The Annals of Statistics
1995, Vol. 23, No. 5, 1823–1847

M-ESTIMATES OF RIGID BODY MOTION ON THE SPHERE AND IN EUCLIDEAN SPACE[1]

BY TED CHANG AND DAIJIN KO

University of Virginia and Virginia Commonwealth University

This paper calculates the influence functions and asymptotic distributions of M-estimators of the rotation A in a spherical regression model on the unit sphere in p dimensions with isotropic errors. The problem arises in the reconstruction of the motion of a rigid body on the surface of the sphere.

The comparable model for p-dimensional Euclidean space data is that $(v_1, \dots, v_n)$ are independent with v_i symmetrically distributed around $\gamma A \cdot u_i + b$, u_i known, where the real constant $\gamma > 0$, $p \times p$ rotation matrix A and p-vector b are the parameters to be estimated. This paper also calculates the influence functions and asymptotic distributions of M-estimators for γ, A and b. Besides rigid body motion, this problem arises in image registration from landmark data.

Particular attention is paid to how the geometry of the rigid body or landmarks affects the statistical properties of the estimators.

1. Introduction. If a rigid body moves on the surface of a unit sphere Ω_p in Euclidean p-dimensional space, the position of a point x on the rigid body is given, as a function of time, by $A(t) \cdot x$ for some curve $A(t)$ in $\mathrm{SO}(p)$, where $\mathrm{SO}(p)$ is the group of $p \times p$ matrices A such that $AA^t = 1$ and $\det A = 1$. Thus if one is trying to reconstruct the movement of the rigid body between two fixed points in time, a natural model to consider is the spherical regression model: the data consists of n pairs of unit vectors (u_i, v_i), $i = 1, \dots, n$, such that, for fixed u_i, the density of v_i with respect to uniform measure on Ω_p is of the form $f(v_i^t A u_i)$ for some unknown A in $\mathrm{SO}(p)$.

If a rigid body moves in Euclidean space R^p, and if x and y are the positions of a point on the body at two specific times, we have $y = A \cdot x + b$, where A is in $\mathrm{SO}(p)$ and b in R^p. We will also consider scaled Euclidean motions $\gamma A \cdot x + b$, where $\gamma > 0$ is a real number. This paper deals with robust estimation of the rotation matrix A of the spherical regression model or the triple (A, b, γ) in the Euclidean motion model. The problems considered are the asymptotic distribution of the estimators and the detection of influential observations.

Received March 1993; revised February 1995.

[1]Supported by NSF Grants DMS-91-01568 and DMS-91-01725, NSA Grant MDA-904-91-H-0031 and ONR Grant N00014-92-J-1009.

AMS 1991 *subject classifications*. Primary 62F10, 62F12, 62F35, 62H12, 86A60.

Key words and phrases. Spherical regression, Procrustes analysis, tectonic plates, rigid body motion, image registration, shape theory, M-estimates, robustness, directional data, standardized influence function, residual analysis.

One of the most important applications of the spherical regression model is the determination of the errors in the fitted reconstruction of the position of one tectonic plate relative to another. A survey of some of the statistical issues in that problem can be found in Chang (1993). In the tectonic context, outliers are not uncommon. Rivest (1989) noticed more outliers in a Gulf of Aden data set than a Fisher error model would predict. Shaw and Cande (1990) noticed that the distribution of fracture zone and magnetic anomaly data for the South Atlantic is extremely heavy tailed. However, techniques for finding such outliers and influential observations have not been fully studied in the spherical regression context.

The Euclidean motion model arises in many contexts. One context is in the problem of "image registration." In this problem, one is trying to correlate two images of the same two- or three-dimensional object from landmarks on the object. Since any nonrigidity in the object is modelled in the statistical error term, long-tailed error distributions and outliers are not uncommon. Depending upon the factual situation, the difference between the two images is often expressible as a scaled Euclidean motion. Although (A, b, γ) is usually not of direct interest, it must be estimated, and one should use an estimator with good statistical properties. Furthermore, if the registration is not satisfactory, an analysis of influential landmarks would indicate which landmarks should be reexamined for better registration.

The same problem arises in shape theory. For example, motivated by applications in biology, Siegel and Benson (1982) propose a repeated median-type method for the reconciliation of two shapes in the plane. Residual analysis is of specific interest in that paper.

We summarize below some of the qualitative conclusions of this paper.

For spherical regression, the shape of an asymptotic confidence region for A depends only upon the geometry of the points u_i; its size depends only upon the objective function minimized and upon the underlying density f. Furthermore, the efficiency of an M-estimator of A on Ω_p is the same as the efficiency of the corresponding M-estimator of a location on Ω_p. In particular, for $p = 3$, the efficiency of a median-type estimator of A, at a von Mises–Fisher distribution with concentration parameter κ, is a decreasing function of κ and approaches $\pi/4$ as $\kappa \to \infty$.

For Euclidean motions, similar behavior is observed. Thus, an L_1-estimator will have the same efficiency as the spatial median. This efficiency is $\pi/4$ and $8/(3\pi)$ for $p = 2$ and 3, respectively, at an isotropic normal distribution. For $p = 3$, the use of an L_1-estimator will guard against outliers and long tails at a cost of only $(3\pi/8)^{1/2} - 1 = 8\%$ in standard error when, in fact, a least squares estimator would have been optimal.

In the spherical regression case, Rivest (1989) used a version of Cook's D to diagnose outliers; but it is well known that this approach suffers from a masking effect for multiple outliers and/or influential points. In contrast to Rivest's approach, we use a robust spherical regression and then use the fit to discover the influential data points. We use a standardized influence function to identify leverage points as well as outliers. Despite the assumed symmetry in the errors hypothesized by the spherical regression model, the influence of a data point depends not only on the length of the residual and the leverage of the design point, but also on the direction of the residual vector relative to the design point and the overall geometry of the u_i. Similar behavior is observed for the rotation A in Euclidean motions. In addition, for a given design point and residual length, an observation which is influential for $\hat{A}$ will tend to be unimportant for $\hat{\gamma}$ and vice versa.

In some applications, such as that given by Chapman, Chen and Kim (1995), the location of the u_i can be controlled. In this case, placing u_i uniformly on Ω_3 will minimize the influence of any observation. Similarly, for Euclidean motions, it is optimal to place the landmarks uniformly on the surface of the largest feasible sphere.

2. *M*-estimators for spherical regression and their standardized influence functions. Under the assumption that the v_i are independently distributed with a density of the form $f(v_i^t A u_i)$, the most natural family of M-estimators of A are those which minimize an expression of the form

$$\frac{1}{n}\sum \rho\left(v_i^t A u_i\right). \tag{2.1}$$

To get a differential expression which is implied by (2.1), one needs to recall that if $L(\mathrm{O}(p))$ is the collection of skew-symmetric matrices H (i.e., matrices which satisfy $H + H^t = 0$), then the map Φ: $L(\mathrm{O}(p)) \to \mathrm{O}(p)$ defined by $\Phi(H) = \sum_{i \geq 0} H^i/i!$ is a 1–1 nonsingular differentiable map of a neighborhood 0 in $L\mathrm{O}(p)$ onto a neighborhood of I in $\mathrm{O}(p)$. For obvious reasons, Φ is usually called the exponential map. Since $\mathrm{O}(p)$ has two connected components, one of which is $\mathrm{SO}(p)$, Φ also parameterizes a neighborhood of I in $\mathrm{SO}(p)$.

Thus if $A = \hat{A}$ minimizes (2.1), either over $\mathrm{O}(p)$ or over $\mathrm{SO}(p)$, for any skew-symmetric H, we have

$$\begin{aligned} 0 &= \frac{d}{dt}\bigg|_{t=0} \frac{1}{n}\sum \rho\left(v_i^t \hat{A}\Phi(tH)u_i\right) = \frac{1}{n}\sum \rho'\left(v_i^t \hat{A}\, u_i\right) v_i^t \hat{A}\, H u_i \\ &= \mathrm{Tr}\left[\left(\frac{1}{n}\sum \rho'\left(v_i^t \hat{A}\, u_i\right) u_i v_i^t \hat{A}\right) H\right]. \end{aligned}$$

This is equivalent to the matrix $(1/n)\sum \rho'(v_i^t \hat{A}\, u_i) u_i v_i^t \hat{A}$ being symmetric. We are lead to the family of M-estimators, each member of which satisfies an equation of the form

$$0 = \frac{1}{n}\sum \psi\left(v_i^t \hat{A}\, u_i\right)\left(u_i v_i^t \hat{A} - \hat{A}^t v_i u_i^t\right), \tag{2.2}$$

where $\psi(t) = -\rho'(t)$. We will always assume that $\psi(t) \geq 0$ for all t in $[0, 1]$ and $\psi(t) > 0$ for some t in $(0, 1)$.

The derivation of (2.2) from (2.1) can be described as "performing calculus in the tangent plane to $\mathrm{O}(p)$ [or $\mathrm{SO}(p)$] at $\hat{A}$." Mathematically, similar reasoning was used in Ko and Chang (1993). That paper studied M-estimators of a modal vector μ of a distribution on the unit sphere Ω_p of R^p using calculus in the tangent plane to Ω_p at $\hat{\mu}$. Most of the proofs here are similar to the proofs given there, and we will content ourselves here with giving some representative calculations. We trust that once the mathematics of our approach becomes clear, the reader will be able to modify the proofs of Ko and Chang (1993) to the present situation.

Again, since asymptotics is essentially local, there is no difference in the theory of estimating a matrix A in $\mathrm{O}(p)$ from estimating such a matrix in $\mathrm{SO}(p)$. Below we will state theorems in terms of estimation of a matrix in $\mathrm{SO}(p)$; however, these theorems hold without change if in fact one is interested in estimation of a matrix in $\mathrm{O}(p)$.

For a specific ψ and distribution F, write $\hat{A}_\psi(F)$ for the matrix in $\mathrm{SO}(p)$ which satisfies

$$0 = E_F\Big[\psi\big(v^t\hat{A}_\psi(F)u\big)\big(uv^t\hat{A}_\psi(F) - \hat{A}_\psi(F)^t vu^t\big)\Big], \tag{2.3}$$

or, equivalently, $E_F[\psi(v^t\hat{A}_\psi(F)u)uv^t\hat{A}_\psi(F)]$ is symmetric. When the context is clear, we will often suppress the subscript ψ or the argument F. We recall that if F represents an empirical distribution, (2.2) becomes a special case of (2.3).

