Statistical Science
2004, Vol. 19, No. 4, 624–635
DOI 10.1214/088342304000000567

Spatial Statistics

Ted Chang

Abstract. When the distribution of $\mathbf{X} \in \mathbb{R}^p$ depends only on its distance to some $\theta_0 \in \mathbb{R}^p$, we discuss results from Hössjer and Croux and Neeman and Chang on rank score statistics. Similar results from Neeman and Chang are also given when $\mathbf{X}$ and θ_0 are constrained to lie on the sphere in $\mathbb{R}^p$. Results from Ko and Chang on M estimation for spatial models in Euclidean space and the sphere are also discussed. Finally we discuss a regression type model: the image registration problem. We have landmarks $\mathbf{u}_i$ on one image and corresponding landmarks $\mathbf{V}_i$ on a second image. It is desired to bring the two images into closest coincidence through a translation, rotation and scale change. The techniques and principles of this paper are summarized through extensive discussion of an example in three-dimensional image registration and a comparison of the L_1 and L_2 registrations. Two principles are important when working with spatial statistics: (1) Assumptions, such as that the distribution of $\mathbf{X}$ depends only on its distance to θ_0, introduce symmetries to spatial models which, if properly used, greatly simplify statistical calculations. These symmetries can be expressed in a more general setting by using the notion of statistical group models. (2) When working with a non-Euclidean parameter space Θ such as the sphere, techniques of elementary differential geometry can be used to minimize the distortions caused by using a coordinate system to reexpress Θ in Euclidean parameters.

Key words and phrases: Nonparametric statistics, directional statistics, spherical regression, image registration.

1. INTRODUCTION

Spatial statistics arises when the data are points in some Euclidean space, usually $\mathbb{R}^2$ or $\mathbb{R}^3$, or some surface, usually the unit circle or the unit sphere (which we denote by Ω_2 and Ω_3, resp.). For example, we might have a satellite image of an Arctic ice floe and a data point consists of the location $\mathbf{X}$ of a marker previously placed on the floe. Since the ice floe is on the surface of the Earth, in principle $\mathbf{X} \in \Omega_3$. More conventionally, one notes that the ice floe is unlikely to travel away from the Arctic. Using a "flat Earth" approximation to map the Arctic into the plane, we can think of $\mathbf{X} \in \mathbb{R}^2$.

This example illustrates one important class of applications of spatial statistics: the data consist of the measured locations of points. Especially in the earth sciences, the points are on the surface of the Earth and, hence, spherical statistics are important. However, spherical and, more generally, spatial statistics have arisen in a plethora of contexts (see, e.g., Fisher, Lewis and Embleton, 1987).

Suppose the "true" location of $\mathbf{X} \in \mathbb{R}^2$ is $\theta_0 \in \mathbb{R}^2$. We might wish to assume that

$$\text{the distribution of } \mathbf{X} \text{ depends only on its distance from } \theta_0. \tag{1.1}$$

Letting $f(\mathbf{x}; \theta_0)$ denote the density of $\mathbf{X}$, it can be shown that this condition implies

$$f(\mathbf{C}\mathbf{x}; \mathbf{C}\theta_0) = f(\mathbf{x}; \theta_0) \tag{1.2}$$

for any matrix $\mathbf{C}$ of the form

$$\mathbf{C} = \mathbf{R}(\rho) = \begin{bmatrix} \cos(\rho) & -\sin(\rho) \\ \sin(\rho) & \cos(\rho) \end{bmatrix}.$$

Ted Chang is Professor, Department of Statistics, University of Virginia, Charlottesville, Virginia 22903-31991, USA (e-mail: tcc8v@virginia.edu).

Equation (1.2) is also true if

$$\mathbf{C} = \mathbf{R} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

or, more generally, any matrix $\mathbf{C}$ of the form $\mathbf{R}(\rho)\mathbf{R}$.

A second way to understand the condition (1.2) is to think of $\mathbf{C}$ as a change of basis matrix. Matrices of the form $\mathbf{R}(\rho)$ [resp. $\mathbf{R}(\rho)\mathbf{R}$] are exactly those for which the rows form an orthonormal basis [with the same (resp. opposite) orientation as the original basis]. In this interpretation the condition (1.2) is equivalent to

$$\begin{array}{l}\text{the distribution of } \mathbf{X} \text{ does not depend on} \\ \text{which orthonormal basis is used to write it} \\ \text{as a two-dimensional vector.}\end{array} \tag{1.3}$$

Thus spatial statistics is often very different from conventional multivariate statistics. For example, if the components of $\mathbf{X}$ were to be the height (in inches) and weight (in pounds) of an individual, condition (1.3) would be very unnatural. In particular, it would change meaning if centimeters and kilograms were used instead.

The matrices of the form $\mathbf{R}(\rho)$ form a *group*, denoted by $SO(2)$, in the sense of abstract algebra. Similarly, the matrices of the form $\mathbf{R}(\rho)\mathbf{R}^{\delta}$, $\delta = 0, 1$, form a group, conventionally written as $O(2)$. When condition (1.2) holds, we are said to have a *statistical group model*. Spatial statistical models are often statistical group models and when this occurs, we see that the usual asymptotic calculations are usually greatly simplified.

Suppose now that $\mathbf{X}$ and θ_0 lie in the sphere Ω_3 and that condition (1.1) holds. An example of such a distribution is the *Fisher–von Mises–Langevin* distribution on Ω_3 whose density is

$$f(\mathbf{X}; \theta_0) = c(\kappa) \exp(\kappa \mathbf{X}^T \theta_0), \tag{1.4}$$

where κ is a *concentration parameter* and $c(\kappa)$ is a normalizing constant. The mode of (1.4) is θ_0. For large κ, $\mathbf{X}$ is concentrated close to θ_0 and its distribution approaches a (singular) multivariate normal distribution with covariance matrix $\kappa^{-1}(\mathbf{I}_3 - \theta_0\theta_0^T)$.

Since $\mathbf{X}^T\mathbf{X} = \theta_0^T\theta_0 = 1$, $\mathbf{X}$ and θ_0 have only two independent components. For the purpose of doing the usual asymptotic calculations, we are tempted to rewrite them as two-dimensional vectors. Thus we are finding a map $\Phi : \mathbb{R}^2 \to \Omega_3$ and doing our calculus in $\mathbb{R}^2$. For example, if we calculate in latitude α and longitude β, we are implicitly using the map $\Phi(\alpha, \beta) = [\cos(\pi\alpha/180)\cos(\pi\beta/180)\cos(\pi\alpha/180) \cdot \sin(\pi\beta/180)\sin(\pi\alpha/180)]^T$. We refer to such a Φ as a *coordinate system*.

Unfortunately, it is the bane of map makers that there is no map Φ such that the distances on Ω_3 correspond to distances on $\mathbb{R}^2$. Thus if condition (1.1) holds on Ω_3, it will be destroyed by the map Φ. Our calculations will be complicated by a plethora of terms whose sole mathematical purpose is to undo the distortions introduced by the artifical map Φ.

Mathematicians long ago introduced constructions in elementary differential geometry to solve this problem. As we later see, if these constructions are used, much beautiful and simple structure in the distribution theory for the estimators in spatial statistics becomes manifest. The focus of this paper is not to summarize results in spatial statistics that can be found elsewhere (and are cited below), but rather to give a heuristic understanding of the use of the mathematical tools from differential geometry and group theory. Traditionally the tools of differential geometry are explained at a substantially higher level of detail and abstraction, but we take the attitude that, at the level we need them, they are simple generalizations of the constructions of multivariable calculus. The goal of this paper is that the reader will find the citations more natural and less mysterious.

In this paper, Euclidean p-dimensional space is written as $\mathbb{R}^p$ and the unit sphere in $\mathbb{R}^p$ written as Ω_p. Thus Ω_2 is the circle and Ω_3 is the sphere, which can be used to represent the Earth on which we live. Elements of $\mathbb{R}^p$ and Ω_p are represented as p-dimensional vectors. In the mathematical literature Ω_p is usually written as S^{p-1}.

In general dimensions, a $p \times p$ matrix $\mathbf{C}$ satisfies (1.2) exactly when $\mathbf{C}^T\mathbf{C} = \mathbf{I}_p$, where $\mathbf{I}_p$ is a $p \times p$ identity matrix. Such matrices form a group denoted by $O(p)$. If we add the condition $\det(\mathbf{C}) = 1$, we get the group $SO(p)$. Groups $SO(2)$ and $SO(3)$ represent the rotations in $\mathbb{R}^2$ or $\mathbb{R}^3$, respectively, whereas $O(2)$ and $O(3)$ represent the rotations and reflections.

The S-Plus code used in this paper is posted on the website *www.stat.virginia.edu*.

