The Annals of Statistics
2001, Vol. 29, No. 1, 252–265

REGULARITY AND UNIQUENESS FOR CONSTRAINED *M*-ESTIMATES AND REDESCENDING *M*-ESTIMATES

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

University of Leeds and Rutgers University

Constrained M-estimates of multivariate location and scatter are found by finding the global minimum of an objective function subject to a constraint. They are related to redescending M-estimates of multivariate location and scatter since any stationary point of the objective function corresponds to such an M-estimate. Unfortunately, even for the population form of the estimator, that is, the constrained M-functional, the objective function may have multiple stationary points. In this paper, we give conditions under which the objective function is as well behaved as possible, in particular that it has at most one local minimum. To carry out this task, we introduce a class of distributions which we call "regular" distributions with respect to a particular objective function.

1. Introduction and summary. Constrained M-estimators (CM-estimators) of multivariate location and scatter were introduced in Kent and Tyler (1996) to combine the good local robustness properties of M-estimators and the good global robustness properties of S-estimators. The constrained M-estimators are defined via the minimization of an objective function subject to some constraint. As with S-estimators, the critical points of this minimization problem correspond to the solutions of redescending M-estimates of multivariate location and scatter and hence there may be multiple critical points. The purpose of this paper is to give conditions under which the constrained M-minimization problem is as well behaved as possible.

Let $\rho(s)$ be a given function of $s \geq 0$ satisfying

$$\lim_{s\to 0}\rho(s) = \rho(0) = 0,\ \lim_{s\to\infty}\rho(s) = 1 \quad \text{and} \quad \rho(s) \text{ is nondecreasing.} \tag{1.1}$$

Let $c > 0$ be a "tuning parameter", and let $0 < \varepsilon < 1$ be a "breakdown parameter." If F is a nondegenerate distribution in $\mathbb{R}^p$, $p \geq 1$, the "constrained M-functionals" of location and scatter, $\boldsymbol{\mu}(F)$ and $\Sigma(F)$, say, are defined by minimizing

$$L(\boldsymbol{\mu}, \Sigma) = cE\{\rho(S)\} + \tfrac{1}{2}\log\det(\Sigma) \tag{1.2}$$

over $(\boldsymbol{\mu}, \Sigma)$, $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\Sigma(p \times p)$ positive definite, subject to the constraint

$$E\{\rho(S)\} \leq \varepsilon\rho(\infty) = \varepsilon. \tag{1.3}$$

Here $S = (\mathbf{X} - \boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu})$ is the squared Mahalanobis distance between $\mathbf{X}$ and $\boldsymbol{\mu}$, with $\mathbf{X} \sim F$. In practice, $\boldsymbol{\mu}(F)$ and $\Sigma(F)$ will usually be uniquely

Received September 1997; revised July 2000.
[1]Supported in part by NSF Grant IRI-95-30546.
AMS 2000 *subject classifications*. Primary 62F35, 62H10; secondary 60E07.
Key words and phrases. Elliptic symmetry, infinite divisibility, M-estimates, Schur concavity.

determined, but the theory allows for the possibility of ties. The terminology "constrained M-estimator" and the notation $(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$ are used when F is given by an empirical distribution F_n.

We call the parameter ε a "breakdown parameter" since it is equivalent to the asymptotic breakdown point of the constrained M-estimator whenever $\varepsilon \leq 1/2$. The value of ε can be chosen to be 1/2 in practice. We call the parameter c a 'tuning constant' since the value of c greatly affects the influence function and the asymptotic efficiency of the constrained M-estimator. Tuning a CM-estimator does not affect its breakdown point. This is in contrast to the S-estimators: tuning them to obtain desirable properties for the influence function and desirable levels of asymptotic efficiency affects their breakdown point. For a more detailed discussion, see Kent and Tyler (1996).

When ρ is differentiable, the critical points of $L(\boldsymbol{\mu}, \Sigma)$ correspond to redescending M-functionals of multivariate location and scatter. That is, they are solutions to the M-functional equations

$$\boldsymbol{\mu} = E\{u(S)\mathbf{X}\}/E\{u(S)\} \tag{1.4}$$

and

$$\Sigma = cE\{u(S)(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T\}, \tag{1.5}$$

where $u(s) = 2\rho'(s)$ and again $S = (\mathbf{X} - \boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu})$. The forms of the influence functions of location and scatter depend upon $s^{1/2}u(s)$ and $su(s)$, respectively, and these redescend to zero whenever ρ is bounded, provided $u(s)$ is nonincreasing; see (3.2) below. Redescending M-estimates based on the empirical versions of (1.4) and (1.5) have been used successfully by Rocke and Woodruff (1996) to identify outliers in high-dimensional datasets.

Next we consider a class of distributions possessing enough symmetry to determine the solution to (1.2) and (1.3). Tatsuoka and Tyler (2000) define the class $\mathscr{F}_p$, say, to consist of the distributions of random vectors $\mathbf{X} \in \mathbb{R}^p$ satisfying the following two conditions:

1. There exists a nonsingular $p \times p$ matrix A and a p-dimensional vector $\boldsymbol{\mu}_0$ such that $U = A^{-1}(\mathbf{X} - \boldsymbol{\mu}_0)$ has a pdf $f(\mathbf{u})$ which is invariant under permutations and sign changes of the components of $\mathbf{u}$. Write $\Sigma_0 = AA^T$.
2. $f(\exp(v_1), \dots, \exp(v_p))$ is Schur concave in $v \in \mathbb{R}^p$. In particular, this property implies that $f(\mathbf{u})$ is radially nonincreasing.

The class $\mathscr{F}_p$ contains all densities $g(\mathbf{x})$ for which (1) holds and for which the level sets $\{\mathbf{x}\colon g(\mathbf{x}) \geq c\}$ are convex sets. In turn this latter class contains the class $\mathscr{E}_p$, say, of elliptical distributions with pdfs which are radially nonincreasing.

If a random vector $\mathbf{X}$ satisfying (1) happens to have finite second moments, then it must satisfy $E(\mathbf{U}) = 0$, $\operatorname{var}(\mathbf{U}) \propto I$, and so $E(\mathbf{X}) = \boldsymbol{\mu}_0$, $\operatorname{var}(\mathbf{X}) \propto \Sigma_0 = AA^T$. Thus we expect the solutions of (1.2) subject to (1.3) to take the form $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, $\Sigma = \Sigma_0/\kappa$ for some $\kappa > 0$. Tatsuoka and Tyler [(2000), Theorem 4.2] confirm this result for $F \in \mathscr{F}_p$ under the additional mild conditon

(1.6) $\rho(s)$ is strictly increasing, or $f(\mathbf{u})$ is strictly radially decreasing.