Equation (2.3) arises as a necessary condition for $\hat{A}_\psi(F)$ to minimize $E_F[\rho(v^tAu)]$, for ρ such that $\rho'(t) = -\psi(t)$. Thus we expect (2.3) to have multiple solutions corresponding to the other critical points. Indeed, because $\mathrm{SO}(p)$ is compact, $E_F(\rho)$ has at least two critical points, corresponding to its maximum and minimum. This is not an uncommon problem with M-estimators, and the choice of the correct solution will usually be clear either by evaluating the original function ρ or, unless the error distribution F is unusually diffuse, by visually inspecting the fit.

From the names of the corresponding location estimators, we call $\hat{A}_\psi$ the mean rotation (LS-mean rotation), the R-median rotation (median with respect to spherical distance on Ω_p) and the L_1-median rotation (median with respect to the L_1 metric) for $\psi(t) = 1$, $\psi(t) = (1-t^2)^{-1/2}$ and $\psi(t) = (1-t)^{-1/2}$, respectively.

The influence function at an observation $x = (u^*, v^*)$ is defined to be

$$\mathrm{IF}(x; \hat{A}, F) = \lim_{\varepsilon \to 0^+} \frac{\hat{A}\,[(1-\varepsilon)F + \varepsilon\delta_x] - \hat{A}\,(F)}{\varepsilon},$$

where δ_x denotes a point mass at x. It is a tangent vector to $\mathrm{SO}(p)$ at $\hat{A}(F)$. In other words, $\hat{A}(F)^t\,\mathrm{IF}(x;\hat{A},F)$ is a $p\times p$ skew-symmetric matrix.

Letting $I(\hat{A}(F))$ denote the information matrix at the observed model, the squared norm of the standardized influence function (SIF) is defined to be $\mathrm{Vect(IF)}^t \cdot I(\hat{A}(F)) \cdot \mathrm{Vect(IF)}$. The SIF is more informative than the influence function because, as Hampel, Ronchetti, Rousseeuw and Stahel [(1986, page 229] noted, it "compares the bias (as measured by the IF) with the scatter of the maximum likelihood estimator at the estimated model distribution."

We will assume that the u_i are i.i.d. with $\Sigma = E(uu^t)$ a positive semidefinite symmetric matrix with at most one zero eigenvalue. Let $\lambda_1 \geq \cdots \geq \lambda_p$ be the eigenvalues of Σ with corresponding eigenvectors $e_1, \dots, e_p$. We further assume that, conditionally on $(u_1, \dots, u_n)$, $(v_1, \dots, v_n)$ are independent with the conditional density of v_i of the form $f(v_i^t A u_i)$ for some A in SO(p). All densities are with respect to the usual surface measure on Ω_p. Let $g(t) = \log f(t)$. Write $v = tAu + \sqrt{1 - t^2}\, w$, where $t = v^t A u$ and w is a unit vector perpendicular to Au. Since the distribution of v, conditionally on u, is rotationally symmetric around Au, t and w are independent and w is uniformly distributed on the unit sphere in the $(p-1)$-dimensional hyperplane $(Au)^{\perp}$. Thus there exist constants c_1, c_2, k_0, k_1, k_2, k_3 and k_4 such that, conditionally on the u's,

$$
\begin{aligned}
E\left[g'(v^t A u)^2 v v^t | u\right] &= c_1 I + c_2 A u u^t A^t,\\
E\left[\psi(v^t A u) v | u\right] &= k_0 A u,\\
E\left[\psi'(v^t A u) v v^t | u\right] &= k_1 I + k_2 A u u^t A^t,\\
E\left[\psi(v^t A u)^2 v v^t | u\right] &= k_3 I + k_4 A u u^t A^t.
\end{aligned}
\tag{2.4}
$$

Note that if H is a skew-symmetric matrix,

$$
\begin{aligned}
&\operatorname{Vect}\left(\hat{A}(F) H_1\right)^t \cdot I\left(\hat{A}(F)\right) \cdot \operatorname{Vect}\left(\hat{A}(F) H_2\right)\\
&\quad = E\left[\left.\frac{d \log f\left(v^t \hat{A}(F) \Phi(t H_1) u\right)}{dt}\right|_{t=0} \times \left.\frac{d \log f\left(v^t \hat{A}(F) \Phi(s H_2) u\right)}{ds}\right|_{s=0}\right]\\
&\quad = E\left[g'\left(v^t \hat{A}(F) u\right)^2 u^t H_1^t \hat{A}(F)^t v v^t \hat{A}(F) H_2 u\right]\\
&\quad = E\left[u^t H_1^t \hat{A}(F)^t \left(c_1 I + c_2 \hat{A}(F) u u^t \hat{A}(F)^t\right) \hat{A}(F) H_2 u\right]\\
&\quad = c_1 \operatorname{Tr}\left[H_1^t \Sigma H_2\right].
\end{aligned}
\tag{2.5}
$$

Therefore $\|\mathrm{SIF}\|^2 = c_1 \operatorname{Tr}[(\hat{A}(F)^t \mathrm{IF})^t \cdot \Sigma \cdot (\hat{A}(F)^t \mathrm{IF})]$.

Before calculating IF, we first prove two technical lemmas.

LEMMA 1. *Under the above assumptions,* $E[\psi(v^t A u)(u v^t A - A^t v u^t) | u] = 0$.

PROOF. This is an immediate consequence of $E[\psi(v^t A u) v | u] = k_0 A u$. □

LEMMA 2. *Suppose g is differentiable and let $t = v^t A u$. Then* $k_0 - k_1 = E[\psi(t)(1 - t^2) g'(t)]/(p-1)$. *Thus if f is strictly increasing,* $k_0 - k_1 > 0$.

PROOF. Define constants k_5 and k_6 by $E[\psi(t)g'(t)vv^t] = k_5 I + k_6 Auu^tA^t$. Write $v = (1-t^2)^{1/2}\xi + tAu$. Then, conditionally on t and u, $E[\xi\xi^t] = (I - Auu^tA^t)/(p-1)$ and it follows that $k_5 = E[\psi(t)(1-t^2)g'(t)]/(p-1)$. For convenience, and without loss of generality, assume $A = I$.

Let dv denote surface measure on Ω_p. Using Lemma 1, we have that, for all skew-symmetric E and H, and $u \in \Omega_p$,

$$\begin{aligned} 0 &= \frac{d}{dt}\bigg|_{t=0} \operatorname{Tr}\left[E \cdot \int \psi(v^t\Phi(tH)u)f(v^t\Phi(tH)u)uv^t\Phi(tH)\,dv\right] \\ &= \operatorname{Tr} E \cdot \int \big[\psi'(v^tu)(v^tHu)f(v^tu)uv^t + \psi(v^tu)f(v^tu)uv^tH \\ &\qquad\qquad + \psi(v^tu)f'(v^tu)(v^tHu)uv^t\big]\,dv \\ &= \operatorname{Tr} E\big[-uu^tH(k_1I + k_2uu^t) + k_0uu^tH - uu^tH(k_5I + k_6uu^t)\big] \\ &= (k_0 - k_1 - k_5)\operatorname{Tr}[Euu^tH]. \end{aligned}$$

Thus $k_5 = k_0 - k_1$. □

PROPOSITION 1.

(a) *Let H be the unique $p \times p$ skew-symmetric matrix for which*

$$(k_0 - k_1)\Sigma H + \psi\big(v^{*t}\hat{A}(F)u^*\big)u^*v^{*t}\hat{A}(F) \tag{2.6}$$

is symmetric. Then $\operatorname{IF}(x; \hat{A}, F) = \hat{A}(F)H$.

(b) *By change of basis, assume without loss of generality that Σ is diagonal, and write $H = (h_{ij})$, $u^* = (u^1, \dots, u^p)^t$, $A^tv^* = (v^1, \dots, v^p)$, $t_{ij} = u^iv^j - u^jv^i$. Then*

$$\|\mathrm{SIF}\|^2 = \frac{c_1}{(k_1 - k_0)^2}\psi^2(v^{*t}Au^*)\sum_{i<j}(\lambda_i + \lambda_j)^{-1}t_{ij}^2.$$

(c) *The standardized gross error sensitivity (SGES) of $\hat{A}$ is*

$$\mathrm{SGES}^2 = \frac{c_1}{(k_1 - k_0)^2} \cdot \sup_{-1 \le t \le 1} \frac{\psi^2(t)(1-t^2)}{\lambda_{p-1} + \lambda_p}.$$

PROOF. Let $F(x, \varepsilon) = (1-\varepsilon)F + \varepsilon\delta_x$. The standard calculation is to substitute $F(x, \varepsilon)$ for F in equation (2.3) and to differentiate, thus calculating $\operatorname{IF}(x; \hat{A}, F)$ by implicit differentiation. We will follow this basic outline.

Assuming for the moment that $\operatorname{IF}(x; \hat{A}, F)$ exists, we can write $H = \hat{A}(F)^t\operatorname{IF}(x; \hat{A}, F)$. Then H will be skew symmetric and $\hat{A}(F(x, \varepsilon)) = \hat{A}(F)\Phi(\varepsilon H + o(\varepsilon))$. If G is skew symmetric,

$$0 = \operatorname{Tr}\left[G\int \psi\big(v^t\hat{A}(F(x,\varepsilon))u\big)uv^t\hat{A}(F(x,\varepsilon))\,dF(x,\varepsilon)\right].$$

Differentiating,

$$0 = \operatorname{Tr} G\Big[\int \psi'\big(v^t\hat{A}(F)u\big)\big(v^t\hat{A}(F)Hu\big)uv^t\hat{A}(F)\,dF$$

$$+\int \psi\big(v^t\hat{A}(F)u\big)uv^t\hat{A}(F)H\,dF$$

$$+\int \psi\big(v^t\hat{A}(F)u\big)uv^t\hat{A}(F)(-dF + d\delta_x)\Big]$$

$$= \operatorname{Tr} G\Big[-\int \psi'\big(v^t\hat{A}(F)u\big)uu^tH\hat{A}(F)^t vv^t\hat{A}(F)\,dF$$

$$+\int k_0 uu^tH\,dF + \psi\big(v^{*t}\hat{A}(F)u^*\big)u^*v^{*t}\hat{A}(F)\Big]$$

$$= \operatorname{Tr} G\Big[-\int uu^tH\big(k_1I + k_2uu^t\big)\,dF + k_0\Sigma H + \psi\big(v^{*t}\hat{A}(F)u^*\big)u^*v^{*t}\hat{A}(F)\Big]$$

$$= \operatorname{Tr} G\big[(k_0 - k_1)\Sigma H + \psi\big(v^{*t}\hat{A}(F)u^*\big)u^*v^{*t}\hat{A}(F)\big].$$

Here we have use that $u^tHu = 0$. Since the above equality is true for all skew-symmetric G, (2.6) must be symmetric.