2. RANK SCORE STATISTICS ON EUCLIDEAN SPACES AND SPHERES

For $X_1, \ldots, X_n \in \mathbb{R}^1$, the *Wilcoxon rank score* statistic is defined by

$$W(\theta) = \sum_i R(|X_i - \theta|) S(X_i - \theta), \tag{2.1}$$

where $R(|X_i - \theta|)$ is the rank of $|X_i - \theta|$ among $|X_1 - \theta|, \dots, |X_n - \theta|$ and $S(X - \theta)$ is the sign of $X - \theta$. If the X_i are i.i.d. with a density $f(x)$ which depends only on $|x - \theta_0|$ for some $\theta_0 \in \mathbb{R}^1$, then:

$$\begin{aligned} &S(X_1 - \theta_0), \dots, S(X_n - \theta_0) \text{ are i.i.d. with} \\ &\Pr(S(X_i - \theta_0) = 1) \\ &\quad = \Pr(S(X_i - \theta_0) = -1) = 0.5. \end{aligned} \tag{2.2}$$

$$\begin{aligned} &\text{All permutations of} \\ &R(|X_1 - \theta_0|), \dots, R(|X_n - \theta_0|) \\ &\text{are equally likely.} \end{aligned} \tag{2.3}$$

$$\begin{aligned} &R(|X_1 - \theta_0|), \dots, R(|X_n - \theta_0|) \\ &\text{are independent of} \\ &S(X_1 - \theta_0), \dots, S(X_n - \theta_0). \end{aligned} \tag{2.4}$$

Properties (2.2)–(2.4) are sufficient to derive the distribution of $W(\theta_0)$; see, for example, Hettmansperger and McKean (1998).

To see how to generalize the Wilcoxon (or more generally rank score statistics) to arbitrary Euclidean spaces, note that $\Omega_1 = \{x \in \mathbb{R}^1 \mid |x| = 1\} = \{\pm 1\}$. For $\mathbf{X}, \theta \in \mathbb{R}^p$ define

$$\mathbf{S}(\mathbf{X} - \theta) = (\mathbf{X} - \theta)/\|\mathbf{X} - \theta\| \in \Omega_p, \tag{2.5}$$

where $\|\mathbf{X} - \theta\| = \sqrt{(\mathbf{X} - \theta)^T(\mathbf{X} - \theta)}$. Under the assumption that $\mathbf{X}_1, \dots, \mathbf{X}_n$ are i.i.d. with a distribution which satisfies (1.1), (2.2) becomes

$$\begin{aligned} &\mathbf{S}(\mathbf{X}_1 - \theta_0), \dots, \mathbf{S}(\mathbf{X}_n - \theta_0) \\ &\quad \text{are i.i.d. uniformly distributed on } \Omega_p. \end{aligned} \tag{2.6}$$

Under the same assumptions, if we let $R(\|\mathbf{X}_i - \theta\|)$ be the rank of $\|\mathbf{X}_i - \theta\|$ among $\|\mathbf{X}_1 - \theta\|, \dots, \|\mathbf{X}_n - \theta\|$, then (2.3) and (2.4) hold with only the most minor change of notation. Under this reinterpretation, the Wilcoxon statistic (2.1), $\mathbf{W}(\theta_0) \in \mathbb{R}^p$, and the properties of its null distribution can be easily derived. For example, we have the following theorem whose proof follows readily from Lemma 2.2:

THEOREM 2.1. *Suppose $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{R}^p$ are i.i.d. with a distribution which satisfies* (1.1). *Then*

$$\mathbf{E}[\mathbf{W}(\theta_0)] = \mathbf{0},$$
$$\mathrm{Cov}[\mathbf{W}(\theta_0)] = \frac{n(n+1)(2n+1)}{6p}\mathbf{I}_p.$$

LEMMA 2.2. *Suppose $\mathbf{S}$ is uniformly distributed on Ω_p. Then $\mathbf{E}[\mathbf{S}] = \mathbf{0}$ and $\mathrm{Cov}[\mathbf{S}] = p^{-1}\mathbf{I}_p$.*

PROOF. Clearly $\mathbf{E}[\mathbf{S}] = \mathbf{0}$ and $\mathrm{Cov}[\mathbf{S}]$ is a multiple of $\mathbf{I}_p$. Now $\mathrm{Tr}\,\mathrm{Cov}[\mathbf{S}] = \mathrm{Tr}\,\mathbf{E}[\mathbf{S}\mathbf{S}^T] = \mathbf{E}[\mathrm{Tr}\,\mathbf{S}\mathbf{S}^T] = \mathbf{E}[\mathbf{S}^T\mathbf{S}] = 1$. □

Additional results on $\mathbf{W}(\theta_0)$ as well as other rank score statistics are given in Neeman (1995), Neeman and Chang (2001) and Hössjer and Croux (1995). Möttönen and Oja (1995) developed a Wilcoxon statistic for Euclidean space models which satisfy (1.2) for the group $\{\pm\mathbf{I}\}$.

For spherical data, we use the following lemma, a formal proof of which can be found, for example, in Watson (1983).

LEMMA 2.3. *Suppose $\mathbf{X}, \theta_0 \in \Omega_p$ and the distribution of $\mathbf{X}$ satisfies* (1.1). *Let $t = \mathbf{X}^T\theta_0$ and*

$$\mathbf{S}(\mathbf{X}; \theta_0) = \frac{\mathbf{X} - t\theta_0}{\sqrt{1 - t^2}}. \tag{2.7}$$

Then t and $\mathbf{S}(\mathbf{X}, \theta_0)$ are independent and $\mathbf{S}(\mathbf{X}, \theta_0)$ is uniformly distributed on $\Omega_{p-1}(\theta_0^\perp) = \{\mathbf{v} \in \mathbb{R}^p \mid \mathbf{v}^T\mathbf{v} = 1, \mathbf{v}^T\theta_0 = 0\}$.

Notice that $\theta_0^\perp = \{\mathbf{x} \mid \mathbf{x}^T\theta_0 = 0\}$ defines a $(p-1)$-dimensional hyperplane of $\mathbb{R}^p$, $\Omega_{p-1}(\theta_0^\perp)$ is the unit sphere in that hyperplane and $\mathbf{X} - t\theta_0$ is the projection of $\mathbf{X}$ onto $\theta_0^\perp$.

Some insight into the geometric reasonableness of Lemma 2.3 can be obtained by considering the Earth, represented as Ω_3. First of all, the spherical distance between $\mathbf{X}$ and θ_0 is $\rho = \arccos(t)$. If we are located at θ_0 we can specify a point $\mathbf{X} \in \Omega_3$ by giving its distance and direction from θ_0. By custom, directions are given in terms of North and East, but this description does not work when θ_0 is the North Pole (and all directions point South) or when θ_0 is the South Pole (and all directions point North). If we were to stand at θ_0 and point in the direction of $\mathbf{X}$, we would be specifying a unit length vector $\mathbf{v}$ which is perpendicular to θ_0, that is, a vector $\mathbf{v} \in \Omega_{p-1}(\theta_0^\perp)$, and $\mathbf{S}(\mathbf{X}; \theta_0)$ is exactly the unit length vector in the direction of $\mathbf{X}$ from θ_0. Indeed

$$\begin{aligned} \mathbf{X} &= \sqrt{1 - t^2}\,\mathbf{S}(\mathbf{X}; \theta_0) + t\theta_0 \\ &= \sin(\rho)\mathbf{S}(\mathbf{X}; \theta_0) + \cos(\rho)\theta_0. \end{aligned} \tag{2.8}$$

By assumption (1.1), the distribution of $\mathbf{X}$ depends only on ρ and all directions are equally likely. This is what Lemma 2.3 says.

The correct definition for the spherical Wilcoxon is now reasonably clear. Given $\mathbf{X}_1, \dots, \mathbf{X}_n \in \Omega_p$, we use $\mathbf{S}(\mathbf{X}_i; \theta_0)$ as the "sign" of $\mathbf{X}_i$ from θ_0. It is tempting to

TABLE 1
Sample spherical Wilcoxon calculations

$\boldsymbol{X}_i$			$\boldsymbol{S}(\boldsymbol{X}_i;\boldsymbol{\theta}_0)$			$\boldsymbol{T}_i$	$\boldsymbol{R}_i$
0.3993	0.8968	−0.1908	0.3614	0.3891	0.8474	0.9434	13
0.2506	0.8197	−0.5150	−0.9356	0.1420	−0.3231	0.9988	1
0.1906	0.7646	−0.6157	−0.6209	−0.2337	−0.7482	0.9865	6
0.1869	0.8792	−0.4384	−0.7536	0.5208	0.4012	0.9900	3
0.0646	0.5260	−0.8480	−0.3941	−0.3727	−0.8401	0.8712	15
−0.3406	0.5451	−0.7660	−0.8076	−0.0661	−0.5860	0.7257	16
0.8794	0.4101	−0.2419	0.9551	−0.2460	0.1655	0.7151	17
0.6012	0.5413	−0.5878	0.8090	−0.4920	−0.3215	0.9125	14
0.5091	0.6287	−0.5878	0.7729	−0.5118	−0.3751	0.9563	11
0.5624	0.7199	−0.4067	0.9500	−0.1971	0.2420	0.9558	12
0.4636	0.6873	−0.5592	0.8049	−0.4944	−0.3280	0.9762	9
0.3971	0.7164	−0.5736	0.6629	−0.5520	−0.5058	0.9875	5
0.4238	0.8318	−0.3584	0.6981	0.1729	0.6949	0.9817	8
0.4147	0.8140	−0.4067	0.8105	0.0627	0.5823	0.9886	4
0.3676	0.7538	−0.5446	0.7071	−0.5388	−0.4580	0.9947	2
0.1355	0.8554	−0.5000	−0.9468	0.3190	−0.0416	0.9862	7
0.1144	0.7223	−0.6820	−0.6314	−0.2260	−0.7418	0.9627	10