If (2) does not hold in the definition of $\mathscr{F}_p$, then Tatsuoka and Tyler (2000) give examples for which the solution of (1.2) subject to (1.3) is not of the form $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, $\Sigma = \Sigma_0/\kappa$.

Thus, for a constrained M-functional to be uniquely defined at $F \in \mathscr{F}_p$, it is necessary and sufficient that the "scaling parameter" κ be uniquely defined. Although the scaling parameter is often viewed as a nuisance parameter, the uniqueness of κ is essential in obtaining the influence functions of the constrained M-functionals and the asymptotic distributions of the constrained M-estimates; see Kent and Tyler (1996) for details. Further, the parameter κ is important for summarizing the concentration of the distribution. The value of κ for the constrained M-functional corresponds to the minimum of the "profile objective function"

$$l(\kappa) = cE\{\rho(\kappa S_0)\} - \frac{p}{2}\log\kappa, \tag{1.7}$$

where $S_0 = (\mathbf{X} - \boldsymbol{\mu}_0)^T \Sigma_0^{-1} (\mathbf{X} - \boldsymbol{\mu}_0)$, subject to the constraint (1.3). Note that a constant term $\frac{1}{2}\log\det(\Sigma_0)$ has been dropped in (1.7). The constraint (1.3) can be reexpressed as $\kappa \le \kappa_0$ with κ_0 being the solution to

$$E\{\rho(\kappa_0 S_0)\} = \varepsilon. \tag{1.8}$$

In general, for a given ρ function and a given distribution F, the uniqueness of κ can be checked numerically since $l(\kappa)$ is a univariate function. One of the main goals of this paper is to give a more detailed theoretical study of the scaling parameter κ.

We first note that $E\{\rho(\kappa S_0)\}$ is bounded and increasing in κ whereas $-\log\kappa$ decreases from $+\infty$ to $-\infty$. Hence, $l(\kappa)$ is dominated by $-(p/2)\log(\kappa)$ for small and large κ. However, for large enough c, $l(\kappa)$ will be increasing on at least one interval of κ values.

Hence, the following definition describes the simplest possible behavior of $l(\kappa)$ and its dependence on c. The underlying randomness can be specified either in terms of the distribution of the random vector $\mathbf{X}$, or equivalently in terms of the distribution of the squared Mahalanobis distance $S_0 = (\mathbf{X} - \boldsymbol{\mu}_0)^T \Sigma_0^{-1} (\mathbf{X} - \boldsymbol{\mu}_0)$.

DEFINITION 1.1. The distribution of a random variable $S_0 > 0$ is said to be "regular" with respect to $\rho(\cdot)$ if there exists $c_0 > 0$ such that:

(a) For $c < c_0$, $l(\kappa)$ has no local maxima or minima.
(b) For $c = c_0$, $l(\kappa)$ has just one critical point (which is also an inflection point).
(c) For $c > c_0$, $l(\kappa)$ has just two critical points, one local maximum and one local minimum.

Regularity of S_0 is clearly a sufficient but not necessary condition for $l(\kappa)$ to have a unique minimum over $\kappa \leq \kappa_0$, except for the rare case whenever $c > c_0$ and the local minimum of $l(\kappa)$ is equal to $l(\kappa_0)$. In this case, the optimal value of the scaling parameter κ corresponds to κ_0 as well as to the value of κ which produces the local minimum of $l(\kappa)$. Consequently, except for this special case, regularity of S_0 insures the uniqueness of the constrained M-functional at $F \in \mathscr{F}_p$.

The regularity of $S_0 \sim \chi_p^2$, that is, when $\mathbf{X}$ has a normal distribution, was established in an earlier unpublished version of Kent and Tyler (1996) in all dimensions $p \geq 1$ for two choices of ρ function: the biweight ρ function and the exponentially weighted ρ function. The biweight ρ function is so named since its weight function $u(s) = 2\rho'(s)$ corresponds to a Tukey biweight function. The exponentially weighted ρ function,

$$\rho(s) = 1 - e^{-s}, \tag{1.9}$$

is so named since its corresponding weight function is $u(s) = 2e^{-s}$. This weight function was first proposed by Dennis and Welsch (1976) within the regression setting and independently proposed within the computer vision literature as a tractable alternative to the biweight function; see Li (1995). The extension of regularity to a wider class of distributions beyond the chi-squared distribution is quite delicate and so far analytic progress has been made only with the exponentially weighted ρ function. The main result of this paper, given in Section 2, establishes the regularity of a certain class of infinitely divisible distributions for S_0 with respect to (1.9). This class of regular distributions for S_0 includes both long- and short-tailed distributions for the corresponding distribution of $\mathbf{X}$.

The uniqueness of the constrained M-functionals as discussed so far refers to the uniqueness of the global minimum of $L(\boldsymbol{\mu}, \Sigma)$ under the constraint (1.3). This leaves open the question of the nature of the critical points of $L(\boldsymbol{\mu}, \Sigma)$, that is, the nature of the solutions of (1.4) and (1.5). We address this question in Section 3 for the class $\mathscr{E}_p$ of elliptically symmetric distributions with nonincreasing radial pdfs. The method of proof does not seem to generalize to the wider class $\mathscr{F}_p$. In particular, we show that if $F \in \mathscr{E}_p$ and further mild regularity conditions hold, then any critical points of $L(\boldsymbol{\mu}, \Sigma)$ must be of the form $\boldsymbol{\mu}(F) = \boldsymbol{\mu}_0$ and $\Sigma(F) = \Sigma_0/\kappa$. Moreover, any critical point of $L(\boldsymbol{\mu}, \Sigma)$ must be either a local minimum or a saddlepoint. So, when the distribution of S_0 is regular and case (c) holds, $L(\boldsymbol{\mu}, \Sigma)$ has one local minimum and one saddlepoint. When case (b) holds, it has one saddlepoint but no local minimum, and when case (a) holds, it has no critical points. Some implications of these results for redescending M-estimates are discussed in Section 3.

In a limited sense, regularity carries over to finite samples. Suppose the population objective functional $L(\boldsymbol{\mu}, \Sigma)$ possesses exactly k critical points, and that at these critical points the second derivative matrix of L is nonsingular. Let C be a compact region in parameter space whose interior contains these critical points. Then for large enough n, $L_n(\boldsymbol{\mu}, \Sigma)$ will also have exactly k critical points on C. This result is a simple consequence of the fact that the first two derivatives of L_n converge to those of L uniformly on C.