In the basis of eigenvectors of Σ, the condition that (2.6) be symmetric becomes

$$(k_0 - k_1)(\lambda_i + \lambda_j)h_{ij} + \psi\big(v^{*t}\hat{A}(F)u^*\big)(u_iv_j - u_jv_i) = 0.$$

By Lemma 2 and the assumption that Σ is positive semidefinite with at most one zero eigenvalue, we have $(k_0 - k_1)(\lambda_i + \lambda_j) > 0$. Thus, by the implicit function theorem, the derivative $\mathrm{IF}(x, \hat{A}, F)$ exists, H is unique and

$$c_1 \operatorname{Tr}(H^t\Sigma H) = c_1 \sum_{i<j} (\lambda_i + \lambda_j)h_{ij}^2$$

$$= \frac{c_1}{(k_1 - k_0)^2}\psi^2(v^{*t}Au^*)\sum_{i<j}(\lambda_i + \lambda_j)^{-1}t_{ij}^2.$$

This establishes (a) and (b). Let $t = v^{*t}Au^*$. Then $\Sigma_{i<j}t_{ij}^2 = 1 - t^2$. This easily yields part (c). □

For $p = 3$, the usual case of practical interest, (b) can be geometrically interpreted as follows: Suppose the length of the residual vector, or, equivalently, $t = v^t\hat{A}u$, is fixed. Then $\hat{A}(F)^tv = tu + (1 - t^2)^{1/2}w$ with $\|w\| = 1$ and $w \perp u$. Let the vector cross product $u \times w = [x_1, x_2, x_3]^t$. Then

$$(2.7)\quad \|\mathrm{SIF}\|^2 = \frac{c_1}{(k_1 - k_0)^2}\psi^2(t)(1 - t^2)\left\{\frac{x_1^2}{\lambda_2 + \lambda_3} + \frac{x_2^2}{\lambda_1 + \lambda_3} + \frac{x_3^2}{\lambda_1 + \lambda_2}\right\}.$$

Note that $\|u \times w\| = 0$ if $|t| = 1$ and is 1 otherwise. Thus we conclude the following:

1. The standardized influence of an observation is not only determined by the length of the residual but also by the relative location of the design point u with respect to eigenvectors of Σ and the direction of w. For a given t,

$$\max_{v^t \hat{A} u = t} \|\text{SIF}\|^2 = \frac{c_1}{(k_1 - k_0)^2} \psi^2(t)(1 - t^2)(\lambda_2 + \lambda_3)^{-1},$$

which occurs if and only if u is perpendicular to the dominant eigenvector e_1 of Σ and $w = \pm(u \times e_1)$. Furthermore,

$$\min_{v^t \hat{A} u = t} \|\text{SIF}\|^2 = \frac{c_1}{(k_1 - k_0)^2} \psi^2(t)(1 - t^2)(\lambda_1 + \lambda_2)^{-1},$$

which occurs if and only if $u \perp e_3$ and $w = \pm(u \times e_3)$.

Notice that if $u = e_2$, then, depending upon the direction of w, $\|\text{SIF}\|^2$ can achieve both its maximum and minimum values. Thus even at a point u of high leverage, namely, when $u \perp e_1$, the actual influence could be very small, depending on the direction of w and the conditioning of the matrix Σ.

We comment that because v is constrained to lie on the sphere Ω_3, $v - \hat{A}u$ is constrained to a two-dimensional surface. We believe that the projection $(1 - t^2)^{1/2}\hat{A}w = v - (v^t\hat{A}u)\hat{A}u$ of v onto the two-dimensional hyperplane of tangent vectors to Ω_3 at $\hat{A}u$ is a more appropriate notion of residual vector in this situation than the simple difference $v - \hat{A}u$.

2. The maximum influence of an observation is minimized when $\lambda_1 = \lambda_2 = \lambda_3 = \frac{1}{3}$. In this case the influence of an observation depends only upon t. We also conclude that the choice of design points (the u_i) which makes the influence of any single point as small as possible is a uniform dispersion of the u_i. Thus, when the location of the points u_i can be controlled, they should be spread uniformly on Ω_3.
3. The standardized gross error sensitivity SGES of $\hat{A}$ at a von Mises distribution on Ω_p is

$$\text{SGES}^2 = (p - 1)\frac{A_p(\kappa)}{\kappa} \frac{\sup \psi^2(s)(1 - s^2)}{E^2(\psi(t)(1 - t^2))},$$

which is identical to the corresponding result [equation (3.12) of Ko and Chang (1993)] for an M-estimator in the spherical location problem. Thus the R-median rotation and the L_1-median rotation are SB-robust at the family of von Mises–Fisher distributions, that is $\sup_\kappa$ SGES is bounded. On the other hand, the mean (MLE) rotation is not SB-robust at the von Mises–Fisher family. Theorem 3.4 of Ko and Chang (1993), which gives necessary and sufficient conditions for an M-estimator of a spherical location to be SB-robust, also generalizes to M-estimated rotations without change.

3. Asymptotics and optimality. We now give results on the asymptotic distribution of the $\hat{A}_n$ which satisfies (2.2). The proof of Proposition 2 will not be given as it is quite similar to the proof of its cousin in Ko and Chang (1993). We continue with the same assumptions on the underlying distribution of the v_i. We also continue to assume that $\psi(t) \geq 0$ for all t in $[0, 1]$ and $\psi(t) > 0$ for some t in $(0, 1)$.

PROPOSITION 2. *Suppose ρ satisfies the following*:

(A1) *ρ is differentiable and $\rho'(t)(1 - t)^{1/2}$ is bounded for t in* $[-1, 1]$;
(A2) *$E\rho(v^t Bu)$ has a unique minimum at $B = A$.*

Then if $\hat{A}_n$ minimizes (2.1), *$\hat{A}_n \to A$ a.s. as $n \to \infty$.*

We remark that (A2) will hold whenever $\rho(\cdot)$ is monotonically decreasing and $f(\cdot)$ is a strictly increasing function. In particular, the examples in the previous section are all consistent under the assumption of strong unimodality of the underlying density $f(v^t Au)$.

PROPOSITION 3. *Suppose ψ satisfies the following*:

(AN0) $E\psi(v^t Au)(uv^t A - A^t vu^t) = 0$;
(AN1) *$\psi(t)$ is continuous*;
(AN2) *$\psi(t)$ is differentiable except at a finite number of points $t = t_0, \dots, t_r$*;
(AN3) *$|\psi'(t)| \leq M$ for $t \neq t_0, \dots, t_r$*;
(AN4) $k_0 - k_1 \neq 0$.

Suppose $\hat{A}_n$ satisfying (2.2) *is consistent. Write $\hat{A}_n = A\Phi(H_n)$. Then H_n has an asymptotic multivariate normal distribution with density proportional to* $\exp[-n(k_0 - k_1)^2 \operatorname{Tr}(H^t \Sigma H)/(2k_3)]$.

If $H_n = (h_{ij})$ when reexpressed in terms of a basis of eigenvectors of Σ, Proposition 3 asserts that the $n^{1/2}h_{ij}$, for $i < j$, are asymptotically independent and normally distributed with the variance of $n^{1/2}h_{ij}$ equal to $k_3(\lambda_i + \lambda_j)^{-1}(k_0 - k_1)^{-2}$.

PROOF OF PROPOSITION 3. Using a proof similar to that given in Ko and Chang (1993), it can be shown that (AN0)–(AN4) imply the hypotheses of the corollary to Theorem 3 in Huber (1967) are satisfied. By replacing v_i with $A^t v_i$, we can further assume that $A = I$. With this simplification, for H skew symmetric, let $\lambda(H)$ be the skew-symmetric matrix defined by

$$\lambda(H) = E\Big[\psi\big(v^t\Phi(H)u\big)\big(uv^t\Phi(H) - \Phi(H)^t vu^t\big)\Big].$$

The derivative of λ at $\mathbf{0}$ [corresponding to $\Phi(\mathbf{0}) = I = A$] is a linear transformation from the skew-symmetric matrices into themselves which satisfies

$$\begin{aligned}\lambda'(\mathbf{0})\cdot H &= \frac{d}{dt}\bigg|_{t=0} E\big[\psi(v^t\Phi(tH)u)\big(uv^t\Phi(tH) - \Phi(tH)^t vu^t\big)\big]\\ &= E\big[\psi'(v^tu)(v^tHu)(uv^t - vu^t) + \psi(v^tu)(uv^tH + Hvu^t)\big]\\ &= E\big[-\psi'(v^tu)(uu^tHvv^t + vv^tHuu^t) + \psi(v^tu)(uv^tH + Hvu^t)\big]\\ &= E\big[-uu^tH\big(k_1I + k_2uu^t\big) - \big(k_1I + k_2uu^t\big)Huu^t\\ &\qquad\qquad + k_0(uu^tH + Huu^t)\big]\\ &= (k_0 - k_1)(\Sigma H + H\Sigma).\end{aligned}$$

[We note that the lemma preceding Theorem 4.2 in Ko and Chang (1993) justifies the above interchange of the derivative and expectation.]

Thus

$$n^{-1/2}\sum\psi\big(v_i^tu_i\big)\big(u_iv_i^t - v_iu_i^t\big) + n^{1/2}(k_0 - k_1)(\Sigma H_n + H_n\Sigma) \to 0 \tag{3.1}$$

in probability. If E and H are skew symmetric $\mathrm{Tr}(\Sigma HE) = \mathrm{Tr}(E\Sigma H) = \mathrm{Tr}(H^t\Sigma E^t) = \mathrm{Tr}(H\Sigma E)$. Thus (3.1) implies that

$$n^{-1/2}\sum\psi\big(v_i^tu_i\big)v_i^tEu_i + n^{1/2}(k_0 - k_1)\mathrm{Tr}(H_n\Sigma E) \to 0 \tag{3.2}$$

in probability for all skew-symmetric E. Now

$$\begin{aligned}\mathrm{Var}\big[\psi(v^tu)v^tEu\big] &= E\Big[\psi(v^tu)^2u^tE^tvv^tEu\Big]\\ &= E\big[-u^tE\big(k_3I + k_4uu^t\big)Eu\big] = -k_3\,\mathrm{Tr}(E\Sigma E).\end{aligned} \tag{3.3}$$

We establish below that $k_3 > 0$. When (3.2) is reexpressed using a basis of eigenvectors of Σ, we get

$$n^{-1/2}\sum_i\psi\big(v_i^tu_i\big)(v_{ik}u_{il} - v_{il}u_{ik}) - n^{1/2}(k_0 - k_1)(\lambda_k + \lambda_l)h_{kl} \to 0, \tag{3.4}$$

for all $k < l$, where $u_i = [u_{i1}, \dots, u_{ip}]^t$ and similarly for v_i. Equation (3.3) implies that when such a basis is used the first terms of (3.4) for different $k < l$ are independent with variances $k_3(\lambda_k + \lambda_l)$. This implies the desired form for the asymptotic distribution of H_n. [Alternatively, the basis-free argument given in Chang (1986), page 913, can be used.] □

LEMMA 3. *Let* $t = v^tAu$. *Then* $k_3 = E[\psi(t)^2(1 - t^2)]/(p - 1) > 0$.

