The Annals of Statistics
2001, Vol. 29, No. 3, 784–814

M-ESTIMATION FOR LOCATION AND REGRESSION PARAMETERS IN GROUP MODELS: A CASE STUDY USING STIEFEL MANIFOLDS[1]

By Ted Chang and Louis-Paul Rivest

University of Virginia and Université Laval

We discuss here a general approach to the calculation of the asymptotic covariance of M-estimates for location parameters in statistical group models when an invariant objective function is used. The calculation reduces to standard tools in group representation theory and the calculation of some constants. Only the constants depend upon the precise forms of the density or of the objective function. If the group is sufficiently large this represents a major simplification in the computation of the asymptotic covariance.

Following the approach of Chang and Tsai we define a regression model for group models and derive the asymptotic distribution of estimates in the regression model from the corresponding distribution theory for the location model. The location model is not, in general, a subcase of the regression model.

We illustrate these techniques using Stiefel manifolds. The Stiefel manifold $\mathscr{V}_{p,m}$ is the collection of $p \times m$ matrices $\mathbf{X}$ which satisfy the condition $\mathbf{X}^T\mathbf{X} = \mathbf{I}_m$ where $m \le p$. Under the assumption that $\mathbf{X}$ has a distribution proportional to $\exp(\mathrm{Tr}(\mathbf{F}^T\mathbf{X}))$, for some $p \times m$ matrix $\mathbf{F}$, Downs (1972) gives approximations to maximum likelihood estimation of $\mathbf{F}$. In this paper, we consider a somewhat different location problem: under the assumption that $\mathbf{X}$ has a distribution of the form $f(\mathrm{Tr}(\boldsymbol{\theta}_0^T\mathbf{X}))$ for some $\boldsymbol{\theta}_0 \in \mathscr{V}_{p,m}$, we calculate the asymptotic distribution of M-estimates which minimize an objective function of the form $\sum_i \rho(\mathrm{Tr}(\boldsymbol{\theta}^T\mathbf{X}_i))$. The assumptions on the form of the density and the objective function can be relaxed to a more general invariant form. In this case, the calculation of the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ reduces to the calculation of four constants and we present consistent estimators for these constants.

Prentice (1989) introduced a regression model for Stiefel manifolds. In the Prentice model, $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n \in \mathscr{V}_{p,m}$ are fixed, $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_n \in \mathscr{V}_{p,m}$ are independent random so that the distribution of $\mathbf{V}_i$ depends only upon $\mathrm{Tr}(\mathbf{V}_i^T\mathbf{A}_2\mathbf{u}_i\mathbf{A}_1^T)$ for unknown $(\mathbf{A}_1, \mathbf{A}_2) \in SO(m) \times SO(p)$. We discuss here M-estimation of $\mathbf{A}_1$ and $\mathbf{A}_2$ under general invariance conditions for both the density and the objective function.

Using a well-studied example on vector cardiograms we discuss the physical interpretation of the invariance assumption as well as of the parameters $(\mathbf{A}_1, \mathbf{A}_2)$ in the Prentice regression model. In particular, $\mathbf{A}_1$ represents a rotation of the $\mathbf{u}$'s to the $\mathbf{V}$'s in a coordinate system relative to the $\mathbf{u}$'s and $\mathbf{A}_2$ represents a rotation of the $\mathbf{u}$'s to the $\mathbf{V}$'s in a coordinate system fixed to the external world.

Received January 2000; revised December 2000.

[1]Supported in part by the National Sciences and Engineering Research Council of Canada and by the Programme pour la formation des chercheurs of Québec.

AMS 2000 *subject classifications*. Primary 62F12, 62H11, 62P10.

Key words and phrases. Group models, equivariance, directional statistics, rotations, orientations, rigid body motion.

1. Introduction. Statistical group models and invariance arguments arise quite naturally in problems of inference, especially in the physical sciences. Often it is desirable that inferences not depend upon the units of measurement, or, for spatial data, on the coordinate system which we use to translate such data into vectors and matrices. This notion of invariance under change of scale or coordinate system can usually be mathematically formulated using statistical group models.

When a statistical group model is justified, the requirement of invariance can greatly restrict the range of statistical procedures to be considered. This viewpoint is developed in Lehmann (1983) and Eaton (1989). Such models occur naturally in the modelling of directional data [see Mardia and Jupp (2000), page 33]. In addition, the use of invariance arguments can greatly simplify statistical calculations; see, for example, Farrell (1985) where the distributions of important multivariate statistics are calculated using invariance arguments.

Similarly, we will show that the asymptotic distributions of invariant M-estimators for location and regression parameters in group models are often determined up to a few constants by standard techniques in group representation theory. Except for these constants, the precise forms of the density and the objective functions are irrelevant.

We will illustrate our techniques using Stiefel manifolds. The Stiefel manifold $\mathscr{V}_{p,m}$, $m \le p$ is the collection of $p \times m$ matrices $\mathbf{X}$ which satisfy the condition $\mathbf{X}^T\mathbf{X} = \mathbf{I}_m$. Special and important examples are the $p-1$-dimensional sphere $\mathscr{V}_{p,1} = \Omega_p \subset R^p$ and the orthogonal group $\mathscr{O}(p) = \{\mathbf{X}_{p\times p} | \mathbf{X}^T\mathbf{X} = \mathbf{I}_p\}$. We are interested in M-estimation for both location and regression parameters on $\mathscr{V}_{p,m}$. Because the orthogonal group $\mathscr{O}(p)$ has two connected components, one of which is the *special orthogonal group* $SO(p) = \{\mathbf{X} \in \mathscr{O}(p) | \det \mathbf{X} = 1\}$, our asymptotic results for $\mathscr{O}(p)$ apply equally well to $SO(p)$. In fact, by convention $\mathscr{V}_{p,p} = SO(p)$.

M-estimation for location problems on the sphere Ω_p were studied by Ko and Chang (1993). Writing the general element $\mathbf{X}$ of Ω_p as a length p column vector, Ko and Chang (1993) assumed that the distribution of $\mathbf{X}$ is of the form $f(\mathbf{X}^T\boldsymbol{\theta}_0)$ for some unknown modal direction $\boldsymbol{\theta}_0$. They estimated $\boldsymbol{\theta}_0$ by minimizing an objective function of the form $\rho(\mathbf{X}^T\boldsymbol{\theta})$ and calculated the asymptotic distribution of the resulting estimator $\hat{\boldsymbol{\theta}}$.

Motivated by a problem in plate tectonics, Chang (1986) introduced the *spherical regression* model: $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n \in \Omega_p$ are fixed, $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_n \in \Omega_p$ are independent random so that the distribution of $\mathbf{V}_i$ depends only upon $\mathbf{V}_i^T\mathbf{A}_0\mathbf{u}_i$ for some unknown $\mathbf{A}_0 \in SO(p)$. The asymptotic distribution of M-estimators $\widehat{\mathbf{A}}$ was derived by Chang and Ko (1995).

Location problems in Stiefel manifolds were studied by Downs (1972) and Khatri and Mardia (1977). The application considered there was to vector cardiograms. Downs, Khatri and Mardia assumed that the distribution of $\mathbf{X}$ is matrix Fisher, so that its density is proportional to $\exp(\mathrm{Tr}(\mathbf{F}^T\mathbf{X}))$ for some $p \times m$ matrix $\mathbf{F}$ and gave approximations to the maximum likelihood estimation of $\mathbf{F}$. Downs gave the following interpretation of the parameter matrix $\mathbf{F}$: using the singular value decomposition, we can write $\mathbf{F} = \boldsymbol{\theta}_0\mathbf{K}$ where $\boldsymbol{\theta}_0 \in \mathscr{V}_{p,m}$ is a modal matrix and $\mathbf{K}$ is an $m \times m$ symmetric matrix. $\mathbf{K}$ plays the role of a concentration parameter. If $\mathbf{K}$ is not a multiple of the identity, the distribution of $\mathbf{X}$ is anisotropic.

Chikuse (1993) derived large p asymptotic expansions for the matrix Fisher distribution on $\mathscr{V}_{p,m}$. Her work can be applied to approximate the conditional distribution of the maximum likelihood estimator of $\boldsymbol{\theta}_0$, with the conditioning on a statistic ancillary to $\boldsymbol{\theta}_0$. Chikuse's work is not asymptotic in sample size, but rather in p, and she envisions applications to compositional data where p represents the number of components of the composition.

Motivated by a problem of matched pairs of vector cardiograms, Prentice (1989) generalized the spherical regression model to a regression model on $\mathscr{V}_{p,m}$. In the Prentice model, $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n \in \mathscr{V}_{p,m}$ are fixed, $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_n \in \mathscr{V}_{p,m}$ are independent random so that the distribution of $\mathbf{V}_i$ is of the form $f_0(\mathrm{Tr}(\mathbf{V}_i^T\mathbf{A}_2\mathbf{u}_i\mathbf{A}_1^T))$ for unknown $(\mathbf{A}_1, \mathbf{A}_2) \in SO(m) \times SO(p)$. Prentice estimated $(\mathbf{A}_1, \mathbf{A}_2)$ by maximizing an objective function of the form $\sum_i \mathrm{Tr}(\mathbf{V}_i^T\mathbf{A}_2 \mathbf{u}_i\mathbf{A}_1^T)$. Unfortunately, as discussed below, there is an error in Prentice's proof and result, and we will revise his result in what follows.

The special case of $SO(3)$ has enjoyed a resurgence of interest due to the study of human motion. For estimating a location parameter $\boldsymbol{\theta}_0 \in SO(3)$, Rancourt, Rivest and Asselin (2000) derive the asymptotic distribution of an estimator $\hat{\boldsymbol{\theta}}$ which maximizes $\sum_i \mathrm{Tr}(\boldsymbol{\theta}^T\mathbf{X}_i)$ under rather general assumptions about the distribution of the $\mathbf{X}_i$. Similarly, for the Prentice regression model on $SO(3)$, Rivest and Chang (2000) calculate the asymptotic distribution of Prentice's estimators $(\widehat{\mathbf{A}}_1, \widehat{\mathbf{A}}_2)$ under equally general distributional assumptions on the distribution of the $\mathbf{V}_i$. Rivest and Chang (2000) also develop a test of correlation, based upon the Prentice regression model, for the $\mathbf{V}_i$ and the $\mathbf{u}_i$.

In this paper, we will consider M-estimation of location and regression on $\mathscr{V}_{p,m}$. For the location problem,

$$\begin{gathered}\mathbf{X}_1, \dots, \mathbf{X}_n \text{ are i.i.d. with density } f(\mathbf{X}, \boldsymbol{\theta}_0),\\ \text{satisfying } f(\mathbf{C}_2\mathbf{X}\mathbf{C}_1^T, \mathbf{C}_2\boldsymbol{\theta}_0\mathbf{C}_1^T) = f(\mathbf{X}, \boldsymbol{\theta}_0)\end{gathered} \tag{1}$$

for any $(\mathbf{C}_1, \mathbf{C}_2) \in SO(m) \times SO(p)$. Densities satisfying (1) include all densities of the form

$$f(\mathbf{X}, \boldsymbol{\theta}_0) = f_0(\mathrm{Tr}(\boldsymbol{\theta}_0^T\mathbf{X})). \tag{2}$$

We derive the asymptotic distribution of M-estimators which minimize an objective function

$$\sum_i \rho(\mathbf{X}_i, \boldsymbol{\theta}), \tag{3}$$

where ρ satisfies a condition similar to (1),

$$\rho(\mathbf{C}_2\mathbf{X}\mathbf{C}_1^T, \mathbf{C}_2\boldsymbol{\theta}\mathbf{C}_1^T) = \rho(\mathbf{X}, \boldsymbol{\theta}). \tag{4}$$

In particular, all ρ of the form

$$\rho(\mathbf{X}, \boldsymbol{\theta}) = \rho_0(\mathrm{Tr}(\boldsymbol{\theta}^T\mathbf{X})) \tag{5}$$

satisfy (4).

Thus, let X_{ij} and θ_{ij} denote the entries of $\mathbf{X}$ and $\boldsymbol{\theta}$, respectively. Then

$$\rho(\mathbf{X}, \boldsymbol{\theta}) = \sum_{i,j}(X_{ij} - \theta_{ij})^2 = \mathrm{Tr}[(\mathbf{X} - \boldsymbol{\theta})^T(\mathbf{X} - \boldsymbol{\theta})] = 2m - 2\ \mathrm{Tr}(\boldsymbol{\theta}^T\mathbf{X}) \tag{6}$$

yields an L^2 estimator. An L^1 estimator can be arrived at using

$$\rho(\mathbf{X}, \boldsymbol{\theta}) = \sqrt{2m - 2\ \mathrm{Tr}(\boldsymbol{\theta}^T\mathbf{X})}.$$

More generally, any ρ which is a function of the eigenvalues of $\mathbf{X}^T\boldsymbol{\theta} + \boldsymbol{\theta}^T\mathbf{X}$ satisfies (4). In particular our results apply if

$$\rho(\mathbf{X}, \boldsymbol{\theta}) = \|(\mathbf{X} - \boldsymbol{\theta})^T(\mathbf{X} - \boldsymbol{\theta})\|_{op}, \tag{7}$$

where $\|\mathbf{A}\|_{op}$ denotes the maximum eigenvalue of the positive semidefinite matrix $\mathbf{A}$.

Thus these results include as a corollary those of Ko and Chang (1993). The results of Downs (1972) and Khatri and Mardia (1977) are included whenever the concentration matrix $\mathbf{K}$ is a multiple of the identity. Similarly, the results of Rancourt, Rivest and Asselin (2000) are included under the additional symmetry condition (1).

For regression on $\mathbf{V}_{p,m}$, we use a generalization of the Prentice regression model,

$$\begin{aligned} &\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n \in \mathscr{V}_{p,m} \text{ are fixed,} \\ &\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_n \in \mathscr{V}_{p,m} \text{ are independent random,} \\ &\text{the distribution of } \mathbf{V}_i \text{ is of the form } f(\mathbf{V}_i, \mathbf{A}_2\mathbf{u}_i\mathbf{A}_1^T), \end{aligned} \tag{8}$$

where f satisfies (1) and $(\mathbf{A}_1, \mathbf{A}_2) \in SO(m) \times SO(p)$ is unknown. We will calculate the asymptotic distribution of estimators $(\widehat{\mathbf{A}_1}, \widehat{\mathbf{A}_2})$ which minimize an objective function of the form

$$\sum_i \rho(\mathbf{V}_i, \mathbf{A}_2\mathbf{u}_i\mathbf{A}_1^T), \tag{9}$$

where ρ satisfies the condition (4). Thus, besides revising Prentice (1989), our results include those of Chang (1988), Chang and Ko (1995) and, when (6) is true, those of Rivest and Chang (2000).

We will relate the asymptotic distribution of $(\widehat{\mathbf{A}}_1, \widehat{\mathbf{A}}_2)$ in the regression model (8) with objective function (9) to the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ in the location model (1) and objective function (3) using the same f and ρ.

Our proofs rely heavily on the statistical group model implied by (1) and (4). In Section 2 we review the notion of a statistical group model and related standard constructions in differential geometry and apply these to the calculation of the asymptotic distribution of M-estimators for location parameters.

In Section 3, we define a regression model for statistical group models and show how the asymptotic distribution of M-estimators for regression parameters can be derived from the asymptotic distribution of M-estimators for location parameters. We note that the location model is usually not a submodel of the regression model.

The asymptotic approximations in Sections 2 and 3 are large sample approximations. In many physical applications, concentrated error approximations are more relevant. Examples where these occur are Rivest (1989), Chang (1988), Rancourt, Rivest and Asselin (2000) and Rivest and Chang (2000). In Section 4, we show how the techniques of Sections 2 and 3 can be adapted to concentrated error asymptotics.

Sections 5 and 6 apply the general theory in detail to location and regression parameters, respectively, in $\mathscr{V}_{p,m}$. The results of Sections 2 and 3 reduce the calculation of the asymptotic distribution of M-estimators to the estimation of a small number of constants (at most four in the case of $\mathscr{V}_{p,m}$). In Section 7, we give consistent estimators for these constants. Section 8 revisits the vector cardiogram example considered by Downs (1972), Khatri and Mardia (1977) and Prentice (1989). It will be seen that the invariance conditions (1) and (4) are physically natural and that the regression parameters $(\mathbf{A}_1, \mathbf{A}_2)$ have important physical interpretations. Section 9 gives the proofs of the mathematical tools used in Sections 2 and 3.

2. The basic approach for group models. Let $\mathscr{X}$ denote a sample space, $\boldsymbol{\Theta}$ a parameter space, $f(\mathbf{x}, \boldsymbol{\theta})$ a family of densities and $\rho(\mathbf{x}, \boldsymbol{\theta})$ an objective function.

Suppose, temporarily, $\boldsymbol{\Theta} = R^q$, that $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ is a sample from $f(\mathbf{x}; \boldsymbol{\theta}_0)$ and that $\hat{\boldsymbol{\theta}}$ minimizes $\sum_i \rho(\mathbf{X}_i, \boldsymbol{\theta})$. This leads to the estimating equation $\mathbf{S}(\mathbf{X}, \boldsymbol{\theta}) = 0$ where $\mathbf{S}(\mathbf{X}, \boldsymbol{\theta}) = \sum_i \frac{\partial \rho(\mathbf{X}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$. Let $\boldsymbol{\nu}_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}} \mathbf{S}(\mathbf{X}, \boldsymbol{\theta}_0)$ and write $\boldsymbol{\nu}'_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0) = \frac{\partial}{\partial \boldsymbol{\theta}}\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \boldsymbol{\nu}_{\boldsymbol{\theta}_0}(\boldsymbol{\theta})$. It can be shown [see, e.g., Brown (1985)] that, under certain regularity conditions, $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is asymptotically multinormal with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{B}^{-1}\mathbf{A}\mathbf{B}^{-1^T}$ where, the $q \times q$ matrices $\mathbf{B}$ and $\mathbf{A}$ are $\mathbf{A} = \lim_{n\to\infty} \operatorname{Cov}_{\boldsymbol{\theta}_0}(\mathbf{S}(\mathbf{X}, \boldsymbol{\theta}_0))/n$ and $\mathbf{B} = \lim_{n\to\infty} \boldsymbol{\nu}'_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0)/n$.