2. The main regularity theorem. In this section we look at the critical points of the profile objective function (1.7). If we let

$$\chi(\kappa) = E\{\rho(\kappa S_0)\} \tag{2.1}$$

denote the "ρ-transform" of S_0, then (1.7) takes the form

$$l(\kappa) = c\chi(\kappa) - \frac{p}{2}\log\kappa \tag{2.2}$$

so the critical points satisfy

$$\kappa\chi'(\kappa) = \frac{p}{2c}. \tag{2.3}$$

It has already been assumed that $\rho(s)$ is a bounded nondecreasing function. If, in addition, $u(s) = 2\rho'(s)$ is a bounded nonincreasing function, it follows that $u(s) \to 0$ and $su(s) \to 0$ as $s \to \infty$.

Then it is easy to show that $\kappa\,\chi'(\kappa) \to 0$ as $\kappa \to 0$ and as $\kappa \to \infty$. Further, the local maxima and minima of $l(\kappa)$ come in pairs. Hence, S_0 will have a regular distribution (see Definition 1.1) if and only if any horizontal line intersects the graph of $\kappa\,\chi'(\kappa)$ versus κ in at most two places, which is true if and only if $\kappa\,\chi'(\kappa)$ is strictly unimodal.

When we specialize to $\rho(s) = 1 - e^{-s}$, this characterization can be described in terms of the Laplace transform (LT) of S_0,

$$\theta(\kappa) = E\{e^{-\kappa S_0}\}. \tag{2.4}$$

In this case $\chi(\kappa) = 1 - \theta(\kappa)$ and (2.3) becomes

$$-\kappa\,\theta'(\kappa) = p/2c, \tag{2.5}$$

with S_0 being regular if $-\kappa\,\theta'(\kappa)$ is strictly unimodal.

THEOREM 2.1. *The following classes of distributions for S_0 are regular with respect to $\rho(s) = 1 - e^{-s}$:*

(a) *The Bondesson power class $\mathscr{B}^*$;*
(b) *The stable distributions with $\theta(\kappa) = \exp(-c\kappa^a)$, $0 < a \le 1$, $c > 0$.*

For any class of infinitely divisible distributions $\mathscr{D}$, the "power class" $\mathscr{D}^*$ is defined here to mean all convolution powers of distributions in $\mathscr{D}$ of order $\lambda \ge 1$. If $f(s)$ is a density in $\mathscr{D}$ with LT $\theta(\kappa)$, then the convolution power $f^{*\lambda}(s)$, say, has LT $\{\theta(\kappa)\}^\lambda$. The fact that $\mathscr{D}$ is infinitely divisible ensures that the convolution powers are well-defined probability distributions for all real $\lambda \ge 0$, though here we are only interested in $\lambda \ge 1$. Note that the class of stable distributions is closed under convolutions, so equals its own power class. For a brief summary of the stable distributions on $(0, \infty)$, see, for example, Feller (1966), page 424].

Before proving the theorem we give a brief description of the Bondesson class. For further details see Bondesson [(1992), especially pages 73, 29, 30, 71, 51, 69] where this class is denoted by both $\mathscr{B}$ and $\mathscr{C}$. Let S be a random variable on $(0, \infty)$ with pdf $f(s)$ and Laplace transform $\theta(\kappa)$.

1. The class $\mathscr{B}$ consists of all probability densities of the form

$$f(s) = Cs^{\beta-1} \prod_{j=1}^{m} (1 + c_j s)^{-\gamma_j}, \qquad s > 0, \tag{2.6}$$

   where $m \geq 1, \beta > 0, c_j > 0, \gamma_j > 0$, and their weak limits. For simplicity of exposition, limiting distributions concentrated at a single point are excluded from $\mathscr{B}$. All densities in $\mathscr{B}$ have support $(0, \infty)$.
2. Another useful class of distributions on $(0, \infty)$ is the "Thorin class" $\mathscr{T}$ of generalized gamma convolutions. A random variable S lies in $\mathscr{T}$ if and only if

$$\theta(\kappa) = \exp\left\{ - b\kappa - \int \log(1 + \kappa/y) U(dy) \right\}, \tag{2.7}$$

   where $U(dy)$ is a measure on $(0, \infty)$ satisfying

$$\int_{(0,\,1)} |\log y| U(dy) < \infty, \qquad \int_{[1,\,\infty)} y^{-1} U(dy) < \infty.$$

   The parameter $b \geq 0$ gives the left-hand endpoint of the support of S.
3. Thus $\mathscr{T}$ can be described as the class of weak limits of convolutions of gamma distributions, and hence all distributions in $\mathscr{T}$ are infinitely divisible. It can be shown that $\mathscr{B}$ is a proper subclass of $\mathscr{T}$ with $b = 0$ in (2.7).
4. If $f(s) \in \mathscr{B}$, let $\beta = \int_0^\infty U(dy)$ in (2.7) with $0 < \beta \leq \infty$. Then β can be alternatively defined by $\beta = \sup\{\delta\colon \lim_{s \downarrow 0} f(s)/s^{\delta-1} = 0\}$. Say that β is the "order of $f(s)$."
5. If $f(s) \in \mathscr{B}$ with LT $\theta(\kappa)$, then $-s f(s)e^{-as}/\theta'(a) \in \mathscr{B}$ for all $a > 0$ [including $a = 0$ if $-\theta'(0+) < \infty$] with LT $\theta'(\kappa + a)/\theta'(a)$. Note that if $f(s)$ is of order β, then $-s f(s)e^{-as}/\theta'(a)$ is of order $\beta + 1$.
6. If $S \in \mathscr{B}$, then $S^\delta \in \mathscr{B}$ for all real $|\delta| \geq 1$. Here S^δ denotes the ordinary power of S, not a convolution power.
7. If S and T are independent random variables in β, then $ST \in \beta$. In particular, if $T \sim \chi^2_\nu$ with $\nu > 0$, then $ST \in \beta$.

Properties 6 and 7 are rather unusual for a class of infinitely divisible distributions. Further, $\mathscr{B}$ is *not* closed under convolution or convolution roots.

Bondesson's class $\mathscr{B}$ should not be confused with the larger class of infinitely divisible distributions $\mathscr{B}_0$, say, of "generalized convolutions of mixtures of exponential distributions," also studied in Bondesson [(1992), page 137].