PROOF. The proof is similar to part of the proof of Lemma 2. □

LEMMA 4. *Assume that the following hold*:

(a) *there is no hyperplane* H (*through the origin*) *so that* $u \in H$ *with probability* 1;

(b) ρ *is decreasing*;
(c) f *is strictly increasing.*

Then condition (A2) *is satisfied.*

PROOF. Temporarily fix u. Let $\mu = Au$ and suppose $\mu' \neq \mu$. We will first show that $E[\rho(v^t\mu)|u] < E[\rho(v^t\mu')|u]$. Once this claim is established, if $B \neq A$, it follows from (a) that $E[\rho(v^tAu)] < E[\rho(v^tBu)]$. Thus (A2) would be established.

To demonstrate the claim, let $n = (\mu - \mu')/\|\mu - \mu'\|$, and let $R(v) = v - 2(v^tn)n$ be reflection through n. Then we have $v^t\mu = R(v)^t\mu'$ and, whenever $v^tn > 0$, $v^t\mu > R(v)^t\mu$. Now using dv to denote surface measure on Ω_p,

$$\begin{aligned}
&E\left[\rho(v^t\mu')|u\right] - E\left[\rho(v^t\mu)|u\right]\\
&\quad= \int_{v^tn>0} f(v^t\mu)\left(\rho(v^t\mu') - \rho(v^t\mu)\right) dv\\
&\qquad+ \int_{v^tn<0} f(v^t\mu)\left(\rho(v^t\mu') - \rho(v^t\mu)\right) dv\\
&\quad= \int_{v^tn>0} \Big[f(v^t\mu)\left(\rho(v^t\mu') - \rho(v^t\mu)\right)\\
&\qquad\qquad + f\left(R(v)^t\mu\right)\left(\rho\left(R(v)^t\mu'\right) - \rho\left(R(v)^t\mu\right)\right)\Big]\, dv\\
&\quad= \int_{v^t\mu > v^t\mu'} \left[f(v^t\mu) - f\left(R(v)^t\mu\right)\right]\left[\rho(v^t\mu') - \rho(v^t\mu)\right] dv > 0. \qquad \square
\end{aligned}$$

The following proposition is the robust spherical regression analog of Hampel's Lemma 5.

PROPOSITION 4. *Suppose f is strictly increasing. Let*

$$\tilde{\psi}(s) = \begin{cases} g'(s), & \text{if } g'(s)\sqrt{1-s^2} \le b,\\ b/\sqrt{1-s^2}, & \text{if } g'(s)\sqrt{1-s^2} > b. \end{cases} \tag{3.5}$$

Let $d = E[g'(t)\tilde{\psi}(t)(1-t^2)]/(p-1)$. Writing, for any ψ, $\hat{A}_\psi = A\Phi(H_\psi)$, $\tilde{\psi}$ minimizes $-\lim_{n\to\infty} nE[\mathrm{Tr}(H_\psi \Sigma H_\psi)]$ over all ψ which satisfy $E[g'(t)\psi(t)(1-t^2)] > 0$ and $\mathrm{SGES}(\psi) \le bc_1^{1/2}/(d(\lambda_{p-1} + \lambda_p)^{1/2})$.

When $\hat{A} = A\Phi(H)$, H is a deviation vector of $\hat{A}$ from A. Using equation (2.5), the length of H, standardized using Fisher information, is proportional to $\mathrm{Tr}(H^t\Sigma H)$. If f is strictly increasing, $\psi(t) \ge 0$ for all t in $[0,1]$ and $\psi(t) > 0$ for some t in (0,1), Lemma 2 gives that $E[g'(t)\psi(t)(1-t^2)] > 0$. Since d depends upon b, $\mathrm{SGES}(\tilde{\psi})$ is a function of b. This proposition shows that, for a given limit on $\mathrm{SGES}(\psi)$, the optimal choice of objective function has the form (3.5).

In principle, for a given bound on SGES we can solve for the appropriate b and hence choose the optimal $\tilde{\psi}$. For the Fisher distribution on the sphere Ω_3 in R^3,

$$c_1 = \kappa \coth \kappa - 1,$$

$$d = \frac{\kappa^2}{3}\left(1 - \sqrt{1 - \frac{b^2}{\kappa^2}}\right) + \frac{5b^2}{6}\sqrt{1 - \frac{b^2}{\kappa^2}} + \frac{\kappa b}{4}\arcsin\sqrt{1 - \frac{b^2}{\kappa^2}}\,.$$

Presumably, solving for b in terms of a bound on SGES would be done numerically.

PROOF OF PROPOSITION 4. Using Propositions 1 and 3 and Lemmas 2 and 3,

$$\begin{aligned}-\lim_{n\to\infty} nE\big[\mathrm{Tr}(H_\psi \Sigma H_\psi)\big] &= \frac{p(p-1)}{2}\cdot\frac{k_2}{(k_0 - k_1)^2}\\ &= \frac{pE\big[\psi(t)^2(1-t^2)\big]}{2E^2\big[\psi(t)(1-t^2)g'(t)/(p-1)\big]},\end{aligned}$$

$$\mathrm{SGES}(\psi) = \frac{c_1^{1/2}}{(\lambda_{p-1}+\lambda_p)^{1/2}}\cdot\frac{\sup|\psi(s)|(1-s^2)^{1/2}}{E\big[\psi(t)(1-t^2)g'(t)/(p-1)\big]}.$$

By rescaling ψ, we can assume $E[\psi(t)(1-t^2)g'(t)/(p-1)] = d$. Thus we need to minimize $E[\psi(t)^2(1-t^2)]$ subject to $|\psi(s)|(1-s^2)^{1/2} \le b$ for all s.

Since

$$\begin{aligned}&E\Big[g'(t)(1-t^2)^{1/2} - \psi(t)(1-t^2)^{1/2}\Big]^2\\ &\quad = E\big[g'(t)^2(1-t^2)\big] - 2d(p-1) + E\big[\psi(t)^2(1-t^2)\big],\end{aligned}$$

it suffices to minimize the left-hand side. This is done by choosing, for each s, $\psi(s)$ to minimize $(g'(s) - \psi(s))^2$. Since $g'(s) \ge 0$ and $|\psi(s)|(1-s^2)^{1/2} \le b$, the optimal choice is $\psi(s) = \tilde{\psi}(s)$. □

Using Lemma 2, for the von Mises–Fisher distribution $VF_p(\mu, \kappa)$,

$$k_0 - k_1 = \kappa E\big[\psi(t)(1-t^2)/(p-1)\big]$$

and the distribution of t depends only on κ; and $k_0 - k_1$ and k_3 are functions of κ and ψ. In particular, for the R-median rotation,

$$k_3 = 1/(p-1)$$

and

$$(3.6)\quad k_0 - k_1 = \kappa E\left[\frac{(1-t^2)^{1/2}}{(p-1)}\right] = \frac{\sqrt{2}\kappa}{p-1}\cdot\frac{\Gamma(p/2)}{\Gamma((p-1)/2)}\cdot\frac{I_{p/2-1/2}(\kappa)}{I_{p/2-1}(\kappa)}.$$

Let

$$c_M = \frac{k_3}{(k_0 - k_1)^2} = \frac{p-1}{2\kappa} \cdot \frac{\Gamma^2((p-1)/2)}{\Gamma^2(p/2)} \cdot \frac{I^2_{p/2-1}(\kappa)}{I^2_{p/2-1/2}(\kappa)}.$$

For large κ, $I_V(\kappa) = (2\pi)^{-1/2}\kappa^{-1/2}e^{\kappa}\{1 - O(\kappa^{-1})\}$ so that

$$c_M \approx \frac{p-1}{2\kappa} \cdot \frac{\Gamma^2((p-1)/2)}{\Gamma^2(p/2)} = \begin{cases} \dfrac{4}{\pi}\kappa^{-1}, & \text{for } p = 3, \\ \dfrac{\pi}{2}\kappa^{-1}, & \text{for } p = 2. \end{cases}$$

For small κ,

$$c_M = \begin{cases} \dfrac{32}{\pi^2}\kappa^{-2}, & \text{for } p = 3, \\ \left(\dfrac{\pi}{2}\right)^2 \kappa^{-2}, & \text{for } p = 2. \end{cases}$$

For the mean rotation,

$$c = \frac{1}{\kappa A_p(\kappa)} \approx \begin{cases} \kappa^{-1}, & \text{for large } \kappa, \\ p\kappa^{-2}, & \text{for small } \kappa. \end{cases}$$

The efficiency of the ψ-rotation $\hat{A}_\psi$ (relative to the mean rotation) may be defined as c/c_ψ. With this definition, the efficiency of the median rotation is, as $\kappa \to \infty$,

$$\frac{2}{p-1} \cdot \frac{\Gamma^2(p/2)}{\Gamma^2((p-1)/2)},$$

or $2/\pi$ for $p = 2$ and $\pi/4$ for $p = 3$. As $\kappa \to 0$, the efficiencies are $8/\pi^2$ for $p = 2$ and $3\pi^2/32$ for $p = 3$. Fisher (1985) showed that the efficiency of the spherical median relative to the directional mean (MLE) has the same limiting values as $\kappa \to 0, \infty$ at $p = 3$. Brown (1983) showed that

$$\frac{2}{p-1} \cdot \frac{\Gamma^2(p/2)}{\Gamma^2((p-1)/2)}$$

is, for general p, the asymptotic relative efficiency of the spatial median relative to the mean at an isotropic $(p-1)$-variate normal distribution. The efficiency of the R-median rotation for $p = 3$ is very high (0.7854 for large κ and 0.9253 for small κ). (See Figure 1.)

