}$ *is bounded away from* 0 *and* $\lambda_{p,k} \to 0$.

PROOF. See Section 5.

Part (iii) is needed in the next section. The existence of a minimum to (2.1) now follows from parts (i) and (ii).

THEOREM 2.1. *Under Condition* D_0, *there exists an* $\hat{A} \in \mathscr{P}_p$ *such that* $L(\hat{A}) \leq L(A)$ *for all* $A \in \mathscr{P}_p$.

If $\rho(s)$ continuously differentiable for $s \geq 0$, with $u(s) = 2\rho'(s)$ and $\psi(s) = su(s)$, then setting the derivative of $L(A)$ to 0 yields the estimating equation

$$\hat{A} = \operatorname{ave}\left\{u\left(\mathbf{y}_i'\hat{A}^{-1}\mathbf{y}_i\right)\mathbf{y}_i\mathbf{y}_i'\right\}, \tag{2.5}$$

which must be satisfied, in particular, by any global minimum in Theorem 2.1.

If ρ is bounded above and below, then $a_0 = 0$ and so Theorem 2.1 does not apply. However, the problem of minimizing $L(A)$ in this case does not have a solution since $\log|A| \to -\infty$ and hence $L(A) \to -\infty$ as $A \in \mathscr{P}_p$ approaches a singular matrix. This case includes strong redescenders, since if $\rho(s)$ is also differentiable with $u(s) = 0$ for large s and $u(s) \geq 0$ for all $s \geq 0$, then $\rho(s)$ must be bounded above and below.

In general, it is not possible to establish the existence of a solution to (2.5) without some condition such as Condition D_0 relating the function u to the sample. When $u(s)$ is continuous and nonnegative, a necessary condition for existence is that for any subspace $V \subset \mathbb{R}^p$ with $\dim(V) < p$,

$$P_n(V) \leq 1 - \{p - \dim(V)\}/K, \tag{2.6}$$

where $K = \sup\{\psi(s) | s \geq 0\}$. The proof of this result corresponds to the proofs of (2.2) and (2.3) in Tyler (1988). Note that if $a_0 = K$, which occurs when $\psi(s)$ is nondecreasing, then the only difference between the sufficient condition D_0 and the necessary condition (2.6) is the strict inequality.

More stringent conditions on $u(s)$, namely monotonicity conditions, are imposed in order to establish the uniqueness of a solution to (2.5), and hence the uniqueness of a global minimum to (2.1). The uniqueness theorem itself is a subtle variation of Theorem 2.2 in Tyler (1988). A brief proof is given in Section 5.

CONDITION M. (i) For $s \geq 0$, $u(s) \geq 0$ and $u(s)$ is continuous and nonincreasing.
(ii) For $s \geq 0$, $\psi(s) = su(s)$ is strictly increasing.

THEOREM 2.2. *If Conditions* D_0 *and* M *hold, then there exists a unique solution* $\hat{A} \in \mathscr{P}_p$ *to* (2.5).

Attention is now directed to the convergence of the following iterative reweighting algorithm. Given an initial estimate $A^{(0)} \in \mathscr{P}_p$, define

$$(2.7) \qquad A^{(m+1)} = \operatorname{ave}\left[u\left\{\mathbf{y}_i'\left(A^{(m)}\right)^{-1}\mathbf{y}_i\right\}\mathbf{y}_i\mathbf{y}_i'\right].$$

The conditions for uniqueness are sufficient for proving that this algorithm is well defined and converges regardless of the initial estimate.

THEOREM 2.3. *If Conditions* D_0 *and* M *hold, then for any initial* $A^{(0)} \in \mathscr{P}_p$, $A^{(m)} \in \mathscr{P}_p$ *and converges to the unique solution* $\hat{A} \in \mathscr{P}_p$ *of* (2.5).

PROOF. See Section 5.

The monotonicity condition on $\psi(s) = su(s)$ can be dropped without affecting the convergence of the sequence $A^{(m)}$, provided the initial estimate $A^{(0)}$ satisfies

$$(2.8) \qquad 0 < \operatorname{ave}\left[\psi\left\{\mathbf{y}_i'\left(A^{(0)}\right)^{-1}\mathbf{y}_i\right\}\right] \leq 1.$$

Note that under Conditions D_0 and M(i), it is always possible to find such an initial estimate since for any $A \in \mathscr{P}_p$ the function $\Psi(c) = \operatorname{ave}\{\psi(c\mathbf{y}_i'A^{-1}\mathbf{y}_i)\}$ satisfies $\Psi(0) = 0$ and $\Psi(\infty) = \{1 - P_n(0)\}a_0$.

THEOREM 2.4. *If Conditions* D_0 *and* M(i) *hold and* $A^{(0)} \in \mathscr{P}_p$ *satisfies* (2.8), *then the algorithm given by* (2.7) *converges to a solution* $\hat{A} \in \mathscr{P}_p$ *of* (2.5).

PROOF. See Section 5.

A comparison of the results and approach of this section with other work is now briefly discussed. The proof of the existence of a solution to (2.5) given by Maronna (1976) and by Huber (1981) essentially consists of finding conditions under which the algorithm (2.5) converges to an element in $\mathscr{P}_p$. Their condition on the sample is stronger than Condition D_0, namely that for any subspace $V \subset \mathbb{R}^p$ with $\dim(V) < p$, $P_n(V) < 1 - p/a_0$. For a_0 close to p, this condition is much stronger than needed. The case a_0 close to p is discussed in more detail in Tyler (1988), where more relaxed conditions for existence are given which are only slightly stronger than Condition D_0. The proof of existence of a solution to (2.5) given by Tyler (1988) consists of showing that an algorithm which is more complicated than (2.7) converges. In light of the results of this section, the former algorithm is not necessary. The proof of Theorem 2.4 can also serve as an alternative proof of existence under Conditions D_0 and M(i). The proof itself is simply a tightening of the arguments used by Huber (1981).

3. Location-scatter problem. Results on the existence of a solution $\hat{\boldsymbol{\mu}} \in \mathbb{R}^p$ and $\hat{\boldsymbol{\Sigma}} \in \mathscr{P}_p$ can be obtained by embedding the problem into the scatter-only problem in $(p+1)$ dimension. Given an observation $\mathbf{y} \in \mathbb{R}^p$, define $\mathbf{z} = (\mathbf{y}', 1)' \in \mathbb{R}^{p+1}$. Also, given $\boldsymbol{\Sigma} \in \mathscr{P}_p$, $\boldsymbol{\mu} \in \mathbb{R}^p$ and a scalar $\lambda > 0$, define $A \in \mathscr{P}_{p+1}$ by

$$A = \begin{bmatrix} \boldsymbol{\Sigma} + \lambda^{-1}\boldsymbol{\mu}\boldsymbol{\mu}' & \lambda^{-1}\boldsymbol{\mu} \\ \lambda^{-1}\boldsymbol{\mu}' & \lambda^{-1} \end{bmatrix}. \tag{3.1}$$

Note that any $A \in \mathscr{P}_{p+1}$ can be decomposed into this form. The main results of this section are based on the reparameterization (3.1) and the algebraic identity

$$\mathbf{z}'A^{-1}\mathbf{z} = \lambda + (\mathbf{y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu}). \tag{3.2}$$

This identity is a special case of the standard identity for a partitioned quadratic form.