PROOF OF THEOREM 2.1. (a) Let $f(s) \in \mathscr{B}$ with LT $\theta(\kappa)$. We wish to show that the convolution power $f^{*\lambda}(s)$, say, for $\lambda \geq 1$ is regular. Write the LT of $f^{*\lambda}(s)$ as $\{\theta(\kappa)\}^{\lambda} = \theta_{\lambda}(\kappa)$ for simplicity. Since $f(s) \in \mathscr{B}$ so is $s\ f(s)$ (up to a constant factor), with LT $-\theta'(\kappa)$. [This statement assumes $-\theta'(0+) < \infty$. If not, work with $-\theta'(\kappa + a)$ below and use a limiting argument as $a \to 0$ to deduce regularity.] Since $\mathscr{B} \subset \mathscr{T}$, we can express $\theta(\kappa)$ in the form (2.7) with $b = 0$, and $\theta'(\kappa)$ in the form

$$-\theta'(\kappa) = C_1 \cdot \exp\Big\{-\int \log(1+\kappa/y)V(dy)\Big\}$$

for some measure $V(dy)$. Hence

$$\begin{aligned} -\theta_{\lambda}'(\kappa) &= -\theta'(\kappa)\,\theta(\kappa)^{\lambda-1} \\ &= C_2 \cdot \exp\Big\{-\int \log(1+\kappa/y)V_{\lambda}(dy)\Big\}, \end{aligned}$$

where $V_{\lambda}(dy) = V(dy) + (\lambda - 1)U(dy)$. Differentiating $\theta_{\lambda}'(\kappa)$ yields

$$\theta_{\lambda}''(\kappa) = -\Big\{\int \frac{1}{\kappa + y} V_{\lambda}(dy)\Big\}\theta_{\lambda}'(\kappa).$$

Setting $\psi(\kappa) = -\kappa\theta_{\lambda}'(\kappa)$, we find

$$\begin{aligned} \psi'(\kappa) &= -\theta_{\lambda}'(\kappa) - \kappa\theta_{\lambda}''(\kappa) \\ &= -\theta_{\lambda}'(\kappa)\Big[1 - \int \frac{\kappa}{\kappa + y} V_{\lambda}(dy)\Big] \\ &= -\theta_{\lambda}'(\kappa)[1 - b(\kappa)] \quad \text{say.} \end{aligned}$$

Note that $b(\kappa)$ is strictly monotone increasing with $b(0) = 0$ and $b(\kappa) \to \lambda\beta + 1$ as $\kappa \to \infty$. Here β is the order of $f(s)$ and $\beta + 1 = \int V(dy)$ is the order of $sf(s)$, $0 < \beta \leq \infty$. If we let κ_0 denote the unique value of κ satisfying $b(\kappa_0) = 1$, then $\psi'(\kappa) > 0$ for $\kappa < \kappa_0$ and $\psi'(\kappa) < 0$ for $\kappa > \kappa_0$; that is, $\psi(\kappa)$ is strictly unimodal as required.

(b) For the stable case, let $\theta(\kappa) = \exp\{-c\kappa^{a}\}, 0 < a \leq 1$. Then $\psi(\kappa) = -\kappa\theta'(\kappa) = ac\,\kappa^{a}\exp\{-c\kappa^{a}\}$ which is easily seen to be strictly unimodal.

COMMENTS. 1. *Spherically symmetric distributions.* The class $\mathscr{B}$ contains a large number of well-known distributions including gamma, F, Pareto, log-normal, and generalized inverse Gaussian. Further, if $\mathbf{X} \in \mathbb{R}^p$ can be written as a scale mixture of multivariate normals, $\mathbf{X} = \mathbf{Z}\sqrt{T}$, with $\mathbf{Z} \sim N_p(\mathbf{0}, I)$ independent of $T \in \mathscr{B}$, then $\mathbf{Z}^T\mathbf{Z} \sim \chi_p^2$ and $S = \mathbf{X}^T\mathbf{X} = T\mathbf{Z}^T\mathbf{Z} \in \mathscr{B}$ also. In particular, if $\mathbf{X}$ follows a multivariate t distribution with $\nu > 0$ degress of freedom, then $S \in \mathscr{B}$. The multivariate t distribution is the single most common example for robustness studies, and in particular includes long-tailed distributions for ν near 0.

2. *Distributions with i.i.d. components.* It is also possible to construct vectors $\mathbf{X}$ of i.i.d. components such that $S = \mathbf{X}^T\mathbf{X}$ satisfies the assumptions of Theorem 2.1. Suppose each component X_i is symmetrically distributed such that X_i^2 lies in the class of distributions in either part (a) or (b) of Theorem 2.1. Then $S = \sum_{i=1}^p X_i^2$ also lies in the class of part (a) or (b), respectively, thus ensuring the regularity of S. For example, if each X_i follows a t-distribution with $\nu > 0$ degrees of freedom, then X_i^2 follows an $F_{1,\nu}$ distribution which lies in $\mathscr{B}$, and S lies in $\mathscr{B}^*$. In addition Tatsuoka and Tyler (2000) show that this distribution for $\mathbf{X}$ lies in $\mathscr{F}_p$.

3. Some, but not all the stable distributions lie in $\mathscr{B}$. Bondesson [(1992), pages 85, 88] gives more details. The limiting case $a = 1$, $\theta(\kappa) = e^{-\kappa}$ corresponds to a point mass at $s = 1$. However, the assumption of a stable law for S does not seem very useful for robustness. For example if $a = 1/2$, the stable density for S takes the form $f(s) \propto s^{-3/2}\exp(-1/s)$, which dies away exponentially as $s \to 0$. Thus, under either an assumption of elliptic symmetry or i.i.d. components for $\mathbf{X}$, the density of $f(\mathbf{x})$ will vanish at $\mathbf{x} = \mathbf{0}$, which violates the assumption used in earlier sections that the density of a symmetric multivariate distribution decreases radially from the origin.

4. Not all distributions on $(0, \infty)$ are regular. Here is an example in $p = 2$ dimensions. Let q take two values $q_1 = 1$ and $q_2 = 50$, each with probability $1/2$ and let $\mathbf{X}|q \sim N_p(\mathbf{0}, qI)$. Then $S = \mathbf{X}^T\mathbf{X}$ is a mixture of two scaled χ_p^2 variates with Laplace transform

$$\theta(\kappa) = \tfrac{1}{2}\sum_{j=1}^{2}(1 + 2q_j\kappa)^{-p/2}$$

and with $\psi(\kappa) = -\kappa\theta'(\kappa)$ given by

$$\psi(\kappa) = \tfrac{1}{2}p\kappa\sum_{j=1}^{2}q_j(1 + 2q_j\kappa)^{-p/2-1}.$$

For these particular parameters it can be checked numerically that $\psi(\kappa)$ has two modes, at $\kappa = 0.012$ and $\kappa = 0.42$ of roughly equal height. A plot of $\psi(\kappa)$ versus $\log\kappa$ is given in Figure 1 and a plot of $l(\kappa)$ versus $\log\kappa$ is given in Figure 2 for $c = 8.3$. Note that $l(\kappa)$ has two local minima and two local maxima. If $\varepsilon = \frac{1}{2}$ the constraint (1.3) takes the form $\kappa \le 0.0707$ ($\log\kappa \le -2.65$) which implies the first local minimum of $l(\kappa)$ is the global constrained minimum.