In this paper, we shall always assume that the regularity conditions hold and concern ourselves with efficient calculation of $\mathbf{A}$ and $\mathbf{B}$.

Chang and Tsai (1999) have reformulated $\mathbf{A}$ and $\mathbf{B}$ in a coordinate-free manner for a differentiable manifold (without boundary) $\boldsymbol{\Theta}$. We describe below this reformulation in informal terms; a rigorous reformulation is given in Chang and Tsai (1999). In Euclidean space R^q if $\gamma = (\gamma_1, \dots, \gamma_q)$: $R^1 \to R^q$ is any curve and f: $R^q \to R^1$ is any function, the chain rule implies

$$(f \circ \gamma)'(0) = \sum_i \frac{\partial f}{\partial x_i}(\gamma(0))\gamma_i'(0).$$

Thus $(f \circ \gamma)'(0)$ depends only upon a basepoint $\gamma(0)$ and a "tangent" vector $\gamma'(0)$.

Viewed in this way, the notion of tangent vector generalizes to manifolds: a *tangent vector* to $\boldsymbol{\Theta}$ at $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is an equivalence class of curves $\gamma: R^1 \to \boldsymbol{\Theta}$ satisfying $\gamma(0) = \boldsymbol{\theta}$ where two curves, γ_1 and γ_2, are considered to be equivalent [define the same tangent vector at $\boldsymbol{\theta}$] if $(f \circ \gamma_1)'(0) = (f \circ \gamma_2)'(0)$ for any $f: \boldsymbol{\Theta} \to \mathrm{R}^1$. The collection of tangent vectors at $\boldsymbol{\theta}$ form a vector space denoted by $T_{\boldsymbol{\theta}}\boldsymbol{\Theta}$.

Chang and Tsai (1999) reformulated $\mathbf{A}$ for manifolds as a family of (possibly indefinite) inner products, one inner product on each $T_{\boldsymbol{\theta}}\boldsymbol{\Theta}$, as follows:

$$(10) \quad \langle \gamma_1'(0), \gamma_2'(0) \rangle_A = \mathrm{Cov}_{\boldsymbol{\theta}}\left[\left(\frac{d}{dt}\Big|_{t=0} \rho(\mathbf{X}, \gamma_1(t))\right)\left(\frac{d}{ds}\Big|_{s=0} \rho(\mathbf{X}, \gamma_2(s))\right)\right],$$

where γ_1 and γ_2 are curves with $\gamma_1(0) = \gamma_2(0) = \boldsymbol{\theta}$. In standard mathematical terminology, $\mathbf{A}$ is said to be a (possibly indefinite) *Riemannian metric* on $\boldsymbol{\Theta}$.

Similarly, $\mathbf{B}$ is reformulated as a family of bilinear forms, one on each $T_{\boldsymbol{\theta}}\boldsymbol{\Theta}$, by

$$(11) \quad \langle \gamma_1'(0), \gamma_2'(0) \rangle_B = \mathrm{E}_{\boldsymbol{\theta}}\left[\left(\frac{d}{dt}\Big|_{t=0} \rho(\mathbf{X}, \gamma_2(t))\right)\left(\frac{d}{ds}\Big|_{s=0} \log f(\mathbf{X}, \gamma_1(s))\right)\right].$$

In general, $\langle , \rangle_B$ is not symmetric; $\langle \gamma_1'(0), \gamma_2'(0) \rangle_B$ is not in general $\langle \gamma_2'(0), \gamma_1'(0) \rangle_B$. Lemma 1 of Chang and Tsai (1999), however, shows that $\langle , \rangle_B$ will be symmetric whenever $\tau_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}_0}(\rho(\mathbf{X}; \boldsymbol{\theta}))$ has a critical point at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ for any $\boldsymbol{\theta}_0$. In this case $\langle , \rangle_B$ has the alternative form

$$(12) \qquad \langle \gamma_1'(0), \gamma_2'(0) \rangle_B = -\mathrm{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2}{\partial s t}\Big|_{(s,\,t)=(0,\,0)} \rho(\mathbf{X}, \gamma(s, t))\right],$$

where $\gamma: R^2 \to \boldsymbol{\Theta}$ satisfies $\gamma(0, 0) = \boldsymbol{\theta}$, $\gamma_1(s) = \gamma(s, 0)$ and $\gamma_2(t) = \gamma(0, t)$. The condition that $\tau_{\boldsymbol{\theta}_0}(\boldsymbol{\theta})$ has a critical point at $\boldsymbol{\theta}_0$, for each $\boldsymbol{\theta}_0$, ensures that the right-hand side of (12) depends only upon $(\gamma_1'(0), \gamma_2'(0))$ and not upon the particular choice of γ.

Lemma 1 in Section 9 gives useful conditions to ensure that this requirement on $\tau_{\boldsymbol{\theta}_0}(\boldsymbol{\theta})$ holds. We note that it is essential to assume here that $\boldsymbol{\Theta}$ is a manifold without boundary, or at least that $\boldsymbol{\theta}_0$ be in the interior of $\boldsymbol{\Theta}$ to ensure that the derivatives of $\tau_{\boldsymbol{\theta}_0}(\boldsymbol{\theta})$ vanish whenever $\tau_{\boldsymbol{\theta}_0}(\boldsymbol{\theta})$ has a minimum at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Suppose now we have a Lie group $\mathscr{G}$ which *acts* on $\mathscr{X}$ and on $\boldsymbol{\Theta}$; that is, there are differentiable maps $\mathscr{G} \times \mathscr{X} \to \mathscr{X}$ and $\mathscr{G} \times \boldsymbol{\Theta} \to \boldsymbol{\Theta}$ denoted by $(g, \mathbf{x}) \to g \cdot \mathbf{x}$ and $(g, \boldsymbol{\theta}) \to g \cdot \boldsymbol{\theta}$, respectively, such that $1 \cdot \mathbf{x} = \mathbf{x}$, $g \cdot (h \cdot \mathbf{x}) = (gh) \cdot \mathbf{x}$, and similarly for $\boldsymbol{\Theta}$. In what follows gx will generally indicate group or matrix multiplication and $g \cdot x$ will indicate the action of g on x.

If, with respect to a $\mathscr{G}$-invariant measure, the density functions $f(\mathbf{x}; \boldsymbol{\theta})$ satisfy

$$(13) \qquad f(g \cdot \mathbf{x}, g \cdot \boldsymbol{\theta}) = f(\mathbf{x}; \boldsymbol{\theta}),$$

the triple $(\mathscr{X}, \boldsymbol{\Theta}, \mathscr{G})$ is said to be a *(differentiable) statistical group model*. We will also assume that all objective functions ρ are $\mathscr{G}$-invariant,

$$(14) \qquad \rho(g \cdot \mathbf{x}, g \cdot \boldsymbol{\theta}) = \rho(\mathbf{x}; \boldsymbol{\theta}).$$

Finally, we assume that $\mathscr{G}$ acts *transitively* on the parameter space $\mathbf{\Theta}$; that is, given $\boldsymbol{\theta}, \boldsymbol{\theta}_1 \in \mathbf{\Theta}$, there is a $g \in \mathscr{G}$ such that $\boldsymbol{\theta}_1 = g \cdot \boldsymbol{\theta}$.

EXAMPLE. The prototypical group action is a group of matrices acting on a space of vectors. Thus let $\mathscr{G} = SO(p)$ act on Ω_p by $\mathbf{A} \cdot \mathbf{x} = \mathbf{A}\mathbf{x}$. Under this action, ordinary Lebesgue surface measure is $\mathscr{G}$ invariant. Let $\mathscr{X} = \mathbf{\Theta} = \Omega_p$. In this case, the invariance conditions (13) and (14) are equivalent to

$$f(\mathbf{x}, \boldsymbol{\theta}) = f_0(\mathbf{x}^T\boldsymbol{\theta}) \quad \text{and} \quad \rho(\mathbf{x}, \boldsymbol{\theta}) = \rho_0(\mathbf{x}^T\boldsymbol{\theta}), \tag{15}$$

respectively. In other words, f and ρ depend only upon the distance from $\mathbf{x}$ to $\boldsymbol{\theta}$. Thus (15) is the form of the general $SO(p)$-invariant function on Ω_p. This $\mathscr{G}$ action is transitive.

More generally, let $\mathscr{X} = \mathbf{\Theta} = \mathscr{V}_{p,m}$ and let $\mathscr{G} = SO(m) \times SO(p)$ act on $\mathscr{X}$ (and on $\mathbf{\Theta}$) by $(\mathbf{A}_1, \mathbf{A}_2) \cdot \mathbf{X} = \mathbf{A}_2\mathbf{X}\mathbf{A}_1^T$. The invariance conditions (13) and (14) are simply (1) and (4) above. This $\mathscr{G}$ action is also transitive: it suffices to show that any $\boldsymbol{\theta} \in \mathscr{V}_{p,m}$ is of the form $\mathbf{A}_2\boldsymbol{\theta}_1\mathbf{A}_1^T$ where $\boldsymbol{\theta}_1 = [\mathbf{I}_m \;\; \mathbf{0}]^T$. To see this, we can let $\mathbf{A}_1 = \mathbf{I}_m$ and $\mathbf{A}_2 = [\boldsymbol{\theta} \;\; \boldsymbol{\theta}_\perp]$ where $\boldsymbol{\theta}_\perp$ is any $p \times (p - m)$ matrix so that $\mathbf{A}_2 \in SO(p)$.

For technical reasons, to be given later, we require that when $m = 2$, $p = 4$, (1) and (4) are true for $(\mathbf{C}_1, \mathbf{C}_2) \in \mathscr{O}(2) \times \mathscr{O}(4)$. These conditions will automatically hold whenever (2), (5) or (7) are true.

If $\gamma(t)$ is a curve on $\mathbf{\Theta}$, so is $g \cdot \gamma(t)$ and hence each $g \in \mathscr{G}$ induces a linear map $T_{\boldsymbol{\theta}}\mathbf{\Theta} \to T_{g\cdot\boldsymbol{\theta}}\mathbf{\Theta}$. If $\gamma'(0) \in T_{\gamma(0)}\mathbf{\Theta}$, we will denote its image under g by $g \cdot \gamma'(0)$. Chang and Tsai (1999) show that $\langle , \rangle_A$ and $\langle , \rangle_B$ are invariant under the action of $\mathscr{G}$. Note that if $\mathscr{H} = \mathscr{G}_{\boldsymbol{\theta}_0} = \{g \in \mathscr{G} | g \cdot \boldsymbol{\theta}_0 = \boldsymbol{\theta}_0\}$, then each $h \in \mathscr{H}$ takes $T_{\boldsymbol{\theta}_0}\mathbf{\Theta}$ into itself; that is, $T_{\boldsymbol{\theta}_0}\mathbf{\Theta}$ is a representation of $\mathscr{H}$ and $\langle , \rangle_A$ and $\langle , \rangle_B$ are invariant under this representation.

We assume that $\mathscr{H}$ is compact. This assumption is essential in what follows; it avoids topological difficulties and is essential for the representation theory used herein.

Lemma 1, stated and proved in Section 9, shows that if there does not exist a nonzero $\mathbf{v} \in T_{\boldsymbol{\theta}_0}\mathbf{\Theta}$ such that $h \cdot \mathbf{v} = \mathbf{v}$ for all $h \in \mathscr{H}$, then $\tau_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}_0}(\rho(\mathbf{X}; \boldsymbol{\theta}))$ has a critical point at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ *for any invariant f and any invariant ρ*. Notice that the assumptions of Lemma 1 have nothing to do with f or ρ besides their invariance. Thus, for example, it applies to $\mathscr{V}_{p,m}$ using the objective function (7).

Because of the assumed transitivity of the action of $\mathscr{G}$ on $\mathbf{\Theta}$, it suffices to perform these calculations at any convenient $\boldsymbol{\theta}_0$.

Thus, under very general conditions, $\langle , \rangle_A$ and $\langle , \rangle_B$ are both $\mathscr{H}$-invariant symmetric bilinear forms. This puts tremendous constraints on them. Indeed we have the following.

PROPOSITION 1. *Suppose the compact Lie group $\mathscr{H}$ is represented on the real vector space $\mathscr{V}$. Write $\mathscr{V} = \oplus\mathscr{V}_i$ as a direct sum of minimally invariant subspaces. Suppose $\langle , \rangle_0$ is an $\mathscr{H}$-invariant positive definite inner product and $\langle , \rangle$*

an $\mathscr{H}$-invariant symmetric bilinear form on $\mathscr{V}$. Then:

(a) *There exist constants c_i such that $\langle , \rangle = c_i \langle , \rangle_0$ on $\mathscr{V}_i$.*

(b) *If $\mathscr{V}_i$ and $\mathscr{V}_j$ are inequivalent as representations of $\mathscr{H}$, they are orthogonal under $\langle , \rangle$ (and $\langle , \rangle_0$).*

Proposition 1 is well known for complex representations. It can be extended to real representations using the relationships between a real representation and its complexification. We outline this proof in Section 9.

Thus, write $T_{\boldsymbol{\theta}_0}\boldsymbol{\Theta} = \oplus \mathscr{V}_i$ as a direct sum of minimally invariant subspaces and suppose the $\mathscr{V}_i$ are all inequivalent. Then there are constants c_i and d_i such that

$$
\begin{aligned}
\langle \boldsymbol{\delta}, \boldsymbol{\delta} \rangle_A &= \sum_i c_i \langle \boldsymbol{\delta}_i, \boldsymbol{\delta}_i \rangle_0, \\
\langle \boldsymbol{\delta}, \boldsymbol{\delta} \rangle_B &= \sum_i d_i \langle \boldsymbol{\delta}_i, \boldsymbol{\delta}_i \rangle_0, \qquad (16)\\
\delta &= \sum_i \boldsymbol{\delta}_i, \qquad \boldsymbol{\delta}_i \in \mathscr{V}_i.
\end{aligned}
$$

This process constructs an asymptotic distribution in $T_{\boldsymbol{\theta}_0}\boldsymbol{\Theta}$, but $\hat{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$. Let $\Phi_{\boldsymbol{\theta}_0}\colon T_{\boldsymbol{\theta}_0}\boldsymbol{\Theta} \to \boldsymbol{\Theta}$ be any map such that $\Phi_{\boldsymbol{\theta}_0}(\mathbf{0}) = \boldsymbol{\theta}_0$ and such that the derivative of $\Phi_{\boldsymbol{\theta}_0}$ at $\mathbf{0}$ is the identity map. This latter condition means that if $\mathbf{v} \in T_{\boldsymbol{\theta}_0}\boldsymbol{\Theta}$, then $\frac{d}{dt}\big|_{t=0}\Phi_{\boldsymbol{\theta}_0}(t\mathbf{v}) = \mathbf{v}$. Brown's theorem becomes the following.

PROPOSITION 2. *Suppose $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ is a sample from $f(\mathbf{x}; \boldsymbol{\theta}_0)$ and that $\hat{\boldsymbol{\theta}}$ minimizes $\sum_i \rho(\mathbf{X}_i, \boldsymbol{\theta})$. Let $\hat{\boldsymbol{\theta}} = \Phi_{\boldsymbol{\theta}_0}(\hat{\mathbf{h}})$ and $\hat{\mathbf{h}} = \sum_{i=1}^{i=r} \hat{\mathbf{h}}_i$ where $\hat{\mathbf{h}}_i \in \mathscr{V}_i$. Then the asymptotic distribution of $n^{1/2}\hat{\mathbf{h}}$ is multivariate normal with density proportional to*

$$
\exp\biggl(-\frac{n}{2}\sum_i \frac{d_i^2}{c_i}\langle \mathbf{h}_i, \mathbf{h}_i \rangle_0\biggr), \ \mathbf{h}_i \in \mathscr{V}_i, \ i = 1, \dots, r.
$$

In particular,

$$
n \sum_i \frac{d_i^2}{c_i} \langle \hat{\mathbf{h}}_i, \hat{\mathbf{h}}_i \rangle_0
$$

is asymptotically $\chi^2(\dim \boldsymbol{\Theta})$.

We note that the proposition is only marginally dependent upon the choice of $\Phi_{\boldsymbol{\theta}_0}$. If $\tilde{\Phi}_{\boldsymbol{\theta}_0}$ is a different choice and $\hat{\boldsymbol{\theta}} = \widetilde{\Phi}_{\boldsymbol{\theta}_0}(\tilde{\mathbf{h}})$, then $\tilde{\mathbf{h}} = \mathbf{h} + o(\|\mathbf{h}\|)$, so that asymptotically, there is no difference.

EXAMPLE. We continue with the example $\mathscr{X} = \boldsymbol{\Theta} = \Omega_p$ and $\mathscr{G} = SO(p)$. In what follows we consistently use $\boldsymbol{\theta}$ to denote the generic element of $\boldsymbol{\Theta}$, $\boldsymbol{\theta}_0$ to denote its "true" value and $\boldsymbol{\theta}_1$ to denote a computationally convenient value of $\boldsymbol{\theta}_0$.

Let $\boldsymbol{\theta}_1 = [1\ 0 \cdots 0]^T$. Then

$$\mathscr{H} = \mathscr{G}_{\boldsymbol{\theta}_1} = \{\mathbf{A} = \text{block diag } (1, \mathbf{A}_1) | \mathbf{A}_1 \in SO(p-1)\}.$$

$T_{\boldsymbol{\theta}_1}\boldsymbol{\Theta} = \{\mathbf{h} = [\mathbf{0}\ \mathbf{h}_1^T]^T | \mathbf{h}_1 \in R^{p-1}\}$ and the representation of $\mathscr{H}$ on $T_{\boldsymbol{\theta}_1}\boldsymbol{\Theta}$ is $\mathbf{A} \cdot \mathbf{h} = \mathbf{A}\mathbf{h}$.