For a given $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\boldsymbol{\Sigma} \in \mathscr{P}_p$, let $A_0 \in \mathscr{P}_{p+1}$ be defined as in (3.1) with $\lambda = 1$. By using identity (3.2) and noting $|A_0| = |\boldsymbol{\Sigma}|$, the objective function defined by (1.1) can be expressed as

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = L_0(A_0) = \sum_{i=1}^{n} \rho_0\big(\mathbf{z}_i' A_0^{-1} \mathbf{z}_i\big) + \tfrac{1}{2} n \log|A_0|, \tag{3.3}$$

where $\rho_0(s) = \rho(s-1)$ for $s \geq 1$, and $\mathbf{z}_i = (\mathbf{y}_i', 1)'$. Thus the problem of minimizing $L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ over $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\boldsymbol{\Sigma} \in \mathscr{P}_p$ is equivalent to minimizing $L_0(A_0)$ over $A_0 \in \mathscr{P}_{p+1}$ with the restriction that the $(p+1, p+1)$ element of A_0 be 1. Since the largest root of A_0 is no smaller than any diagonal element, it is bounded below by 1. The existence of a minimum $\hat{\boldsymbol{\mu}} \in \mathbb{R}^p$ and $\hat{\boldsymbol{\Sigma}} \in \mathscr{P}_p$ to $L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ thus follows from Lemma 2.2(i) and (iii) with $q = 1$. In applying Lemma 2.2, note that the value of a_0 defined by (2.2) is the same for ρ and ρ_0. Also, a subset $\{\mathbf{z}_j;\ j \in J\}$ lies in a linear subspace $V \subset \mathbb{R}^{p+1}$ with $1 \leq \dim(V) < p+1$ if and only if $\{\mathbf{y}_j;\ j \in J\}$ lies in a hyperplane $H \subset \mathbb{R}^p$ with $\dim(H) = \dim(V) - 1$. Condition D_1 and Lemma 2.1 for $\{\mathbf{z}_i;\ 1 \leq i \leq n\} \subset \mathbb{R}^{p+1}$ can be recast as follows for $\{\mathbf{y}_i;\ 1 \leq i \leq n\} \subset \mathbb{R}^p$.

CONDITION D*. Let $\rho(s)$ have sill $a_0 > 0$ and suppose that for all hyperplanes $H \subset \mathbb{R}^p$ with $0 \leq \dim(H) \leq p - 1$,

$$P_n(H) < 1 - \{p - \dim(H)\}/a_0. \tag{3.4}$$

LEMMA 3.1. (i) *Under Condition* D*, *it follows that* $a_0 > p$ *and* $n \geq p + 1$.

(ii) *If* $a_0 > p + 1$, $n \geq p + 1$ *and if* $\{\mathbf{y}_i;\ 1 \leq i \leq n\}$ *lies in general position, then Condition* D* *holds.*

(iii) *If* $p < a_0 \leq p + 1$, $n > a_0/(a_0 - p)$ *and if* $\{\mathbf{y}_i;\ 1 \leq i \leq n\}$ *lies in general position, then Condition* D* *holds.*

For a proof of part (iii), the only new conclusion, see Section 5. Recall that the data points lying in general position mean that every proper hyperplane H

contains at most $\dim(H)+1$ points. In particular, a data set lies in general position with probability 1 when sampling from a continuous population.

Summarizing, we obtain the following existence theorem for the location-scatter problem.

THEOREM 3.1. *Under Condition* D*, *there exists a* $\hat{\boldsymbol{\mu}} \in \mathbb{R}^p$ *and* $\hat{\boldsymbol{\Sigma}} \in \mathscr{P}_p$ *such that* $L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) \le L(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *for all* $\boldsymbol{\mu} \in \mathbb{R}^p$, $\boldsymbol{\Sigma} \in \mathscr{P}_p$.

If ρ is continuously differentiable, then $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ will be a solution of the estimating equations (1.2) and (1.3).

Since minimizing (3.3) involves a restricted search over $A_0 \in \mathscr{P}_{p+1}$, the results from Section 2 on uniqueness and computation cannot be applied immediately to the location-scatter problem. By employing the reparameterization (3.1), the implicit equations (1.2) and (1.3) can be expressed equivalently as

$$\hat{A} = \begin{bmatrix} \hat{\boldsymbol{\Sigma}} + \hat{\lambda}^{-1}\hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}' & \hat{\lambda}^{-1}\hat{\boldsymbol{\mu}} \\ \hat{\lambda}^{-1}\hat{\boldsymbol{\mu}}' & \hat{\lambda}^{-1} \end{bmatrix} = \operatorname{ave}\{u(s_i)\mathbf{z}_i\mathbf{z}_i'\}, \tag{3.5}$$

where $s_i = (\mathbf{y}_i - \hat{\boldsymbol{\mu}})'\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\mu}})$ and $\hat{\lambda}$ is defined by

$$\hat{\lambda}^{-1} = \operatorname{ave}\{u(s_i)\}. \tag{3.6}$$

Using (3.2), expression (3.5) can then be rewritten as

$$\hat{A} = \operatorname{ave}\left\{u^*\left(\mathbf{z}_i'\hat{A}^{-1}\mathbf{z}_i; \hat{\lambda}\right)\mathbf{z}_i\mathbf{z}_i'\right\}, \tag{3.7}$$

where $u^*(s;\lambda) = u(s-\lambda)$ for $s \ge \lambda$ and $u^*(s;\lambda) = u(0)$ for $s < \lambda$. The key observation here is that (3.7) looks similar to an implicit equation for an M-estimate of scatter only; see (2.5).

The difference between (2.5) and (3.7) is that the weight function is fixed in the former, whereas in the latter it depends on $\hat{\lambda}$. Since $\hat{\lambda}$ may depend on the data, uniqueness does not follow by applying Theorem 2.2 to (3.7) directly. Some additional arguments are needed. The complete proof is given in Section 5.

THEOREM 3.2. *If Conditions* D* *and* M *hold, then there exists a solution* $\hat{\boldsymbol{\mu}} \in \mathbb{R}^p$ *and* $\hat{\boldsymbol{\Sigma}} \in \mathscr{P}_p$ *to* (1.2) *and* (1.3). *Further, this solution is unique if* $(s+\hat{\lambda})u(s)$ *is strictly increasing in* s, *where* $\hat{\lambda}$ *is defined by* (3.6).