If $p = 2$, then S follows a mixture of exponential distributions, a class which is known to be infinitely divisible. Hence this example also shows that not all infinitely divisible distributions are regular.

3. Stationary points under elliptical symmetry. In this section we study the nature of the solutions of the M-functional equations (1.4) and (1.5) when the distribution of $\mathbf{X}$ lies in $\mathscr{E}_p$. Note that the M-functionals are affine equivariant in the sense that if $\mathbf{X}$ is transformed to $\mathbf{X}^* = A\mathbf{X} + \mathbf{b}$ where $A(p \times p)$ is nonsingular and $\mathbf{b} \in \mathbb{R}^p$, then $\boldsymbol{\mu}(F^*) = \boldsymbol{\mu}(F) + \mathbf{b}$, $\Sigma(F^*) = A\Sigma(F)A^T$. Hence, without loss of generality we may restrict attention to the case where

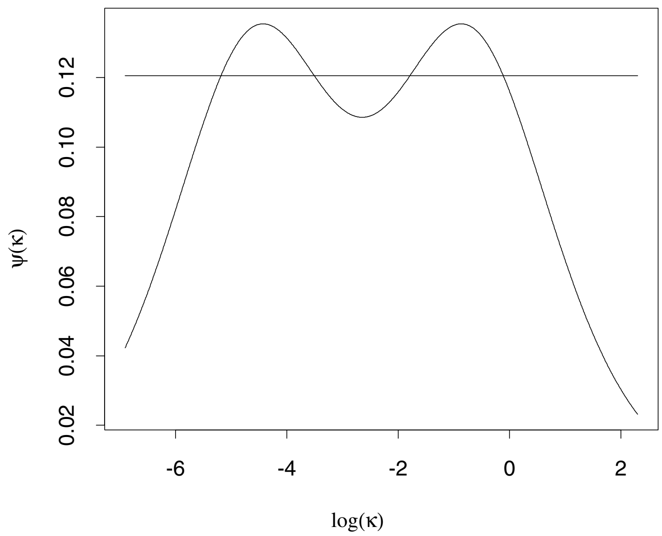

FIG. 1. *A plot of $\psi(\kappa)$ for the example in Comment* 4 *of Section* 2. *Note the bimodality of $\psi(\kappa)$. The horizontal line has a value of $\psi = p/2c = 2/(2 \cdot 8.3) = 0.12$ and intersects this curve in four places which are the critical points of $l(\kappa)$.*

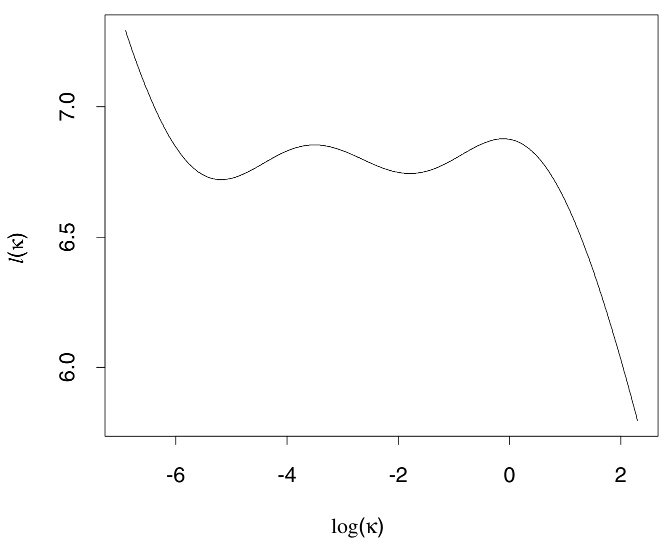

FIG. 2. *A plot of $l(\kappa)$ for the example in Comment* 4 *of Section* 2 *with $c = 8.3$. Note the presence of four critical points, two of which are local minima, thus demonstrating the lack of regularity.*

$\mathbf{X}$ is spherically symmetric about the origin with density $f(\mathbf{x}) = f_0(s)$ say, where $s = \mathbf{x}^T\mathbf{x}$.

The assumptions of Section 1 already imply that:

1. $\rho(s)$ is nondecreasing and $f_0(s)$ is nonincreasing in $s > 0$.

For this section we need the following additional mild regularity conditions, where $u(s) = 2\rho'(s)$:

2. $\rho(s)$ is continuous in $s > 0$ and is continuously differentiable for all but a finite set of s values.
3. $s^{1/2}u(s)$ is bounded over $s > 0$.
4. $s\,u(s)$ is bounded over $s > 0$.
5. In some interval $0 < s < s^*$, $\rho(s)$ is strictly increasing and $f_0(s)$ is strictly decreasing.
6. $f_0(s)$ is continuous in $s > 0$ and is continuously differentiable for all but a finite set of s values.

The last condition is just used in Proposition 3.2 below; the other conditions are used in both propositions.

For derivative calculations it is convenient to parameterize $V = \Sigma^{-1}$ in three pieces,

$$V = \kappa\Gamma\Delta\Gamma^T, \tag{3.1}$$

where $\kappa > 0, \Gamma$ is an orthogonal matrix, and $\Delta = \mathrm{diag}(e^{\delta_i})$ is a diagonal matrix with $\sum \delta_i = 0$ so that $\det(\Delta) = 1$. Let $\nabla_{\boldsymbol{\mu}} L(\boldsymbol{\mu}, \Sigma)$ and $\nabla_{\boldsymbol{\delta}} L(\boldsymbol{\mu}, \Sigma)$ denote the gradient vectors of L with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\delta} = (\delta_1, \dots, \delta_p)^T$, ignoring the constraint on $\boldsymbol{\delta}$. This constraint can be accommodated by considering directional derivatives $\boldsymbol{\beta}^T\nabla_{\boldsymbol{\delta}} L$ where $\sum \beta_i = 0$. If $H((p-1)\times p)$ is a rank $p-1$ matrix whose row sums all vanish, then the vector $H\nabla_{\boldsymbol{\delta}} L$ summarizes all the directional derivative information. Similarly, let $\nabla_{\boldsymbol{\mu}}\nabla_{\boldsymbol{\mu}}^T L$ and $\nabla_{\boldsymbol{\delta}}\nabla_{\boldsymbol{\delta}}^T L$ denote the $(p \times p)$ matrices of second partial derivatives.