Notice that the shape of the confidence region based on any M-estimator $\hat{A}_\psi$ is determined by Σ; its size depends only upon the underlying error distribution and on ψ. We conclude that H_n is least constrained if the axis of $\Phi(H_n)$ lies in the direction of the largest eigenvector of Σ. Notice also that, for any ψ and any density $f(t)$, the efficiency of $\hat{A}_\psi$ relative to the mean rotation agrees with that found by Ko and Chang (1993) for the corresponding

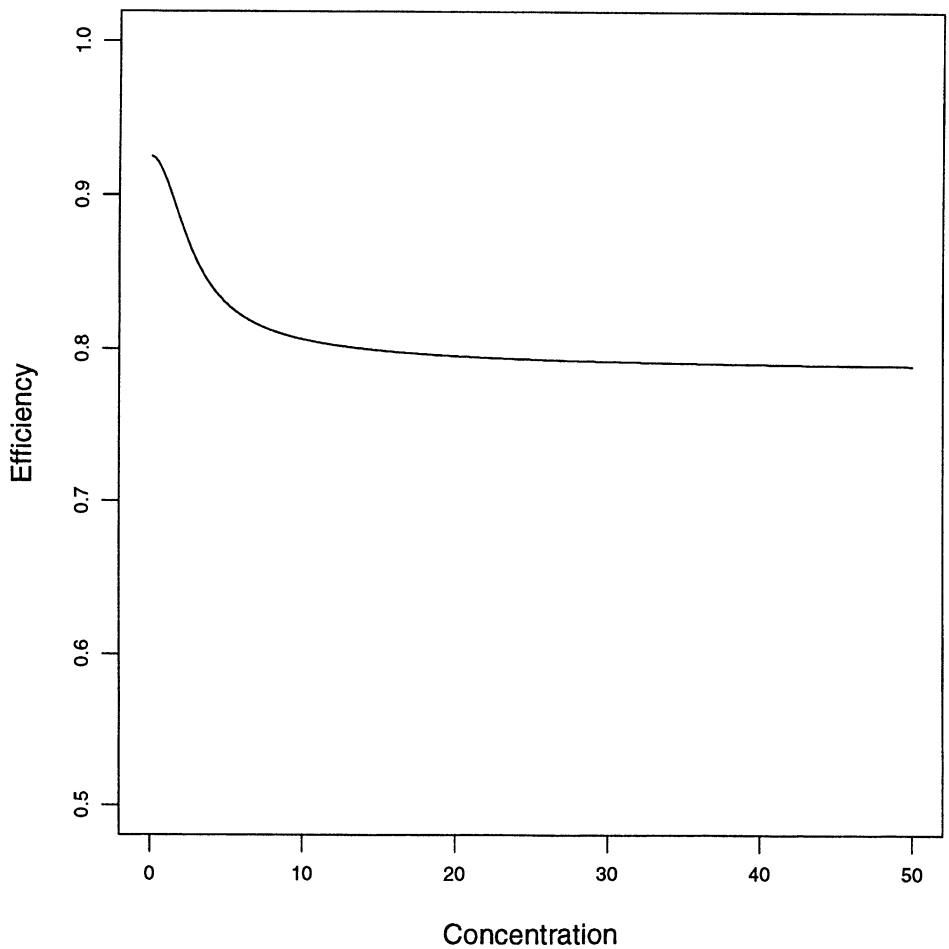

FIG. 1. *Efficiency of R-median rotation relative to mean rotation at a Fisher–von Mises distribution on Ω_3 as a function of the concentration parameter κ.*

M-estimator of θ from i.i.d. x_i which have a density of the form $f(t_i)$, where $t_i = x_i^t\theta$. Proposition 4 also mirrors its counterpart for the spherical location problem.

4. An example. MacKenzie (1957) gave an algorithm to calculate an estimate $\hat{A}$ which maximizes (2.1) over O(p) when $\rho(t) = t$. His algorithm uses the singular-value decomposition as follows: Write $\Sigma u_i v_i^t = O_1 \Lambda O_2^t$, where O_1 and O_2 are in O(p) and Λ is diagonal with entries $\lambda_1 \geq \cdots \geq \lambda_{p-1} \geq \lambda_p \geq 0$. Then $\hat{A} = O_2 O_1^t$. This solution is unique if $\lambda_p > 0$.

Stephens (1979) modified the MacKenzie algorithm to find an $\hat{A}$ which maximizes (2.1) over SO(p) when $\rho(t) = t$. Instead, one uses a modified singular-value decomposition: $\Sigma u_i v_i^t = O_1 \Lambda O_2^t$, where O_1 and O_2 are in

SO(p) and Λ is diagonal with entries $\lambda_1 \geq \cdots \geq \lambda_{p-1} \geq |\lambda_p|$. Then again $\hat{A} = O_2 O_1^t$. This solution is unique if $\lambda_{p-1} > -\lambda_p$.

When $\psi(t) \geq 0$ is even and nondecreasing on $[0, 1]$ (but not necessarily strictly so), we propose the following iterative algorithm to solve (2.2): Suppose $\hat{A}_k$ is a guess for $\hat{A}$. We pick $\hat{A}_{k+1}$ so that $\Sigma\psi(v_i^t \hat{A}_k u_i) u_i v_i^t \hat{A}_{k+1}$ is symmetric. There are at least 2^{p-1} choices for $\hat{A}_{k+1}$, but the choices are exactly the critical points of the function $f(A) = \Sigma\psi(v_i^t \hat{A}_k u_i) v_i^t A u_i$. Since ψ is even and positive nondecreasing on $[0, 1]$, we pick the $\hat{A}_{k+1}$ which maximizes $f(A)$. Then $\hat{A}_{k+1} = O_2 O_1^t$, where O_1 and O_2 are defined using the standard singular-value decomposition of $\Sigma\psi(v_i^t \hat{A}_k u_i) u_i v_i^t$ if a solution in O(p) is desired and a modified singular decomposition if the solution is to be restricted to SO(p). For an initial guess $\hat{A}_0$, we suggest the R-median rotation, which may be estimated via a Nelder–Mead-type algorithm from (2.1).

The asymptotic distribution of $n^{1/2} H_n$ depends upon the parameters $k_0 - k_1$ and k_3. Motivated by equations (2.4), possible estimates are

$$\hat{k}_1 - \hat{k}_0 = \sum \psi'(t_i)(1 - t_i)^2/(np - n) - \sum \psi(t_i) t_i / n$$

and

$$\hat{k}_3 = \sum \psi^2(t_i)(1 - t_i^2)/(np - n),$$

where $t_i = v_i^t \hat{A}_\psi u_i$.

Alternatively, we can assume a Fisher–von Mises error distribution and use an estimate of the concentration parameter κ to estimate $k_1 - k_0$ and k_3 using equations (3.6).

Nonrobustness of the MLE of κ is well known. Ko (1992) suggested that we use an estimate $\hat{\kappa}_{MC}$ based upon median absolute deviation. In the spherical regression context it is defined as

$$\hat{\kappa}_{MC} = C_p^{-1}\Big(\operatorname*{med}_i u_i^t \hat{A}_\psi^t v_i\Big)$$

where $\hat{A}_\psi$ is the R-median rotation and $C_p(\kappa)$ is the median of $\mu^t X$ when X is $VF_p(\mu, \kappa)$ distributed. For $p = 3$ and $\kappa \gg 0$,

$$\hat{\kappa}_{MC} \approx \frac{\ln 2}{1 - \operatorname{med}_i u_i^t \hat{A}_\psi^t v_i}.$$

We refer the reader to Ko (1992) for further details.

We use an updated version of the Central Atlantic data set compiled by Klitgord and Schouten (1986) to illustrate these methods. The data consists of identifications of fracture zone and magnetic anomalies and cannot be analyzed using a spherical regression model [see Chang (1993) for a discussion of the statistical analysis of this type of data]. To construct data suitable for spherical regression modelling, we have replaced each segment by its midpoint. A sequel will discuss influence diagnostics applicable to the data as it is originally collected. A map of the data is shown in Figure 2.

The errors in this data set are so concentrated that the approximation for a large concentration parameter is quite appropriate in estimating the concentration parameter. The estimated mean rotation is a rotation of $-5.39°$

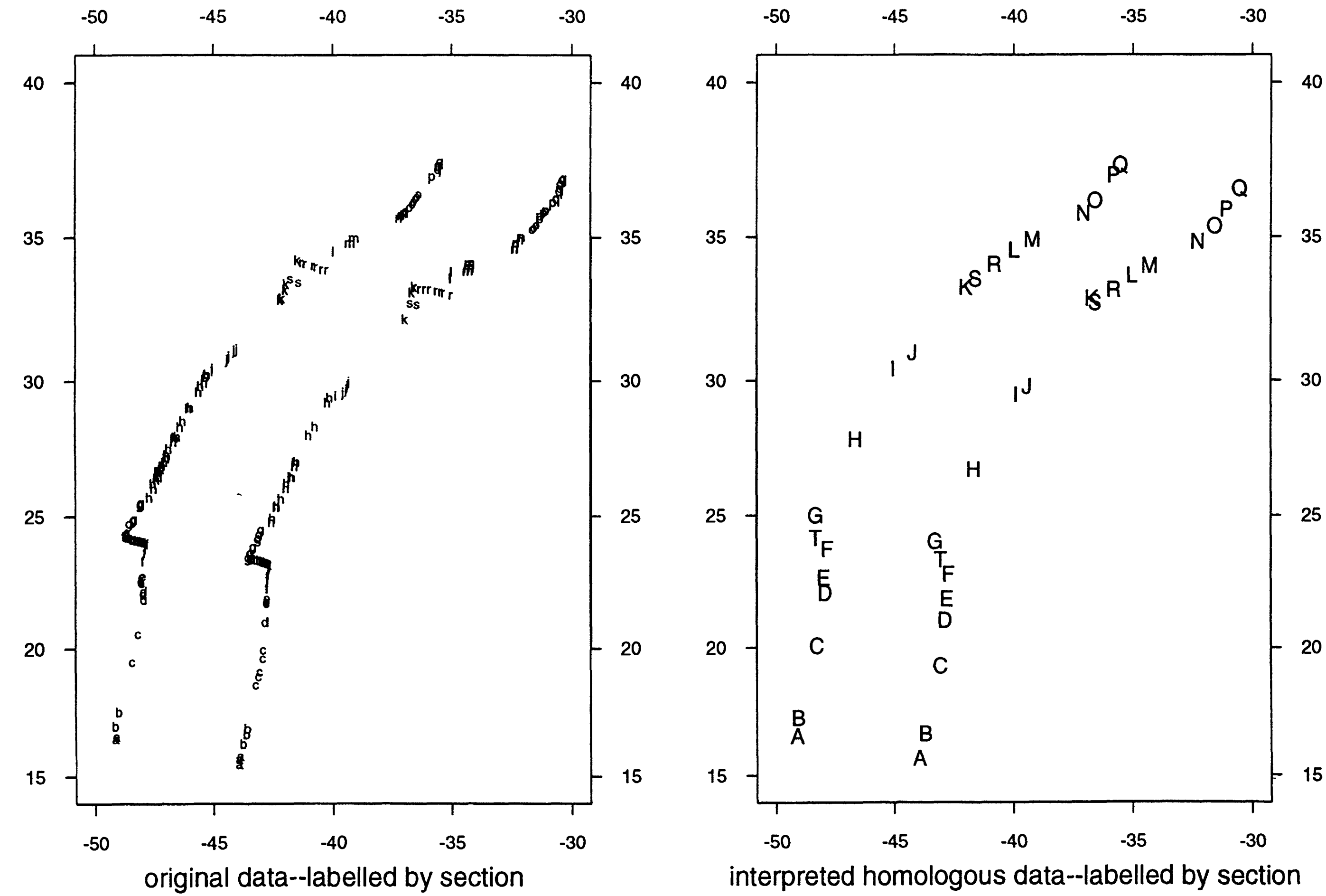

FIG. 2. *Central Atlantic Anomaly 6 data set: original identifications and* (u, v) *pairs used in this paper. (Data courtesy of Kim Klitgord and Hans Schouten.)*

around an axis through 79.59°N latitude and 24.22°E longitude and $\hat{\kappa} = 1.56 \times 10^6$. We obtained the R-median rotation as a rotation of $-5.28°$ around an axis through 80.29°N and 40.08°E and $\hat{\kappa}_{MC} = 3.35 \times 10^6$. The two estimates of the rotation axes are, in fact, only 2.85° apart.