Theorem 3.2 may be difficult to apply in practice because the value of $\hat{\lambda}$ depends on $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$. Given a solution $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$, though, it does give conditions which guarantee that no other solutions exist. To obtain a condition which is easier to check we can look for an upper bound λ_u for λ since, if $\lambda_u \ge \lambda$ and $(s+\lambda_u)u(s)$ is strictly increasing and Condition M holds, then $(s+\lambda)u(s)$ is also strictly increasing.

An upper bound λ_u can be derived as follows. By multiplying (1.3) by $\hat{\Sigma}^{-1}$ and taking the trace, we obtain $\text{ave}\{u(s_i)s_i\} = p$. Thus, for all $b > 0$, $p = \text{ave}\{(s_i + b)u(s_i)\} - b\hat{\lambda}^{-1}$ which implies $\lambda_l \le \hat{\lambda} \le \lambda_u$, where

$$\lambda_u^{-1} = \sup_{b>0} \inf_{s>0} \{(s + b)u(s) - p\}/b, \tag{3.8}$$

$$\lambda_l^{-1} = \inf_{b>0} \sup_{s>0} \{(s + b)u(s) - p\}/b. \tag{3.9}$$

Hence we obtain the following uniqueness theorem.

THEOREM 3.3. *If Conditions* D* *and* M *hold and* $(s + \lambda_u)u(s)$ *is strictly increasing, where* λ_u *is defined by* (3.10), *then there exists a unique solution* $\hat{\mu} \in \mathbb{R}^p$, $\hat{\Sigma} \in \mathscr{P}_p$ *to* (1.2) *and* (1.3).

An even simpler condition to check can be obtained by noting that if $(s + b)u(s)$ is strictly increasing for $b = (p + 1)/u(0)$, then $\lambda_u^{-1} \ge \{bu(0) - p\}/b = b^{-1}$ and so $(s + \lambda_u)u(s)$ is strictly increasing provided Condition M holds. Hence the following result holds.

THEOREM 3.4. *If Conditions* D* *and* M *hold and* $\{s + (p + 1)/u(0)\}u(s)$ *is strictly increasing, then there exists a unique solution* $\hat{\mu} \in \mathbb{R}^p$, $\hat{\Sigma} \in \mathscr{P}_p$ *to* (1.2) *and* (1.3).

For a given weight function $u(s)$ satisfying Condition M, consider the class of weight functions $u(s; c) = cu(s)$ which is indexed by the "tuning" constant $c > 0$. If $(s + \varepsilon)u(s)$ is strictly increasing for some $\varepsilon > 0$ but not for $\varepsilon = (p + 1)/u(0)$, then the weight function $u(s)$ can be "tuned," that is, replaced by $u(s; c)$ for some $c > 1$, so that $\{s + (p + 1)/u(0; c)\}u(s; c)$ is strictly increasing. This is done by choosing $c \ge (p + 1)/\{\varepsilon u(0)\}$. Theorem 3.4 on uniqueness then applies to $u(s; c)$ provided Condition D* holds for $u(s; c)$. Note that if Condition D* holds for $u(s)$, then it holds for $u(s; c)$ for any $c > 1$ since the sill corresponding to $u(s; c)$ is ca_0, where a_0 is the sill corresponding to $u(s)$.

As an example, consider the class of weight functions which are normal in the middle and have t-type tails; that is, define for $s \ge 0$, $c > 0$,

$$u(s; c) = c\begin{cases} 1, & s \le s_0, \\ (\nu + s_0)/(\nu + s), & s > s_0, \end{cases} \tag{3.10}$$

where $\nu > 0$ and $s_0 \ge 0$. The function $u(s; c)$ has corresponding sill $c(\nu + s_0)$ and satisfies Condition M. In addition, $\{s + (p + 1)/u(0; c)\}u(s; c)$ is strictly increasing in s for $c > (p + 1)/(s_0 + \nu)$.

Another example is the class of weight functions which can be represented as the ratio of two polynomials with alternating roots; that is, define for $s \geq 0$, $c > 0$,

$$u(s;c) = c\left\{\prod_{i=1}^{k}(s+\alpha_i)\right\}\Big/\left\{\prod_{j=1}^{k+1}(s+\beta_j)\right\}, \tag{3.11}$$

where $k \geq 1$, $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_k > 0$, $\beta_1 \geq \beta_2 \geq \cdots \geq \beta_{k+1} > 0$ and with $\beta_i \geq \alpha_i \geq \beta_{i+1}$ for $i = 1,\dots,k$. The weight function $u(s;c)$ has corresponding sill equal to c and satisfies Condition M. Furthermore, $\{s + (p+1)/u(0;c)\}u(s;c)$ is strictly increasing in s for $c > (p+1)\prod_{i=1}^{k}(\alpha_i/\beta_i)$. Both examples (3.10) and (3.11) include the multivariate t weight functions (1.6) as special cases. The multivariate t M-estimates are discussed in more detail in the next section.

Concerning computation, guaranteed-convergent iterative reweighting algorithms are more difficult to construct for location-scatter problems than for scatter-only problems, even when the solutions to (1.2) and (1.3) are known to be unique. Such algorithms can be constructed for the multivariate t M-estimates. These are briefly discussed in the next section. In general, though, the problem remains open.

Finally, we note that only minor relaxations of Condition D* may be possible to ensure the existence of $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$. Necessary conditions for the existence of solutions $\hat{\boldsymbol{\mu}} \in \mathbb{R}^p$, $\hat{\boldsymbol{\Sigma}} \in \mathscr{P}_p$ to (1.2) and (1.3) can be obtained by using the representation (3.7) and noting that although $\hat{\lambda}$ may depend on the data, this does not affect the application of (2.6) to (3.7).

LEMMA 3.2. *If a solution* $\hat{\boldsymbol{\mu}} \in \mathbb{R}^p$ *and* $\hat{\boldsymbol{\Sigma}} \in \mathscr{P}_p$ *exists to* (1.2) *and* (1.3) *and* $\hat{\lambda}$ *is defined by* (3.6), *then*:

(i) $K(\hat{\lambda}) = \sup\{(s+\hat{\lambda})u(s) | s \geq 0\} \geq p+1$, *and*

(ii) *for any hyperplane* $H \subset \mathbb{R}^p$ *with* $0 \leq \dim(H) \leq p-1$, $P_n(H) \leq 1 - \{p - \dim(H)\}/K(\hat{\lambda})$.