To see this we note that $\mathscr{H}$ acts on $\boldsymbol{\Theta}$ and its representation on $T_{\boldsymbol{\theta}_1}\boldsymbol{\Theta}$ is defined using the derivative. Formally, it is calculated as follows: if $\gamma(t)$ is a curve in Ω_p with $\gamma(0) = \boldsymbol{\theta}_1$ then the action of $\mathbf{A}$ on $\gamma'(0)$ is defined to be

$$\mathbf{A} \cdot \gamma'(0) = \left.\frac{d}{dt}\right|_{t=0} \mathbf{A} \cdot \gamma(t) = \left.\frac{d}{dt}\right|_{t=0} \mathbf{A}\gamma(t) = \mathbf{A}\gamma'(0),$$

where the last equality is due to the linearity of matrix multiplication. We see that, at least when matrices and vectors are involved, the representation of $\mathscr{H}$ on $T_{\boldsymbol{\theta}_1}\boldsymbol{\Theta}$ is usually clear.

The $\mathscr{H}$ representation on $T_{\boldsymbol{\theta}_1}\boldsymbol{\Theta}$ is irreducible. Clearly also $\mathbf{A}\mathbf{h} = \mathbf{h}$ for all $\mathbf{A} \in \mathscr{H}$ can only occur for $\mathbf{h} = \mathbf{0}$. Hence, using Lemma 1, $\langle , \rangle_B$ is symmetric and there exist constants c_2 and d_2 (the addition of the subscript is to make the notation consistent with later sections) such that

$$\langle \mathbf{h}_1, \mathbf{h}_2 \rangle_A = c_2 \mathbf{h}_1^T \mathbf{h}_2,$$
$$\langle \mathbf{h}_1, \mathbf{h}_2 \rangle_B = d_2 \mathbf{h}_1^T \mathbf{h}_2.$$

Using transitivity and invariance, we can deduce the general $\boldsymbol{\theta}_0$ from the specific case $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1$. Pick $\mathbf{A} \in SO(p)$ so that $\boldsymbol{\theta}_1 = \mathbf{A}\boldsymbol{\theta}_0$. If $\mathbf{h} \in T_{\boldsymbol{\theta}_0}\Omega_p$, then $\mathbf{A} \cdot \mathbf{h} = \mathbf{A}\mathbf{h} \in T_{\boldsymbol{\theta}_1}\Omega_p$. By invariance of $\langle , \rangle_A$, for $\mathbf{h}_1, \mathbf{h}_2 \in T_{\boldsymbol{\theta}_0}\Omega_p$,

$$\langle \mathbf{h}_1, \mathbf{h}_2 \rangle_A = \langle \mathbf{A} \cdot \mathbf{h}_1, \mathbf{A} \cdot \mathbf{h}_2 \rangle_A = \langle \mathbf{A}\mathbf{h}_1, \mathbf{A}\mathbf{h}_2 \rangle_A = c_2(\mathbf{A}\mathbf{h}_1)^T \mathbf{A}\mathbf{h}_2 = c_2 \mathbf{h}_1^T \mathbf{h}_2.$$

A convenient $\Phi_{\boldsymbol{\theta}_0}$ to use is $\Phi_{\boldsymbol{\theta}_0} = (1 - \mathbf{h}^T\mathbf{h})^{1/2}\boldsymbol{\theta}_0 + \mathbf{h}$ (for $\mathbf{h} \in T_{\boldsymbol{\theta}_0}\Omega_p$, $\mathbf{h}^T\mathbf{h} < 1$). Proposition 2 then becomes: write $\hat{\boldsymbol{\theta}}$ as $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}^T\boldsymbol{\theta}_0)\boldsymbol{\theta}_0 + \hat{\mathbf{h}}$ where $\hat{\mathbf{h}}^T\boldsymbol{\theta}_0 = 0$. Then $\hat{\mathbf{h}}$ has multivariate normally distribution, of dimension $p-1$ and supported on $T_{\boldsymbol{\theta}_0}\Omega_p$, with a density proportional to $\exp(-\frac{n}{2}\frac{d_2^2}{c_2}\mathbf{h}^T\mathbf{h})$.

In essence then, what we need to do is to decompose the $\mathscr{H}$ representation on $T_{\boldsymbol{\theta}_0}\boldsymbol{\Theta}$ into its irreducible components and to find consistent estimators for the dispersion constants c_i and d_i. The latter is the only place that the specific form of f and ρ plays a role.

We also see the point of the coordinate free approach. A parameterization of $\boldsymbol{\Theta}$ will usually destroy the symmetries induced by the action of $\mathscr{G}$. For example, any map of $\Omega_3 \to R^2$ destroys spherical distance which is invariant under $SO(3)$. Thus calculations we do in a specific parameterization will usually contain correction terms whose sole purpose is to correct for the distortions caused by the parameterization.

3. *M*-estimation for regression parameters in statistical group models. We will say that a statistical group model is a *location* model if $\mathbf{\Theta} = \mathscr{X}$ (with the same $\mathscr{G}$ actions). Given a location model, Chang and Tsai defined an associated regression model,

$$\begin{aligned} &\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n \in \mathscr{X} \text{ are fixed,} \\ &\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_n \in \mathscr{X} \text{ are independent random,} \\ &\text{the distribution of } \mathbf{V}_i \text{ is of the form } f(\mathbf{V}_i, g_0 \cdot \mathbf{u}_i), \qquad g_0 \in \mathscr{G}, \end{aligned} \tag{17}$$

where f satisfies the invariance condition (13). Here g_0 is unknown and the parameter of interest. We will estimate g_0 by minimizing the objective function $\sum_i \rho(\mathbf{V}_i, g \cdot \mathbf{u}_i)$ where ρ satisfies the invariance condition (14).

The regression model (17) is a statistical group model with $\mathbf{\Theta} = \mathscr{G}$ and $\mathscr{G}$ acting on itself by left multiplication. In principle, the results of the previous section apply. However, for the regression model $\mathscr{H} = \mathscr{G}_{g_0} = \{g | g g_0 = g_0\} = \{1\}$, so that Propositions 1 and 2 yield no useful information about a regression group model.

Let $\langle , \rangle_A$ and $\langle , \rangle_B$ denote the Riemannian metrics on $\mathscr{X}$ derived from the location model. Let $\langle , \rangle_{AG}$ and $\langle , \rangle_{BG}$ denote the Riemannian metrics on $\mathscr{G}$ derived from the associated regression model *using the same f and ρ*. Lemma 3 of Chang and Tsai (1999) shows how $\langle , \rangle_{AG}$ and $\langle , \rangle_{BG}$ can be calculated from $\langle , \rangle_A$ and $\langle , \rangle_B$. Thus our approach for regression models is to use the preceding section on $\langle , \rangle_A$ and $\langle , \rangle_B$ and then to calculate $\langle , \rangle_{AG}$ and $\langle , \rangle_{BG}$ from $\langle , \rangle_A$ and $\langle , \rangle_B$.

Let $\mathbf{u} \in \mathscr{X}$ and define an associated map $r_{\mathbf{u}}: \mathscr{G} \to \mathscr{X}$ by $r_{\mathbf{u}}(g) = g \cdot \mathbf{u}$. If $\gamma(t)$ is a curve in $\mathscr{G}$ with $\gamma(0) = g_0$, then $\tilde{\gamma}(t) = \gamma(t) \cdot \mathbf{u}$ is a curve in $\mathscr{X}$ with basepoint $\tilde{\gamma}(0) = g_0 \cdot u$. Hence the derivative of $r_{\mathbf{u}}$ is a map $R_{\mathbf{u}}: T_{g_0}\mathscr{G} \to T_{g_0 \cdot \mathbf{u}}\mathscr{X}$ such that $\tilde{\gamma}'(0) = R_{\mathbf{u}}(\gamma'(0))$. It follows from (10) and (11) that

$$\langle \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\delta}} \rangle_{AG} = n^{-1} \sum_i \langle R_{\mathbf{u}_i}(\tilde{\boldsymbol{\delta}}), R_{\mathbf{u}_i}(\tilde{\boldsymbol{\delta}}) \rangle_A \tag{18}$$

for $\tilde{\boldsymbol{\delta}} \in T_{\mathbf{g}_0}\mathscr{G}$. A similar equation holds for $\langle , \rangle_{BG}$.

Notice that in the right-hand side of (18), $R_{\mathbf{u}_i}(\tilde{\boldsymbol{\delta}})$ is tangent to $\mathscr{X}$ at $g_0 \cdot \mathbf{u}_i$. Thus, in fact, different inner products in the family $\langle , \rangle_A$ of inner products are being applied for the summands of (18). Chang and Tsai (1999) use invariance and transitivity to eliminate this awkwardness by consolidating the basepoints (18) to $g_0 \in \mathscr{G}$ on the left-hand side and $\boldsymbol{\theta}_1 \in \mathscr{X}$ on the right-hand side.

In particular, for $g \in \mathscr{G}$, let $l_g: \mathscr{G} \to \mathscr{G}$ be defined by $l_g(h) = gh$. Define $ad_g: \mathscr{G} \to \mathscr{G}$ by $ad_g(h) = ghg^{-1}$. We will use L_g and Ad_g, respectively, to denote the derivatives of l_g and ad_g, respectively. Given $\boldsymbol{\theta}_1 \in \mathscr{X}$ and $g_0 \in \mathscr{G}$, choose $\tilde{g}_i \in \mathscr{G}$ so that $\boldsymbol{\theta}_1 = (\tilde{g}_i g_0) \cdot \mathbf{u}_i$. Given $\tilde{\boldsymbol{\delta}} \in T_{g_0}\mathscr{G}$, let $\boldsymbol{\delta}_i = R_{\boldsymbol{\theta}_1} Ad_{\tilde{g}_i g_0} L_{g_0^{-1}}(\tilde{\boldsymbol{\delta}})$, $i = 1, \ldots, n$. $\boldsymbol{\delta}_i \in T_{\boldsymbol{\theta}_1}\mathscr{X}$ and we have

$$\langle \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\delta}} \rangle_{AG} = n^{-1} \sum_i \langle \boldsymbol{\delta}_i, \boldsymbol{\delta}_i \rangle_A \tag{19}$$

with a similar result for $\langle , \rangle_{BG}$.

Recall that for $\mathbf{H}$ a skew-symmetric $r \times r$ matrix (usually $r = m$ or $r = p$), $\exp(\mathbf{H})$ is defined to be the $r \times r$ matrix $\sum_i \mathbf{H}^i / i!$. $\exp(\mathbf{H}) \in SO(r)$, $\exp(\mathbf{0}) = \mathbf{I}_r$ and $\frac{d}{dt}|_{t=0} \exp(t\mathbf{H}) = \mathbf{H}$.

EXAMPLE. We continue with the example $\mathscr{X} = \Omega_p$ and $\mathscr{G} = SO(p)$ and derive the asymptotic distribution of M-estimators for spherical regression. We use at $\mathbf{A}_0 \in SO(p)$ the local coordinate chart $\Phi_{\mathbf{A}_0}$: $T_{\mathbf{A}_0} SO(p) \to SO(p)$ defined by

$$\Phi_{\mathbf{A}_0}(\tilde{\boldsymbol{\delta}}) = \mathbf{A}_0 \exp(L_{\mathbf{A}_0}^{-1}(\tilde{\boldsymbol{\delta}})), \qquad \tilde{\boldsymbol{\delta}} \in T_{\mathbf{A}_0} SO(p).$$

Notice that if $\widehat{\mathbf{A}} = \Phi_{\mathbf{A}_0}(\tilde{\boldsymbol{\delta}})$ then $\widehat{\mathbf{A}} = \mathbf{A}_0 \exp(\mathbf{H})$ where $\mathbf{H} = L_{\mathbf{A}_0}^{-1}(\tilde{\boldsymbol{\delta}}) \in T_{\mathbf{I}} SO(p)$ is skew symmetric $p \times p$.

Following (19), let $\tilde{g}_i \in SO(p)$ be such that $\boldsymbol{\theta}_1 = (\tilde{g}_i \mathbf{A}_0)\mathbf{u}_i$, where $\boldsymbol{\theta}_1 = [1\ 0 \cdots 0]^T$. Now we define

$$\begin{aligned}
\boldsymbol{\delta}_i &= R_{\boldsymbol{\theta}_1} Ad_{\tilde{g}_i \mathbf{A}_0} L_{\mathbf{A}_0}^{-1}(\tilde{\boldsymbol{\delta}}) \\
&= R_{\boldsymbol{\theta}_1} Ad_{\tilde{g}_i \mathbf{A}_0} \mathbf{H} = R_{\boldsymbol{\theta}_1} \frac{d}{dt}\bigg|_{t=0} ad_{\tilde{g}_i \mathbf{A}_0}(\exp(t\mathbf{H})) \\
&= \frac{d}{dt}\bigg|_{t=0} r_{\boldsymbol{\theta}_1}(\tilde{g}_i \mathbf{A}_0 \exp(t\mathbf{H}) \mathbf{A}_0^T \tilde{g}_i^T) \\
&= \frac{d}{dt}\bigg|_{t=0} \tilde{g}_i \mathbf{A}_0 \exp(t\mathbf{H}) \mathbf{A}_0^T \tilde{g}_i^T \boldsymbol{\theta}_1 = \tilde{g}_i \mathbf{A}_0 \mathbf{H} \mathbf{A}_0^T \tilde{g}_i^T \boldsymbol{\theta}_1 = \tilde{g}_i \mathbf{A}_0 \mathbf{H} \mathbf{u}_i.
\end{aligned}$$

Thus

$$\boldsymbol{\delta}_i^T \boldsymbol{\delta}_i = \mathbf{u}_i^T \mathbf{H}^T \mathbf{H} \mathbf{u}_i = -\mathrm{Tr}(\mathbf{H}^2 \mathbf{u}_i \mathbf{u}_i^T)$$

and using (19),

$$\begin{aligned}
\langle \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\delta}} \rangle_{AG} &= n^{-1} \sum_i \langle \boldsymbol{\delta}_i, \boldsymbol{\delta}_i \rangle_A = -c_2 n^{-1} \sum_i \boldsymbol{\delta}_i^T \boldsymbol{\delta}_i \\
&= -c_2 n^{-1} \sum_i \mathrm{Tr}(\mathbf{H}^2 \mathbf{u}_i \mathbf{u}_i^T), \\
\langle \tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\delta}} \rangle_{BG} &= -d_2 n^{-1} \sum_i \mathrm{Tr}(\mathbf{H}^2 \mathbf{u}_i \mathbf{u}_i^T).
\end{aligned}$$

Taking the asymptotic limit we get the following. Let $\Sigma = \lim_{n\to\infty} \frac{1}{n} \sum_i \mathbf{u}_i \mathbf{u}_i^T$ and write the M-estimate $\widehat{A}$ in the form $\widehat{A} = A_0 \exp(\widehat{\mathbf{H}})$. Then asymptotically $\widehat{\mathbf{H}}$ is multivariate normal with density proportional to $\exp(\frac{n}{2} \frac{d_2^2}{c_2} \mathrm{Tr}(\mathbf{H}^2 \Sigma))$.

Notice that c_2 and d_2 are the same for a spherical location model and its corresponding spherical regression model. Sample estimates of these constants are given in Section 7 [see equations (39)].

4. Concentrated error asymptotic approximations. The Fisher–von Mises Langevin distribution $F(\boldsymbol{\theta}_0, \kappa)$ on Ω_p has density proportional to exp $(\kappa \mathbf{x}^T \boldsymbol{\theta}_0)$ where $\kappa > 0$ is a *concentration* parameter and $\boldsymbol{\theta}_0 \in \Omega_p$. If $\mathbf{X}$ is distributed $F(\boldsymbol{\theta}_0, \kappa)$, write $\mathbf{X} = \mathbf{X}^T \boldsymbol{\theta}_0 \boldsymbol{\theta}_0 + \mathbf{Y}$. $\mathbf{Y} \in T_{\boldsymbol{\theta}_0}\Omega_p$ has the property that $\kappa^{1/2}\mathbf{Y}$ approaches a multivariate normal distribution $N_{p-1}(\mathbf{0}, \mathbf{I})$ supported on the $p-1$-dimensional subspace $T_{\boldsymbol{\theta}_0}\Omega_p$ of R^p as $\kappa \to \infty$. Rivest (1989) analyzed spherical regressions using an asymptotic $\kappa \to \infty$ approximation (with a fixed sample size).

Thus for large κ asymptotics in location families, we hypothesize a family of densities $f_\kappa(\mathbf{x}, \boldsymbol{\theta}_0)$ such that if $\mathbf{X}$ is distributed $f_\kappa(\mathbf{x}, \boldsymbol{\theta}_0)$, and we write $\mathbf{X} = \Phi_{\boldsymbol{\theta}_0}(\mathbf{Y})$, then $\kappa^{1/2}\mathbf{Y}$ has a limiting normal distribution as $\kappa \to \infty$. Here, as before, $\Phi_{\boldsymbol{\theta}_0}\colon T_{\boldsymbol{\theta}_0}\boldsymbol{\Theta} \to \boldsymbol{\Theta}$ is any map such that $\Phi_{\boldsymbol{\theta}_0}(\mathbf{0}) = \boldsymbol{\theta}_0$ and such that the derivative of $\Phi_{\boldsymbol{\theta}_0}$ at $\mathbf{0}$ is the identitity map. This approach was used by Chang (1988), Rancourt, Rivest and Asselin (2000) and Rivest and Chang (2000).

Assuming, temporarily, that $\boldsymbol{\Theta} = R^q$ and examining the usual Taylor series,

$$\begin{aligned} 0 &= (\kappa/n)^{1/2}\mathbf{S}(\mathbf{X}, \hat{\boldsymbol{\theta}}) \\ &= (\kappa/n)^{1/2}\mathbf{S}(\mathbf{X}, \boldsymbol{\theta}_0) + \left[\frac{1}{n}\sum_i \frac{\partial^2}{\partial \boldsymbol{\theta}^2}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \rho(\mathbf{X}_i, \boldsymbol{\theta})\right](n\kappa)^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o(1). \end{aligned}$$

As $\kappa \to \infty$,

$$\frac{1}{n}\sum_i \frac{\partial^2}{\partial \boldsymbol{\theta}^2}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \rho(\mathbf{X}_i, \boldsymbol{\theta}) \to \frac{\partial^2}{\partial \boldsymbol{\theta}^2}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \rho(\boldsymbol{\theta}_0, \boldsymbol{\theta}).$$

In general, second derivatives do not have coordinate free descriptions. They do, however, at critical points. Thus, if $\tilde{\tau}_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}) = \rho(\boldsymbol{\theta}_0, \boldsymbol{\theta})$ has a critical point at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, then $\langle , \rangle_B$ can be defined as

$$\langle \boldsymbol{\delta}_1, \boldsymbol{\delta}_2 \rangle_B = -\frac{\partial^2}{\partial s \partial t}\bigg|_{(s,t)=(0,0)} \rho(\boldsymbol{\theta}_0, \gamma(s,t)), \tag{20}$$

where $\gamma\colon R^2 \to \boldsymbol{\Theta}$ satisfies $\gamma(0,0) = \boldsymbol{\theta}_0$, $\frac{\partial}{\partial s}|_{(s,t)=(0,0)}\gamma(s,t) = \boldsymbol{\delta}_1$, and $\frac{\partial}{\partial t}|_{(s,t)=(0,0)}$ $\gamma(s,t) = \boldsymbol{\delta}_2$. The condition that $\tilde{\tau}_{\boldsymbol{\theta}_0}(\boldsymbol{\theta})$ have a critical point at $\boldsymbol{\theta}_0$ ensures that the right-hand side of (20) depends only upon $(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$ and not upon the particular choice of γ.