rank the spherical distances $\arccos(\mathbf{X}_i^T\theta_0)$. However, if $\rho = \pi$ in (2.8), $\mathbf{X} = -\theta_0$ for any choice of $\mathbf{S}(\mathbf{X};\theta_0)$. It follows that if $\mathbf{X}_i$ is close to $-\theta_0$ and we assign a high rank to such an $\mathbf{X}_i$, the corresponding Wilcoxon will be unstable with small changes in $\mathbf{X}_i$. For this reason, Neeman (1995) and Neeman and Chang (2001) defined the spherical Wilcoxon for $\theta \in \Omega_p$ as

$$\mathbf{W}_S(\theta) = \sum_i R\bigl(\arccos(|\mathbf{X}_i^T\theta|)\bigr)\mathbf{S}(\mathbf{X}_i;\theta) \tag{2.9}$$

and we note that $\mathbf{W}_S(\theta) \in \theta^\perp$. In essence they ranked a point $\mathbf{X}_i$ by its distance to the closer of θ_0 and $-\theta_0$. We now have the following theorem:

THEOREM 2.4. *Suppose* $\mathbf{X}_1, \dots, \mathbf{X}_n \in \Omega^p$ *are i.i.d. with a distribution which satisfies* (1.1). *Then*

$$\mathbf{E}[\mathbf{W}_S(\theta_0)] = \mathbf{0},$$
$$\mathrm{Cov}[\mathbf{W}_S(\theta_0)] = \frac{n(n+1)(2n+1)}{6(p-1)}[\mathbf{I}_p - \theta_0\theta_0^T].$$

For example, Fisher, Lewis and Embleton (1987) gave 17 measurements of magnetic remanence from specimens collected from the Tumblagooda Sandstone in Western Australia. The data, converted into Euclidean coordinates, are given in Table 1. We test if $\theta_0 = (0.2962, 0.8138, -0.5000)$, which is the Euclidean coordinates for a declination of 70° and an inclination of −30° (θ_0 has been arbitrarily chosen for illustrative purposes only).

Table 1 also gives the values of the sign $\mathbf{S}(\mathbf{X}_i;\theta_0)$, $T_i = \mathbf{X}_i^T\theta_0$ and the corresponding ranks R_i. Notice we rank $\arccos(|T_i|)$, not T_i. Summing, $\mathbf{W}_S = \sum_i R_i\mathbf{S}(\mathbf{X}_i;\theta_0) = (34.2676, -27.0342, -23.7008)$. Using Theorem 2.4, $\frac{10710}{12}\|\mathbf{W}_S\|^2 = 2.764$ should be compared to a χ_2^2 distribution, and we fail to reject the null hypothesis at $\alpha = 0.05$.

The set of θ which are not rejected can be used to produce a 95% confidence region. The resulting region, together with the data, is shown in Figure 1.

We refer the reader to Neeman (1995) and Neeman and Chang (2001) for further discussion of rank score statistics for spherical data. The example is discussed more exhaustively in Neeman and Chang (2001).

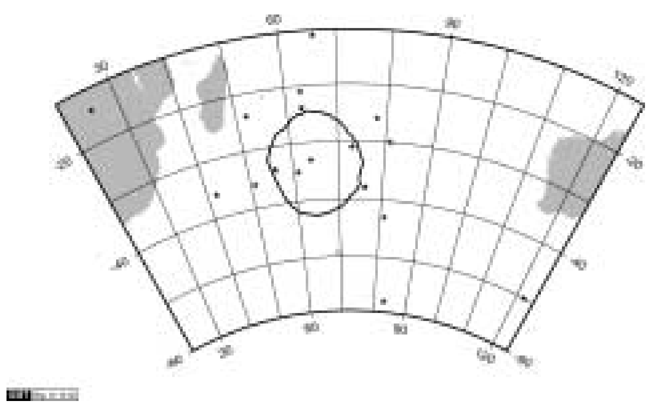

FIG. 1. *The* 95% *confidence region for modal direction.*

3. M ESTIMATION IN STATISTICAL MANIFOLDS

From the viewpoint of a directional data analyst, when spheres have been conquered it is natural to look at other surfaces (or *manifolds*) in $\mathbb{R}^p$. Downs (1972), Rancourt, Rivest and Asselin (2000) and Jupp and Mardia (1989) gave some examples of this type of data.

The precise mathematical definition of a manifold is somewhat abstruse. Somewhat loosely, a d-dimensional manifold $\mathcal{M}$ is a subset of $\mathbb{R}^p$, for some p, which can be written as a union of open sets $\mathcal{U}_i$ such that each $\mathcal{U}_i$ has a 1–1 bicontinuous map $\phi_i : \mathcal{U}_i \to \mathbb{R}^d$ (i.e., ϕ_i and ϕ_i^{-1} are continuous). For example, we can map any portion of the Earth bicontinuously onto $\mathbb{R}^2$ and this implies that Ω_3 is a two-dimensional manifold in $\mathbb{R}^2$. Notice, however, that any map of the entire Earth must cut the Earth somewhere (such as at the North and South Poles and the International Date Line for the most common projection), and hence at least two $\mathcal{U}_i$'s are needed.

Lemma 2.3 is the basis for the definition of the spherical Wilcoxon and other rank score statistics on Ω_p. For reasons given in Chang and Tsai (2003), I believe (but have not proven) that Lemma 2.3 cannot be generalized to more than a very small collection of manifolds, principally Euclidean spaces and spheres. For other manifolds, fully nonparametric inference may not be possible and M estimation offers a useful alternative.

So suppose $\mathbf{X}_1, \dots, \mathbf{X}_n$ in the sample space $\mathcal{X}$ are i.i.d. with a density $f(\mathbf{x}; \theta)$ for some θ in the manifold Θ. Given an objective function $\rho_0(\mathbf{x}; \theta)$, the M *estimate* $\widehat{\theta}$ minimizes the objective function

$$\rho(\theta) = \sum_i \rho_0(\mathbf{X}_i; \theta). \tag{3.1}$$

For example, suppose $\mathcal{X} = \Theta = \Omega_p$. Then we could use the sum of the spherical (great circle) distances $\rho_0(\mathbf{x}; \theta) = \arccos(\mathbf{x}^T\theta)$. The resulting $\widehat{\theta}$ is called the *spherical median* and was introduced by Fisher (1985). Alternatively, we could use the sum of the linear (through the Earth) distances $\rho_0(\mathbf{x}; \theta) = \|\mathbf{x} - \theta\| = \sqrt{2 - 2\mathbf{x}^T\theta}$, which yields the so-called *normalized spatial median*. Another common choice is the L_2 estimator, which uses $\rho_0(\mathbf{x}; \theta) = \|\mathbf{x} - \theta\|^2$. In this case, $\widehat{\theta} = \overline{\mathbf{X}}/\|\overline{\mathbf{X}}\|$ is the *spherical mean*.

For M estimators when $\Theta = \mathbb{R}^p$ and under regularity conditions, Brown (1985) showed that $\sqrt{n}(\widehat{\theta} - \theta_0)$ is asymptotically multivariate normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{B}(\theta_0)^{-1}\mathbf{A}(\theta_0)\mathbf{B}(\theta_0)^{-1}$, where the matrices are defined by

$$\mathbf{A}(\theta_0)_{ij} = \mathrm{Cov}_{\theta_0}\left[\frac{\partial}{\partial\theta^i}\bigg|_{\theta=\theta_0} \rho(\mathbf{X}; \theta) \cdot \frac{\partial}{\partial\theta^j}\bigg|_{\theta=\theta_0} \rho(\mathbf{X}; \theta)\right], \tag{3.2}$$

$$\mathbf{B}(\theta_0)_{ij} = E_{\theta_0}\left[\frac{\partial}{\partial\theta^i}\bigg|_{\theta=\theta_0} \rho(\mathbf{X}; \theta) \cdot \frac{\partial}{\partial\theta^j}\bigg|_{\theta=\theta_0} \log(f(\mathbf{X}; \theta))\right], \tag{3.3}$$

where $\theta = (\theta^1, \dots, \theta^p)$. [There are several forms for $\mathbf{B}$ which are all equivalent when $E_{\theta_0}\rho(\mathbf{X}; \theta)$ has a critical point at $\theta = \theta_0$. Equation (3.3) is a slight generalization of the third form given by Hettmansperger and McKean (1998), equation (6.1.2). Under the same condition, $\mathbf{B}$ is symmetric.]