The first part of the lemma can be justified by multiplying (3.7) by $\hat{A}^{-1}$ and taking the trace. This gives

$$p + 1 = \operatorname{ave}\left\{\mathbf{z}_i'\hat{A}^{-1}\mathbf{z}_i u^*\left(\mathbf{z}_i'\hat{A}^{-1}\mathbf{z}_i;\hat{\lambda}\right)\right\} \leq \sup\left\{su(s;\hat{\lambda})\middle| s \geq 0\right\} = K(\hat{\lambda}).$$

Part (ii) is obtained from (2.6) by again making use of the relationship that a subset $\{\mathbf{z}_j;\ j \in J\}$ lies in a linear subspace $V \subset \mathbb{R}^{p+1}$ with $1 \leq \dim(V) < p+1$ if and only if $\{\mathbf{y}_j;\ j \in J\}$ lies in a hyperplane $H \subset \mathbb{R}^p$ with $\dim(H) = \dim(V) - 1$.

If $(s + \hat{\lambda})u(s)$ is nondecreasing in s, then $K(\hat{\lambda}) = a_0$ and so the only difference between the sufficient condition D* and the necessary condition given by Lemma 3.2(ii) is the strict inequality. For the existence of a minimum to (1.1) in general, it is fairly easy to show that Condition D* essentially cannot be relaxed in the univariate case, and we conjecture that a similar result is true for $p \geq 2$. A proof of the following lemma is given in Section 5.

LEMMA 3.3. *Suppose* $\rho(s)$ *has sill* $a_0 > 0$ *and that for any* $a > a_0$, $s^{a/2}\exp\{-\rho(s)\} \to \infty$ *as* $s \to \infty$. *For* $p = 1$, *if there exists a solution* $\hat{\boldsymbol{\mu}} \in \mathbb{R}$ *and* $\hat{\boldsymbol{\Sigma}} \in \mathbb{R}^+$ *to minimizing* (1.1), *then for any* $\mathbf{y} \in \mathbb{R}$, $P_n(\mathbf{y}) \leq 1 - 1/a_0$.

As noted previously, when ρ is differentiable a solution to (1.2) and (1.3) may exist even if (1.1) does not have a minimum.

4. The multivariate *t* *M*-estimates. This section discusses the multivariate t M-estimates. Clearly, $u_\nu(s)$ defined by (1.6) satisfies Condition M for all $\nu > 0$. Also, $su_\nu(s)$ is bounded, and hence $s^{1/2}u_\nu(s)$ is redescending. The sill is easily seen to be $a_{0,\nu} = \nu + p$.

First we discuss the scatter-only problem. For any $\nu > 0$, Theorems 2.2 and 2.3 on uniqueness and computation can be applied provided Condition D_0 holds for the sample. Condition D_0 for this example can be stated as follows: For any subspace $V \subset \mathbb{R}^p$ with $0 \leq \dim(V) \leq p - 1$, $P_n(V) < \{\nu + \dim(V)\}/(\nu + p)$. This condition becomes a necessary condition for the existence of a solution to (2.5) when $u = u_\nu$ if the strict inequality is replaced by $\leq$; see (2.6). Condition D_0 is most restrictive for ν near 0. In particular, if any data point equals 0, then the condition cannot be met for sufficiently small ν. Nevertheless, for any $\nu > 0$, Condition D_0 holds with probability 1 for random samples of size $n \geq p$ from any continuous distribution.

Next we turn to the location-scatter problem. Let $\hat{\boldsymbol{\mu}}_\nu \in \mathbb{R}^p$ and $\hat{\boldsymbol{\Sigma}}_\nu \in \mathscr{P}_p$ represent a solution to (1.2) and (1.3) when $u = u_\nu$. Such solutions exist provided Condition D^* holds for the sample, which for this example can be stated as follows.

CONDITION D_ν^*. For any hyperplane $H \in \mathbb{R}^p$ with $0 \leq \dim(H) \leq p - 1$, $P_n(H) < \{\nu + \dim(H)\}/(p + \nu)$.

For $\nu > 0$, any solution $(\hat{\boldsymbol{\mu}}_\nu, \hat{\boldsymbol{\Sigma}}_\nu)$ gives

$$\hat{\lambda}_\nu^{-1} = \mathrm{ave}\{u_\nu(s_i)\} = 1, \tag{4.1}$$

where $s_i = (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_\nu)'\hat{\boldsymbol{\Sigma}}_\nu^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_\nu)$. Statement (4.1) can be verified by evaluating the expressions in (3.8) and (3.9) at $b = \nu$, $s = 0$ and $b = \nu$, $s = \infty$, respectively. One important consequence of $\hat{\lambda}_\nu$ not being dependent on the data is that any method for finding a solution in the scatter-only problem, such as the iterative reweighting algorithm (2.7), can be applied to the location-scatter problem through (3.7).

Consider now the case $\nu > 1$. When random sampling from continuous distributions in $\mathbb{R}^p$, Condition D_ν^* holds with probability 1 for samples of size $n \geq p + 1$. We note that Condition D_ν^* can be made a necessary condition for the existence of $(\hat{\boldsymbol{\mu}}_\nu, \hat{\boldsymbol{\Sigma}}_\nu)$ by replacing the strict inequality by a simple inequality. Theorem 3.2 on uniqueness holds since $(s + 1)u_\nu(s)$ is strictly increasing.

Alternatively, since $\hat{\lambda}_\nu = 1$ regardless of the data, the uniqueness of $(\hat{\boldsymbol{\mu}}_\nu, \hat{\boldsymbol{\Sigma}}_\nu)$ follows directly by applying Theorem 2.2 to (3.7).

The special case $\nu = 1$, the Cauchy case, requires separate consideration. Note that $(s+1)u_1(s) = p+1$ and so is not strictly increasing. Equation (2.5) has been recently considered by Kent and Tyler (1988) for $u(s) = p/s$. Assume the following conditions on the sample hold: (i) $P_n\{0\} = 0$ and (ii) for any subspace $V \in \mathbb{R}^p$ with $0 < \dim(V) < p$, $P_n(V) < \dim(V)/p$. Under these conditions, a solution $\hat{A}$ to (2.5) exists and if $\hat{A}_0$ is any other solution, then $\hat{A} = \gamma \hat{A}_0$ for some $\gamma > 0$. This result is given in Theorem 1 of Kent and Tyler (1988). Results on uniqueness for $\nu = 1$ follow from applying the above result to (3.5). For this case, (3.5) can be expressed as

$$\hat{A} = \operatorname{ave}\left\{u_0\left(\mathbf{z}_i'\hat{A}^{-1}\mathbf{z}_i\right)\mathbf{z}_i\mathbf{z}_i'\right\}, \tag{4.2}$$

where $u_0(s) = (p+1)/s$, $\mathbf{z}_i' = (\mathbf{y}_i', 1)$ and for any $\gamma > 0$,

$$\hat{A} = \gamma \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_1 + \hat{\boldsymbol{\mu}}_1\hat{\boldsymbol{\mu}}_1' & \hat{\boldsymbol{\mu}}_1 \\ \hat{\boldsymbol{\mu}}_1' & 1 \end{bmatrix}. \tag{4.3}$$

Now for the data set $\{\mathbf{z}_i \in \mathbb{R}^{p+1} | 1 \le i \le n\}$, Condition (i) above obviously holds since $\mathbf{z}_i \neq 0$. Condition (ii) is equivalent to Condition D_1^*. Since the indeterminancy of γ does not affect $\hat{\boldsymbol{\mu}}_1$ and $\hat{\boldsymbol{\Sigma}}_1$, the estimates are uniquely defined when Condition D_1^* holds.