The main results of this section are given by the following two properties.

PROPOSITION 3.1. *Under conditions* (1)–(5):

(a) $\nabla_{\boldsymbol{\mu}} E\{\rho(\kappa(\mathbf{X}-\boldsymbol{\mu})^T\Delta(\mathbf{X}-\boldsymbol{\mu}))\} = \mathbf{0}$ *if and only if* $\boldsymbol{\mu} = \mathbf{0}$.
(b) $H\nabla_{\boldsymbol{\delta}} E\{\rho(\kappa\mathbf{X}^T\Delta\mathbf{X})\} = \mathbf{0}$ *if and only if* $\boldsymbol{\delta} = \mathbf{0}$, *that is*, $\Delta = I$.

PROPOSITION 3.2. *Under conditions* (1)–(6):

(a) $\nabla_{\boldsymbol{\mu}}\nabla_{\boldsymbol{\mu}}^T E\{\rho(\kappa(\mathbf{X}-\boldsymbol{\mu})^T(\mathbf{X}-\boldsymbol{\mu}))\}|_{\boldsymbol{\mu}=\mathbf{0}}$ *is positive definite.*
(b) $H\nabla_{\boldsymbol{\delta}}\nabla_{\boldsymbol{\delta}}^T E\{\rho(\kappa\mathbf{X}^T\Delta\mathbf{X})H^T\}|_{\boldsymbol{\delta}=\mathbf{0}}$ *is positive definite.*
(c) $H\nabla_{\boldsymbol{\delta}}\nabla_{\boldsymbol{\mu}}^T E\{\rho(\kappa(\mathbf{X}-\boldsymbol{\mu})^T\Delta(\mathbf{X}-\boldsymbol{\mu}))\}|_{\boldsymbol{\mu}=\mathbf{0},\,\boldsymbol{\delta}=\mathbf{0}} = 0$.
(d) $(\partial/\partial\kappa)\nabla_{\boldsymbol{\mu}} E\{\rho(\kappa(\mathbf{X}-\boldsymbol{\mu})^T(\mathbf{X}-\boldsymbol{\mu}))\}|_{\boldsymbol{\mu}=\mathbf{0}} = 0$.
(e) $(\partial/\partial\kappa)H\nabla_{\boldsymbol{\delta}} E\{\rho(\kappa\mathbf{X}^T\Delta\mathbf{X})\}|_{\boldsymbol{\delta}=\mathbf{0}} = \mathbf{0}$.

COMMENTS. 1. The proofs of these propositions are given at the end of this section. For the moment we focus on their implications about solutions to the M-functional equations (1.4) and (1.5), respectively.

2. By spherical symmetry, $L(\boldsymbol{\mu}, \Sigma)$ does not vary with Γ. Hence critical points of L are determined by $\nabla_{\boldsymbol{\mu}} L = \mathbf{0}, \partial L/\partial\kappa = 0$ and $H^T \nabla_{\boldsymbol{\delta}} L = \mathbf{0}$. Since $\log\det\Sigma = -p\log\kappa$ does not depend on $\boldsymbol{\mu}$ or $\boldsymbol{\delta}$, $\nabla_{\boldsymbol{\mu}} \log\det(\Sigma) = \mathbf{0}$ and $H^T\nabla_{\boldsymbol{\delta}} \log\det(\Sigma) = \mathbf{0}$. Hence from Proposition 3.1, any critical point of $L(\boldsymbol{\mu}, \Sigma)$ must satisfy $\boldsymbol{\mu} = \mathbf{0}$ and $\Sigma = I/\kappa$ for some $\kappa > 0$. A similar conclusion is reached by Hampel, Ronchetti, Rousseeuw and Stahel [(1986), Theorem 2, page 288] under a somewhat more general framework. However, they do not include the results of Proposition 3.2, and they left open the question of uniqueness for κ, which is addressed in Section 2.

3. The nature of these critical points is determined by Proposition 3.2. Note that $\nabla_{\boldsymbol{\mu}}\nabla_{\boldsymbol{\mu}}^T L(\boldsymbol{\mu}, \Sigma) > 0$ at $\boldsymbol{\mu} = \mathbf{0}$ and $H\nabla_{\boldsymbol{\delta}}\nabla_{\boldsymbol{\delta}}^T \log\det(\Sigma) H^T > 0$ at $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\delta} = \mathbf{0}$. Hence $L(\boldsymbol{\mu}, \Sigma)$ always has a local minimum with respect to $\boldsymbol{\mu}$ and $\boldsymbol{\delta}$, respectively, at $\boldsymbol{\mu} = \mathbf{0}$, $\boldsymbol{\delta} = \mathbf{0}$. Therefore, with respect to all the parameters, the critical points of L can never be local maxima; they can only be local minima or saddle points, depending on the sign of $\partial^2 L/\partial\kappa^2$.

4. Conditions (1)–(6) are sufficient in the following proofs to ensure that the integrals converge, that the interchange of differentiation and integration is justified and that the strict inequalities hold where required. Note that in general (3) is more restrictive than (4) for s near 0, and vice versa for s near ∞. In practice $u(s)$ will usually be a bounded nonincreasing function, possibly with some discontinuities. In this case (3) and (4) hold automatically. In particular, $s\,u(s) \to 0$ as $s \to \infty$ in this case because

$$\tfrac{1}{2} s\, u(s) \le \int_{s/2}^{s} u(t)\, dt = 2\{\rho(s) - \rho(s/2)\}, \tag{3.2}$$

and the right-hand side tends to 0 as $s \to \infty$ since ρ is bounded and monotone. The fact that $f(\mathbf{x})$ is a pdf on $\mathbb{R}^p$ implies that $\int_0^\infty s^{(p-2)/2} f_0(s)\, ds < \infty$. Under (1) and (6), a result in Feller [(1966), page 148] based on integration by parts shows that this bound is equivalent to $-\int_0^\infty s^{p/2} f_0'(s)\, ds < \infty$. Hence $\mathbf{x}^T\mathbf{x} f'(\mathbf{x})$ is integrable over $\mathbb{R}^p$.