Our reformulation $\mathbf{A}$ and $\mathbf{B}$ for manifolds is similar to previous reformulations of Fisher information. These reformulations require the notion of a *tangent vector* to the manifold.

Equation (2.8) can be used to understand how tangent vectors can be defined for general manifolds (at least those embedded in some Euclidean space). Suppose we fix $\mathbf{v} = \mathbf{S}(\mathbf{X}; \theta_0)$ and let $s = \rho$ vary in (2.8). We get a curve

$$\gamma(s) = \sin(s)\mathbf{v} + \cos(s)\theta_0 \tag{3.4}$$

which satisfies $\gamma(s) \in \Omega_p \subseteq \mathbb{R}^3$ for all s, $\gamma(0) = \theta_0$ and $\gamma'(0) = \mathbf{v}$. Here, $\gamma'(0)$ is the derivative of γ as a map $\mathbb{R}^1 \to \mathbb{R}^3$. So for a manifold Θ and a point $\theta_0 \in \Theta$, we define the *tangent space* to Θ at θ_0 to be $\{\gamma'(0)\}$, where $\gamma(s)$ is a curve in $\Theta \subseteq \mathbb{R}^p$ [so that $\gamma(s) \in \Theta$ for all s] with $\gamma(0) = \theta_0$.

Here is a simple lemma which establishes that for spheres Ω_p, the new definition (in terms of derivatives of curves) of a tangent vector at $\theta_0 \in \Omega_p$ coincides with the old definition (a vector $\mathbf{v}$ such that $\mathbf{v}^T\theta_0 = 0$).

LEMMA 3.1. *Let $\gamma(s)$ be a curve in Ω_p. Then $\gamma'(0)^T\gamma(0) = 0$. Conversely, if $\mathbf{v} \in \theta_0^\perp$, then $\mathbf{v} = \gamma'(0)$ for some curve $\gamma(s)$ in Ω_p with $\gamma(0) = \theta_0$.*

PROOF. Since $\gamma(s) \in \Omega_p$ for all s, $1 = \gamma(s)^T\gamma(s)$. Therefore, $0 = \frac{d}{ds}\gamma(s)^T\gamma(s) = \gamma'(s)^T\gamma(s) + \gamma(s)^T \cdot \gamma'(s) = 2\gamma'(s)^T\gamma(s)$, where the last equality follows since both $\gamma(s)$ and $\gamma'(s)$ are column vectors. The converse assertion is established by (3.4). $\square$

To reformulate Fisher information suppose, temporarily, $\Theta = \mathbb{R}^p$. The Fisher information matrix $I(\theta_0)$

is the matrix whose (i, j)th entry is

$$I(\theta_0)_{ij} = E_{\theta_0}\left[\frac{\partial}{\partial\theta^i}\bigg|_{\theta=\theta_0} \log(f(\mathbf{X};\theta)) \cdot \frac{\partial}{\partial\theta^j}\bigg|_{\theta=\theta_0} \log(f(\mathbf{X};\theta))\right]. \tag{3.5}$$

Notice that if γ_1, γ_2 are curves in $\mathbb{R}^p$ with $\gamma_1(0) = \gamma_2(0) = \theta_0$, then

$$\begin{aligned}(\gamma_1'(0))^T I(\theta_0)(\gamma_2'(0)) &= E_{\theta_0}\left[\frac{d}{ds}\bigg|_{s=0} \log\bigl(f(\mathbf{X}, \gamma_1(s))\bigr) \cdot \frac{d}{dt}\bigg|_{t=0} \log\bigl(f(\mathbf{X}, \gamma_2(t))\bigr)\right],\end{aligned} \tag{3.6}$$

where the left-hand side of (3.6) is matrix multiplication. Equation (3.6) follows from (3.5) and the chain rule

$$\frac{d}{dt}\bigg|_{t=0} g(\gamma(t)) = \sum_{i=1}^{p} \frac{\partial}{\partial\theta^i}\bigg|_{\theta=\theta_0} g(\theta) v_i,$$

where $g:\mathbb{R}^p \to \mathbb{R}^1$ and $\gamma(t)$ is any curve in $\mathbb{R}^p$ with $\gamma(0) = \theta_0$ and $\gamma'(0) = (v_1, \dots, v_p)$.

Notice that the right-hand side of (3.6) makes sense for any manifold. Thus we can define, for general Θ and $\theta_0 \in \Theta$, Fisher information at θ_0 to be an inner product, defined on the tangent space to Θ at θ_0 by

$$\langle \mathbf{v}, \mathbf{w}\rangle_{I(\theta_0)} = E_{\theta_0}\left[\frac{d}{ds}\bigg|_{s=0} \log\bigl(f(\mathbf{X}, \gamma_1(s))\bigr) \cdot \frac{d}{dt}\bigg|_{t=0} \log\bigl(f(\mathbf{X}, \gamma_2(t))\bigr)\right] \tag{3.7}$$

for any curves γ_1, γ_2 in Θ with $\gamma_1(0) = \gamma_2(0) = \theta_0$ and $\gamma_1'(0) = \mathbf{v}$, $\gamma_2'(0) = \mathbf{w}$. The notation on the left-hand side of (3.7) is designed to emphasize that we are thinking of Fisher information as a *Riemannian metric* on Θ, that is, a family of inner products, one inner product on each tangent space of Θ.

This approach to Fisher information has long been known. It was used, for example, by Reeds (1975) to explain the nonexistence of variance stabilizing transformations for multivariate parameters.

Similarly we redefine $\mathbf{A}$ and $-\mathbf{B}$ as Riemannian metrics (the conventional definition of $\mathbf{B}$ makes $\mathbf{B}$ negative definite on each tangent space) using

$$\langle \mathbf{v}, \mathbf{w}\rangle_{A(\theta_0)} = \mathrm{Cov}_{\theta_0}\left[\frac{d}{ds}\bigg|_{s=0} \rho(\mathbf{X}, \gamma_1(s)) \cdot \frac{d}{dt}\bigg|_{t=0} \rho(\mathbf{X}, \gamma_2(t))\right] \tag{3.8}$$

and

$$\langle \mathbf{v}, \mathbf{w}\rangle_{B(\theta_0)} = E_{\theta_0}\left[\frac{d}{ds}\bigg|_{s=0} \rho(\mathbf{X}, \gamma_1(s)) \cdot \frac{d}{dt}\bigg|_{t=0} \log\bigl(f(\mathbf{X}, \gamma_2(t))\bigr)\right]. \tag{3.9}$$

The advantages of using (3.7)–(3.9) instead of their matrix formulations is that they allow us to compute them without choosing any particular coordinate system. We then apply these calculations using a convenient coordinate system, one that generally depends on θ_0. This is our approach for avoiding the map maker's dilemna discussed in the Introduction. We illustrate this using the sphere Ω_p in the following theorem.

THEOREM 3.2. *Suppose $\mathbf{X}_1, \dots, \mathbf{X}_n \in \Omega_p$ are i.i.d. with a density $f(\mathbf{x}; \theta_0)$, for $\theta_0 \in \Omega_p$, which satifies $\log(f(\mathbf{x};\theta)) = g(\mathbf{x}^T\theta)$. Suppose $\widehat{\theta}$ is the M estimator which minimizes an objective function of the form* (3.1), *where $\rho_0(\mathbf{x};\theta) = \tilde{\rho}_0(\mathbf{x}^T\theta)$. Let $\psi(t) = -\tilde{\rho}_0'(t)$. Then*

$$\langle \mathbf{v}, \mathbf{w}\rangle_{I(\theta_0)} = \frac{E[g'(t)^2(1-t^2)]}{p-1}\mathbf{v}^T\mathbf{w},$$

$$\langle \mathbf{v}, \mathbf{w}\rangle_{A(\theta_0)} = \frac{E[\psi(t)^2(1-t^2)]}{p-1}\mathbf{v}^T\mathbf{w}$$

and

$$\langle \mathbf{v}, \mathbf{w}\rangle_{B(\theta_0)} = -\frac{E[\psi(t)g'(t)(1-t^2)]}{p-1}\mathbf{v}^T\mathbf{w}.$$

Let $\Phi:\theta_0^\perp \to \Omega_p$ be defined by $\Phi(s\mathbf{v}) = \sin(s)\mathbf{v} + \cos(s)\theta_0$, where $\mathbf{v} \in \theta_0^\perp$ has unit length. Let $\widehat{\mathbf{h}} \in \theta_0^\perp$ be defined by $\Phi(\widehat{\mathbf{h}}) = \widehat{\theta}$. Then $\sqrt{n}\,\widehat{\mathbf{h}}$ is asymptotically singular multivariate normal,

$$N_p\left(\mathbf{0}, \frac{(p-1)E[\psi(t)^2(1-t^2)]}{E^2[\psi(t)g'(t)(1-t^2)]}(\mathbf{I}_p - \theta_0\theta_0^T)\right).$$