With this reinterpretation of $\langle , \rangle_B$, the remainder of Sections 2 and 3 are valid without change.

5. *M*-estimation for location parameters in Stiefel manifolds. Let $\boldsymbol{\theta}_1 = [\mathbf{I}_m\ \mathbf{0}]^T$. Rather than apply Proposition 2 to the general $\boldsymbol{\theta}_0$, we will first perform calculations as if $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1$, and then use invariance (see Proposition 3 of Section 9) for the general $\boldsymbol{\theta}_0$.

Let $\gamma(t)$ be a curve in $\mathscr{V}_{p,m}$ with $\gamma(0) = \boldsymbol{\theta}_1$. Write $\gamma(t) = [\gamma_1^T(t)\ \gamma_2^T(t)]^T$. Then $\mathbf{I}_m = \gamma_1(t)^T\gamma_1(t) + \gamma_2(t)^T\gamma_2(t)$, so that $\mathbf{0} = \gamma_1'(0)^T\gamma_1(0) + \gamma_1(0)^T\gamma_1'(0) + \gamma_2'(0)^T\gamma_2(0) + \gamma_2(0)^T\gamma_2'(0) = \gamma_1'(0) + \gamma_1'(0)^T$. Thus every vector in $T_{\boldsymbol{\theta}_1}\mathscr{V}_{p,m}$ has the form $[\mathbf{K}_1^T\ \mathbf{K}_2^T]^T$ where $\mathbf{K}_1$ is $m \times m$ skew symmetric and $\mathbf{K}_2$ is $m \times p - m$ without restriction.

We next show that any vector of the form $[\mathbf{K}_1^T\ \mathbf{K}_2^T]^T$ is in $T_{\boldsymbol{\theta}_1}\mathscr{V}_{p,m}$. Define the curve $g(t) \in SO(m) \times SO(p)$ by

$$g(t) = \left(\exp(-t\mathbf{K}_1), \exp\left(t\begin{bmatrix} \mathbf{0} & -\mathbf{K}_2^T \\ \mathbf{K}_2 & \mathbf{0} \end{bmatrix}\right)\right)$$

and hence

$$\gamma(t) = g(t) \cdot \boldsymbol{\theta}_1 = \exp\left(t\begin{bmatrix} \mathbf{0} & -\mathbf{K}_2^T \\ \mathbf{K}_2 & \mathbf{0} \end{bmatrix}\right)\boldsymbol{\theta}_1 \exp(t\mathbf{K}_1) \tag{21}$$

is a curve in $\mathscr{V}_{p,m}$. Now

$$\gamma'(0) = \boldsymbol{\theta}_1\mathbf{K}_1 + \begin{bmatrix} \mathbf{0} & -\mathbf{K}_2^T \\ \mathbf{K}_2 & \mathbf{0} \end{bmatrix}\boldsymbol{\theta}_1 = \begin{bmatrix} \mathbf{K}_1 \\ \mathbf{K}_2 \end{bmatrix} \tag{22}$$

and this establishes that $[\mathbf{K}_1^T\ \mathbf{K}_2^T]^T \in T_{\boldsymbol{\theta}_1}\mathscr{V}_{p,m}$. In other words,

$$T_{\boldsymbol{\theta}_1}\mathscr{V}_{p,m} = \Big\{[\mathbf{K}_1^T\ \mathbf{K}_2^T]^T \mid \mathbf{K}_1 \text{ is } m \times m \text{ skew symmetric, } \mathbf{K}_2 \text{ is } m \times (p-m) \text{arbitrary}\Big\}.$$

We define a local coordinate chart $\Phi_{\boldsymbol{\theta}_1}\colon T_{\boldsymbol{\theta}_1}\mathscr{V}_{p,m} \to \mathscr{V}_{p,m}$ of $\mathscr{V}_{p,m}$ by

$$\Phi_{\boldsymbol{\theta}_1}\left(\begin{bmatrix} \mathbf{K}_1 \\ \mathbf{K}_2 \end{bmatrix}\right) = \exp\left(\begin{bmatrix} \mathbf{0} & -\mathbf{K}_2^T \\ \mathbf{K}_2 & \mathbf{0} \end{bmatrix}\right)\boldsymbol{\theta}_1 \exp(\mathbf{K}_1). \tag{23}$$

Then, using (22),

$$\left.\frac{d}{dt}\right|_{t=0}\Phi_{\boldsymbol{\theta}_1}\left(t\begin{bmatrix} \mathbf{K}_1 \\ \mathbf{K}_2 \end{bmatrix}\right) = \begin{bmatrix} \mathbf{K}_1 \\ \mathbf{K}_2 \end{bmatrix},$$

which is another way of saying that the Jacobian of $\Phi_{\boldsymbol{\theta}_1}$ at $\mathbf{0}$ is the identity map.

It is easily checked that $\boldsymbol{\theta}_1 = \mathbf{A}_2\boldsymbol{\theta}_1\mathbf{A}_1^T$ if and only if

$$\mathbf{A}_2 = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix},$$

where $\mathbf{A}_{22} \in SO(p-m)$. Thus $\mathscr{H} = SO(m) \times SO(p-m)$ [$\mathscr{H} = \mathscr{O}(2) \times \mathscr{O}(2)$ when $p = 4, m = 2$].

To calculate the $\mathscr{H}$ action on $T_{\boldsymbol{\theta}_1}\mathscr{V}_{p,m}$, let $\gamma(t)$ be as in (21). Then using (22),

$$\begin{aligned}(\mathbf{A}_1, \mathbf{A}_{22}) \cdot \begin{bmatrix} \mathbf{K}_1 \\ \mathbf{K}_2 \end{bmatrix} &= \left.\frac{d}{dt}\right|_{t=0}(\mathbf{A}_1, \mathbf{A}_{22}) \cdot \gamma(t) = \left.\frac{d}{dt}\right|_{t=0}\begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix}\gamma(t)\mathbf{A}_1^T \\ &= \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix}\gamma'(0)\mathbf{A}_1^T = \begin{bmatrix} \mathbf{A}_1\mathbf{K}_1\mathbf{A}_1^T \\ \mathbf{A}_{22}\mathbf{K}_2\mathbf{A}_1^T \end{bmatrix}.\end{aligned}$$

It is easily checked that if $\mathbf{A}_1\mathbf{K}_1\mathbf{A}_1^T = \mathbf{K}_1$ for all $\mathbf{A}_1 \in SO(m)$, then $\mathbf{K}_1 = \mathbf{0}$. Similarly, if $\mathbf{A}_{22}\mathbf{K}_2\mathbf{A}_1^T = \mathbf{K}_2$ for all $(\mathbf{A}_1, \mathbf{A}_{22}) \in SO(m) \times SO(p-m)$, then $\mathbf{K}_2 = \mathbf{0}$. Thus Lemma 1 applies and we conclude that $\langle , \rangle_B$ has the coordinate free definition (11).

Let $\mathscr{V}_1 = \{[\mathbf{K}_1^T\ \mathbf{0}]^T\}$ and $\mathscr{V}_2 = \{[\mathbf{0}\ \mathbf{K}_2^T]^T\}$. $\mathscr{V}_1$ and $\mathscr{V}_2$ are invariant under $\mathscr{H}$. If $m = 1$, then $\mathscr{V}_1 = \{\mathbf{0}\}$, so we only have $\mathscr{V}_2$ and it is irreducible. Similarly, if $m = p$, $\mathscr{V}_2 = \{\mathbf{0}\}$, so we only have $\mathscr{V}_1$ and it is also irreducible. If $2 \le m \le p-1$, $T_{\boldsymbol{\theta}_1}\mathscr{V}_{p,m} = \mathscr{V}_1 \oplus \mathscr{V}_2$ is a decomposition of $T_{\boldsymbol{\theta}_1}\mathscr{V}_{p,m}$ into minimal invariant subspaces.

Note that when $p = 4, m = 2$, $\mathscr{V}_2$ is irreducible under the action of $\mathscr{H} = \mathscr{O}(2) \times \mathscr{O}(2)$ but breaks up into two irreducible components under the action of $SO(2) \times SO(2)$. For this reason we have required invariance under $\mathscr{G} = \mathscr{O}(2) \times \mathscr{O}(4)$ for $p = 4$, $m = 2$.

There is an obvious $\mathscr{H}$ invariant inner product,

$$\left\langle \begin{bmatrix} \mathbf{K}_1 \\ \mathbf{K}_2 \end{bmatrix}, \begin{bmatrix} \tilde{\mathbf{K}}_1 \\ \tilde{\mathbf{K}}_2 \end{bmatrix} \right\rangle_0 = \operatorname{Tr}(\mathbf{K}_1^T\tilde{\mathbf{K}}_1) + \operatorname{Tr}(\mathbf{K}_2^T\tilde{\mathbf{K}}_2). \tag{24}$$

Recall that for $\boldsymbol{\theta}_0 \in \mathscr{V}_{p,m}$, $\boldsymbol{\theta}_{0\perp}$ is any $p\times(p-m)$ matrix so that $[\boldsymbol{\theta}_0\ \boldsymbol{\theta}_{0\perp}] \in SO(p)$.

PROPOSITION 4. *Suppose* $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ *is a sample from a density* $f(\mathbf{x}, \boldsymbol{\theta}_0)$ *and that* $\hat{\boldsymbol{\theta}}$ *minimizes an objective function* $\sum_i \rho(\mathbf{X}_i, \boldsymbol{\theta})$ *where* f *and* ρ *satisfy* (13) *and* (14).

If $1 < m < p$, *write*

$$[\boldsymbol{\theta}_0\ \boldsymbol{\theta}_{0\perp}]^T\hat{\boldsymbol{\theta}} = \exp\left(\begin{bmatrix} 0 & -\hat{\mathbf{h}}_2^T \\ \hat{\mathbf{h}}_2 & 0 \end{bmatrix}\right)\begin{bmatrix} \mathbf{I}_m \\ \mathbf{0} \end{bmatrix}\exp(\hat{\mathbf{h}}_1). \tag{25}$$

Then there are constants c_1, c_2, d_1, d_2 *such that* $n^{1/2}(\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2)$ *is asymptotically multivariate normal with a density proportional to*

$$\exp\left(-\frac{n}{2}\left(\frac{d_1^2}{c_1}\operatorname{Tr}(\mathbf{h}_1^T\mathbf{h}_1) + \frac{d_2^2}{c_2}\operatorname{Tr}(\mathbf{h}_2^T\mathbf{h}_2)\right)\right).$$

If $m = p$, *write*

$$\boldsymbol{\theta}_0^T\hat{\boldsymbol{\theta}} = \exp(\hat{\mathbf{h}}_1).$$

Then $n^{1/2}\hat{\mathbf{h}}_1$ *is asymptotically normal with a density proportional to* $\exp\left(-\frac{n}{2}\frac{d_1^2}{c_1}\operatorname{Tr}(\mathbf{h}_1^T\mathbf{h}_1)\right)$.

PROOF. If $\boldsymbol{\theta}_0 = \boldsymbol{\theta}_1$ this is simply Proposition 2 using the local coordinate chart (23). For general $\boldsymbol{\theta}_0$, let $g = (\mathbf{I}_m, [\boldsymbol{\theta}_0\ \boldsymbol{\theta}_{0\perp}])$. Then $g \cdot \boldsymbol{\theta}_1 = \boldsymbol{\theta}_0$ and Proposition 4 is a result of Proposition 3 (see Section 9). □

In Section 7, we will characterize the constants c_1, c_2, d_1 and d_2 and provide consistent estimators for them. In addition, we show there that if f and ρ have the form (2) and (5), $c_1 = c_2$ and $d_1 = d_2$.

Proposition 4 can easily be used to construct tests of a null hypothesis of the form $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Using part (b) of Proposition 3 of Section 9, the resulting test procedure does not depend upon the nonunique choice of $\boldsymbol{\theta}_{0\perp}$.

Proposition 4 does not readily lead to confidence regions. However, we have the following asymptotic inversion.

COROLLARY 1. *Let χ^2 be the upper αth quantile of a χ^2 distribution with $m(2p-m-1)/2$ degrees of freedom. An asymptotic size $1-\alpha$ confidence region $\mathscr{C}$ for $\boldsymbol{\theta}_0$ can be calculated as follows*:

If $1 < m < p$, let

$$\mathscr{C} = \left\{\boldsymbol{\theta} \mid \boldsymbol{\theta} = [\hat{\boldsymbol{\theta}}\ \hat{\boldsymbol{\theta}}_\perp] \exp\left(\begin{bmatrix} \mathbf{0} & -\mathbf{h}_2^T \\ \mathbf{h}_2 & \mathbf{0} \end{bmatrix}\right)\begin{bmatrix} \mathbf{I}_m \\ \mathbf{0} \end{bmatrix} \exp(\mathbf{h}_1) \right.$$
$$\left. \textit{such that } n\left(\frac{d_1^2}{c_1}\mathrm{Tr}(\mathbf{h}_1^T\mathbf{h}_1) + \frac{d_2^2}{c_2}\mathrm{Tr}(\mathbf{h}_2^T\mathbf{h}_2)\right) < \chi^2\right\}.$$

If $m = p$, let

$$\mathscr{C} = \left\{\boldsymbol{\theta} | \boldsymbol{\theta} = \hat{\boldsymbol{\theta}} \exp(\mathbf{h}_1) \textit{ such that } n\frac{d_1^2}{c_1}\mathrm{Tr}(\mathbf{h}_1^T\mathbf{h}_1) < \chi^2\right\}.$$

PROOF. We prove the case $1 < m < p$. If (25) is true, then

$$[\boldsymbol{\theta}_0\ \boldsymbol{\theta}_{0\perp}]^T\hat{\boldsymbol{\theta}} = \left(\mathbf{I}_p + \begin{bmatrix} \mathbf{0} & -\hat{\mathbf{h}}_2^T \\ \hat{\mathbf{h}}_2 & \mathbf{0} \end{bmatrix} + O(|\hat{\mathbf{h}}_2|^2)\right)\begin{bmatrix} \mathbf{I}_m \\ \mathbf{0} \end{bmatrix}(\mathbf{I}_m + \hat{\mathbf{h}}_1 + O(|\hat{\mathbf{h}}_1|^2)).$$

Therefore up to terms $O_p(n^{-1})$,

$$\begin{aligned} \boldsymbol{\theta}_0^T\hat{\boldsymbol{\theta}} &= \mathbf{I}_m + \hat{\mathbf{h}}_1, \\ \boldsymbol{\theta}_{0\perp}^T\hat{\boldsymbol{\theta}} &= \hat{\mathbf{h}}_2. \end{aligned} \tag{26}$$

Note that

$$\mathbf{I}_p = [\boldsymbol{\theta}_0\boldsymbol{\theta}_{0\perp}]\begin{bmatrix} \boldsymbol{\theta}_0^T \\ \boldsymbol{\theta}_{0\perp}^T \end{bmatrix},$$
$$\mathbf{I}_p = \boldsymbol{\theta}_0\boldsymbol{\theta}_0^T + \boldsymbol{\theta}_{0\perp}\boldsymbol{\theta}_{0\perp}^T.$$

Now write

$$[\hat{\boldsymbol{\theta}}\ \hat{\boldsymbol{\theta}}_\perp]^T\boldsymbol{\theta}_0 = \exp\left(\begin{bmatrix} \mathbf{0} & -\mathbf{h}_2^T \\ \mathbf{h}_2 & \mathbf{0} \end{bmatrix}\right)\begin{bmatrix} \mathbf{I}_m \\ \mathbf{0} \end{bmatrix}\exp(\mathbf{h}_1).$$

Then (26) implies $\mathbf{h}_1 = -\hat{\mathbf{h}}_1 + O_p(n^{-1})$ and

$$\begin{aligned} \mathrm{Tr}(\hat{\mathbf{h}}_2\hat{\mathbf{h}}_2^T) + O_p(n^{-3/2}) &= \mathrm{Tr}(\boldsymbol{\theta}_{0\perp}^T\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^T\boldsymbol{\theta}_{0\perp}) = \mathrm{Tr}(\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^T(\mathbf{I}_p - \boldsymbol{\theta}_0\boldsymbol{\theta}_0^T)) \\ &= \mathrm{Tr}(\hat{\boldsymbol{\theta}}^T\hat{\boldsymbol{\theta}}) - \mathrm{Tr}(\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^T\boldsymbol{\theta}_0\boldsymbol{\theta}_0^T) = m - \mathrm{Tr}(\boldsymbol{\theta}_0\boldsymbol{\theta}_0^T\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^T) \\ &= \mathrm{Tr}(\mathbf{h}_2\mathbf{h}_2^T) + O_p(n^{-3/2}). \end{aligned}$$

Hence,

$$n\left(\frac{d_1^2}{c_1}\mathrm{Tr}(\hat{\mathbf{h}}_1^T\hat{\mathbf{h}}_1) + \frac{d_2^2}{c_2}\mathrm{Tr}(\hat{\mathbf{h}}_2^T\hat{\mathbf{h}}_2)\right) = n\left(\frac{d_1^2}{c_1}\mathrm{Tr}(\mathbf{h}_1^T\mathbf{h}_1) + \frac{d_2^2}{c_2}\mathrm{Tr}(\mathbf{h}_2^T\mathbf{h}_2)\right) + O_p(n^{-1/2})$$

and the corollary follows. □

6. Regression on Stiefel manifolds. We now turn to M-estimation of $\mathbf{A}_1$ and $\mathbf{A}_2$ where the data satisfy the regression model (8) and the estimates $\widehat{\mathbf{A}}_1$ and $\widehat{\mathbf{A}}_2$ minimize (9) where f and ρ satisfy the invariance condition (1) and (4). Major simplifications occur when f and ρ satisfy (2) and (5), respectively.