We calculated the 95% confidence regions for the rotation matrix based on mean rotation and m.l.e. $\hat{\kappa}$ and R-median rotation and $\hat{\kappa}_{MC}$. Figure 3 shows the projection of the resulting 95% confidence regions onto the axis latitude–longitude space. Under the Fisher error model, for any large κ, the confidence region based on the mean rotation should have each linear dimension $(\pi/4)^{1/2} = 89\%$ of the corresponding dimension of a confidence region based on the R-median rotation. However, the region based on the mean rotation and m.l.e. $\hat{\kappa}$ is 130% in each lineal dimension of the size of the region based upon the R-median rotation and $\hat{\kappa}_{MC}$. This indicates that some data points do not follow the Fisher error model.

Although the regions are centered at different rotations, the shape of the confidence region is the same for both the mean and median estimators.

Influence diagnostics for the least squares procedure were performed using the standardized influence function for each point. Since $c_1/(k_0 - k_1)^2 = 2\kappa$ for large κ,

$$\|\mathrm{SIF}\|^2 = \{2\kappa(1-t)\} \cdot \left\{2\left[\frac{x_1^2}{\lambda_2+\lambda_3} + \frac{x_2^2}{\lambda_1+\lambda_3} + \frac{x_3^2}{\lambda_1+\lambda_2}\right]\right\},$$

where $t = v^t\hat{A}u$ and x_1, x_2 and x_3 are as in equation (2.7). Under von Mises–Fisher errors, $2\kappa(1-t)$, which is $\kappa\|v - \hat{A}u\|^2$, is asymptotically distributed as $\chi_2{}^2$ for large κ [Watson (1984)]. So $\|\mathrm{SIF}\|^2$ may be separated into two parts, namely, the residual part $2\kappa(1-t)$ and the leverage–direction part

$$2\left[\frac{x_1^2}{\lambda_2+\lambda_3} + \frac{x_2^2}{\lambda_1+\lambda_3} + \frac{x_3^2}{\lambda_1+\lambda_2}\right].$$

Note that the leverage–direction part measures both the leverage of the design point u and the influence of the direction of the residual vector.

For this data set we used the R-median rotation and $\hat{\kappa}_{MC}$ in estimating the residual part. In estimating the leverage–direction part, we used the

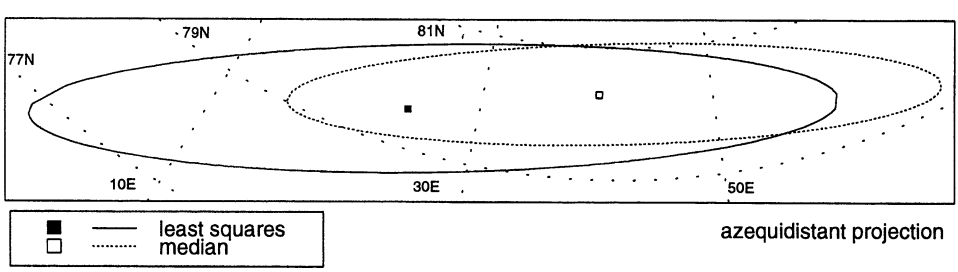

FIG. 3. *Projection of 95% confidence region for the reconstruction of Africa to North America at anomaly 6′.*

eigenvector decomposition of Σ_{-i}, the matrix defined analogously to Σ but without using the ith point u_i. Note that (u_i, v_i) should be expressed in a coordinate system which diagonalizes Σ_{-i}. (See Table 1.)

Comparing the $\|\mathrm{SIF}\|^2$ with $\chi_3{}^2$, we may identify points with high influence on least squares estimate of the rotation matrix. Outlying data points are identified by comparing residual part with a $\chi_2{}^2$ distribution. The points $i = B, J, K, P$ have high influence due to the large residuals at the 0.01 level and $i = H$ at the 0.05 level. The point A, whose standardized influence function has a large component due to leverage and direction of residual vector, does not have a large residual. Thus this point has low influence. To a lesser degree, the same behavior is exhibited by points L, Q and R. Cook's delete-one procedures, as adapted by Rivest (1989) (the t_i^2 and D_i columns in Table 1) identify only one point $i = K$ at the 0.01 level and another point $i = J$ at the 0.05 level.

5. Euclidean space models. For Euclidean space models, a rigid transformation is of the form $x \to Ax + b$, where $A \in \mathrm{O}(p)$ [or $\mathrm{SO}(p)$ if the transformation is orientation preserving] and $b \in R^p$. The corresponding M-estimators are obtained by minimizing an expression of the form $(1/n)\Sigma\,\rho(\|v_i - A(u_i - \bar{u}) - \beta\|^2)$, where $\beta = b + A\bar{u}$. However, in many applications, especially in shape analysis and problems of image registration, transformations of the form $x \to \gamma Ax + b$ are preferred. The additional parameter $\gamma \in R^1$ is interpreted as a scale change.

TABLE 1
Influence statistics for Central Atlantic 6 data set

	t_i^2	D_i	Residual part	Leverage direction	$\|\mathrm{SIF}\|^2$
A	0.03	0.00	0.00	6.01	0.00
B	2.65	0.10	9.63	2.63	25.38
C	0.48	0.02	1.51	2.01	3.03
D	0.48	0.02	3.02	2.00	6.06
E	0.32	0.01	0.84	2.03	1.71
F	0.04	0.00	0.51	2.04	1.03
G	0.19	0.01	1.67	2.02	3.37
H	1.38	0.05	7.88	2.06	16.23
I	0.18	0.01	0.43	2.03	0.88
J	4.35	0.13	19.52	2.08	40.62
K	8.87	0.22	25.50	2.01	51.37
L	0.04	0.00	1.26	3.25	4.09
M	0.06	0.00	1.26	2.88	3.62
N	0.41	0.02	3.45	2.49	8.57
O	0.24	0.02	0.05	2.81	0.14
P	2.52	0.09	11.52	2.00	23.05
Q	0.24	0.01	0.62	3.46	2.15
R	0.02	0.00	0.98	3.13	3.05
S	0.09	0.00	1.19	2.87	3.41
T	0.57	0.02	1.68	2.20	3.70

Letting $\psi(s) = \rho'(s)$, we get

$$0 = \Sigma\psi(\hat{s}_i)\Big[\hat{A}^t\big(v_i - \hat{\gamma}\hat{A}(u_i - \bar{u}) - \hat{\beta}\big)(u_i - \bar{u})^t - (u_i - \bar{u})\big(v_i - \hat{\gamma}\hat{A}(u_i - \bar{u}) - \hat{\beta}\big)^t\hat{A}\Big], \tag{5.1}$$

$$0 = \Sigma\psi(\hat{s}_1)\big(v_i - \hat{\gamma}\hat{A}(u_i - \bar{u}) - \hat{\beta}\big), \tag{5.2}$$

$$0 = \Sigma\psi(\hat{s}_i)\big(v_i - \hat{\gamma}\hat{A}(u_i - \bar{u}) - \hat{\beta}\big)^t\hat{A}(u_i - \bar{u}), \tag{5.3}$$

where $\hat{s}_i = \|v_i - \hat{\gamma}\hat{A}(u_i - \bar{u}) - \hat{\beta}\|^2$. If scale changes are not allowed, we get equations (5.1) and (5.2) with $\hat{\gamma} \equiv 1$.

Let $\bar{u} = E(u)$, $\Sigma = E[(u - \bar{u})(u - \bar{u})^t]$ and $s = \|v - \gamma A(u - \bar{u}) - \beta\|^2$. We assume that the distribution of v, conditional on u, is of the form $f(s)$, and we let $g(s) = \log f(s)$. Define positive constants d_1 and d_2 as $d_1 = E[4g'(s)^2 s/p]$ and $d_2 = -E[\psi(s)g'(s)s]$. Analogously to Lemma 2, it can be shown that $d_2 = E[\psi'(s)s + (p/2)\psi(s)]$. Proceeding as in (2.5), the information matrix $I[\hat{A}(F), \hat{\beta}(F), \hat{\delta}(F)]$ satisfies

$$\begin{aligned}&\operatorname{Vect}\big[\hat{A}(F)H_1, h_1, g_1\big]^t \\ &\quad\cdot I\big[\hat{A}(F), \hat{\beta}(F), \hat{\gamma}(F)\big]\cdot\operatorname{Vect}\big[\hat{A}(F)H_2, h_2, g_2\big] \\ &\quad = d_1\big[\hat{\gamma}(F)^2\operatorname{Tr}\big(H_1^t\Sigma H_2\big) + h_1^t h_2 + g_1 g_2\operatorname{Tr}(\Sigma)\big].\end{aligned} \tag{5.4}$$

Here H_1 and H_2 are skew $p \times p$ matrices; h_1 and h_2 are vectors in R^p; and g_1 and g_2 are real numbers.

Because of (5.4), the SIF of $\hat{A}$, $\hat{\beta}$ and $\hat{\gamma}$ can be considered separately. In light of Lemma 2, the following proposition is easily seen to be the Euclidean space analog of Proposition 1.