PROOF. Let $\gamma(s)$ be a curve in Ω_p with $\gamma(0) = \theta_0$ and $\gamma'(0) = \mathbf{v}$. Then

$$\begin{aligned}\frac{d}{ds}\bigg|_{s=0} &\log\bigl(f(\mathbf{X};\gamma(s))\bigr) \\ &= \frac{d}{ds}\bigg|_{s=0} g(\mathbf{X}^T\gamma(s)) \\ &= g'(\mathbf{X}^T\gamma(0))\mathbf{X}^T\gamma'(0) \\ &= g'(t)\mathbf{v}^T\bigl(\mathbf{X} - (\mathbf{X}^T\theta_0)\theta_0\bigr),\end{aligned}$$

where we have used above that $\mathbf{v} \in \theta_0^\perp$. Therefore, using Lemma 2.3 and then Lemma 2.2,

$$\begin{aligned}
&\langle \mathbf{v}, \mathbf{w}\rangle_{I(\theta_0)} \\
&= \mathbf{v}^T E\biggl[\bigl(g'(t)^2(1-t^2)\bigr)\biggl(\frac{\mathbf{X}-(\mathbf{X}^T\theta_0)\theta_0}{\sqrt{1-t^2}}\biggr) \\
&\qquad\qquad \cdot \biggl(\frac{\mathbf{X}-(\mathbf{X}^T\theta_0)\theta_0}{\sqrt{1-t^2}}\biggr)^T\biggr]\mathbf{w} \\
&= E[g'(t)^2(1-t^2)]\mathbf{v}^T\frac{\mathbf{I}_p-\theta_0\theta_0^T}{p-1}\mathbf{w} \\
&= \frac{E[g'(t)^2(1-t^2)]}{p-1}\mathbf{v}^T\mathbf{w}.
\end{aligned}$$

The calculations of **A** and **B** follow similarly. Now

$$\Phi(\mathbf{0}) = \theta_0, \quad \frac{d}{ds}\bigg|_{s=0}\Phi(s\mathbf{v}) = \mathbf{v} \tag{3.10}$$

for all $\mathbf{v} \in \theta_0^\perp$. Write $\theta = \Phi(\mathbf{h})$ for $\mathbf{h} \in \theta_0^\perp$, $f^\Phi(\mathbf{x};\mathbf{h}) = f(\mathbf{x};\Phi(\mathbf{h}))$ and $\rho_0^\Phi(\mathbf{x},\mathbf{h}) = \rho_0(\mathbf{x};\Phi(\mathbf{h}))$. Then for $\mathbf{v},\mathbf{w} \in \theta_0^\perp$,

$$\begin{aligned}
&\langle \mathbf{v}, \mathbf{w}\rangle_{A^\Phi(\mathbf{0})} \\
&= \mathrm{Cov}_{\mathbf{0}}^{\Phi}\biggl[\frac{d}{ds}\bigg|_{s=0}\rho^\Phi(\mathbf{X}, s\mathbf{v})\frac{d}{dt}\bigg|_{t=0}\rho^\Phi(\mathbf{X}, t\mathbf{w})\biggr] \\
&= \mathrm{Cov}_{\theta_0}\biggl[\frac{d}{ds}\bigg|_{s=0}\rho(\mathbf{X},\Phi(s\mathbf{v}))\frac{d}{dt}\bigg|_{t=0}\rho(\mathbf{X},\Phi(t\mathbf{w}))\biggr] \\
&= \langle \mathbf{v}, \mathbf{w}\rangle_{A(\theta_0)},
\end{aligned}$$

where the last equality follows from (3.10). The remainder of the theorem follows by applying Brown's (1985) results to f^Φ and ρ_0^Φ as a density and objective function on the Euclidean space $\theta_0^\perp$. $\square$

Several points about Theorem 3.2 should be made. First, $\widehat{\mathbf{h}}$ measures the deviation of the estimate $\widehat{\theta}$ from θ_0. Letting $\mathbf{v} = \widehat{\mathbf{h}}/\|\widehat{\mathbf{h}}\|$ and recalling the discussion of Equation (3.4), the curve $\gamma(s) = \Phi(s\mathbf{v})$ describes a great circle connecting θ_0 to $\gamma(\|\widehat{\mathbf{h}}\|) = \Phi(\widehat{\mathbf{h}}) = \widehat{\theta}$. Thus $\|\widehat{\mathbf{h}}\|$ is the spherical distance of $\widehat{\theta}$ from θ_0 and $\widehat{\mathbf{h}}/\|\widehat{\mathbf{h}}\|$ represents the direction of the deviation of $\widehat{\theta}$ from θ_0. In Euclidean space, we represent the deviation of $\widehat{\theta}$ from θ_0 by the Euclidean vector $\widehat{\theta} - \theta_0$; for manifolds, we need to use a tangent vector $\widehat{\mathbf{h}}$ at θ_0.

Second, instead of focusing on the distribution of $\widehat{\theta}$, we focus on the distribution of the *deviation* (as measured by $\widehat{\mathbf{h}}$) of $\widehat{\theta}$ from θ_0. This is sufficient for statistical inference. In the first problem of elementary statistics, it is analogous to constructing confidence intervals by using an estimate $\overline{X}$ and noting that the distribution of $\overline{X} - \mu$ is $N(0, \sigma^2/n)$. It is also consistent with the approach used in nonlinear regression.

Finally, although no distance preserving coordinate system is possible everywhere on Ω_p, the map Φ is, to first order, nondistorting in the region of interest, namely the θ close to θ_0. The precise required conditions on Φ are given by (3.10). The map Φ is essentially a "polar projection" centered at θ_0. This is our solution to the map maker's dilemma: To map Australia, a polar projection centered at the North Pole would be highly deceptive, whereas centering the projection at Ayer's Rock would be excellent. The coordinate system to be used depends on the area to be mapped.

The map Φ approximates Ω_p close to θ_0 by its tangent space $\theta_0^\perp$. This is quite natural: Calculus teaches us to linearly approximate functions using tangent lines.

Kirkwood, Royer, Chang and Gordon (1999) gave an example (from paleomagnetism) in which it is shown, by simulation, that the curvature effects inherent in nonlinear regression are lower when the coordinate system Φ is used on Ω_3 instead of latitude and longitude. Chang (1993) gave an example from plate tectonics in which $\Theta = SO(3)$, the group of rotations of $\mathbb{R}^3$ [note that the position of a rigid body moving on Ω_3 relative to its past position at fixed time is given by an element of $SO(3)$]. It is shown there that if a coordinate system Ψ on $SO(3)$ which depends on θ_0 (in a similar way to dependence of Φ) is used, vastly simpler formulas result than when a popular fixed coordinate system for $SO(3)$ is used.

The constant $(p-1)^{-1}E[\psi(t)^2(1-t^2)]$, and hence **A**, can be estimated from a sample by $(np - n)^{-1}\sum_i \psi(\widehat{t_i})^2(1-\widehat{t}_i^2)$, where $\widehat{t_i} = \mathbf{X}_i^T\widehat{\theta}$. Notice that this estimate does not require choice of a particular form for f. The following proposition from Chang (1986) allows for "nonparametric" estimation of **B**.

PROPOSITION 3.3. *We have*

$$\begin{aligned}
&E[\psi(t)g'(t)(1-t^2)] \\
&\quad = (p-1)E[\psi(t)t] - E[\psi'(t)(1-t^2)].
\end{aligned}$$

Recall that the *asymptotic relative efficiency* (ARE) of two estimators $\widehat{\theta}_1$ and $\widehat{\theta}_2$ is

$$\mathrm{ARE}(\widehat{\theta}_1, \widehat{\theta}_2) = \lim_{n\to\infty} \mathrm{Cov}(\widehat{\theta}_2)/\mathrm{Cov}(\widehat{\theta}_1) \tag{3.11}$$

whenever the right-hand side of (3.11) makes sense. For example, under the assumptions of Theorem 3.2, if $\mathbf{A}_i$ and $\mathbf{B}_i$ are the **A** and **B** matrices of $\widehat{\theta}_i$, then (in any coordinate system of Ω_p) $\mathbf{A}_1, \mathbf{A}_2, \mathbf{B}_1$ and $\mathbf{B}_2$ are all multiples of each other. It follows that $\mathrm{Cov}(\widehat{\theta}_2)$ is a multiple of $\mathrm{Cov}(\widehat{\theta}_1)$ and the following corollary holds.