Let $g_0 = (\mathbf{A}_1, \mathbf{A}_2)$ and $\tilde{g}_i = (\mathbf{I}_m, [\mathbf{A}_2\mathbf{u}_i\mathbf{A}_1^T \quad \mathbf{A}_2\mathbf{u}_{i\perp}]^T)$. Then $\tilde{g}_i g_0 = (\mathbf{A}_1, [\mathbf{u}_i\mathbf{A}_1^T \quad \mathbf{u}_{i\perp}]^T)$ and $(\tilde{g}_i g_0)\cdot \mathbf{u}_i = \boldsymbol{\theta}_1$.

The general element $\tilde{\boldsymbol{\delta}} \in T_{g_0}(SO(m)\times SO(p))$ has the form $\tilde{\boldsymbol{\delta}} = L_{g_0}(\mathbf{H},\mathbf{K})$ where $\mathbf{H}$ and $\mathbf{K}$ are skew symmetric matrices of size $m\times m$ and $p\times p$, respectively.

We have, by the usual trick of starting with the curve $[\exp(t\mathbf{H}), \exp(t\mathbf{K})]$,

$$\begin{aligned} Ad_{\tilde{g}_i g_0}(\mathbf{H},\mathbf{K}) &= \left(\mathbf{A}_1\mathbf{H}\mathbf{A}_1^T, \begin{bmatrix}\mathbf{A}_1\mathbf{u}_i^T\\ \mathbf{u}_{i\perp}^T\end{bmatrix}\mathbf{K}\begin{bmatrix}\mathbf{A}_1\mathbf{u}_i^T\\ \mathbf{u}_{i\perp}^T\end{bmatrix}^T\right)\\ &= \left(\mathbf{A}_1\mathbf{H}\mathbf{A}_1^T, \begin{bmatrix}\mathbf{A}_1\mathbf{u}_i^T\mathbf{K}\mathbf{u}_i\mathbf{A}_1^T & \mathbf{A}_1\mathbf{u}_i^T\mathbf{K}\mathbf{u}_{i\perp}\\ \mathbf{u}_{i\perp}^T\mathbf{K}\mathbf{u}_i\mathbf{A}_1^T & \mathbf{u}_{i\perp}^T\mathbf{K}\mathbf{u}_{i\perp}\end{bmatrix}\right)\end{aligned}$$

and for arbitrary skew symmetric matrices $\widetilde{\mathbf{H}}$ and $\widetilde{\mathbf{K}}$ of sizes $m\times m$ and $p\times p$, respectively,

$$\begin{aligned} R_{\boldsymbol{\theta}_1}(\widetilde{\mathbf{H}},\widetilde{\mathbf{K}}) &= \frac{d}{dt}\bigg|_{t=0}(\exp(t\widetilde{\mathbf{H}}),\exp(t\widetilde{\mathbf{K}}))\cdot\boldsymbol{\theta}_1 = \frac{d}{dt}\bigg|_{t=0}\exp(t\widetilde{\mathbf{K}})\boldsymbol{\theta}_1\exp(-t\widetilde{\mathbf{H}})\\ &= \widetilde{\mathbf{K}}\boldsymbol{\theta}_1 - \boldsymbol{\theta}_1\widetilde{\mathbf{H}} = \begin{bmatrix}\widetilde{\mathbf{K}}_{11}-\widetilde{\mathbf{H}}\\ \widetilde{\mathbf{K}}_{21}\end{bmatrix}.\end{aligned}$$

Here $\widetilde{\mathbf{K}}_{11}$ and $\widetilde{\mathbf{K}}_{21}$ represent the upper left $m\times m$ and lower left $p-m\times m$ submatrices of $\widetilde{\mathbf{K}}$, respectively. Therefore,

$$\boldsymbol{\delta}_i = R_{\boldsymbol{\theta}_1}Ad_{\tilde{g}_i g_0}L_{g_0^{-1}}(\tilde{\boldsymbol{\delta}}) = \begin{bmatrix}\mathbf{A}_1(\mathbf{u}_i^T\mathbf{K}\mathbf{u}_i - \mathbf{H})\mathbf{A}_1^T\\ \mathbf{u}_{i\perp}^T\mathbf{K}\mathbf{u}_i\mathbf{A}_1^T\end{bmatrix}.$$

Using (19),

$$\begin{aligned} n\langle\tilde{\boldsymbol{\delta}},\tilde{\boldsymbol{\delta}}\rangle_{AG} &= \sum_i\langle\boldsymbol{\delta}_i,\boldsymbol{\delta}_i\rangle_A = c_1\sum_i\mathrm{Tr}[(\mathbf{H}-\mathbf{u}_i^T\mathbf{K}\mathbf{u}_i)(\mathbf{H}-\mathbf{u}_i^T\mathbf{K}\mathbf{u}_i)^T]\\ &\quad + c_2\sum_i\mathrm{Tr}[(\mathbf{u}_{i\perp}^T\mathbf{K}\mathbf{u}_i)(\mathbf{u}_{i\perp}^T\mathbf{K}\mathbf{u}_i)^T],\\ n\langle\tilde{\boldsymbol{\delta}},\tilde{\boldsymbol{\delta}}\rangle_{BG} &= \sum_i\langle\boldsymbol{\delta}_i,\boldsymbol{\delta}_i\rangle_B = d_1\sum_i\mathrm{Tr}[(\mathbf{H}-\mathbf{u}_i^T\mathbf{K}\mathbf{u}_i)(\mathbf{H}-\mathbf{u}_i^T\mathbf{K}\mathbf{u}_i)^T]\\ &\quad + d_2\sum_i\mathrm{Tr}[(\mathbf{u}_{i\perp}^T\mathbf{K}\mathbf{u}_i)(\mathbf{u}_{i\perp}^T\mathbf{K}\mathbf{u}_i)^T].\end{aligned} \tag{27}$$

The apparent dependence of (27) on the somewhat arbitrary choices of the $\mathbf{u}_{i\perp}$ can be removed as follows. Recall that $\mathbf{u}_{i\perp}\mathbf{u}_{i\perp}^T = \mathbf{I}_p - \mathbf{u}_i\mathbf{u}_i^T$. Hence

$$\begin{aligned} n\langle\tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\delta}}\rangle_{AG} = {} & -c_1\Big[n\mathrm{Tr}(\mathbf{H}^2) - 2\mathrm{Tr}\,\Big(\sum_i \mathbf{H}\mathbf{u}_i^T\mathbf{K}\mathbf{u}_i\Big)\Big] \\ & - c_2\Big[\mathrm{Tr}\Big(\sum_i \mathbf{K}^2\mathbf{u}_i\mathbf{u}_i^T\Big)\Big] - (c_1 - c_2)\Big[\mathrm{Tr}\Big(\sum_i \mathbf{K}\mathbf{u}_i\mathbf{u}_i^T\mathbf{K}\mathbf{u}_i\mathbf{u}_i^T\Big)\Big] \end{aligned} \tag{28}$$

with a similar expression for $\langle\tilde{\boldsymbol{\delta}}, \tilde{\boldsymbol{\delta}}\rangle_{BG}$.

It is important to note that the constants c_1, c_2, d_1, d_2 are the same for location and regression models as long as the same f and ρ apply.

We note that the simplicity of the spherical regression case is due to the lack of terms involving c_1 and d_1 and to the fact that $\mathbf{u}_i^T\mathbf{K}\mathbf{u}_i = 0$ when $m = 1$.

The cases $1 < m < p$ are quite messy except for the important case in which $c_1 = c_2$ and $d_1 = d_2$. This case occurs, for example, when f and ρ are as in (2) and (5) (see the Corollary to Proposition 7 of the next section). In this case and when $m = p$, we have the following proposition (we have used in the proof that if $m = p$, $\mathbf{u}_i\mathbf{u}_i^T = \mathbf{I}_p$).

PROPOSITION 5. *Suppose $1 < m < p$ and that $c_1 = c_2 = c$ and $d_1 = d_2 = d$ or suppose that $m = p$ and let $c = c_1$ and $d = d_1$. Write $(\widehat{\mathbf{A}}_1, \widehat{\mathbf{A}}_2) = (\mathbf{A}_1\exp(\widehat{\mathbf{H}}), \mathbf{A}_2\exp(\widehat{\mathbf{K}}))$ where $\widehat{\mathbf{H}}$ and $\widehat{\mathbf{K}}$ are skew symmetric of sizes $m\times m$ and $p \times p$, respectively. Then $(\widehat{\mathbf{H}}, \widehat{\mathbf{K}})$ is asymptotically multivariate normal with a density proportional to*

$$\exp\Big[\frac{nd^2}{2c}\Big(\mathrm{Tr}(\mathbf{H}^2) + \mathrm{Tr}(\mathbf{K}^2\Sigma) - \frac{2}{n}\mathrm{Tr}\sum_i(\mathbf{H}\mathbf{u}_i^T\mathbf{K}\mathbf{u}_i)\Big)\Big]$$

where $\Sigma = \lim_{n\to\infty} n^{-1}\sum_i u_iu_i^T$.

REMARK. The main theorem of Prentice (1989) does not contain the cross-product term $-\frac{2}{n}\mathrm{Tr}\sum_i(\mathbf{H}\mathbf{u}_i^T\mathbf{K}\mathbf{u}_i)$. The authors believe that the discrepancy is due to an error in Prentice's proofs. In that paper, to apply Chang (1986), one needs to assume that when $\mathbf{V}_i$ is written as a vector $\mathrm{vec}(\mathbf{V}_i)$ of length mp, it will have a spherically symmetric distribution around $\mathrm{vec}(\mathbf{A}_2\mathbf{u}_i\mathbf{A}_1^T)$. However $\mathrm{vec}(\mathbf{V}_i)$ is constrained to lie in a $m(2p-m-1)/2$-dimensional submanifold of Ω_{pm}. Thus Prentice's application of the spherical regression results of Chang (1986) cannot be justified.

We now consider the general case when either $c_1 \neq c_2$ or $d_1 \neq d_2$. For $1 \leq j < k \leq m$, let $\mathbf{H}_{jk}$ be the $m \times m$ skew symmetric matrix whose entries are all zero except for the (j, k)th and (k, j)th entries which are $+1$ and -1, respectively. Similarly define the $p \times p$ skew symmetric matrices $\mathbf{K}_{qr}$ for $1 \leq q < r \leq p$. Let the index set $\mathscr{I} = \{(1, j, k) | 1 \leq j < k \leq m\} \cup \{(2, q, r) \mid 1 \leq q < r \leq p\}$. Let $\mathbf{A}_n$ be the $[m(m-1)/2 + p(p-1)/2] \times [m(m-1)/2 + p(p-1)/2]$

matrix of the quadratic form (28). Its entries are

$$
\begin{aligned}
n\mathbf{A}_{n\iota_1\iota_2} &= n\mathbf{A}_{n\iota_2\iota_1} \\
&= -c_1 n\mathrm{Tr}(\mathbf{H}_{jk}\mathbf{H}_{j'k'}), \qquad \iota_1 = (1, j, k), \qquad \iota_2 = (1, j', k') \\
&= c_1\mathrm{Tr}\Big(\sum_i \mathbf{H}_{jk}\mathbf{u}_i^T\mathbf{K}_{qr}\mathbf{u}_i\Big), \qquad \iota_1 = (1, j, k), \qquad \iota_2 = (2, q, r) \\
&= -c_2\ \mathrm{Tr}\Big(\sum_i \mathbf{K}_{qr}\mathbf{K}_{q'r'}\mathbf{u}_i\mathbf{u}_i^T\Big) - (c_1 - c_2)\mathrm{Tr}\Big(\sum_i \mathbf{K}_{qr}\mathbf{u}_i\mathbf{u}_i^T\mathbf{K}_{q'r'}\mathbf{u}_i\mathbf{u}_i^T\Big), \\
&\qquad\qquad\qquad\qquad\qquad\qquad \iota_1 = (2, q, r), \quad \iota_2 = (2, q', r').
\end{aligned} \tag{29}
$$

Similarly, define the matrix $\mathbf{B}_n$ by replacing, in (29), c_1 and c_2 by d_1 and d_2. Let $\mathbf{A} = \lim_{n\to\infty}\mathbf{A}_n$ and $\mathbf{B} = \lim_{n\to\infty}\mathbf{B}_n$.

PROPOSITION 6. *Suppose* $1 < m < p$. *Write* $(\widehat{\mathbf{A}}_1, \widehat{\mathbf{A}}_2) = (\mathbf{A}_1\exp(\widehat{\mathbf{H}}), \mathbf{A}_2\times \exp(\widehat{\mathbf{K}}))$ *where* $\widehat{\mathbf{H}}$ *and* $\widehat{\mathbf{K}}$ *are skew symmetric of sizes* $m\times m$ *and* $p\times p$, *respectively. Write* $\widehat{\mathbf{H}} = \sum_{1\le j<k\le m}\hat{a}_{jk}\mathbf{H}_{jk}$ *and* $\widehat{\mathbf{K}} = \sum_{1\le q<r\le p}\hat{b}_{qr}\mathbf{K}_{qr}$. *Then* $\sqrt{n}(\hat{a}_{12}, \ldots, \hat{a}_{(m-1)m}, \hat{b}_{12}, \ldots, \hat{b}_{(p-1)p})$ *is asymptotically normally distributed with mean zero and covariance matrix* $\mathbf{B}^{-1}\mathbf{A}\mathbf{B}^{-1}$.

7. Estimation of the constants c_1, c_2, d_1, d_2. Consider first the location model for $2\le m\le p-1$ and general invariant f and ρ satisfying (13) and (14). Let $\gamma(s)$ be a curve in $\mathbf{\Theta}$ with $\gamma(0) = \boldsymbol{\theta}_0$. Using Lemma 1 (see Section 9),

$$
E_{\boldsymbol{\theta}_0}\bigg[\frac{d}{ds}\bigg|_{s=0}\rho(\mathbf{X}, \gamma(s))\bigg] = 0. \tag{30}
$$

Write $\gamma'(0)$ in the form $[\boldsymbol{\theta}_0\ \boldsymbol{\theta}_{0\perp}][\mathbf{K}_1^T\ \mathbf{K}_2^T]^T$ where, by appropriate choice of γ, $\mathbf{K}_1$ can be any $m\times m$ skew symmetric matrix and $\mathbf{K}_2$ any $p-m\times m$ matrix without restriction. Using (10), (12) and (30) it follows that

$$
\begin{aligned}
\langle\gamma'(0), \gamma'(0)\rangle_A &= \mathrm{E}_{\boldsymbol{\theta}_0}\bigg[\bigg(\frac{d}{ds}\bigg|_{s=0}\rho(\mathbf{X}, \gamma(s))\bigg)^2\bigg], \\
\langle\gamma'(0), \gamma'(0)\rangle_B &= -\mathrm{E}_{\boldsymbol{\theta}_0}\bigg[\frac{d^2}{ds^2}\bigg|_{s=0}\rho(\mathbf{X}, \gamma(s))\bigg].
\end{aligned}
$$

On the other hand, $\gamma'(0) = g\cdot[\mathbf{K}_1^T\ \mathbf{K}_2^T]^T$ where $g = (\mathbf{I}_m, [\boldsymbol{\theta}_0\boldsymbol{\theta}_{0\perp}])\in SO(m)\times SO(p)$. Therefore, since $\langle,\rangle_A$ and $\langle,\rangle_B$ are invariant,

$$
\begin{aligned}
\langle\gamma'(0), \gamma'(0)\rangle_A &= \langle[\mathbf{K}_1^T\ \mathbf{K}_2^T]^T, [\mathbf{K}_1^T\ \mathbf{K}_2^T]^T\rangle_A = c_1\ \mathrm{Tr}(\mathbf{K}_1^T\mathbf{K}_1) + c_2\ \mathrm{Tr}(\mathbf{K}_2^T\mathbf{K}_2), \\
\langle\gamma'(0), \gamma'(0)\rangle_B &= \langle[\mathbf{K}_1^T\ \mathbf{K}_2^T]^T, [\mathbf{K}_1^T\ \mathbf{K}_2^T]^T\rangle_B = d_1\mathrm{Tr}(\mathbf{K}_1^T\mathbf{K}_1) + d_2\ \mathrm{Tr}(\mathbf{K}_2^T\mathbf{K}_2).
\end{aligned}
$$

Suppose $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ are curves such that $\tilde{\gamma}_1(0) = \tilde{\gamma}_2(0) = \boldsymbol{\theta}_0$, $\tilde{\gamma}_1'(0) = [\boldsymbol{\theta}_0\ \boldsymbol{\theta}_{0\perp}][\mathbf{K}_1^T\ \mathbf{0}]^T$, and $\tilde{\gamma}_2'(0) = [\boldsymbol{\theta}_0\, \boldsymbol{\theta}_{0\perp}][\mathbf{0}\, \mathbf{K}_2^T]^T$. Then

$$\begin{aligned}
c_1 &= \mathrm{E}_{\boldsymbol{\theta}_0}\left[\left(\frac{d}{ds}\bigg|_{s=0} \rho(\mathbf{X}, \tilde{\gamma}_1(s))\right)^2\right] \bigg/ \mathrm{Tr}(\mathbf{K}_1^T\mathbf{K}_1),\\
c_2 &= \mathrm{E}_{\boldsymbol{\theta}_0}\left[\left(\frac{d}{ds}\bigg|_{s=0} \rho(\mathbf{X}, \tilde{\gamma}_2(s))\right)^2\right] \bigg/ \mathrm{Tr}(\mathbf{K}_2^T\mathbf{K}_2),\\
d_1 &= -\mathrm{E}_{\boldsymbol{\theta}_0}\left[\frac{d^2}{ds^2}\bigg|_{s=0} \rho(\mathbf{X}, \tilde{\gamma}_1(s))\right] \bigg/ \mathrm{Tr}(\mathbf{K}_1^T\mathbf{K}_1),\\
d_2 &= -\mathrm{E}_{\boldsymbol{\theta}_0}\left[\frac{d^2}{ds^2}\bigg|_{s=0} \rho(\mathbf{X}, \tilde{\gamma}_2(s))\right] \bigg/ \mathrm{Tr}(\mathbf{K}_2^T\mathbf{K}_2).
\end{aligned} \tag{31}$$

These equations do not depend upon the choice of $\mathbf{K}_1$, $\mathbf{K}_2$, $\tilde{\gamma}_1$ or $\tilde{\gamma}_2$.