PROPOSITION 5. *Let $H = \hat{A}(F)^t \mathrm{IF}(x; \hat{A}, F)$. Then H is the unique $p \times p$ skew-symmetric matrix for which*

$$2\hat{\gamma}(F)\frac{d_2}{p}\Sigma\cdot H + \psi(s^*)[u^* - \bar{u}] \times\big[v^* - \hat{\gamma}(F)\hat{A}(F)(u^* - \bar{u}) - \hat{\beta}(F)\big]^t\hat{A}(F) \tag{5.5}$$

is symmetric;

$$\mathrm{IF}\big(x; \hat{\beta}, F\big) = \frac{p}{2d_2}\psi(s^*)[v^* - \hat{\gamma}(F)\hat{A}(F)(u^* - \bar{u}) - \hat{\beta}(F)], \tag{5.6}$$

$$\begin{aligned}\mathrm{IF}(x; \hat{\gamma}, F) &= \frac{p}{2d_2\cdot\operatorname{Tr}(\Sigma)}\psi(s^*) \\ &\quad\times\big[v^* - \hat{\gamma}(F)\hat{A}(F)(u^* - \bar{u}) - \hat{\beta}(F)\big]^t\hat{A}(F)(u^* - \bar{u}).\end{aligned} \tag{5.7}$$

In a model without scale change parameter γ, equations (5.5) *and* (5.6) *hold when $\hat{\gamma}(F) \equiv 1$.*

The proof uses the same ideas as that of Proposition 1 and is omitted.

As in the spherical regression case, suppose $p = 3$, and let $\lambda_1 \geq \lambda_2 \geq \lambda_3$ be the eigenvalues of Σ. For $\hat{A}$, we have (for models with or without scale change γ)

$$\|\mathrm{SIF}(\hat{A})\|^2 = \frac{p^2 d_1}{4d_2^2}\psi^2(s)s\|u - \bar{u}\|^2\left\{\frac{x_1^2}{\lambda_2 + \lambda_3} + \frac{x_2^2}{\lambda_1 + \lambda_3} + \frac{x_3^2}{\lambda_1 + \lambda_2}\right\}, \tag{5.8}$$

where $w = A^t(v - \gamma A(u - \bar{u}) - \beta)/s^{1/2}$ with $s = \|v - \gamma A(u - \bar{u}) - \beta\|^2$ and, in a coordinate system of eigenvectors of Σ,

$$\frac{u - \bar{u}}{\|u - \bar{u}\|} \times w = [x_1, x_2, x_3]^t.$$

Thus, as before, we conclude the following:

1. The standardized influence of an observation is not only determined by the length of the residual and the length of the centered design point $u - \bar{u}$, but also by the relative location of $u - \bar{u}$ with respect to eigenvectors of Σ and the direction of the residual. The maximum influence for given lengths of residual and $u - \bar{u}$ occurs if and only if $u - \bar{u}$ is perpendicular to the dominant eigenvector e_1 of Σ and $w = \pm((u - \bar{u}) \times e_1)/\|u - \bar{u}\|$. The minimum influence for any u is 0 and occurs if and only if $w = \pm(u - \bar{u})/\|u - \bar{u}\|$.
2. Notice that the maximum influence is minimized for fixed $\mathrm{Tr}(\Sigma)$ by $\lambda_1 = \lambda_2 = \lambda_3$. Thus the choice of design which would make the influence of any single point small is a spherical symmetric distribution of the u_i about $\bar{u}$ concentrated on a sphere of large radius.
3. The standardized influence of (u_i, v_i) on $\hat{\beta}$ depends only upon the length of the residual vector. Indeed (for models with or without scale change),

$$\|\mathrm{SIF}(\hat{\beta})\|^2 = \frac{p^2 d_1}{4d_2^2}\psi^2(s)s. \tag{5.9}$$

4. Finally,

$$\|\mathrm{SIF}(\hat{\gamma})\|^2 = \frac{p^2 d_1}{4d_2^2 \cdot \mathrm{Tr}(\Sigma)}\psi^2(s)s((u - \bar{u}) \cdot w)^2 \tag{5.10}$$

   Thus, when $p = 3$, for a given length of residual, the influence of an observation on $\hat{\gamma}$ will be maximized when the residual, after back-transformation by A^t, is parallel to $u - \bar{u}$. In this event, its influence on $\hat{A}$ will be zero. Similarly, for any fixed u, an observation with maximal influence on $\hat{A}$ will have zero influence on $\hat{\gamma}$.

For rigid body motion in the plane, so that $p = 2$, equation (5.8) becomes

$$\|\mathrm{SIF}(\hat{A})\|^2 = \frac{d_1}{d_2^2}\psi^2(s)s(\lambda_1 + \lambda_2)^{-1}\|(u - \bar{u}) \times w\|^2. \tag{5.11}$$

Thus a point u_i has high leverage on $\hat{A}$ if it is far from $\bar{u}$. However, even if u_i has high leverage and the length of the residual is large, the pair (u_i, v_i) will have little influence on $\hat{A}$ if the back-transformed residual $A^t(v - \gamma A(u - \bar{u}) - \beta)$ is close to parallel to $u - \bar{u}$. In models with scale change γ, equations (5.10) and (5.11) yield

$$\|\mathrm{SIF}(\hat{A})\|^2 + \|\mathrm{SIF}(\hat{\gamma})\|^2 = \frac{d_1}{d_2^2 \cdot \mathrm{Tr}(\Sigma)} \gamma^2(s) s \|u - \bar{u}\|^2.$$

Thus there is a direct trade-off between the influence of an observation on $\hat{A}$ and its influence on $\hat{\gamma}$.

The asymptotic covariance of $\hat{A}$, $\hat{\beta}$ and $\hat{\gamma}$ is derived in much the same way as in Proposition 3. Let $\hat{A}_n = \hat{A}(F_n)$, $A = \hat{A}(F)$ and write $\hat{A}_n = A\Phi(H_n)$ for a skew-symmetric matrix H_n. The proofs in Propositions 2 and 3 of consistency and asymptotic normality do not apply as compactness arguments were used there. However, it is easy to show, assuming the consistency of $\hat{A}_n$, $\hat{\beta}_n = \hat{\beta}(F_n)$, $\hat{\gamma}_n = \hat{\gamma}(F_n)$ and the joint asymptotic normality of H_n, $\hat{\beta}_n$ and $\hat{\gamma}_n$, that the asymptotic distribution of H_n has a density proportional to $\exp[-2nd_2^2\gamma^2 \,\mathrm{Tr}(H^t\Sigma H)/(pd_3)]$, where $d_3 = E[\psi^2(s)s]$. Similarly, $\hat{\beta}_n$ is asymptotically multivariate $N(\beta, pd_3 I_p/(4nd_2^2))$, where I_p is a $p \times p$ identity matrix and $\hat{\gamma}_n$ will be asymptotically $N(\gamma, pd_3/\{4nd_2^2 \,\mathrm{Tr}(\Sigma)\})$. Asymptotically, H_n, $\hat{\beta}_n$ and $\hat{\gamma}_n$ will be independent.

Thus the efficiency of the median estimator of A, β and γ relative to the mean estimator at a p-variate isotropic normal distribution is $2 \cdot \Gamma((p+1)/2)^2/\{p \cdot \Gamma(p/2)^2\}$, or $2/\pi = 0.6366$, $\pi/4 = 0.7854$ and $8/(3\pi) = 0.8488$ for $p = 1, 2, 3$. Brown (1983) has shown that these are the same as the efficiencies of the spatial median to the mean vector. As $p \to \infty$, this efficiency increases to 1.

More generally, if x_i are i.i.d. with a density of the form $f(s_i)$, where $s_i = \|x_i - \theta\|^2$, and if θ is estimated by an equation of the form $\Sigma\psi(\hat{s}_i)(x_i - \hat{\theta}) = 0$, then the asymptotic distribution of $\hat{\theta}$ is multivariate normal with mean θ and covariance matrix $pd_3 I_p/(nd_2^2)$. It follows that the efficiency relative to the mean estimators of any M-estimator of $\hat{A}$, $\hat{\beta}$ and $\hat{\gamma}$ satisfying equations of the form (5.1)–(5.3) will be the same as the corresponding M-estimator in the location problem. Also, the shape of a confidence region of A, β and γ will depend only upon Σ; its size will depend upon ψ and the underlying error distribution. Similar phenomena occur in univariate multiple linear regression, as can be seen in Huber [(1981), equation (5.2)].

6. An example in three dimensions. Table 2 contains the digitized locations of 12 points on the left and right hands (with fingers fully spread apart) of one of the authors. We consider the problem of bringing them into coincidence using a transformation of the form $x \to \gamma AX + b$. In this case, the matrix A must be orthogonal with determinant -1, that is, it represents both a rotation and a reflection. The parameter γ allows for the possibility of slightly different-sized hands or different spread.

TABLE 2
Digitized locations on hand

	Left hand			Right hand			
A	5.17	11.30	16.18	5.91	11.16	16.55	*Top of little finger*
B	7.40	12.36	17.5	8.63	10.62	18.33	*Top of ring finger*
C	8.56	12.59	17.87	10.09	10.60	18.64	*Top of middle finger*
D	9.75	13.62	17.01	10.89	10.95	17.90	*Top of forefinger*
E	11.46	14.55	12.96	12.97	10.13	13.88	*Top of thumb*
F	7.10	13.12	12.56	8.79	11.21	13.17	*Gap between thumb and forefinger*
G	8.85	13.82	12.60	10.70	11.10	13.42	*Center of palm*
H	6.77	13.07	10.32	8.47	11.09	11.35	*Base of palm*
I	6.26	11.62	13.34	7.28	12.52	14.04	*Little finger knuckle*
J	6.83	12.00	13.83	8.05	12.42	14.56	*Ring finger knuckle*
K	7.94	12.29	13.84	9.07	12.39	14.86	*Middle finger knuckle*
L	8.68	12.71	13.67	10.15	12.17	14.44	*Forefinger knuckle*

The least squares solution [with $\psi(s) = 1$] to equations (5.1)–(5.3) has a closed form. Writing $O_1 \Lambda O_2^t$ for the singular-value decomposition of the matrix $\Sigma(u_i - \bar{u})(v_i - \bar{v})^t$, it is well known that [see Goodall (1991)] $\hat{A} = O_2 O_1^t$, $\hat{\gamma} = \Sigma(v_i - \bar{v})^t \hat{A}(u_i - \bar{u}) / \Sigma(u_i - \bar{u})^t(u_i - \bar{u})$ and $\hat{\beta} = \bar{v}$.