THEOREM 4.1. *If Condition* D_1^* *holds, then there exists a unique solution* $\hat{\boldsymbol{\mu}}_1 \in \mathbb{R}^p$ *and* $\hat{\boldsymbol{\Sigma}}_1 \in \mathscr{P}_p$ *to* (1.2) *and* (1.3) *for the weight function* $u_1(s) = (p+1)/(s+1)$.

Finally, consider the case $0 < \nu < 1$. Condition D_1^* is still sufficient for ensuring the existence of $(\hat{\boldsymbol{\mu}}_\nu, \hat{\boldsymbol{\Sigma}}_\nu)$ for $0 < \nu < 1$. However, this condition can only be satisfied if $n > (p+\nu)/\nu$, which goes to ∞ as $\nu \to 0$. Also, there is a large gap between the sufficient Condition D_ν^* and the necessary condition given by Lemma 3.2 when ν is near 0. For $0 < \nu < 1$, $(s+1)u_\nu(s)$ is decreasing rather than increasing and so the necessary condition of Lemma 3.2 becomes: For any hyperplane H with $0 \le \dim(H) \le p-1$,

$$P_n(H) \le 1 - \nu\{p - \dim(H)\}/(p+\nu). \tag{4.4}$$

This necessary condition is very weak for ν near 0, since the upper bound is close to 1. Another consequence of $(s+1)u_\nu(s)$ being decreasing for $0 < \nu < 1$ is that the results on uniqueness from Section 3 do not apply.

The $p = 1$-dimensional case has been well studied. As noted in Section 1, uniqueness in the univariate t setting was shown by Copas (1975) for $\nu = 1$ and for $\nu > 1$ by Märkeläinen, Schmidt and Styan (1981). Uniqueness is known not to hold in general for the case $\nu < 1$; see Gabrielson (1982). For $\nu > 0$ in the univariate case, Condition D_ν^* simply states that no more than $\nu n/(1+\nu)$ observations can coincide. This is the same condition imposed by Copas (1975) and by Märkeläinen, Schmidt and Styan (1981). This condition is the same as conditions (ii) and (iii) of Lemma 4 in Dempster, Laird and Rubin (1980) when applied to a univariate t-distribution on $\nu > 0$ degrees of freedom. Under this condition, they prove the convergence of an iteratively reweighted least squares algorithm to a solution of the likelihood equation associated with the univariate t-distribution. Rubin (1983) generalizes their algorithm to the multivariate t setting, but does not explicitly give conditions on the sample for ensuring its convergence.

5. Proofs.

PROOF OF (2.3). The proof is by contradiction. Let $b_0 = \lim_{s\to\infty} su(s)$, $-\infty \le b_0 \le \infty$. First, suppose there exists an $a < b_0$ such that as $s \to \infty$, $\limsup s^{a/2} \exp\{-\rho(s)\} = \infty$ or equivalently $\limsup H(s) = \infty$, where $H(s) = a(\log s) - 2\rho(s)$. Since $H(s)$ is differentiable, by the mean value theorem there exists a sequence $s_j \to \infty$ such that $H'(s_j) = a/s_j - 2\rho'(s_j) = a/s_j - u(s_j) > 0$ or equivalently $s_j u(s_j) < a$. This contradicts the definition of b_0 and so for $a < b_0$ it follows that $s^{a/2}\exp\{-\rho(s)\} \to 0$ as $s \to \infty$.

Next, suppose there exists an $a > b_0$ such that as $s \to \infty$,

$$s^{a/2}\exp\{-\rho(s)\} \to 0$$

or equivalently $H(s) \to -\infty$. This implies there exists a sequence $s_j \to \infty$ such that $H'(s_j) = a/s_j - u(s_j) < 0$ or $s_j u(s_j) > a$. Again this contradicts the definition of b_0. Thus b_0 equals the sill of $\rho(s)$ and (2.3) holds. □

PROOF OF LEMMA 2.1. (i) Let V be any q-dimensional subspace so that $P_n(V) \ge 0$. Then (2.4) implies $a_0 > p - q$.

Next suppose $n < p$ and let V be any $(p-1)$-dimensional subspace containing all the data, so that $P_n(V) = 1$. But (2.4) implies $P_n(V) < 1$, giving a contradiction. Hence $n \ge p$.

(ii) Let V be any proper subspace and set $\dim(V) = r$ say, $r \le p - 1$. Under the stated conditions,

$$P_n(V) \le \frac{r}{n} \le \frac{r}{p} = 1 - \frac{(p-r)}{p} < 1 - \frac{(p-r)}{a_0}.$$

That is, Condition D_0 holds. □

PROOF OF LEMMA 2.2. Define $g(s) = \exp\{-\rho(s)\}$ for $s \ge 0$ and $G(A) = \exp\{-L(A)\}$ for $A \in \mathscr{P}_p$. Since $a_0 > 0$, $\rho(s) \to \infty$ as $s \to \infty$. Also, since ρ is continuous, this implies $d = \inf\{\rho(s) | s \ge 0\} > -\infty$. Thus

$$G(A) = |A|^{-n/2} \prod_{i=1}^{n} g(\mathbf{y}_i' A^{-1} \mathbf{y}_i) \tag{5.1}$$

goes to 0 if and only if $L(A)$ goes to ∞. Clearly, if $\lambda_{1,k} \to \infty$ while $\lambda_{p,k}$ remains bounded away from 0, then $G(A_k) \to 0$ since $|A_k|^{-1} \to 0$ and $g(s) \le e^{-d}$. Thus part (i) is proven. The proofs of parts (ii) and (iii) can be combined. For an arbitrary $q = 0, 1, \dots, p-1$, suppose part (ii) holds when $q = 0$ and part (iii) holds when $q \neq 0$. Define r, $1 \le r \le p$, so that $\lambda_{j,k} \to 0$ for $r \le j \le p$ and $\lambda_{j,k} \ge c > 0$ for $1 \le j < r$. It follows that $r > q$. For clarity, the proof is broken into several steps.