5. The results of this section can also be applied to the CM-functionals which minimize (1.2) subject to (1.3). For a given value of κ, the optimal choice of $\boldsymbol{\mu}$ and $\boldsymbol{\delta}$ [in the sense of minimizing (1.2) and making it easiest to satisfy the constraint (1.3)] must take the form $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\delta} = \mathbf{0}$ by Proposition 3.1. Hence, under the regularity conditions of this section, Propositions 3.1 and 3.2 yield an alternative proof that for $F \in \mathscr{E}_p$, the CM-functionals must always take the form $\boldsymbol{\mu} = \mathbf{0}$, $\Sigma = I/\kappa$. The original proof given in Kent and Tyler (1996) is based on uniqueness results for S-functionals given by Davies (1987).

PROOF OF PROPOSITION 3.1. (a) Let $S = \kappa(\mathbf{X}-\boldsymbol{\mu})\Delta(\mathbf{X}-\boldsymbol{\mu})$. Then

$$E\{\rho(S)\} = \int \rho(s) f_0(\mathbf{x}^T\mathbf{x})\, d\mathbf{x}$$

with derivative components, $j = 1, \dots, p$, given by

$$\begin{aligned}[\nabla_{\boldsymbol{\mu}} E\{\rho(S)\}]_j &= \kappa\left[\int \Delta(\boldsymbol{\mu}-\mathbf{x})\rho'(s) f_0(\mathbf{x}^T\mathbf{x})\, d\mathbf{x}\right]_j \\ &= \kappa e^{\delta_j}\int (\mu_j - x_j)\rho'(s) f_0(\mathbf{x}^T\mathbf{x})\, d\mathbf{x} \qquad (3.3)\\ &= -\kappa e^{\delta_j}\int y_j \rho'\Big(\kappa \sum e^{\delta_i} y_i^2\Big) f_0\Big(\sum (y_i+\mu_i)^2\Big)\, d\mathbf{y},\end{aligned}$$

where $y_i = x_i - \mu_i$. If $\boldsymbol{\mu} = \mathbf{0}$ this integral vanishes by symmetry. If not, choose a component j for which $\mu_j \neq 0$, and split the integral between $y_j > 0$ and $y_j < 0$ to get

$$[\nabla_{\boldsymbol{\mu}} E\{\rho(S)\}]_j = -\kappa e^{\delta_j}\int_{y_j>0} y_j \rho'\Big(\kappa \sum e^{\delta_i} y_i\Big)\{f_0(u_1) - f_0(u_2)\}\, d\mathbf{y},$$

where

$$u_1 = (y_j+\mu_j)^2 + \sum_{i\neq j}(y_i+\mu_i)^2, \qquad u_2 = (-y_j+\mu_j)^2 + \sum_{i\neq j}(y_i+\mu_i)^2.$$

If $\mu_j > 0$, then $u_1 > u_2$ and $f_0(u_1) \le f_0(u_2)$, with strict inequality for $\mathbf{y}$ near $\mathbf{0}$. Further, $\rho'(s) \ge 0$ for $s > 0$ with strict inequality for s near 0. Thus the integrand is nonnegative everywhere and strictly positive for $\mathbf{y}$ near $\mathbf{0}$. Thus the integral is positive, and conversely if $\mu_j < 0$. In either case we have established that $[\nabla_{\boldsymbol{\mu}} E\{\rho(S)\}]_j \neq 0$.

(b) When $\boldsymbol{\mu} = \mathbf{0}$,

$$E\{\rho(S)\} = \int \rho\Big(\kappa \sum e^{\delta_i} x_i^2\Big) f_0\Big(\sum x_i^2\Big)\, d\mathbf{x}$$

$$\begin{aligned}\big[\nabla_{\boldsymbol{\delta}} E\{\rho(S)\}\big]_j &= \kappa e^{\delta_j}\int x_j^2 \rho'\Big(\kappa \sum e^{\delta_i} x_i^2\Big) f_0\Big(\sum x_i^2\Big)\, d\mathbf{x} \qquad (3.4)\\ &= \kappa e^{-\sum \delta_i/2}\int y_j^2 \rho'\Big(\kappa \sum y_i^2\Big) f_0\Big(\sum e^{-\delta_i} y_i^2\Big)\, d\mathbf{y},\end{aligned}$$

where $y_i = e^{\delta_i/2} x_i$. If $\boldsymbol{\delta} = \mathbf{0}$ the value of this integral does not depend on j; hence $\boldsymbol{\beta}^T \nabla_{\boldsymbol{\delta}} E\{\rho(S)\} = 0$ wherever $\sum \beta_i = 0$ and so $H\nabla_{\boldsymbol{\delta}} E\{\rho(S)\} = \mathbf{0}$.

If $\boldsymbol{\delta} \neq \mathbf{0}$ and $\sum \delta_i = 0$, then there are at least two distinct components of $\boldsymbol{\delta}$, say $\delta_1 > \delta_2$ for simplicity. Let $\beta_1 = 1$, $\beta_2 = -1$, $\beta_i = 0$ for $i \neq 1, 2$, and split the integral in between $y_1^2 > y_2^2$ and $y_1^2 < y_2^2$ to get

$$\begin{aligned}\boldsymbol{\beta}^T \nabla_{\boldsymbol{\delta}} E\{\rho(S)\} &= \kappa \int (y_1^2 - y_2^2)\rho'\Big(\kappa \sum y_i^2\Big) f_0\Big(\sum e^{-\delta_i} y_i^2\Big)\, d\mathbf{y} \\ &= \kappa \int_{y_1^2 > y_2^2} (y_1^2 - y_2^2)\rho'\Big(\kappa \sum y_i^2\Big)\{f_0(u_1) - f_0(u_2)\}\, d\mathbf{y},\end{aligned}$$

where $u_1 = e^{\delta_1} y_1^2 + e^{\delta_2} y_2^2 + \sum_{i=3}^p e^{\delta_i} y_i^2$, $u_2 = e^{\delta_1} y_2^2 + e^{\delta_2} y_1^2 + \sum_{i=3}^p y_i^2$. Since $y_1^2 > y_2^2$, it follows that $u_1 > u_2$, so that $f_0(u_1) \le f_0(u_2)$ with strict inequality near $\mathbf{y} = \mathbf{0}$. Thus the integral is positive, so $H^T \nabla_{\boldsymbol\delta} E\{\rho(S)\} \ne \mathbf{0}$.