To solve equations (5.1)–(5.3) for the L_1 (median) solution [with $\psi(s) = s^{-1/2}$], the following iterative scheme was selected. The least squares solution was chosen as initial guess. Using the values of $\hat{s}_i$ obtained from a given iterate, equations (5.2) and (5.3) were used to calculate the next iterates for $\hat{\beta}$ and $\hat{\gamma}$, respectively. Equation (5.1) is equivalent to the condition that $\Sigma\psi(\hat{s}_i)(u_i - \bar{u})(v_i - \hat{\beta})^t \hat{A}$ be symmetric. If the singular values of the matrix $\Sigma\psi(\hat{s}_i)(u_i - \bar{u})(v_i - \hat{\beta})^t$ are distinct, there are eight solutions, for fixed $\hat{s}_i$, in O(3) for the next iterate of $\hat{A}$. The correct solution is obtained by finding the singular-value decomposition $O_1 \Lambda O_2^t$ of $\Sigma\psi(\hat{s}_i)(u_i - \bar{u})(v_i - \hat{\beta})^t$ and choosing the next iterative to be $O_2 O_1^t$.

The influence of an observation on $\hat{\beta}$ is proportional to s for the least squares solution and constant for the L_1 solution. Using equations (5.8) and (5.10), the influence on $\hat{A}$ and $\hat{\gamma}$ of each observation (for each of the two solutions, L_1 and least squares) was calculated. To control the effect of individual observations, the L_1 solutions for the population values of A, β and γ were used to estimate s and, when the effect of the ith observation was calculated, u_i was not used in the calculation of Σ. We consider here the question of which observations are most influential on $\hat{A}$ and $\hat{\gamma}$, without considering whether the observations are unreasonably influential. Thus the values of $\|\mathrm{SIF}\|^2$ were renormalized to make the sum of the influences of the 12 observations of $\hat{\gamma}$ and $\hat{A}$ equal to 1. The results are plotted in Figure 4 together with the lengths $s^{1/2}$ of the residuals.

Point E (top of thumb) has both the longest residual and the largest value of $\|u - \bar{u}\|$. It is the most influential observation for both the least squares and L_1 estimates of A, but its influence can be greatly tamed by using an L_1 estimator. Note that for the L_1 estimator, $\psi^2(s)s = 1$ and hence the length of

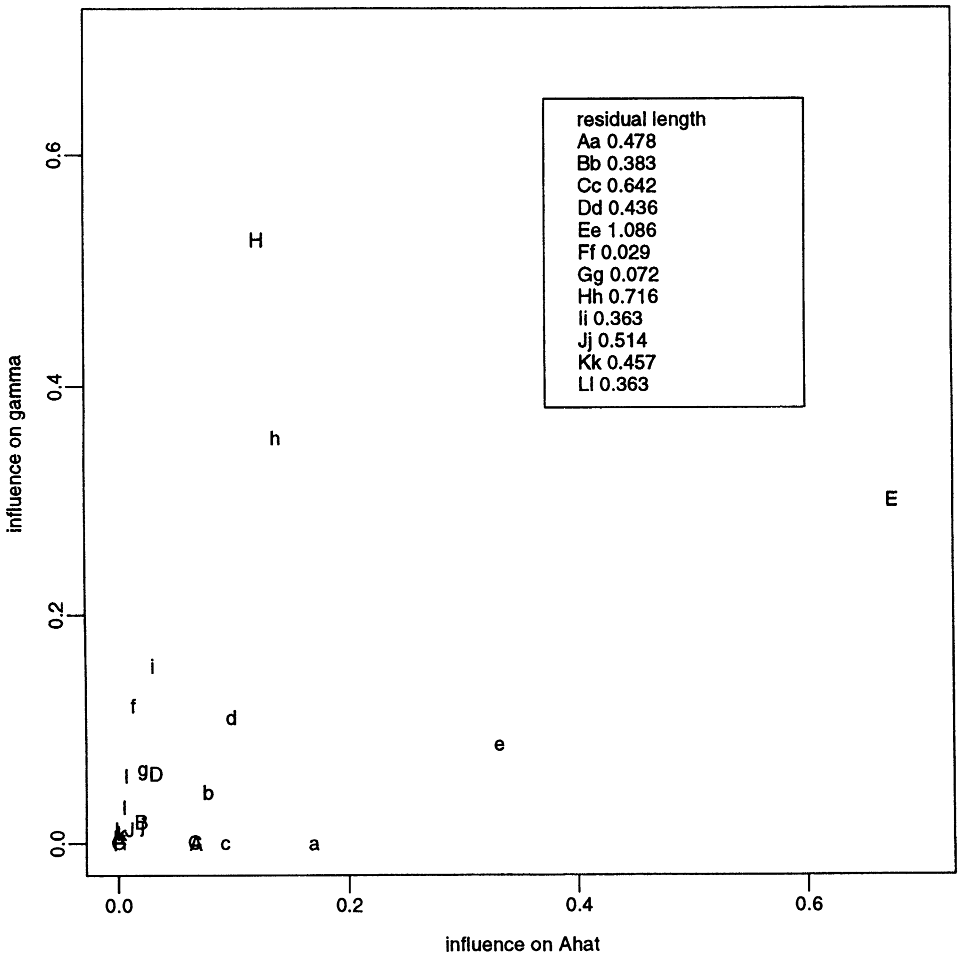

FIG. 4. *Relative influence on mean (upper case) and median (lower case) estimates of A and* γ *for the hands data of Table* 2.

the residual is not a factor in the influence of E on an L_1 estimator. Thus the influence of E on an L_1 estimator $\hat{A}$, must be due to the direction of $u_5 - \bar{u}$ and w. If the point E is not used in the calculation of Σ, $u_5 - \bar{u}$ forms an angle of 13° with e_2, and the residual forms an angle of 12° with e_3. Note that if $u - \bar{u} = e_2$ and $w = e_3$, the observation will have maximum influence for a given length $\|u - \bar{u}\|$.

The most striking point is H (base of the palm): H is by far the most influential point on the scale estimate; H has a substantially shorter residual than E and the L_1 estimator is less effective in ameliorating its influence. Points H and G (center of palm) are the least well defined points, but H defines the length of the hand. Thus H's influence on the scale parameter $\hat{\gamma}$ is not surprising.

An interesting comparison is between observation H and C (top of the middle finger). The least squares estimate of γ is 0.9925 and the L_1 estimate 1.0086. If C is deleted, these estimates become 0.9895 and 1.0047. If H is deleted, these estimates are changed more dramatically: to 1.0110 and 1.0262, respectively. Notice that H is far more influential than C.

After centering by $\bar{u}$, and in a coordinate system consisting of the eigenvectors of Σ, the coordinates of H are $[-3.98, 1.15, 0.33]$. The coordinates of C are $[3.55, -0.77, -0.03]$. Notice that the coordinates of C are close to minus the coordinates of H. Since equation (5.10) for the SIF on $\hat{\gamma}$ remains unchanged if $u - \bar{u}$ is replaced by its negative, location of the design point u cannot account for the difference in the effect of H and C on $\hat{\gamma}$. Also, the lengths of the residual at C and H are fairly close, so this also cannot account for the difference in the influence of the two points on $\hat{\gamma}$. However, the residual at C forms an angle of 88° with $u_3 - \bar{u}$ and hence the influence of C on $\hat{\gamma}$ is almost negligible. The residual at H forms an angle of 124° with $u_8 - \bar{u}$, accounting for the greater influence of H on $\hat{\gamma}$.

Thus if the registration is unsatisfactory, the point E should be reexamined when $\hat{A}$ or, if least squares is used, $\hat{\beta}$ is suspect and the point H should be reexamined when $\hat{\gamma}$ is suspect. Because of the normalization to make the sum of the influences equal to 1, this analysis does not depend upon the underlying probability density f.

Acknowledgments. The authors wish to express their deep appreciation to Peter Molnar, Joann Stock, Kim Klitgord and Hans Schouten for their help in obtaining the data used here. We wish to thank the referees for their close reading of the manuscript and for many helpful suggestions.

We especially wish to thank Dr. Nell Sedransk for the masterpiece of bureaucratic coordination that our four support awards represent.

REFERENCES

Brown, B. M. (1983). Statistical uses of the spatial median. *J. Roy. Statist. Soc. Ser. B* **45** 25–30.

Chang, T. (1986). Spherical regression. *Ann. Statist.* **14** 907–924.

Chang, T. (1987). On the statistical properties of estimated rotations. *Journal of Geophysical Research B* **92** 6319–6329.

Chang, T. (1993). Spherical regression and the statistics of tectonic plate reconstructions. *Internat. Statist. Rev.* **61** 299–316.

Chapman, C. R., Chen, G. and Kim, P. T. (1995). Assessing geometric integrity through spherical regression techniques. *Statist. Sinica* **5** 173–220.

Fisher, N. I. (1985). Spatial medians. *J. Roy. Statist. Soc. Ser. B* **47** 341–348.

Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *J. Roy. Statist. Soc. Ser. B* **53** 285–339.

Hampel, F. R., Ronchetti, E. M., Rouseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

Huber, P. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233. Univ. California Press, Berkeley.

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

KLITGORD, K. and SCHOUTEN H. (1986). Plate kinematics of the central Atlantic. In *Geology of North America* **M** (P. R. Vogt and B. E. Tucholke, eds.) 351–358. Geological Society of America, Boulder, CO.

KO, D. (1992). Robust estimation of the concentration parameter of a von Mises Fisher distribution. *Ann. Statist.* **20** 917–928.

KO, D. and CHANG, T. (1993). Robust M-estimators on spheres. *J. Multivariate Anal.* **45** 104–136.

MACKENZIE, J. K. (1957). The estimation of an orientation relationship. *Acta Cryst.* **10** 61–62.

RIVEST, L. (1989). Spherical regression for concentrated Fisher–von Mises distributions. *Ann. Statist.* **17** 307–317.

SHAW, P. R. and CANDE, S. C. (1990). High-resolution inversion for South Atlantic plate kinematics using joint altimeter and magnetic anomaly data. *Journal of Geophysical Research B* **95** 2625–2644.

SIEGEL, A. F. and BENSON, R. H. (1982). A robust comparison of biological shapes. *Biometrics* **38** 341–350.

STEPHENS, M. A. (1979). Vector correlation. *Biometrika* **66** 41–48.

WATSON, G. S. (1984). The theory of concentrated Langevin distributions. *J. Multivariate Anal.* **14** 74–82.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VIRGINIA 22903-3199

DEPARTMENT OF BIOSTATISTICS
MEDICAL COLLEGE OF VIRGINIA
VIRGINIA COMMONWEALTH UNIVERSITY
RICHMOND, VIRGINIA 23298-0032