Step 1. Let $E_k = \{\mathbf{e}_{1,k}, \dots, \mathbf{e}_{p,k}\}$ be an orthonormal set of eigenvectors of A_k so that $A_k \mathbf{e}_{j,k} = \lambda_{j,k}\mathbf{e}_{j,k}$, $1 \le j \le p$. Without loss of generality, it can be assumed that $E_k \to E = \{\mathbf{e}_1, \dots, \mathbf{e}_p\}$, an orthonormal set of vectors. This can be assumed since any sequence has a subsequence satisfying this property, and so if convergence of $G(A_k)$ to 0 is proven under such sequences, then it holds in general. For each $j = 1, \dots, p$, let S_j denote the space spanned by $\{\mathbf{e}_1, \dots, \mathbf{e}_j\}$ and $D_j = S_j \setminus S_{j-1} = \{\mathbf{y} \in \mathbb{R}^p | \mathbf{y} \in S_j,\ \mathbf{y} \notin S_{j-1}\}$ with $S_0 = \{0\}$ and $D_0 = \{0\}$.

Step 2. Express $G(A_k) = \prod_{j=0}^{p} G_j(A_k)$, where for $j = 1, \dots, p$,

$$G_j(A_k) = \lambda_{j,k}^{-n/2} \prod_j g(\mathbf{y}_i' A_k^{-1} \mathbf{y}_i) \tag{5.2}$$

with the product $\prod_j$ being taken over $\mathbf{y}_i \in D_j$, and $G_0(A_k) = \{g(0)\}^{nP_n(S_0)}$. Note that $\prod_{j=0}^{r-1} G_j(A_k)$ is bounded since $g(s)$ is bounded and for $1 \le j \le r-1$, $\lambda_{j,k}^{-1} < c^{-1}$. Thus it only remains to show that $\prod_{j=r}^{p} G_j(A_k) \to 0$.

Set $\alpha_l = \{1 - (p - l + 1)/a\} - P_n(S_{l-1})$. Choose $a < a_0$ close enough to a_0 so that $\alpha_l > 0$, $l = r, \dots, p$. Condition D_q guarantees that these inequalities can be met because $\dim(S_{l-1}) = l - 1 \ge r - 1 \ge q$. Then it will follow that $\prod_{j=r}^{p} G_j(A_k) \to 0$ if we can show that

$$\prod_{j=l}^{p} G_j(A_k) = o(\lambda_{l,k}^{na\alpha_l/2}) \tag{5.3}$$

for $r \le l \le p$.

Step 3. Before showing (5.3), an intermediate result is established. Let $r \le h \le p$ and suppose $\mathbf{y} \in D_h$. Since $\mathbf{y}'A_k^{-1}\mathbf{y} = \sum_{j=1}^{p} \lambda_{j,k}^{-1}(\mathbf{e}_{j,k}'\mathbf{y})^2$ and $(\mathbf{e}_h'\mathbf{y})^2 > 0$, it follows that

$$\liminf \lambda_{h,k} \mathbf{y}'A_k^{-1}\mathbf{y} \ge (\mathbf{e}_h'\mathbf{y})^2 > 0. \tag{5.4}$$

Now since $\lambda_{h,k} \to 0$, it must be that $\mathbf{y}'A_k^{-1}\mathbf{y} \to \infty$ and by definition for any $a < a_0$, $(\mathbf{y}'A_k^{-1}\mathbf{y})^{a/2} g(\mathbf{y}'A_k^{-1}\mathbf{y}) \to 0$. Together with (5.4), this gives

$$g(\mathbf{y}'A_k^{-1}\mathbf{y}) = o(\lambda_{h,k}^{a/2}). \tag{5.5}$$

Step 4. The proof of (5.3) now proceeds by reverse induction. Since $P_n(D_p) = 1 - P_n(S_{p-1})$ and $\dim(S_{p-1}) = p - 1$, it follows immediately from (5.5) that (5.3) holds for $l = p$.

Next, suppose (5.3) holds for some l, $r < l \le p$. Together with (5.5), this implies

$$\prod_{j=l-1}^{p} G_j(A_k) = o(\lambda_{l-1,k}^{n\{aP_n(D_{l-1}) - 1\}/2}) o(\lambda_{l,k}^{na\alpha_l/2}). \tag{5.6}$$

However, since $P_n(D_{l-1}) = P_n(S_{l-1}) - P_n(S_{l-2})$, it follows that $\alpha_{l-1} = \alpha_l + \{P_n(D_{l-1})\} - 1/a$. Furthermore, since $\lambda_{l,k} \le \lambda_{l-1,k}$, (5.5) holds with α_l replaced by α_{l-1}. Thus (5.5) and consequently Lemma 2.2(ii) and (iii) are proven. □

Proof of Theorem 2.2. Without loss of generality, suppose both $\hat{A} = I$ and $\hat{A} = A$ satisfy (1.3). Let δ_1 be the largest eigenvalue of A, and suppose $\delta_1 > 1$. Since $su(s)$ is strictly increasing and $u(s)$ is nonincreasing, it follows that for $\mathbf{y} \ne 0$, $u(\mathbf{y}'A^{-1}\mathbf{y}) \le u(\delta_1^{-1}\mathbf{y}'\mathbf{y}) < \delta_1 u(\mathbf{y}'\mathbf{y})$. By Condition M, this implies $A = \operatorname{ave}\{u(\mathbf{y}_i'A^{-1}\mathbf{y}_i)\mathbf{y}_i\mathbf{y}_i'\} < \delta_1 \operatorname{ave}\{u(\mathbf{y}_i'\mathbf{y}_i)\mathbf{y}_i\mathbf{y}_i'\} = \delta_1 I$, where $A_1 < A_2$ means $A_2 - A_1$ is positive definite. This gives the contradiction $\delta_1 < \delta_1$, and so $\delta_1 \le 1$. A similar argument implies that the smallest eigenvalue of A is $\delta_p \ge 1$. Thus $A = I$. □

Proof of Theorem 2.3. Since the conditions of the theorem ensure $\hat{A}$ exists and since the transformation $\{\mathbf{y}_i\} \to \{B\mathbf{y}_i\}$ for any nonsingular B induces the transformation $\hat{A} \to B\hat{A}B'$, it can be assumed without loss of generality that $\hat{A} = I$ is a solution to (2.5). Again, for clarity the proof is given in several steps.

Step 1. Let $\delta_{1,m}$ and $\delta_{p,m}$ denote the largest and smallest eigenvalue of $A^{(m)}$, $m = 1, 2, \dots$. In this step, the following results are established: (i) $\delta_{1,m} \le 1 \Rightarrow \delta_{1,m+1} \le 1$, (ii) $\delta_{1,m} > 1 \Rightarrow \delta_{1,m+1} < \delta_{1,m}$, (iii) $\delta_{p,m} \ge 1 \Rightarrow \delta_{p,m+1} \ge 1$ and (iv) $\delta_{p,m} < 1 \Rightarrow \delta_{p,m+1} > \delta_{p,m}$. Note that (iii) and (iv) imply $A^{(m+1)} \in \mathscr{P}_p$ whenever $A^{(m)} \in \mathscr{P}_p$.