PROOF OF PROPOSITION 3.2. (a) Differentiating (3.3) and setting $\boldsymbol\mu = \mathbf{0}$, $\boldsymbol\delta = \mathbf{0}$, yields

$$\nabla_{\boldsymbol\mu} \nabla_{\boldsymbol\mu}^T E\{\rho(S)\} = -\kappa \int \mathbf{y}\mathbf{y}^T \rho'\Big(\kappa \sum y_i^2\Big) f_0'\Big(\sum y_i^2\Big)\, d\mathbf{y}.$$

Since $\rho' \ge 0$, $f_0' \le 0$, with strict inequality near $\mathbf{y} = \mathbf{0}$, the resulting integral is a positive definite matrix.

(b) Differentiating (3.4) with respect to $\boldsymbol\delta$ and setting $\boldsymbol\delta = \mathbf{0}$ yields

$$\nabla_{\boldsymbol\delta} \nabla_{\boldsymbol\delta}^T E\{\rho(S)\} = -\kappa \int \mathbf{z}\mathbf{z}^T \rho'\Big(\kappa \sum y_i^2\Big) f_0'\Big(\sum y_i^2\Big)\, d\mathbf{y}$$
$$-\tfrac{1}{2}\kappa \mathbf{1}_p \mathbf{1}_p^T \int y_1^2 \rho'\Big(\kappa \sum y_i^2\Big) f_0\Big(\sum y_i^2\Big)\, d\mathbf{y},$$

where $\mathbf{z}$ is a p-vector with entries $z_i = y_i^2$, and $\mathbf{1}_p$ is a p-vector of ones. Since $\rho' \ge 0$ and $f_0' \le 0$ with strict inequality near $\mathbf{y} = \mathbf{0}$, the first integral is positive definite. Thus, since $H\mathbf{1}_p = \mathbf{0}$, $H \nabla_{\boldsymbol\delta} \nabla_{\boldsymbol\delta}^T E\{\rho(S)\} H^T$ is positive definite.

(c)–(e) To show the mixed partials vanish, use symmetry arguments. Differentiate (3.3) with respect to $\boldsymbol\delta$ or κ, set $\boldsymbol\mu = \mathbf{0}$ and $\boldsymbol\delta = \mathbf{0}$ and note that the integral is an odd function of $\mathbf{y}$ to conclude that the integral is 0. Similarly, differentiate (3.4) with respect to κ, set $\boldsymbol\delta = \mathbf{0}$ and note that the answer does not depend on j. Hence $H(d/d\kappa) \nabla_{\boldsymbol\delta} E\{\rho(\kappa \mathbf{X}^T \Delta \mathbf{X})\}|_{\boldsymbol\delta=\mathbf{0}} \propto H\mathbf{1}_p = \mathbf{0}$.

4. Regression. The results of the paper also apply to the constrained M-estimators of regression given in Mendes and Tyler (1996). Consider the regression equation $y = \mathbf{x}'\boldsymbol\beta_0 + e$, and let F represent the joint distribution of $(y, \mathbf{x})$. The constrained M-functionals of regression and scale, say $\boldsymbol\beta(F)$ and $\sigma(F)$, respectively, are defined by minimizing

$$L(\boldsymbol\beta, \sigma) = cE\{\rho(r^2/\sigma^2)\} + \log\ \det\,(\sigma) \tag{4.1}$$

over $(\boldsymbol\beta, \sigma)$, $\boldsymbol\beta \in \mathbb{R}^q$ and $\sigma > 0$, subject to the constraint

$$E\{\rho(r^2/\sigma^2)\} \le \varepsilon, \tag{4.2}$$

with $r = y - \mathbf{x}'\boldsymbol\beta$.

The class $\mathscr{F}_1$ corresponds to the unimodal symmetric distributions. It is shown in Mendes and Tyler (1996) that if the distribution of e lies in $\mathscr{F}_1$, then the constrained M-functional of regression is uniquely given by $\boldsymbol\beta(F) = \boldsymbol\beta_0$.

The asymptotic results for $\boldsymbol\beta(F_n)$ and the influence function of $\boldsymbol\beta(F)$ given in Mendes and Tyler (1996) require that $\sigma(F)$ also be uniquely defined. The value of $\sigma^2(F)$ corresponds to the minimum of $cE\{\rho(r_0^2/\sigma^2)\} - \log\sigma$, where $r_0 = y - \mathbf{x}'\boldsymbol\beta_0$, subject to the constraint $\sigma \ge \sigma_0$ with σ_0 being the solution to $E\{\rho(r_0^2/\sigma_0^2)\} = \varepsilon$. This is equivalent to minimizing the "profile likelihood equation" (1.7) subject to the constraint (1.8) if we set $S_0 = r_0^2$, $p = 1$, $\Sigma_0 = 1$

and $\kappa = \sigma^{-2}$. Thus, the study of the uniqueness of $\sigma(F)$ is identical to the study of the uniqueness of κ for the case $p = 1$.

REFERENCES

BONDESSON, L. (1992). *Generalized Gamma Convolutions and Related Classes of Distributions and Densities*. Springer, New York.

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

DENNIS, J. E. and WELSCH, R. E. (1976). Techniques for nonlinear least squares and robust regression. In *Proceedings of the Statistical Computing Section* 83–87. Amer. Statist. Assoc., Washington, D.C.

FELLER, W. (1966). *An Introduction to Probability Theory and Its Applications*. Wiley, New York.

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.

KENT, J. T. and TYLER, D. E. (1996). Constrained M-estimation for multivariate location and scatter. *Ann. Statist.* **24** 1346–1370.

LI, S. Z. (1995). Discontinuous MRF prior and robust statistics: a comparative study. *Image and Vision Computing* **13** 227–233.

MENDEZ, X. and TYLER, D. E. (1996). Constrained M-estimation for regression. In *Robust Statistics, Data Analysis and Computer Intensive Methods: In Honor of Peter Huber's 60th Birthday. Lecture Notes in Math.* **109** 299–320. Springer, New York.

ROCKE, D. M. and WOODRUFF, D. L. (1996). Identification of outliers in multivariate data. *J. Amer. Statist. Assoc.* **91** 1047–1061.

TATSUOKA, K. and TYLER, D. E. (2000). On the uniqueness of S-functionals and M-functionals under nonelliptical distributions. *Ann. Statist.* **28** 1219–1243.

DEPARTMENT OF STATISTICS
UNIVERSITY OF LEEDS
LEEDS, LS2 9JT
UNITED KINGDOM
E-MAIL: j.t.kent@leeds.ac.uk

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
110 FRELINGHUYSEN ROAD
PISCATAWAY, NEW JERSEY 08854-8019
E-MAIL: dtyler@rci.rutgers.edu