REMARK. It is worthwhile to comment here about the mysterious appearance of the matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ in (31).

We reason by analogy to the sphere. Suppose $\mathbf{X} \in \Omega_p$, written as a column vector, has a rotationally symmetric distribution around $\boldsymbol{\theta}_0 \in \Omega_p$. Chang (1986) has expressed this condition in the form

$$\mathrm{E}_{\boldsymbol{\theta}_0}(\mathbf{X}\mathbf{X}^T) = k_1\boldsymbol{\theta}_0\boldsymbol{\theta}_0^T + k_2\mathbf{I}_p.$$

Equivalently, we can say that for all $\mathbf{K}$ which satisfy $\mathbf{K}^T\boldsymbol{\theta}_0 = 0$,

$$\mathrm{E}_{\boldsymbol{\theta}_0}((\mathbf{X}^T\mathbf{K})^2) = k_2\mathbf{K}^T\mathbf{K}. \tag{32}$$

Since $\mathbf{K}^T\boldsymbol{\theta}_0 = 0$, $\mathbf{K}$ represents a tangential direction in which $\mathbf{X}$ can stray from its true value $\boldsymbol{\theta}_0$. Equation (32) expresses the idea that all directions are the same (at least to second-order moments).

The Stiefel manifolds are more complicated than the sphere; they exhibit less symmetry. Essentially we are saying that there are two types of directions: those of the form $[\mathbf{K}_1\ \mathbf{0}]^T$ and those of the form $[\mathbf{0}\ \mathbf{K}_2]^T$. Within each of these two groups, all directions are the same. Notice however that $[\mathbf{K}_1\ \mathbf{0}]^T$ and $[\mathbf{0}\ \mathbf{K}_2]^T$ are not tangent to $\mathscr{V}_{p,m}$ at $\boldsymbol{\theta}_0$. They are tangent at $\boldsymbol{\theta}_1 = [\mathbf{I}_m\ \mathbf{0}]^T$. We have used a transformation $[\boldsymbol{\theta}_0^T\ \boldsymbol{\theta}_{0\perp}^T]$ to bring $\boldsymbol{\theta}_0$ to $\boldsymbol{\theta}_1$. This is analogous to the often used simplification in spherical statistics of assuming that a true modal direction is the North Pole.

One way to produce the curves $\tilde{\gamma}_1$ or $\tilde{\gamma}_2$ is to define $\tilde{\gamma}_1(s) = \gamma_1(s\mathbf{K}_1, \boldsymbol{\theta}_0)$ and $\tilde{\gamma}_2(s) = \gamma_2(s\mathbf{K}_2, \boldsymbol{\theta}_0)$, where γ_1 and γ_2 are defined by

$$\begin{aligned}
\gamma_1(\mathbf{K}_1, \boldsymbol{\theta}) &= \boldsymbol{\theta}\exp(\mathbf{K}_1),\\
\gamma_2(\mathbf{K}_2, \boldsymbol{\theta}) &= \exp(\boldsymbol{\theta}_\perp\mathbf{K}_2\boldsymbol{\theta}^T - \boldsymbol{\theta}\mathbf{K}_2^T\boldsymbol{\theta}_\perp^T)\boldsymbol{\theta}.
\end{aligned}$$

To obtain sample estimates from (31), one can replace the expected values by the obvious sample means, using any convenient $\mathbf{K}_1$ and $\mathbf{K}_2$. However, the $m \times m$ skew symmetric matrices $\mathbf{K}_1$ span a vector space of dimension $m(m-1)/2$ and a more reasonable approach would be to use a basis of this vector space and average the results over this basis. We propose a similar approach for $\mathbf{K}_2$. Thus let $\mathbf{K}_{1jk}$ be the $m \times m$ skew symmetric matrix which has 1 in the (j, k)th spot, -1 in the (k, j)th spot, and zero elsewhere. Similarly, let $\mathbf{K}_{2jk}$ be the $p - m \times m$ matrix which has 1 in the (j, k)th spot and zero elsewhere.

This leads to the sample estimates

$$
\begin{aligned}
\hat{c}_1 &= \frac{1}{nm(m-1)} \sum_i \sum_{j<k} \left[\frac{d}{ds}\bigg|_{s=0} \rho(\mathbf{X}_i, \gamma_1(s\mathbf{K}_{1jk}, \hat{\boldsymbol{\theta}})) \right]^2, \\
\hat{c}_2 &= \frac{1}{nm(p-m)} \sum_i \sum_{j,k} \left[\frac{d}{ds}\bigg|_{s=0} \rho(\mathbf{X}, \gamma_2(s\mathbf{K}_{2jk}, \hat{\boldsymbol{\theta}})) \right]^2, \\
\hat{d}_1 &= -\frac{1}{nm(m-1)} \sum_i \sum_{j<k} \left[\frac{d^2}{ds^2}\bigg|_{s=0} \rho(\mathbf{X}, \gamma_1(s\mathbf{K}_{1jk}, \hat{\boldsymbol{\theta}})) \right], \\
\hat{d}_2 &= -\frac{1}{nm(p-m)} \sum_i \sum_{j,k} \left[\frac{d^2}{dt^2}\bigg|_{s=0} \rho(\mathbf{X}, \gamma_2(s\mathbf{K}_{2jk}, \hat{\boldsymbol{\theta}})) \right].
\end{aligned}
\tag{33}
$$

An obvious difficulty with (33) is the evaluation of the derivatives when ρ is messy, such as in (7). In this case we can use "central difference" approximations to the derivatives of a function ϕ,

$$
\begin{aligned}
\phi'(0) &= (2\varepsilon)^{-1}(\phi(\varepsilon) - \phi(-\varepsilon)) + O(\varepsilon^2), \\
\phi''(0) &= \varepsilon^{-2}(\phi(\varepsilon) + \phi(-\varepsilon) - 2\phi(0)) + O(\varepsilon^3).
\end{aligned}
\tag{34}
$$

Thus, we can apply (34) within (33) with $\phi(s) = \rho(\mathbf{X}_i, \gamma_1(s\mathbf{K}_{1jk}, \hat{\boldsymbol{\theta}}))$ or with $\phi(s) = \rho(\mathbf{X}_i, \gamma_2(s\mathbf{K}_{2jk}, \hat{\boldsymbol{\theta}}))$ and ε chosen small.

The difficulties with evaluating the derivatives in (31) and (33) do not occur when (5) is true. In this case

$$
\begin{aligned}
&\frac{d}{ds}\bigg|_{s=0} \rho(\mathbf{X}, \gamma_1(s\mathbf{K}_1, \boldsymbol{\theta})) = \rho_0'(t)\mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}\mathbf{K}_1), \\
&\frac{d}{ds}\bigg|_{s=0} \rho(\mathbf{X}, \gamma_2(s\mathbf{K}_2, \boldsymbol{\theta})) = \rho_0'(t)\mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}_\perp\mathbf{K}_2), \\
&\frac{d^2}{ds^2}\bigg|_{s=0} \rho(\mathbf{X}, \gamma_1(s\mathbf{K}_1, \boldsymbol{\theta})) = \rho_0''(t)[\mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}\mathbf{K}_1)]^2 - \rho_0'(t)\mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}\mathbf{K}_1^T\mathbf{K}_1), \\
&\frac{d^2}{ds^2}\bigg|_{s=0} \rho(\mathbf{X}, \gamma_2(s\mathbf{K}_2, \boldsymbol{\theta})) = \rho_0''(t)[\mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}_\perp\mathbf{K}_2)]^2 - \rho_0'(t)\mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}\mathbf{K}_2^T\mathbf{K}_2) \\
&\qquad t = \mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}).
\end{aligned}
\tag{35}
$$

This yields the population parameters

$$
(36)\qquad
\begin{aligned}
c_1 &= \mathrm{E}_{\boldsymbol{\theta}_0}[\rho_0'(t)^2(\mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}_0\mathbf{K}_1))^2]/\mathrm{Tr}(\mathbf{K}_1^T\mathbf{K}_1),\\
c_2 &= \mathrm{E}_{\boldsymbol{\theta}_0}[\rho_0'(t)^2(\mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}_{0\perp}\mathbf{K}_2))^2]/\mathrm{Tr}(\mathbf{K}_2^T\mathbf{K}_2),\\
d_1 &= -\mathrm{E}_{\boldsymbol{\theta}_0}[\rho_0''(t)(\mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}_0\mathbf{K}_1))^2 - \rho_0'(t)\ \mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}_0\mathbf{K}_1^T\mathbf{K}_1)]/\mathrm{Tr}(\mathbf{K}_1^T\mathbf{K}_1),\\
d_2 &= -\mathrm{E}_{\boldsymbol{\theta}_0}[\rho_0''(t)(\mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}_{0\perp}\mathbf{K}_2))^2 - \rho_0'(t)\ \mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}_0\mathbf{K}_2^T\mathbf{K}_2)]/\mathrm{Tr}(\mathbf{K}_2^T\mathbf{K}_2),
\end{aligned}
$$

where $t = \mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}_0)$ and, as before, the right-hand sides of (36) do not depend upon the choices of $\mathbf{K}_1$ or $\mathbf{K}_2$.

We note that if $\mathbf{R} = \mathbf{X}^T\boldsymbol{\theta}$ and $\mathbf{R}_\perp = \mathbf{X}^T\boldsymbol{\theta}_\perp$,

$$
(37)\qquad
\begin{aligned}
\sum_{j<k}\mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}\mathbf{K}_{1jk}^T\mathbf{K}_{1jk}) &= \mathrm{Tr}\Bigl(\mathbf{R}\sum_{j<k}(\mathbf{K}_{1jk}^T\mathbf{K}_{1jk})\Bigr) = (m-1)\mathrm{Tr}(\mathbf{R}),\\
\sum_{j,k}\mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}\mathbf{K}_{2jk}^T\mathbf{K}_{2jk}) &= (p-m)\mathrm{Tr}(\mathbf{R}),\\
\sum_{j<k}(\mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}\mathbf{K}_{1jk}))^2 &= \sum_{j<k}(\mathbf{R}_{kj}-\mathbf{R}_{jk})^2 = -\frac{1}{2}\mathrm{Tr}((\mathbf{R}-\mathbf{R}^T)^2),\\
\sum_{j,k}(\mathrm{Tr}(\mathbf{X}^T\boldsymbol{\theta}_\perp\mathbf{K}_{2jk}))^2 &= \mathrm{Tr}(\mathbf{R}_\perp^T\mathbf{R}_\perp) = \mathrm{Tr}(\mathbf{X}\mathbf{X}^T\boldsymbol{\theta}_\perp\boldsymbol{\theta}_\perp^T),\\
&= \mathrm{Tr}(\mathbf{X}\mathbf{X}^T(\mathbf{I}_m - \boldsymbol{\theta}\boldsymbol{\theta}^T)) = m - \mathrm{Tr}(\mathbf{R}\mathbf{R}^T).
\end{aligned}
$$

Applying (35) and (37) to (33) we get the following estimates:

$$
(38)\qquad
\begin{aligned}
\hat{c}_1 &= -\frac{1}{2nm(m-1)}\sum_i(\rho_0'(\hat{t}_i))^2\mathrm{Tr}((\widehat{\mathbf{R}}_i - \widehat{\mathbf{R}}_i^T)^2),\\
\hat{c}_2 &= \frac{1}{nm(p-m)}\sum_i(\rho_0'(\hat{t}_i))^2(m - \mathrm{Tr}(\widehat{\mathbf{R}}_i\widehat{\mathbf{R}}_i^T)),\\
\hat{d}_1 &= \frac{1}{2nm(m-1)}\sum_i\rho_0''(\hat{t}_i)\mathrm{Tr}((\widehat{\mathbf{R}}_i - \widehat{\mathbf{R}}_i^T)^2) + \frac{1}{nm}\sum_i\rho_0'(\hat{t}_i)\hat{t}_i,\\
\hat{d}_2 &= -\frac{1}{nm(p-m)}\sum_i\rho_0''(\hat{t}_i)(m - \mathrm{Tr}(\widehat{\mathbf{R}}_i\widehat{\mathbf{R}}_i^T)) + \frac{1}{nm}\sum_i\rho_0'(\hat{t}_i)\hat{t}_i,
\end{aligned}
$$

where $\widehat{\mathbf{R}}_i = \mathbf{X}_i^T\hat{\boldsymbol{\theta}}$ and $\hat{t}_i = \mathrm{Tr}\widehat{\mathbf{R}}_i$.

For the important case $m = 1$ of the sphere Ω_p, (38) becomes

$$
(39)\qquad
\begin{aligned}
\hat{c}_2 &= \frac{1}{n(p-1)}\sum_i\rho_0'(\hat{t}_i)^2(1-\hat{t}_i^2),\\
\hat{d}_2 &= -\frac{1}{n(p-1)}\sum_i\rho_0''(\hat{t}_i)(1-\hat{t}_i^2) + \frac{1}{n}\sum_i\rho_0'(\hat{t}_i)\hat{t}_i,
\end{aligned}
$$

where $\hat{t}_i = \mathrm{Tr}(\mathbf{X}_i^T\hat{\boldsymbol{\theta}})$.

If $m = p$, we use the obvious subset of the estimates (33) or (38).

If the density f has the form (2) and ρ has the form (5), we will show at the end of this section that $c_1 = c_2 = c$ and $d_1 = d_2 = d$. In this case we pool the estimates given in (38). Recalling that the $\mathbf{K}_1$ and the $\mathbf{K}_2$ form vector spaces of dimensions $m(m-1)/2$ and $m(p-m)$, respectively, we propose

$$\begin{aligned} \hat{c} &= \frac{m(m-1)}{m(2p-m-1)}\hat{c}_1 + \frac{2m(p-m)}{m(2p-m-1)}\hat{c}_2, \\ \hat{d} &= \frac{m(m-1)}{m(2p-m-1)}\hat{d}_1 + \frac{2m(p-m)}{m(2p-m-1)}\hat{d}_2, \end{aligned} \tag{40}$$

where $\hat{c}_1$, $\hat{c}_2$, $\hat{d}_1$ and $\hat{d}_2$ come from (38).

Recall that the constants are the same for regression and location models when the same f and ρ apply. Thus we can transform (33), (38), (39) and (40) into formulas for regression models by replacing, within the ith summands, each $\mathbf{X}_i$ by $\mathbf{V}_i$ and each $\hat{\boldsymbol{\theta}}$ by $\widehat{\mathbf{A}}_2\mathbf{u}_i\widehat{\mathbf{A}}_1^T$. Note that the $p \times p-m$ matrix $\widehat{\mathbf{A}}_2\mathbf{u}_\perp$ is an orthogonal complement to the $p \times m$ matrix $\widehat{\mathbf{A}}_2\mathbf{u}\widehat{\mathbf{A}}_1^T$.

In this way, for example, (38) and (40) remain valid for regression models when $\widehat{R}_i = \mathbf{V}_i^T\widehat{\mathbf{A}}_2\mathbf{u}_i\widehat{\mathbf{A}}_1^T$.

Similarly, for spherical regressions, (39) remains valid when $\hat{t}_i = \mathbf{V}_i^T\widehat{\mathbf{A}}\mathbf{u}_i$. The resulting estimators coincide with those of Chang and Ko [(1995), page 1837], after correction for a typo there. When $\rho_0(t) = t$, they also coincide with Chang (1986), Proposition 1.

REMARK. It is interesting to note what (36) and (40) become when $\rho_0(t) = t$ and f satisfies (2). In this case, Prentice (1989) described the dispersion in f_0 using three constants which, to avoid conflict with the notation of this paper, we will denote as c_{0p}, c_{1p} and c_{2p}.

Prentice defined c_{0p} by the equation $\mathrm{E}_{\boldsymbol{\theta}_0}(\mathbf{X}) = c_{0p}\boldsymbol{\theta}_0$. It easily follows from (36) that $c_{0p} = d$. From (38) and (40), $\hat{d} = (nm)^{-1}\sum_i \mathrm{Tr}(\mathbf{X}_i^T\hat{\boldsymbol{\theta}})$ for the location model and $\hat{d} = \hat{c}_{0p}$ for the regression model (after correction for a typo in the definition of $\hat{c}_{0p}$).

Prentice defined c_{1p} and c_{2p} using the equation

$$\mathrm{E}_{\boldsymbol{\theta}_0}(\mathbf{X} \otimes \mathbf{X}) = (c_{0p}^2 + c_{1p})\boldsymbol{\theta}_0 \otimes \boldsymbol{\theta}_0 + c_{2p}\mathbf{I}_q, \tag{41}$$

where $\mathbf{I}_q$ represents a $q \times q$ identity matrix with $q = pm$. The authors believe that, except for the case $m = 1$, the left-hand side of (41) does not necessarily have the form of the right-hand side of (41).

Consider, for example, the important case of $p = m = 3$. Without loss of generality, we can assume $\boldsymbol{\theta}_0 = \mathbf{I}_3$. Then, letting x_{ij} denote the entries of $\mathbf{X}$, (41) implies $\mathrm{E}(x_{ii}^2) = c_{0p}^2 + c_{1p} + c_{2p}$, $\mathrm{E}(x_{ij}^2) = c_{2p}$, and $\mathrm{E}(x_{ii}x_{jj}) = c_{0p}^2 + c_{1p}$, $i \neq j$, with all other second moments equal to zero. $\mathbf{X}$ represents a rotation of an angle ϕ around an axis $\mathbf{t} = [t_1\ t_2\ t_3]^T$, and under the assumption (2), ϕ is independent of $\mathbf{t}$ with the latter uniformly distributed on the sphere Ω_3. Then

$$\mathbf{X} = \cos\phi \mathbf{I}_3 + \sin\phi \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix} + (1 - \cos\phi)\mathbf{t}\mathbf{t}^T.$$

Routine calculations then establish

$$\begin{aligned} 15\mathrm{E}(x_{ii}^2) &= 3 + 4\mathrm{E}(\cos\phi) + 8\mathrm{E}(\cos^2\phi), \\ 15\mathrm{E}(x_{ij}^2) &= 6 - 2\mathrm{E}(\cos\phi) - 4\mathrm{E}(\cos^2\phi), \\ 15\mathrm{E}(x_{ii}x_{jj}) &= 1 + 8\mathrm{E}(\cos\phi) + 6\mathrm{E}(\cos^2\phi), \end{aligned}$$

in contradiction to (41).