COROLLARY 3.4. *Suppose $\mathbf{X}_1, \dots, \mathbf{X}_n \in \Omega_p$ are i.i.d. and the assumptions of Theorem* 3.2 *hold. Let $\widehat{\theta}_1$ and $\widehat{\theta}_2$ be M estimators with ψ functions ψ_1 and ψ_2, respectively. Then*

$$\mathrm{ARE}(\widehat{\theta}_1, \widehat{\theta}_2) = \frac{E^2[\psi_1(t)g'(t)(1-t^2)]}{E[\psi_1(t)^2(1-t^2)]} \cdot \frac{E[\psi_2(t)^2(1-t^2)]}{E^2[\psi_2(t)g'(t)(1-t^2)]}.$$

For example, the maximum likelihood estimator for the Fisher–von Mises–Langevin distribution (1.4) is the spherical mean [$\psi(t) = t$]. Comparing it to the spherical median [$\psi(t) = 1/\sqrt{1-t^2}$] and using (2.2.2) from Watson (1983), we get

$$\begin{aligned} &\mathrm{ARE}(\text{spherical median, spherical mean}) \\ &\quad = \frac{E^2[\sqrt{1-t^2}]}{E[1-t^2]} \\ &\quad = \left[\int_{-1}^{1} e^{\kappa t}(1-t^2)^{(p-2)/2}\,dt\right]^2 \\ &\qquad \cdot \left(\left[\int_{-1}^{1} e^{\kappa t}(1-t^2)^{(p-1)/2}\,dt\right]\right. \\ &\qquad \left. \cdot \left[\int_{-1}^{1} e^{\kappa t}(1-t^2)^{(p-3)/2}\,dt\right]\right)^{-1}. \end{aligned} \tag{3.12}$$

When $p = 3$, (3.12) is due to Fisher (1985) and, in this case, (3.12) is a decreasing function of κ and approaches $\pi/4$ as $\kappa \to \infty$. For general p,

$$\begin{aligned} &\lim_{\kappa\to\infty} \mathrm{ARE}(\text{spherical median, spherical mean}) \\ &\quad = \frac{2\Gamma^2(p/2)}{(p-1)\Gamma^2((p-1)/2)}. \end{aligned} \tag{3.13}$$

These results on the asymptotic distribution of M estimators on Ω_p were derived by Ko and Chang (1993). For Euclidean space data we have similar results (see Chang and Ko, 1995):

THEOREM 3.5. *Suppose $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{R}^p$ are i.i.d. with a density $f(\mathbf{x}; \theta_0)$ which satisfies condition* (1.1). *Let $\widehat{\theta}$ be the M estimator which minimizes an objective function of the form* (3.1), *where $\rho_0(\mathbf{x}, \theta_0)$ depends only on $\|\mathbf{x} - \theta_0\|$. Then*

$$\langle \mathbf{v}, \mathbf{w}\rangle_{I(\theta_0)} = \frac{4E[g'(s)^2 s]}{p}\mathbf{v}^T\mathbf{w},$$

$$\langle \mathbf{v}, \mathbf{w}\rangle_{A(\theta_0)} = \frac{4E[\psi(s)^2 s]}{p}\mathbf{v}^T\mathbf{w}$$

and

$$\langle \mathbf{v}, \mathbf{w}\rangle_{B(\theta_0)} = \frac{4E[\psi(s)g'(s)s]}{p}\mathbf{v}^T\mathbf{w},$$

where we write $\log(f(\mathbf{X}, \theta_0)) = g(s)$, $\rho_0(\mathbf{X}, \theta_0) = \tilde{\rho}_0(s)$ and $\psi(s) = \tilde{\rho}_0'(s)$ with $s = \|\mathbf{X} - \theta_0\|^2$. Therefore, $\sqrt{n}(\widehat{\theta} - \theta_0)$ is asymptotically distributed $N_p(\mathbf{0}, k\mathbf{I}_p)$, where

$$k = \frac{pE[\psi(s)^2 s]}{4E^2[\psi(s)g'(s)s]}.$$

PROPOSITION 3.6. *We have*

$$E[\psi(s)g'(s)s] = -E[\psi'(s)s + (p/2)\psi(s)].$$

COROLLARY 3.7. *Suppose $\mathbf{X}_1, \dots, \mathbf{X}_n \in \mathbb{R}^p$ are i.i.d. and the assumptions of Theorem* 3.5 *hold. Let $\widehat{\theta}_1$ and $\widehat{\theta}_2$ be M estimators with ψ functions ψ_1 and ψ_2, respectively. Then*

$$\mathrm{ARE}(\widehat{\theta}_1, \widehat{\theta}_2) = \frac{E^2[\psi_1(s)g'(s)s]}{E[\psi_1(s)^2 s]}\frac{E[\psi_2(s)^2 s]}{E^2[\psi_1(s)g'(s)s]}.$$

The *spatial median*, introduced by Brown (1983), minimizes $\sum_i \|\mathbf{X}_i - \theta\|$. He showed that if the $\mathbf{X}_i$ are distributed $N_p(\mathbf{0}, \sigma^2\mathbf{I}_p)$, then

$$\mathrm{ARE}(\text{spatial median}, \overline{\mathbf{X}}) = \frac{2\Gamma^2((p+1)/2)}{p\Gamma^2(p/2)}, \tag{3.14}$$

a result which also follows from Corollary 3.7.

Notice that the right-hand sides of (3.13) and (3.14) become identical if p is replaced by $p-1$ in (3.14). This is hardly surprising since if $\mathbf{X} \in \Omega_p$ has the distribution (1.4), then as $\kappa \to \infty$, $\mathbf{X}$ becomes increasingly close to θ_0 and hence the curvature of Ω_p becomes irrelevant. We have also noted that the distribution of $\mathbf{X}$ approaches a singular normal distribution $N_p(\theta_0, \kappa^{-1}(\mathbf{I}_p - \theta_0\theta_0^T))$.

Furthermore (3.14) increases with p. Thus, for example, if $p = 3$, the use of the spatial median protects against long tailed distributions at the modest cost (in standard error terms) of $(3\pi/8)^{1/2} - 1 = 8\%$ when, in fact, $\overline{\mathbf{X}}$ would have been optimal.

4. IMAGE REGISTRATION IN EUCLIDEAN SPACE

Suppose we have fixed $\mathbf{u}_i \in \mathbb{R}^p$ and independent random $\mathbf{V}_i \in \mathbb{R}^p$, such that the distribution of $\mathbf{V}_i$ depends only on $\|\mathbf{V}_i - \gamma_0\mathbf{C}_0\mathbf{u}_i - \mathbf{b}_0\|$ for some $\mathbf{C}_0 \in SO(p)$, $\mathbf{b}_0 \in \mathbb{R}^p$ and positive real number γ_0. The number γ_0 is interpreted as a scale change. Although the results outlined below can be generalized to arbitrary dimensions, for the sake of simplicity, we restrict attention to the cases of greatest practical interest $p = 2, 3$.

This model arises in the problem of *image registration*. We have two images (usually at different times) of the same object. The $\mathbf{u}_i$ represent the position of some landmarks on one image and the $\mathbf{V}_i$ represent the position of the same landmarks on the second image. We are interested in estimating the best rotation $\mathbf{C}$ and translation $\mathbf{b}$ to bring the images into alignment.

Alternatively, we might have a prototypical shape, say of a kidney. We have an image of a kidney and we want to find the rotation and translation to bring the imaged kidney into closest alignment with the prototypical kidney shape. This might be a prelude to automatic processing of a large number of kidney images.

Chang and Ko (1995) provided a data set that consists of the digitized locations of 12 pairs of landmarks on the left and right hands of one of the authors. These data are replicated in Table 2. Because of the opposite orientation of the left and right hands, one of the hands must first be reflected (in any plane) before finding the rotation, translation and scale change to best bring the two hands into alignment. This is essentially a three-dimensional image registration.

Let $\mathbf{\Sigma}_{\tilde{V}}$ be the matrix $n^{-1}\sum_i(\tilde{\mathbf{V}}_i - \bar{\tilde{\mathbf{V}}})(\tilde{\mathbf{V}}_i - \bar{\tilde{\mathbf{V}}})^T$. The eigenvalues of $\mathbf{\Sigma}_{\tilde{V}}$ are 4.80, 3.31 and 0.56. The eigenvector that corresponds to the smallest eigenvalue is the direction of the least variation in the measured points of the right hand and is presumably perpendicular to the plane of the palm. We reflect the $\tilde{\mathbf{V}}_i$ around this plane. This turns the right hand into a left hand and produces the new points $\mathbf{V}_i$ shown in Table 2.

For $p = 2$ let $\Psi : \mathbb{R}^1 \to SO(2)$ be defined by $\Psi(h)$ is the rotation of h radians around the z axis for $h \in \mathbb{R}^1$. For $p = 3$, $\Psi : \mathbb{R}^3 \to SO(3)$ is defined by $\Psi(\mathbf{h})$ is right-hand rule rotation of $\|\mathbf{h}\|$ radians around the axis $\mathbf{h}/\|\mathbf{h}\|$ for $\mathbf{h} \in \mathbb{R}^3$. For general p, Ψ is defined using a matrix exponential map.

Chang and Ko (1995) rewrote $\mathbf{C} \in SO(p)$ in the form $\mathbf{C} = \mathbf{C}_0\Psi(\mathbf{h})$. Notice, in this way, $SO(p)$ is parameterized close to $\mathbf{C}_0$ by a coordinate system which depends on $\mathbf{C}_0$. The following theorem was proved in Chang and Ko (1995).