To prove (i), if $\delta_{1,m} \le 1$, then $\mathbf{y}'\{A^{(m)}\}^{-1}\mathbf{y} \ge \delta_{1,m}^{-1}\mathbf{y}'\mathbf{y} \ge \mathbf{y}'\mathbf{y}$ and since $u(s)$ is nonincreasing $u(\mathbf{y}'\{A^{(m)}\}^{-1}\mathbf{y}) \le u(\mathbf{y}'\mathbf{y})$. Given that $\hat{A} = I$ is a solution to (2.5), this implies $A^{(m+1)} \le \operatorname{ave}\{u(\mathbf{y}_i'\mathbf{y}_i)\mathbf{y}_i\mathbf{y}_i'\} = I$, where the partial ordering $A_1 \le A_2$ means $A_2 - A_1$ is positive semidefinite. Thus $\delta_{1,m+1} \le 1$. The proof of (iii) is analogous.

To prove (ii), since $\mathbf{y}'\{A^{(m)}\}^{-1}\mathbf{y} \ge \delta_{1,m}^{-1}\mathbf{y}'\mathbf{y}$, $u(s)$ is nonincreasing and $su(s)$ is strictly increasing, it follows that if $\delta_{1,m} > 1$, then

$$u\left(\mathbf{y}'\{A^{(m)}\}^{-1}\mathbf{y}\right) \le u\left(\delta_{1,m}^{-1}\mathbf{y}'\mathbf{y}\right) \le \delta_{1,m}u(\mathbf{y}'\mathbf{y}), \tag{5.7}$$

with the second inequality strict for $\mathbf{y} \ne 0$. This implies $A^{(m+1)} < \delta_{1,m}\operatorname{ave}\{u(\mathbf{y}_i'\mathbf{y}_i)\mathbf{y}_i\mathbf{y}_i'\} = \delta_{1,m}I$. Thus $\delta_{1,m+1} < \delta_{1,m}$. The proof of (iv) is analogous.

Step 2. In this step we shall show that (i) $\limsup \delta_{1,m} \le 1$ and (ii) $\liminf \delta_{p,m} \ge 1$, from which it follows that $\delta_{1,m} \to 1$, $\delta_{p,m} \to 1$, so that $A^{(m)} \to I$.

First we set up some more notation. Given a symmetric positive definite matrix A, let $\delta_1(A)$ denote its largest eigenvalue and set $\varphi(A) = \operatorname{ave}\{u(\mathbf{y}_i'A^{-1}\mathbf{y}_i)\mathbf{y}_i\mathbf{y}_i'\}$. The logic in Step 1 implies that if $\delta_1(A) > 1$ and $B = \varphi(A)$, then $\delta_1(B) < \delta_1(A)$.

In view of Step 1, statement (i) requires proof only in the case where $\delta_{1,m} = \delta_1(A^{(m)}) > 1$ for all m. Note that $\{\delta_{1,m}\}$ forms a decreasing sequence in this case. Let $\delta^* = \lim \delta_{1,m} \geq 1$ and suppose if possible that $\delta^* > 1$.

From Step 1, the eigenvalues of the $\{A^{(m)}\}$ are bounded away from 0 and ∞. Thus we can find a convergent subsequence $A^{(m_j)} \to B_0$ say, where B_0 is positive definite. Further $A^{(m_j+1)} = \varphi(A^{(m_j)}) \to \varphi(B_0) = B_1$, say. Since $\{\delta_{1,m}\}$ forms a decreasing sequence, $\delta_1(B_0) = \lim \delta_{1,m_j} = \delta^*$ and $\delta_1(B_1) = \lim \delta_{1,m_j+1} = \delta^*$. However, Step 1 implies that $\delta_1(B_1) < \delta_1(B_0)$, giving a contradiction. Hence statement (i) follows.

Statement (ii) can be proved similarly. □

Proof of Theorem 2.4. By the Cauchy–Schwarz inequality, for $\mathbf{a} \in \mathbb{R}^p$,

$$\begin{aligned} \mathbf{a}'A^{(1)}\mathbf{a} &= \operatorname{ave}\left[u\left\{\mathbf{y}_i'(A^{(0)})^{-1}\mathbf{y}_i\right\}(\mathbf{a}'\mathbf{y}_i)^2\right] \\ &\leq \operatorname{ave}\left[\psi\left\{\mathbf{y}_i'(A^{(0)})^{-1}\mathbf{y}_i\right\}\right]\mathbf{a}'A_0\mathbf{a} \leq \mathbf{a}'A_0\mathbf{a}. \end{aligned}$$

Thus $A^{(1)} \leq A^{(0)}$. Also, since $u(s)$ is nonincreasing, if $A^{(m+1)} \leq A^{(m)}$, then

$$\begin{aligned} A^{(m+2)} &= \operatorname{ave}\left[u\left\{\mathbf{y}_i'(A^{(m+1)})^{-1}\mathbf{y}_i\right\}\mathbf{y}_i\mathbf{y}_i'\right] \\ &\leq \operatorname{ave}\left[u\left\{\mathbf{y}_i'(A^{(m)})^{-1}\mathbf{y}_i\right\}\mathbf{y}_i\mathbf{y}_i'\right] = A^{(m+1)}. \end{aligned}$$

So $A^{(m)}$ is a nonincreasing sequence of positive definite matrices and hence has a positive semidefinite limit A. It suffices then to show that A is nonsingular.

Let R represent the orthogonal projection into V, the range space of A, and let $r = \dim(V)$. Let R_m represent the orthogonal projection into V_m, the space spanned by the eigenvectors associated with the r largest eigenvalues of A_m. It follows that for $\mathbf{y} \notin V$,

$$\mathbf{y}'\{A^{(m)}\}^{-1}\mathbf{y} \to \infty \quad \text{and} \quad \frac{\mathbf{y}'(I - R_m)(A^{(m)})^{-1}\mathbf{y}}{\mathbf{y}'(A^{(m)})^{-1}\mathbf{y}} \to 1. \tag{5.8}$$

Also, for large enough m, $\operatorname{trace}(R_m) = r$.