Thus our $\hat{c}$ has no obvious relationship to Prentice's $\hat{c}_{2p}$, due to, we believe, a misspecification in the definition of c_{2p}.

We now establish $c_1 = c_2 = c$ and $d_1 = d_2 = d$ when f has the form (2) and ρ has the form (5). Let $\mu(d\mathbf{x})$ denote a measure on $\mathscr{V}_{p,m}$ which is invariant under $SO(m) \times SO(p)$.

PROPOSITION 7. *Let $\boldsymbol{\theta}_1 = [\mathbf{I}_m\, \mathbf{0}]^T$. Let $\phi(t)$ be continuous. Then there is a constant $c(\phi)$ such that for all $\mathbf{K} = [\mathbf{K}_1^T\, \mathbf{K}_2^T]^T$, with $\mathbf{K}_1$ $m \times m$ skew-symmetric and $\mathbf{K}_2$ $m \times p - m$ arbitrary,*

$$\int_{\mathscr{V}_{p,m}} \phi(\mathrm{Tr}(\mathbf{x}^T\boldsymbol{\theta}_1))[\mathrm{Tr}(\mathbf{x}^T\mathbf{K})]^2\mu(d\mathbf{x}) = c(\phi)\mathrm{Tr}(\mathbf{K}^T\mathbf{K}).$$

PROOF. Let $g(t)$ be arbitrary. We claim that $\int_{\mathscr{V}_{p,m}} g(\mathrm{Tr}(\mathbf{x}^T\boldsymbol{\theta}))\mu(d\mathbf{x})$ does not depend upon $\boldsymbol{\theta}$. To see this, write $\boldsymbol{\theta} = \mathbf{A}_2\boldsymbol{\theta}_1\mathbf{A}_1^T$. Then

$$\int_{\mathscr{V}_{p,m}} g(\mathrm{Tr}(\mathbf{x}^T\boldsymbol{\theta}))\mu(d\mathbf{x}) = \int_{\mathscr{V}_{p,m}} g(\mathrm{Tr}((\mathbf{A}_2^T\mathbf{x}\mathbf{A}_1)^T\boldsymbol{\theta}_1))\mu(d\mathbf{x})$$

and the claim follows from making the substitution $\mathbf{y} = \mathbf{A}_2^T\mathbf{x}\mathbf{A}_1$ and $\mu(d\mathbf{y}) = \mu(d\mathbf{x})$.

Thus if $\gamma(s)$ is any curve in $\mathscr{V}_{p,m}$,

$$\begin{aligned} 0 &= \frac{d}{ds}\int_{\mathscr{V}_{p,m}} g(\mathrm{Tr}(\mathbf{x}^T\gamma(s)))\mu(d\mathbf{x}) \\ &= \int_{\mathscr{V}_{p,m}} g'(\mathrm{Tr}(\mathbf{x}^T\gamma(s)))\mathrm{Tr}(\mathbf{x}^T\gamma'(s))\mu(d\mathbf{x}), \\ &\int_{\mathscr{V}_{p,m}} g''(\mathrm{Tr}(\mathbf{x}^T\gamma(s)))[\mathrm{Tr}(\mathbf{x}^T\gamma'(s))]^2\mu(d\mathbf{x}) \\ &= -\int_{\mathscr{V}_{p,m}} g'(\mathrm{Tr}(\mathbf{x}^T\gamma(s)))\mathrm{Tr}(\mathbf{x}^T\gamma''(s))\mu(d\mathbf{x}). \end{aligned}$$

We apply this identity to the curve

$$\begin{aligned}\gamma(s) &= \exp\left(s\begin{bmatrix}\mathbf{0} & -\mathbf{K}_2^T\\ \mathbf{K}_2 & \mathbf{0}\end{bmatrix}\right)\boldsymbol{\theta}_1\exp(s\mathbf{K}_1)\\ &= \begin{bmatrix}\mathbf{I}_m + s\mathbf{K}_1 - s^2(\mathbf{K}_1^T\mathbf{K}_1 + \mathbf{K}_2^T\mathbf{K}_2)/2\\ s\mathbf{K}_2 + s^2\mathbf{K}_2\mathbf{K}_1\end{bmatrix} + O(s^3),\end{aligned}$$

where we let $t = \mathrm{Tr}(\mathbf{x}^T\boldsymbol{\theta}_1)$. This yields

$$\int_{\mathscr{V}_{p,m}} g''(t)[\mathrm{Tr}(\mathbf{x}^T\mathbf{K})]^2\mu(d\mathbf{x}) = \mathrm{Tr}[\mathbf{C}\mathbf{K}_1^T\mathbf{K}_1] + \mathrm{Tr}[\mathbf{C}\mathbf{K}_2^T\mathbf{K}_2] - 2\,\mathrm{Tr}[\mathbf{D}\mathbf{K}_2\mathbf{K}_1],$$

where the $m \times m$ and $m \times p - m$ matrices $\mathbf{C}$ and $\mathbf{D}$ are $\int g'(t)\mathbf{x}^T\boldsymbol{\theta}_1\mu(d\mathbf{x})$ and $\int g'(t)\mathbf{x}^T\boldsymbol{\theta}_{1\perp}\mu(d\mathbf{x})$, respectively.

It is easily checked that $\mathbf{A}\mathbf{C}\mathbf{A}^T = \mathbf{C}$ for all $\mathbf{A} \in \mathscr{O}(m)$. This implies, by the Schur lemma [see Vinberg (1989)], that $\mathbf{C} = c\mathbf{I}_m$ for some constant c. Writing $\mathbf{x} = [\mathbf{x}_1^T\ \mathbf{x}_2^T]^T$, make the change of variables $\mathbf{y} = [\mathbf{x}_1^T\ \ -\mathbf{x}_2^T]^T$ in $\mathbf{D}$. This yields $\mathbf{D} = -\mathbf{D}$ and completes the proof of the proposition. □

COROLLARY 2. *Suppose the density and objective functions satisfy* (2) *and* (5). *Then* $c_1 = c_2$ *and* $d_1 = d_2$.

PROOF. We apply Proposition 7 to (10) with $\gamma(s) = \gamma_1(s\mathbf{K}_1, \boldsymbol{\theta}_0)$. Letting $\boldsymbol{\theta}_0 = \mathbf{A}_2\boldsymbol{\theta}_1\mathbf{A}_1^T$ and making the substitution $\mathbf{y} = \mathbf{A}_2^T\mathbf{x}\mathbf{A}_1$, we get

$$\begin{aligned}\langle\boldsymbol{\theta}_0\mathbf{K}_1, \boldsymbol{\theta}_0\mathbf{K}_1\rangle_A &= \int_{\mathscr{V}_{p,m}} [\rho_0'(\mathrm{Tr}(\mathbf{x}^T\boldsymbol{\theta}_0))\mathrm{Tr}(\mathbf{x}^T\boldsymbol{\theta}_0\mathbf{K}_1)]^2 f_0(\mathrm{Tr}(\mathbf{x}^T\boldsymbol{\theta}_0))\mu(d\mathbf{x})\\ &= c(\rho_0'^2 f_0)\mathrm{Tr}(\mathbf{K}_1^T\mathbf{K}_1).\end{aligned}$$

Thus $c_1 = c(\rho_0'^2 f_0)$. Similarly we use $\gamma(s) = \gamma_2(s\mathbf{K}_2, \boldsymbol{\theta}_0)$ to establish that $c_2 = c(\rho_0'^2 f_0)$.

The exact same proof using (11) establishes that $d_1 = c(\rho_0' f_0') = d_2$. □

8. A numerical example. Downs, Liebman and MacKay (1974) discuss a data set consisting of 98 matched pairs of "leads of vector cardiograms." Each pair corresponds to one subject; there are 28 male subjects age 2–10, 28 male subjects age 11–19, 17 female subjects age 2–10, and 25 female subjects age 11–19. The leads are referred to as the "Frank" lead and the "McFee" lead.

The leads are close to planar curves in R^3; each curve has a shape close to a cardiod. Thus each lead can be described by a 3×2 matrix whose first column is a unit vector $\mathbf{q}$, in the plane of the curve, in the direction of the vector from the apex of the cardiod to its cusp, and whose second column is a vector $\mathbf{p}$ perpendicular to the plane of the curve. We denote by $(\mathbf{u}_i, \mathbf{V}_i)$ the pair of 3×2 matrices, corresponding to the Frank and McFee leads respectively, for the ith subject.

Downs (1972) and Khatri and Mardia (1977) analyze the McFee leads under the assumption that the density of $\mathbf{V}$ is matrix Fisher,

$$f(\mathbf{v}, \boldsymbol{\theta}, \mathbf{K}) = c(\mathbf{K}) \exp(\mathrm{Tr}(\mathbf{v}\mathbf{K}\boldsymbol{\theta}^T)), \tag{42}$$

where the unknown parameter $\mathbf{K}$ is a 2×2 symmetric positive definite matrix. Downs uses a tangent space approximation at $\boldsymbol{\theta}$ for the distribution of $\mathbf{V}$, and hence his results are asymptotic as the eigenvalues of $\mathbf{K}$ approach infinity. Khatri and Mardia (1977) calculate the large sample asymptotic behavior of the MLE. For one sample problems, these two papers consider primarily the question of testing if $\boldsymbol{\theta}$ is some specified $\boldsymbol{\theta}_0$.

Prentice (1986, 1989) augmented the pairs $(\mathbf{u}_i, \mathbf{V}_i)$ to pairs $(\tilde{\mathbf{u}}_i, \widetilde{\mathbf{V}}_i)$ in $SO(3) \times SO(3)$ by adding third columns to each $\mathbf{u}_i$ and $\mathbf{V}_i$. Thus within each matrix, the third vector is in the plane of the cardiod and is perpendicular to the vector from the apex to its cusp. Prentice (1986), using a large sample nonparametric approach, studied location problems on $SO(3)$. He concluded that the modal matrix of the $\widetilde{\mathbf{V}}_i$ depends upon age, but not gender. Letting $\widetilde{\mathbf{X}}_i = \tilde{\mathbf{u}}_i^T \widetilde{\mathbf{V}}_i$, Prentice (1986) concluded that the modal vector of the $\widetilde{\mathbf{X}}_i$ does not depend upon either gender or age. He did conclude, however, that the distribution of the $\widetilde{\mathbf{X}}_i$ is not rotationally symmetric; this would preclude, for example, its distribution having the form $\tilde{f}_0(\mathrm{Tr}(\widetilde{\mathbf{X}}_i^T \boldsymbol{\theta}))$.

Prentice (1989) used a regression approach to the $(\tilde{\mathbf{u}}_i, \widetilde{\mathbf{V}}_i)$, and assumed that the distribution of $\widetilde{\mathbf{V}}_i$ is the form

$$\tilde{f}(\tilde{\mathbf{v}}_i, \widetilde{\mathbf{A}}_2 \tilde{\mathbf{u}}_i \widetilde{\mathbf{A}}_1^T) = \tilde{f}_0(\mathrm{Tr}(\tilde{\mathbf{v}}_i^T \widetilde{\mathbf{A}}_2 \tilde{\mathbf{u}}_i \widetilde{\mathbf{A}}_1^T)), \tag{43}$$

where $(\widetilde{\mathbf{A}}_1, \widetilde{\mathbf{A}}_2) \in SO(3) \times SO(3)$. As discussed above, we disagree with the correctness of his derived distribution of the estimators $(\widehat{\widetilde{\mathbf{A}}}_1, \widehat{\widetilde{\mathbf{A}}}_2)$.

Prentice (1986, 1989) performed separate analyses for each age and gender group. Unfortunately, the authors have only been able to obtain data without age and gender labels. Emboldened by the results of Prentice (1986), we will assume that a single regression model for the original combined data $(\mathbf{u}_i, \mathbf{V}_i)$ holds. In other words, the density f satisfies the invariance condition (1) which, when rewritten in the regression context, becomes

$$f(\mathbf{C}_2 \mathbf{v}_i \mathbf{C}_1^T, \mathbf{C}_2 \mathbf{A}_2 \mathbf{u}_i \mathbf{A}_1^T \mathbf{C}_1^T) = f(\mathbf{v}_i, \mathbf{A}_2 \mathbf{u}_i \mathbf{A}_1^T), \tag{44}$$

where $(\mathbf{A}_1, \mathbf{A}_2) \in SO(2) \times SO(3)$ is the unknown parameter and $(\mathbf{C}_1, \mathbf{C}_2) \in SO(2) \times SO(3)$ is any pair of matrices.

It is useful to try to envision what the assumption (44) or, equivalently, (1) for the location model, physically means. Suppose the basic physical object is a distance-preserving linear transformation ϕ: $R^m \to R^p$. In this case we can imagine a prototypical cardiod curve in R^2 and the observed space curve is, after centering, a three-dimensional rotation of the prototype, together with some random deformation. In the studies of human motion, $m = p = 3$ and ϕ represents the rotation of one limb relative to the other limb. Let $X(\phi)$ be the matrix of ϕ with respect to some orthonormal bases of R^m and R^p.

The invariance assumption is simply that it does not matter which orthonormal bases are chosen. It is known that for many pairs (m, p), $SO(m) \times SO(p)$ is the largest compact connected Lie group which acts on $\mathscr{V}_{p,m}$ [see Hsiang and Su (1968)], so that invariance under $SO(m) \times SO(p)$ would seem to be mathematically quite natural.

In the studies of human motion, the bases are determined by three-dimensional sensors placed upon each limb, and the invariance assumption is clearly reasonable. In the planar curve randomly oriented in space example, we are assuming that neither the orthonormal basis of R^3 nor the orientation of the prototypical curve in R^2 matters. In the example of the Down's vector cardiogram data, the columns of X are the direction of the axis $\mathbf{q}$ of the cardiod and the normal $\mathbf{p}$ to the plane. Thus the invariance assumption under $SO(2)$ refers to rotating the prototypical curve around the axis $\mathbf{q} \times \mathbf{p}$. The authors believe this invariance assumption is not unreasonable.

The interpretation of the parameters $(\mathbf{A}_1, \mathbf{A}_2)$ in the regression model is a little complicated since $\mathbf{A}_1$ is completely superfluous unless there are at least two pairs of matched curves. [Indeed, Rivest and Chang (2000) establish that at least three, and sometimes four, pairs are needed to fit $\mathbf{A}_1$ and $\mathbf{A}_2$.] An additional complication is that the rotation groups $SO(p)$ are not commutative if $p \geq 3$. Imagine a prototypical cardiod curve in R^2 and that the maps ϕ_u and ϕ_v are maps from $R^2 \to R^3$ which, except for random deformation, generate the curves whose 3×2 matrices are $\mathbf{u}$ and $\mathbf{V}$. Suppose $\phi_v = \mathbf{A}_2 \phi_u \mathbf{A}_1^T$. If $\mathbf{A}_2 = \mathbf{I}_3$, the prototypical cardiod for the *v-cardiod* is obtained from the prototypical cardiod for the *u-cardiod* by the rotation $\mathbf{A}_1$ and the observed configuration is the result of applying ϕ_u to this pair of prototypical cardiods. Thus when $\mathbf{A}_2 = \mathbf{I}_3$, the observed pairs of cardiods are rigid motions of a prototypical pair of cardiods, possibly together with translations of one cardiod relative to the other. In the Prentice reformulation of the data as 3×3 matrices the orientation of the prototypical McFee lead to the prototypical Frank lead is arbitrary. In the original formulation of the data as 3×2 matrices, the regression model, when $\mathbf{A}_2 = \mathbf{I}_3$, would assume that the protypical McFee lead is obtained from the prototypical Frank lead by a rotation about the $\mathbf{q} \times \mathbf{p}$ axis.

On the other hand, if $\mathbf{A}_1$ equals $\mathbf{I}_2$ or $\mathbf{I}_3$ (depending upon whether two columns or three are used to express the data), the orientation of the McFee lead is obtained from that of the Frank lead by rotation around a fixed axis—namely, that of $\mathbf{A}_2$. Presumably in that case the axis of $\mathbf{A}_2$ has some intrinsic physical interest. In other words, $\mathbf{A}_1$ rotates the Frank lead into the McFee lead, using a rotation whose axis is fixed in a coordinate frame fixed to the Frank lead and $\mathbf{A}_2$ rotates the Frank lead into the McFee lead, using a rotation whose axis is fixed in a coordinate frame fixed to the outside. When $\mathbf{A}_2 = \mathbf{I}_3$, the orientation of the McFee lead relative to the Frank lead does not depend upon the orientation of the Frank lead. But when $\mathbf{A}_2 \neq \mathbf{I}_3$, the orientation of the McFee lead relative to the Frank lead depends upon the orientation of the Frank lead relative to the axis of $\mathbf{A}_2$. We again emphasize that one should think of multiple pairs of leads, since with one pair of leads, only one of $\mathbf{A}_1$ or $\mathbf{A}_2$ is needed to rotate the Frank lead into the McFee lead.

Borrowing terminology standard to the kinematics of rigid body motion, we will refer to these coordinate systems as "body" and "space" coordinates, respectively.