THEOREM 4.1. *Let $\mathbf{u}_1, \dots, \mathbf{u}_n \in \mathbb{R}^p$ be fixed. Let $\mathbf{V}_1, \dots, \mathbf{V_n} \in \mathbb{R}^p$ be independent random vectors such that the density of $\mathbf{V}_i$ is of the form $f(\mathbf{v}; \gamma_0\mathbf{C}_0\mathbf{u}_i + \mathbf{b}_0)$ for some $\gamma_0 > 0 \in \mathbb{R}^1$, $\mathbf{C}_0 \in SO(p)$, $\mathbf{b}_0 \in \mathbb{R}^p$. Assume $\log(f(\mathbf{x}; \theta)) = g(s)$ with $s = \|\mathbf{x} - \theta\|^2$. Write $\beta = \mathbf{b} + \gamma\mathbf{C}\bar{\mathbf{u}}$.*

Suppose $\hat{\gamma}$, $\widehat{\mathbf{C}}$ and $\hat{\beta}$ minimize an objective function of the form $\sum_i \rho_0(\mathbf{V}_i; \gamma\mathbf{C}(\mathbf{u}_i - \bar{\mathbf{u}}) + \beta)$, where $\rho_0(\mathbf{x}, \theta) = \tilde{\rho}_0(s)$ with $s = \|\mathbf{x} - \theta\|^2$. Write $\psi(s) = \tilde{\rho}_0'(s)$. Finally, let $\mathbf{\Sigma} = n^{-1}\sum_i(\mathbf{u}_i - \bar{\mathbf{u}})(\mathbf{u}_i - \bar{\mathbf{u}})^T$. Then, asymptotically:

(i) *$\hat{\gamma}$, $\widehat{\mathbf{C}}$, and $\hat{\beta}$ are independent.*

(ii) *$\sqrt{n}(\hat{\beta} - \beta_0)$ is $N_p(0, k\mathbf{I}_p)$.*

(iii) (a) *If $p = 2$, write $\widehat{\mathbf{C}} = \mathbf{C}_0\Psi(\hat{h})$ for $\hat{h} \in \mathbb{R}^1$. Then $\sqrt{n}\,\hat{h}$ is asymptotically $N(0, k/\operatorname{Tr}(\mathbf{\Sigma}))$.*

(iii) (b) *If $p = 3$, write $\widehat{\mathbf{C}} = \mathbf{C}_0\Psi(\widehat{\mathbf{h}})$ for $\widehat{\mathbf{h}} \in \mathbb{R}^3$. Let $\mathbf{\Sigma} = \lambda_1\xi_1\xi_1^T + \lambda_2\xi_2\xi_2^T + \lambda_3\xi_3\xi_3^T$ be the spectral decomposition of $\mathbf{\Sigma}$. Then $\sqrt{n}\,\widehat{\mathbf{h}}$ is asymptotically trivari-*

TABLE 2
Twelve digitized locations on the left and right hands

	Left hand u_i			**Right hand $\tilde{V}_i$**			**Flipped right hand V_i**		
A	5.17	11.30	16.18	5.91	11.16	16.55	6.28	12.72	16.73
B	7.40	12.36	17.50	8.63	10.62	18.33	8.86	11.60	18.44
C	8.56	12.59	17.87	10.09	10.60	18.64	10.16	10.90	18.67
D	9.75	13.62	17.01	10.89	10.95	17.90	10.76	10.40	17.84
E	11.46	14.55	12.96	12.97	10.13	13.88	13.19	11.04	13.98
F	7.10	13.12	12.56	8.79	11.21	13.17	9.00	12.11	13.27
G	8.85	13.82	12.60	10.70	11.10	13.42	10.75	11.30	13.44
H	6.77	13.07	10.32	8.47	11.09	11.35	8.86	12.74	11.54
I	6.26	11.62	13.34	7.28	12.52	14.04	7.03	11.46	13.92
J	6.83	12.00	13.83	8.05	12.42	14.56	7.73	11.09	14.41
K	7.94	12.29	13.84	9.07	12.39	14.86	8.64	10.60	14.66
L	8.68	12.71	13.67	10.15	12.17	14.44	9.73	10.40	14.24

NOTE. A, top of little finger; B, top of ring finger; C, top of middle finger; D, top of forefinger; E, top of thumb; F, gap between thumb and forefinger; G, center of palm; H, base of palm; I, little finger knuckle; J, ring finger knuckle; K, middle finger knuckle; L, forefinger knuckle.

ate normal with mean $\mathbf{0}$ *and covariance matrix* $k[(\lambda_2+\lambda_3)^{-1}\xi_1\xi_1^T + (\lambda_3+\lambda_1)^{-1}\xi_2\xi_2^T + (\lambda_1+\lambda_2)^{-1}\xi_3\xi_3^T]$.

(iv) $\sqrt{n}(\widehat{\gamma}-\gamma_0)$ *is* $N(0, k/\operatorname{Tr}(\mathbf{\Sigma}))$, *where*

$$k = \frac{pE[\psi(s)^2 s]}{4E^2[\psi(s)g'(s)s]} = \frac{pE[\psi(s)^2 s]}{4E^2[\psi'(s)s + (p/2)\psi(s)]}. \tag{4.1}$$

Examining Theorem 4.1, we see that the asymptotic covariance of $(\widehat{\gamma}, \widehat{\mathbf{H}}, \widehat{\beta})$ is determined up to a constant k by the geometry of the $\mathbf{u}_i$ (as summarized by the matrix $\mathbf{\Sigma}$). Only the constant k depends on the density or on the objective function, and this constant k is the same as in Theorem 3.5.

Another consequence of Theorem 4.1 is that the asymptotic relative efficiencies of Corollary 3.7 also apply to the image registration problems. In particular, to protect against outliers (perhaps misidentified landmarks), we could use the L_1 estimator of $\mathbf{C}_0$, $\mathbf{b}_0$ and γ_0, which minimizes $\sum_i \|\mathbf{V}_i - \gamma\mathbf{C}\mathbf{u}_i - \mathbf{b}\|$, instead of the more conventional L_2 estimator, which minimizes $\sum_i \|\mathbf{V}_i - \gamma\mathbf{C}\mathbf{u}_i - \mathbf{b}\|^2$. When $p=1$, this results in a penalty (in standard error) of $\sqrt{\pi/2}-1 = 25\%$ for normal errors when the L_2 estimator is optimal. However, the penalty decreases with p and is only 8% when $p=3$.

Indeed, let $\tilde{s}_i$ and $\widehat{s}_i$ be the square residual lengths for the L_1 and L_2 fits, respectively, and let k_1 and k_2 be the respective constants k in Theorem 4.1. Using (4.1) with $\psi(\tilde{s}_i) = 1/\sqrt{\tilde{s}_i}$ and $\psi(\widehat{s}_i) = 1$, we have for the hands data $\widehat{k}_1 = 0.023$ and $\widehat{k}_2 = 0.086$. Notice that if the errors in the $\mathbf{V}_i$ were normal, Corollary 3.7 would imply that $k_1/k_2 = 3\pi/8 = 1.18$. For the hands data set $\widehat{k}_1/\widehat{k}_2 = 0.26$. A quick and dirty estimate (computed by Taylor linearization) of the standard error in this ratio is 0.23. It is likely that the data come from a distribution which is long tailed relative to the normal distribution.

The L_1 and L_2 estimated values of γ are $\widehat{\gamma}_1 = 1.0086$ and $\widehat{\gamma}_2 = 0.9925$, respectively. Using Theorem 4.1(iv), we calculate that the standard errors of these two estimates are 0.0150 and 0.0293, respectively. Thus we cannot reject the null hypothesis $\gamma = 1$—that the two hands have the same size.

Notice that if $\widehat{\mathbf{C}} = \mathbf{C}_0\Psi(\widehat{\mathbf{h}})$, then $\mathbf{C}_0 = \widehat{\mathbf{C}}\Psi(\widehat{\mathbf{h}})^T = \widehat{\mathbf{C}}\Psi(-\widehat{\mathbf{h}})$. Therefore, using Theorem 4.1(iii)(b), an asymptotic confidence region for $\mathbf{C}$ has the form

$$\begin{aligned}\Big\{\widehat{\mathbf{C}}\Psi(\mathbf{h}) \,\Big|\, \frac{n}{k}\big[&(\lambda_2+\lambda_3)(\mathbf{h}^T\xi_1)^2 \\ &+ (\lambda_3+\lambda_1)(\mathbf{h}^T\xi_2)^2 \\ &+ (\lambda_1+\lambda_2)(\mathbf{h}^T\xi_3)^2\big] < \chi^2_{3,\alpha}\Big\},\end{aligned} \tag{4.2}$$

where $\chi^2_{3,\alpha}$ is an appropriate critical point of a χ^2 distribution with 3 degrees of freedom. Recall that $\Psi(\mathbf{h})$ is the rotation of $\|\mathbf{h}\|$ radians around the axis $\mathbf{h}/\|\mathbf{h}\|$. Thus the region (4.2) expresses the possible $\mathbf{C}$ in the form of a small rotation $\Psi(\mathbf{h})$ followed by the best fit rotation $\widehat{\mathbf{C}}$.