Premultiplying

$$A^{(m)} \geq A^{(m+1)} = \operatorname{ave}\left[u\left\{\mathbf{y}_i'(A^{(m)})^{-1}\mathbf{y}_i\right\}\mathbf{y}_i\mathbf{y}_i'\right]$$

by $(I - R_m)(A^{(m)})^{-1}$ and taking the trace gives for large m,

$$p - r \geq \operatorname{ave}\left[u\left\{\mathbf{y}_i'(A^{(m)})^{-1}\mathbf{y}_i\right\}\mathbf{y}_i'(I - R_m)(A^{(m)})^{-1}\mathbf{y}_i\right]. \tag{5.9}$$

By considering only the sum in (5.9) over $\mathbf{y}_i \notin V$ and taking the limit, the results (5.8) imply $p - r \geq \{1 - P_n(V)\}a_0$. This contradicts Condition D_0 unless $V = \mathbb{R}$. Thus A is nonsingular. □

PROOF OF LEMMA 3.1(iii). Let $p < a_0 \le p + 1$, $n > a_0/(a_0 - p)$, and let H be a k-dimensional hyperplane, $0 \le k \le p - 1$. Since the data points are in general position.

$$P_n(H) \le \frac{k+1}{n} < \frac{(k+1)(a_0 - p)}{a_0} = 1 - \frac{p - k(a_0 - p)}{a_0} \le 1 - \frac{p-k}{a_0}.$$

Hence Condition D* holds. □

PROOF OF THEOREM 3.2. Note that $u^*(s;\lambda)$ is nonincreasing in s and that $su^*(s;\hat{\lambda})$ is assumed to be strictly increasing in s. Suppose $\hat{\boldsymbol{\mu}}_0 \in \mathbb{R}^p$ and $\hat{\boldsymbol{\Sigma}}_0 \in \mathscr{P}_p$ represent another solution to (1.2) and (1.3) and let $\hat{\lambda}_0^{-1} = \text{ave}[u\{(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_0)^{-1}\hat{\boldsymbol{\Sigma}}_0^{-1}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_0)\}]$. If $\hat{\lambda}_0 = \hat{\lambda}$, then $\hat{\boldsymbol{\mu}}_0 = \hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}_0 = \hat{\boldsymbol{\Sigma}}$. Otherwise there would be two distinct solutions to (3.7), which contradicts Theorem 2.2. Thus it only needs to be shown that $\hat{\lambda}_0 = \hat{\lambda}$.

If $\hat{\lambda}_0 < \hat{\lambda}$, then $u^*(s;\hat{\lambda}_0) \le u^*(s;\hat{\lambda})$ since $u(s)$ is nonincreasing. Define a sequence $\hat{A}_k \in \mathscr{P}_p$ starting with the above $\hat{A}_0$ by

$$\hat{A}_{k+1} = \text{ave}\left\{u^*\left(\mathbf{z}_i'\hat{A}_k^{-1}\mathbf{z}_i;\hat{\lambda}\right)\mathbf{z}_i\mathbf{z}_i'\right\}. \tag{5.10}$$

Note that $\hat{A}_1 \ge \text{ave}\{u^*(\mathbf{z}_i'\hat{A}_0^{-1}\mathbf{z}_i;\hat{\lambda}_0)\mathbf{z}_i\mathbf{z}_i'\} = \hat{A}_0$. Also, if $\hat{A}_{k+1} \ge \hat{A}_k$, then $\mathbf{z}'\hat{A}_{k+1}^{-1}\mathbf{z} \le \mathbf{z}'\hat{A}_k^{-1}\mathbf{z}$ and so $u^*(\mathbf{z}'\hat{A}_{k+1}^{-1}\mathbf{z};\hat{\lambda}) \ge u(\mathbf{z}'\hat{A}_k^{-1}\mathbf{z};\hat{\lambda})$ which implies $\hat{A}_{k+2} \ge \hat{A}_{k+1}$. By induction, $\hat{A}_0 \le \hat{A}_1 \le \cdots \le \hat{A}_k \le \cdots$. However, by Theorem 2.3, $\hat{A}_k \to \hat{A}$ and thus $\hat{A}_0 \le \hat{A}$. Since $\hat{\lambda}_0^{-1}$ and $\hat{\lambda}^{-1}$ are the $(p+1, p+1)$ entries of $\hat{A}_0$ and $\hat{A}$, respectively, this implies $\hat{\lambda}_0 \ge \hat{\lambda}$, a contradiction.

A similar contradiction arises if $\hat{\lambda}_0 > \hat{\lambda}$. Thus $\hat{\lambda}_0 = \hat{\lambda}$ and hence $\hat{\boldsymbol{\mu}}_0 = \hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}_0 = \hat{\boldsymbol{\Sigma}}$. □

PROOF OF LEMMA 3.3. Suppose for $y \in \mathbb{R}$, $P_n(y) = m/n > 1 - 1/a_0$. This implies $n > (n-m)a_0$ and so for some $a > a_0$, $n > (n-m)a$. For such a, express

$$\begin{aligned}\exp\{-L(y,\sigma^2)\} &= \sigma^{-\{n-(n-m)a\}}\exp\{-m\rho(0)\} \\ &\quad\times \prod_{y_i \ne y}(\sigma^2)^{a/2}\exp\left[-\rho\left\{(y_i - y)^2/\sigma^2\right\}\right],\end{aligned}$$

which is seen to go to ∞ as $\sigma \to 0$. Thus $L(y,\sigma^2) \to -\infty$ as $\sigma \to 0$, whereas $L(\mu,\sigma^2) > -\infty$ for any $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$. □

Acknowledgment. We wish to thank Yehuda Vardi for helpful discussions and for drawing our attention to the results on the t-distribution given in Dempster, Laird and Rubin (1980).

REFERENCES

BARNDORFF-NIELSEN, O. and BLAESILD, P. (1980). Global maxima and likelihood in linear models. Research Report 57, Dept. Theoretical Statistics, Aarhus Univ.

COPAS, J. B. (1975). On the unimodality of the likelihood for the Cauchy distribution. *Biometrika* **62** 701–704.

DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. (1980). Iteratively reweighted least squares for linear regression where errors are normal independent distributed. In *Multivariate Analysis V* (P. R. Krishnaiah, ed.) 35–37. North-Holland, New York.

GABRIELSON, G. (1982). On the unimodality of the likelihood for the Cauchy distribution: Some comments. *Biometrika* **69** 677–678.

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

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

KENT, J. T. and TYLER, D. E. (1988). Maximum likelihood estimation for the wrapped Cauchy distribution. *Journal of Applied Statistics* **15** 247–254.

MARDIA, K. V., KENT, J. T. and BIBBY, J. M. (1979). *Multivariate Analysis*. Academic, New York.

MÄRKELÄINEN, T., SCHMIDT, K. and STYAN, G. P. H. (1981). On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed sample sizes. *Ann. Statist* **9** 758–767.

MARONNA, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4** 51–67.

REEDS, J. A. (1985). Asymptotic number of roots of Cauchy likelihood equations. *Ann. Statist* **13** 778–784.

RUBIN, D. B. (1983). Iteratively reweighted least squares. In *Encyclopedia of Statistical Sciences* **4** 272–275. Wiley, New York.

TYLER, D. E. (1988). Some results on the existence, uniqueness and computation of the M-estimates of multivariate location and scatter. *SIAM J. Sci. Statist. Comput.* **9** 354–362.

DEPARTMENT OF STATISTICS
UNIVERSITY OF LEEDS
LEEDS LS2 9JT
ENGLAND

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903