Recall that $c_1 = c_2$ if

$$f(\mathbf{v}_i, \mathbf{A}_2\mathbf{u}_i\mathbf{A}_1^T) = f_0(\mathrm{Tr}(\mathbf{v}_i^T\mathbf{A}_2\mathbf{u}_i\mathbf{A}_1^T)) \tag{45}$$

and an objective function of the form $\rho(\mathbf{V}_i, \mathbf{A}_2\mathbf{u}_i\mathbf{A}_1^T) = \rho_0(\mathrm{Tr}(\mathbf{V}_i^T\mathbf{A}_2\mathbf{u}_i\mathbf{A}_1^T))$ is used. Thus a test that the density has the form (45) versus $f(\mathbf{v}_i, \mathbf{A}_2\mathbf{u}_i\mathbf{A}_1^T)$ is a general invariant density (44) can be based upon $\hat{c}_2 - \hat{c}_1$. We will use the least squares objective function (6): $\rho_0(t) = 4 - 2t$. Then using (38),

$$\hat{c}_2 - \hat{c}_1 = \frac{2}{n}\sum_{i=1}^{n} 2 - 2\mathrm{Tr}(\widehat{R}_i\widehat{R}_i^T) + \mathrm{Tr}(\widehat{R}_i^2),$$

$$\widehat{R}_i = \mathbf{V}_i^T\widehat{\mathbf{A}}_2\mathbf{u}_i\widehat{\mathbf{A}}_1^T.$$

For the vector cardiogram data, $n = 98$,

$$\widehat{\mathbf{A}}_1 = \begin{bmatrix} 0.999 & -0.0425 \\ -0.0425 & 0.999 \end{bmatrix},$$

$$\widehat{\mathbf{A}}_2 = \begin{bmatrix} 0.952 & 0.124 & 0.281 \\ -0.153 & 0.985 & 0.0829 \\ -0.266 & -0.122 & 0.956 \end{bmatrix},$$

$\hat{c}_1 = 0.8058$ and $\hat{c}_2 = 1.0306$. Letting $Z_i = 2 - 2\mathrm{Tr}(\widehat{R}_i\widehat{R}_i^T) + \mathrm{Tr}(\widehat{R}_i^2)$, we can estimate the variance of $n^{1/2}(\hat{c}_2 - \hat{c}_1)$ by the sample variance of the Z_i. In this way we estimate the standard error of $\hat{c}_2 - \hat{c}_1$ to be 0.3155 and hence we cannot reject that the density has the form (45).

On the other hand, suppose we redefine the $\mathbf{u}_i$ to be the first and third columns of $\tilde{u}_i$, and similarly redefine $\mathbf{V}_i$. Thus the columns of $\mathbf{u}_i$ and $\mathbf{V}_i$ are now the axis $\mathbf{q}$ of the cardiod and a vector, in the plane of the cardiod, perpendicular to $\mathbf{q}$. Then $\hat{c}_2 - \hat{c}_1 = -2.5420$ with a standard error of 0.5585. Thus in this case we can reject that the density has the form (45).

While this might seem curious, elementary calculations show that if the augmented matrix $\widetilde{\mathbf{V}}_i$ has a density of the form (43) for nonuniform $\tilde{f}_0$, then any two of its columns cannot have a density of the form (45). Similarly if $\mathbf{V}_i$ has a nonuniform density of the form (45), then its augmented matrix $\widetilde{\mathbf{V}}_i$ cannot have a density of the form (43) and the first and third columns of $\widetilde{\mathbf{V}}_i$ cannot have a density of the form (45).

More generally, if $\widetilde{\mathbf{V}}_i$ has a matrix Fisher distribution where the eigenvalues of the concentration parameter $\mathbf{K}$ are all nonzero, no two columns of $\widetilde{\mathbf{V}}_i$ will have a matrix Fisher distribution. On the other hand, if distribution of $\widetilde{\mathbf{V}}_i$ satisfies the invariance condition for the group $SO(3) \times SO(3)$, then the distribution of any two columns of $\widetilde{\mathbf{V}}_i$ will satisfy the invariance condition for $SO(2) \times SO(3)$.

Suppose we desire to test if the joint configuration of the McFee and Frank leads is rigid. That is, if $\mathbf{A}_2 = \mathbf{I}_3$. Since the regression model for the 3×2 matrices would, if $\mathbf{A}_2 = \mathbf{I}_3$, imply a more restricted relationship between the two leads, we will apply this test to the augmented 3×3 matrices $(\tilde{\mathbf{u}}_i, \widetilde{\mathbf{V}}_i)$. As discussed above, it would appear to be prudent to assume the underlying density is a general invariant one rather than a density of the form (43), and we continue to use a least squares objective function. For the mechanics of calculation, assuming only the more general invariance condition makes no difference since $m = p$ and hence the constants c_2 and d_2 do not exist. It would have made a difference in the calculations if we had used the original 3×2 matrices.

For this data,

$$\widehat{\mathbf{A}}_1 = \begin{bmatrix} 0.9971 & -0.0490 & 0.0577 \\ 0.0446 & 0.9961 & 0.0761 \\ -0.0612 & -0.0733 & 0.9954 \end{bmatrix},$$

$$\widehat{\mathbf{A}}_2 = \begin{bmatrix} 0.9677 & 0.1560 & 0.1979 \\ -0.1767 & 0.9800 & 0.0916 \\ -0.1796 & -0.1236 & 0.9759 \end{bmatrix},$$

$$\hat{c}_1 = 0.4015, \ \hat{d}_1 = -1.875.$$

A 3×3 skew symmetric matrix $\mathbf{H}$ can always be put in the form $M(\mathbf{h})$ where

$$M(\mathbf{h}) = \begin{bmatrix} 0 & -h_3 & h_2 \\ h_3 & 0 & -h_1 \\ -h_2 & h_1 & 0 \end{bmatrix}$$

and $\mathbf{h} = [h_1\, h_2\, h_3]^T$. Using Proposition 5, and letting, $(\widehat{\mathbf{A}}_1, \widehat{\mathbf{A}}_2) = (\mathbf{A}_1 \exp M(\mathbf{h}), \mathbf{A}_2 \exp M(\mathbf{k}))$, we calculated the joint precision matrix $\mathbf{P}$ of $(\mathbf{h}, \mathbf{k})$ as a 6×6 matrix with diagonal entries 196 and all other entries zero except for the upper right and lower left 3×3 blocks $\mathbf{P}_{21} = \mathbf{P}_{12}^T$ where

$$\mathbf{P}_{12} = \begin{bmatrix} 94.144 & 114.692 & 91.578 \\ 124.492 & -126.469 & 35.235 \\ 84.232 & 48.933 & -142.727 \end{bmatrix}.$$

Write $\widehat{\mathbf{A}}_2 = \exp(M(\mathbf{k}))$. Under the null hypothesis that $\mathbf{A}_2 = \mathbf{I}_3$, we can obtain from $\mathbf{P}$ the asymptotic covariance matrix of $\mathbf{k}$. Since $\mathbf{k}$ is asymptotically multivariate normal with mean $\mathbf{0}$, we obtain an asymptotic χ_3^2-value of 21.56. Thus $\widehat{\mathbf{A}}_2$ is clearly significantly different from the identity. We conclude that the joint configuration of the McFee and Frank leads is not rigid.

Proceeding similarly, we can test if $\mathbf{A}_1 = \mathbf{I}_3$ and it turns out that p-value of this test is 0.25. We conclude that the orientation of the McFee lead is obtained from that of the Frank lead by the rotation $\mathbf{A}_2$; this rotation is estimated to be a rotation of 0.277 radians around the axis $(-0.393, 0.690, -0.608)$ in space coordinates.

9. Some supplementary results and their proofs. We rely on Vinberg (1989) for background theorems and all references refer to Chapter I of Vinberg.

LEMMA 1. *Let $\mathscr{H} = \mathscr{G}_{\boldsymbol{\theta}_0}$ and suppose that $\mathscr{H}$ is compact. Consider the representation of $\mathscr{H}$ on $T_{\boldsymbol{\theta}_0}\boldsymbol{\Theta}$ and suppose there is no $\mathbf{v} \in T_{\boldsymbol{\theta}_0}\boldsymbol{\Theta}$ such that $h \cdot \mathbf{v} = \mathbf{v}$ for all $h \in \mathscr{H}$. Then $\boldsymbol{\theta}_0$ is a critical point for $\tau_{\boldsymbol{\theta}_0}(\boldsymbol{\theta})$.*

PROOF. Let the densities $f(\mathbf{x}; \boldsymbol{\theta})$ be written with respect to the $\mathscr{H}$ invariant measure $\mu(d\mathbf{x})$. We note that even if $\mu(d\mathbf{x})$ is only $\mathscr{G}$ relatively invariant, it will be $\mathscr{H}$ invariant since $\mathscr{H}$ is compact. Let $\gamma(s)$ be a curve in $\boldsymbol{\Theta}$ such that $\gamma(0) = \boldsymbol{\theta}_0$. Then

$$
\begin{aligned}
\tau'_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0)(h \cdot \gamma'(0)) &= \frac{d}{ds}\bigg|_{s=0} \int_{\mathscr{X}} \rho(\mathbf{x}, h \cdot \gamma(s)) f(\mathbf{x}; \boldsymbol{\theta}_0)\mu(d\mathbf{x}) \\
&= \frac{d}{ds}\bigg|_{s=0} \int_{\mathscr{X}} \rho(\mathbf{x}, h \cdot \gamma(s)) f(\mathbf{x}; h \cdot \boldsymbol{\theta}_0)\mu(d\mathbf{x}) \\
&= \frac{d}{ds}\bigg|_{s=0} \int_{\mathscr{X}} \rho(h^{-1} \cdot \mathbf{x}, \gamma(s)) f(h^{-1} \cdot \mathbf{x}; \boldsymbol{\theta}_0)\mu(d\mathbf{x}) \\
&= \frac{d}{ds}\bigg|_{s=0} \int_{\mathscr{X}} \rho(\mathbf{y}, \gamma(s)) f(\mathbf{y}; \boldsymbol{\theta}_0)\mu(d\mathbf{y}) \\
&= \tau'_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0)(\gamma'(0)),
\end{aligned}
$$

where we have made, in the above, the substitution $\mathbf{y} = h^{-1} \cdot \mathbf{x}$. In other words $\tau'_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0)\colon T_{\boldsymbol{\theta}_0}\boldsymbol{\Theta} \to R$ is a linear transformation which satisfies

$$
\tau'_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0)(h \cdot \mathbf{v}) = \tau'_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0)(\mathbf{v}). \tag{46}
$$

Let $\mathscr{V}$ be the kernel of $\tau'_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0)$. It is an invariant subspace and hence if $\mathscr{V} \neq T_{\boldsymbol{\theta}_0}\boldsymbol{\Theta}$, it has an invariant orthogonal complement $\mathscr{W}$. Then $\tau'_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}_0)\colon \mathscr{W} \to R$ is an isomorphism. But (46) implies that for any $\mathbf{v} \in \mathscr{W}$, $h \cdot \mathbf{v} = \mathbf{v}$ for all $h \in \mathscr{H}$. This contradiction establishes the lemma. □

PROOF OF PROPOSITION 1. We note that Proposition 1 is Theorems 8 and 9 of Vinberg for complex representations and complex inner products (that is, $\langle , \rangle$ is positive definite Hermitian sesquilinear – $\langle \mathbf{y}, \mathbf{x} \rangle = \overline{\langle \mathbf{x}, \mathbf{y} \rangle}$ and $\langle \mathbf{x}, \alpha\mathbf{y} \rangle = \bar{\alpha}\langle \mathbf{x}, \mathbf{y} \rangle$ for complex constants $\alpha \in C$). However careful examination of the proofs of Theorems 8 and 9 will reveal that only $\langle , \rangle_0$ need be positive definite; $\langle , \rangle$ can simply be any Hermitian sesquilinear form.

Let $\mathscr{V}_C = C \otimes_R \mathscr{V}$. The real representation of $\mathscr{H}$ on $\mathscr{V}$ extends to a complex representation on $\mathscr{V}_C$ in the obvious manner. By Vinberg, Theorem 6, if $\mathscr{V}_{iC}$ is not irreducible it can be written $\mathscr{V}_{iC} = \mathscr{U} + \bar{\mathscr{U}}$ where $\mathscr{U}$ and $\bar{\mathscr{U}} = \{\mathbf{x} - i\mathbf{y} \mid \mathbf{x} + i\mathbf{y} \in \mathscr{U}\}$ are both complex irreducible. Using this, it is routine to prove the proposition. □

The following proposition follows easily from the invariance of $\mathbf{A}$ and $\mathbf{B}$.

PROPOSITION 3. *Suppose* $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)$ *is a sample from* $f(\mathbf{x}; \boldsymbol{\theta}_0)$ *and that* $\hat{\boldsymbol{\theta}}$ *minimizes* $\rho(\mathbf{X}, \boldsymbol{\theta}) = \sum_i \rho_0(\mathbf{X}_i, \boldsymbol{\theta})$. *Suppose that for some* $\boldsymbol{\theta}_1$ (*not necessarily the same as* $\boldsymbol{\theta}_0$), $\Phi_{\boldsymbol{\theta}_1}\colon T_{\boldsymbol{\theta}_1}\boldsymbol{\Theta} \to \boldsymbol{\Theta}$, *minimal* $\mathscr{G}_{\boldsymbol{\theta}_1}$-*invariant subspaces* $\mathscr{V}_i$ *of* $T_{\boldsymbol{\theta}_1}\boldsymbol{\Theta} = \oplus_i \mathscr{V}_i$, $\mathscr{G}_{\boldsymbol{\theta}_1}$-*invariant inner product* $\langle , \rangle_0$, *and constants* c_i *and* d_i *have been identified as in Proposition* 2.

(a) *Pick* $g \in \mathscr{G}$ *such that* $g \cdot \boldsymbol{\theta}_1 = \boldsymbol{\theta}_0$. *Let* $\hat{\boldsymbol{\theta}} = g \cdot \Phi_{\boldsymbol{\theta}_1}(\hat{\mathbf{h}})$ *and* $\hat{\mathbf{h}} = \sum_{i=1}^{i=r} \hat{\mathbf{h}}_i$ *where* $\hat{\mathbf{h}}_i \in \mathscr{V}_i$. *Then the asymptotic distribution of* $n^{1/2}\hat{\mathbf{h}}$ *is multivariate normal with density proportional to*

$$\exp\left(-\frac{n}{2}\sum_i \frac{d_i^2}{c_i}\langle \mathbf{h}_i, \mathbf{h}_i\rangle_0\right).$$

In particular,

$$n\sum_i \frac{d_i^2}{c_i}\langle \mathbf{h}_i, \mathbf{h}_i\rangle_0$$

is asymptotically $\chi^2(\dim \boldsymbol{\Theta})$.

(b) *Suppose* $\tilde{g} \in \mathscr{G}$ *also satisfies* $\tilde{g} \cdot \boldsymbol{\theta}_1 = \boldsymbol{\theta}_0$. *Let* $\hat{\boldsymbol{\theta}} = \tilde{g} \cdot \Phi_{\boldsymbol{\theta}_1}(\tilde{\mathbf{h}})$. *Then* $\tilde{\mathbf{h}} = (\tilde{g}g^{-1}) \cdot \hat{\mathbf{h}}$ *where* $\tilde{g}g^{-1} \in \mathscr{G}_{\boldsymbol{\theta}_1}$. *Thus*

$$n\sum_i \frac{d_i^2}{c_i}\langle \tilde{\mathbf{h}}_i, \tilde{\mathbf{h}}_i\rangle_0 = n\sum_i \frac{d_i^2}{c_i}\langle \hat{\mathbf{h}}_i, \hat{\mathbf{h}}_i\rangle_0.$$

REFERENCES

BROWN, B. M. (1985). Multiparameter linearization theorems. *J. Roy. Statist. Soc. B* **47** 323–331.

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

CHANG, T. (1988). On reconstructing tectonic plate motion from ship track crossings. *J. Amer. Statist. Assoc.* **83** 1178–1183.

CHANG, T. and KO, D. (1995). M-estimates of rigid body motion on the sphere and in Euclidean space. *Ann. Statist.* **23** 1823–1847.

CHANG, T. and TSAI, M. T. (1999). Asymptotic relative Pitman efficiency in group models. Unpublished manuscript.

CHIKUSE, Y. (1993). High dimensional asymptotic expansions for the matrix Langevin distributions on the Stiefel manifold. *J. Multivariate Anal.* **44** 82–101.

DOWNS, T. D. (1972). Orientation statistics. *Biometrika* **59** 665–676.

DOWNS, T. D., LIEBMAN, J. and MACKAY, W. (1974). Statistical methods for vector cardiogram orientations. In *Proceedings Eleventh International Vectorcardiogram Symposium* 216–222. North-Holland, Amsterdam.

EATON, M. L. (1989). *Group Invariance Arguments in Statistics*. IMS, Hayward, CA and Amer. Statist. Assoc., Alexandria, VA.

FARRELL, R. (1985). *Multivariate Calculation*. Springer, New York.

HSIANG, W. Y. and SU, J. C. (1968). On the classification of transitive actions on Stiefel manifolds. *Trans. Amer. Math. Soc.* **130** 322–336.

KHATRI, C. G. and MARDIA, K. V. (1977). The von Mises–Fisher matrix distribution in orientation statistics. *J. Roy. Statist. Soc. B* **39** 95–106.

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

LEHMANN, E. L. (1983). *Theory of Point Estimation*. Wiley, New York.

MARDIA, K. V. and JUPP, P. E. (2000). *Directional Statistics*. Wiley, New York.

PRENTICE, M. J. (1986). Orientation statistics without parametric assumptions. *J. Roy. Statist. Soc. B* **48** 214–222.

PRENTICE, M. J. (1989). Spherical regression on matched pairs of orientation statistics. *J. Royal Statist. Soc. B* **51** 241–248.

RANCOURT, D., RIVEST, L.-P. and ASSELIN, J. (2000). Using orientation statistics to investigate variations in human kinetics. *Appl. Statist.* **49** 81–94.

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

RIVEST, L. and CHANG, T. (2000). Regression and correlation for 3×3 rotations. Unpublished manuscript.

VINBERG, E. (1989). *Linear Representations of Groups*. Birkhäuser, Basel.

DEPARTMENT OF STATISTICS
UNIVERSITY OF VIRGINIA
CHARLOTTESVILLE, VIRGINIA 22903-4135
E-MAIL: TCC8v@virginia.edu

DÉPARTEMENT DE MATHÉMATIQUES
ET DE STATISTIQUE
UNIVERSITÉ LAVAL
STE-FOY, QUÉBEC
CANADA G1K 7P4