The eigenvalues of $\Sigma = n^{-1}\sum_i(\mathbf{u}_i - \overline{\mathbf{u}})(\mathbf{u}_i - \overline{\mathbf{u}})^T$ are $\lambda_1 = 5.004$, $\lambda_2 = 3.255$ and $\lambda_3 = 0.105$. The variables ξ_1 and ξ_2 are in the directions of the length and width of the left hand, and ξ_3 is in the direction of the normal to the plane of the left palm. From (4.2), we can see that the covariance of $\mathbf{h}$ is largest in the direction ξ_1 and smallest in the direction ξ_3.

If a rotation of $\phi = \|\mathbf{h}\|$ radians around the axis $\xi = \mathbf{h}/\|\mathbf{h}\|$ is applied to a point $\mathbf{p}_i = \mathbf{u}_i - \overline{\mathbf{u}}$, then $\mathbf{p}_i$ is moved a distance of approximately $\phi \cdot$ (distance of $\mathbf{p}_i$ to ξ). Let $\mathbf{W}_i = \widehat{\gamma}^{-1}\widehat{\mathbf{C}}^T(\mathbf{V}_i - \widehat{\beta})$ be the back transformed $\mathbf{V}_i$ under the estimates $\widehat{\gamma}$, $\widehat{\mathbf{C}}$, $\widehat{\beta}$. If we fix ϕ and vary ξ, we see that fit of the $\mathbf{p}_i$ to the $\mathbf{W}_i$ will be more degraded if the distances of the $\mathbf{p}_i$ to ξ are large than if they are small. Alternatively, if ξ is chosen so that the distances of the $\mathbf{p}_i$ to ξ are small, then, for the same degradation of the best fit, ϕ is less constrained than if the distances of the $\mathbf{p}_i$ to ξ are large. This is why the covariance of $\mathbf{h}$ is largest in the direction ξ_1 (the length of the hand) and smallest in the direction ξ_3 (the normal to the palm).

This example was further explored by Chang and Ko (1995). The emphasis there is to study the influence of the data on the estimates $\widehat{\mathbf{C}}$, $\widehat{\beta}$, and $\widehat{\gamma}$. Their techniques could be used if, for example, the image registration were unsatisfactory in some aspect (rotation, translation or scale change) and one wanted to determine which points should be reinspected.

5. STATISTICAL GROUP MODELS

Recall that if $\mathbf{X}, \theta_0 \in \Omega_p$ satisfy condition (1.1), then (1.2) will follow for all $\mathbf{C}$ in the group $SO(p)$. If $\mathbf{X}, \theta_0 \in \mathbb{R}^p$ satisfy condition (1.1), then $f(\mathbf{C}\mathbf{x} + \mathbf{b}; \mathbf{C}\theta_0 + \mathbf{b}) = f(\mathbf{x}; \theta_0)$ for all $(\mathbf{C}, \mathbf{b})$ with $\mathbf{C} \in SO(p)$ and $\mathbf{b} \in \mathbb{R}^p$. Such $(\mathbf{C}, \mathbf{b})$ form the group of *Euclidean motions*.

Usually spatial statistical models exhibit a great deal of symmetry and such symmetry is expressed using a *statistical group* model. In a statistical group model, we have a group $\mathcal{G}$, a sample space $\mathcal{X}$ and a parameter space Θ, together with maps $\mathcal{G} \times \mathcal{X} \to \mathcal{X}$ and $\mathcal{G} \times \Theta \to \Theta$. The images of $(g, \mathbf{X})$ and (g, θ) under these maps are usually denoted by $g \cdot \mathbf{X}$ and $g \cdot \theta$, and we assume the *group action* conditions $(g_1 g_2) \cdot \mathbf{X} = g_1 \cdot (g_2 \cdot \mathbf{X})$ and $1 \cdot \mathbf{X} = \mathbf{X}$, where $g_1 g_2$ represents multiplication in $\mathcal{G}$ and 1 represents the identity in $\mathcal{G}$. A similar condition is placed on $g \cdot \theta$. We also assume $\Pr_{g \cdot \theta}[g \cdot \mathbf{X} \in g \cdot \mathcal{A}] = \Pr_\theta[\mathbf{X} \in \mathcal{A}]$ for any $\mathcal{A} \subset \mathcal{X}$. Our objective function is also assumed to satisfy an invariance condition: $\rho_0(g \cdot \mathbf{x}, g \cdot \theta) = \rho_0(\mathbf{x}, \theta)$.

Notice that the forms of $\mathbf{I}$, $\mathbf{A}$ and $\mathbf{B}$ in Theorems 3.2 and 3.5 are all of the form $c\mathbf{v}^T\mathbf{w}$ for some constant c. Only the constants c depend on the precise form of the density f or of the objective function ρ.

Chang and Rivest (2001) showed that often the statistical group model condition places strong restrictions on the form of $\mathbf{I}$, $\mathbf{A}$ and $\mathbf{B}$. Heuristically speaking in terms of matrices (these are actually irreducible group representations), $\mathbf{I}$, $\mathbf{A}$ and $\mathbf{B}$ often have a block diagonal form. The form of the blocks is determined up to a constant by the action of $\mathcal{G}$ on Θ. Only the constants depend on the specific form of f and ρ.

In our examples of $SO(p)$ acting on Ω_p or the Euclidean motions acting on $\mathbb{R}^p$, there is only one "block." Chang and Rivest (2001) also gave a numerical example, arising from cardiography, with two "blocks." In this example, the group action determines the asymptotic distribution of $\widehat{\theta}$ up to four constants, two each for $\mathbf{A}$ and $\mathbf{B}$.

Another "coincidence" we have observed is that the asymptotic relative efficiencies of Corollary 3.7 also apply to the image registration problem. This is also a result of the group action: Chang and Tsai (2003) have shown that to any location statistical group model there is a corresponding regression group model, and that the $\mathbf{A}$ and $\mathbf{B}$ Riemannian metrics of the regression group model can be derived from those of the location group model.

6. SUMMARY

Two general principles apply when working with spatial models:

1. Spatial statistical models are often characterized by a great deal of symmetry. This symmetry is expressed using a statistical group model and when the group structure is properly used, many statistical calculations are greatly simplified.
2. Spatial statistical models also often have a parameter space Θ which is a q-dimensional manifold in some Euclidean space $\mathbb{R}^p$. To do statistical computations it is often necessary to reexpress Θ in q coordinates. To avoid distortions caused by the choice of coordinates, it is advisable to choose a coordinate system which depends on the true parameter $\theta_0 \in \Theta$.

By properly applying these two principles, we are able to reach statistical insights which are mathematically natural and elegant, and which lead to a deeper understanding than obtained from simple number crunching.

ACKNOWLEDGMENT

Figure 1 was produced by the Generic Mapping Tools software package written by Paul Wessel and Walter H. F. Smith and available at the URL *gmt.soest.hawaii.edu*.

REFERENCES

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

BROWN, B. M. (1985). Multi-parameter linearization theorems. *J. Roy. Statist. Soc. Ser. B* **47** 323–331.

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

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

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 RIVEST, L.-P. (2001). M-estimation for location and regression parameters in group models: A case study using Stiefel manifolds. *Ann. Statist.* **29** 784–814.

CHANG, T. and TSAI, M.-T. (2003). Asymptotic relative Pitman efficiency in group models. *J. Multivariate Anal.* **85** 395–415.

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

FISHER, N. I. (1985). Spherical medians. *J. Roy. Statist. Soc. Ser. B* **47** 342–348.

FISHER, N. I., LEWIS, T. and EMBLETON, B. J. J. (1987). *Statistical Analysis of Spherical Data.* Cambridge Univ. Press.

HETTMANSPERGER, T. P. and MCKEAN, J. W. (1998). *Robust Nonparametric Statistical Methods*. Wiley, New York.

HÖSSJER, O. and CROUX, C. (1995). Generalizing univariate signed rank statistics for testing and estimating a multivariate location parameter. *J. Nonparametr. Statist.* **4** 293–308.

JUPP, P. E. and MARDIA, K. V. (1989). A unified view of the theory of directional statistics, 1975–1988. *Internat. Statist. Rev.* **57** 261–294.

KIRKWOOD, B., ROYER, J.-Y., CHANG, T. and GORDON, R. (1999). Statistical tools for estimating and combining finite rotations and their uncertainties. *Geophysical J. International* **137** 408–428.

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

MÖTTÖNEN, J. and OJA, H. (1995). Multivariate spatial sign and rank methods. *J. Nonparametr. Statist.* **5** 201–213.

NEEMAN, T. M. (1995). Rank statistics for spherical data. Ph.D. dissertation, Univ. Virginia.

NEEMAN, T. M. and CHANG, T. (2001). Rank score statistics for spherical data. In *Algebraic Methods in Statistics and Probability* (M. A. G. Viana and D. St. P. Richards, eds.) 241–254. Amer. Math. Soc., Providence, RI.

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

REEDS, J. (1975). Discussion of “Defining the curvature of a statistical problem (with applications to second order efficiency),” by B. Efron. *Ann. Statist.* **3** 1234–1238.

WATSON, G. S. (1983). *Statistics on Spheres*. Wiley, New York.