The Annals of Statistics
1995, Vol. 23, No. 6, 1960–1974

SHAPE CHANGES IN THE PLANE FOR LANDMARK DATA

BY MICHAEL J. PRENTICE AND KANTI V. MARDIA

Edinburgh University and Leeds University

This paper deals with the statistical analysis of matched pairs of shapes of configurations of landmarks in the plane. We provide inference procedures on the complex projective plane for a basic measure of shape change in the plane, on observing that shapes of configurations of $(k + 1)$ landmarks in the plane may be represented as points on $\mathbb{C}P^{k-1}$ and that complex rotations are the only maps on $\mathbb{C}S^{k-1}$ which preserve the usual Hermitian inner product. Specifically, if $u_1, \dots, u_n$ are fixed points on $\mathbb{C}P^{k-1}$ represented as $\mathbb{C}S^{k-1}/\mathrm{U}(1)$, and $v_1, \dots, v_n$ are random points on $\mathbb{C}P^{k-1}$ such that the distribution of v_j depends only on $\|v_j^* A u_j\|^2$ for some unknown complex rotation matrix A, then this paper provides asymptotic inference procedures for A. It is demonstrated that shape changes of a kind not detectable as location shifts by standard Euclidean analysis can be found by this frequency domain method. A numerical example is given.

1. Introduction. There are a variety of practical problems which require the statistical analysis of matched pairs of shapes of configurations of landmarks in the plane. For various examples, see Bookstein (1991). We assume that data are available in the form of a random sample of n independent and identically distributed matched pairs $(X_1, Y_1), \dots, (X_n, Y_n)$ of $(k + 1)$ landmarks in the plane, where here each X_j and each Y_j is a complex $(k + 1)$ vector. The shape of each object is the information on the landmarks after all "pose" information has been removed, that is, the effect of location, size and orientation. Thus, as in Mardia and Dryden (1989), for example, each X_j and Y_j is centred, scaled and multiplied, conventionally by the Helmert matrix, to give a random sample of matched pairs of complex unit k-vectors (u_j, v_j), where $u_j^* u_j = 1 = v_j^* v_j$, u_j^* denotes transposed complex conjugate and each u_j and v_j is identified with $u_j \exp(i\theta)$ and $v_j \exp(i\phi)$, respectively, for all real θ, ϕ, and where $i^2 = -1$. Thus each u_j and v_j is a point on the complex projective hyperplane $\mathbb{C}P^{k-1}$ represented as $\mathbb{C}S^{k-1}/\mathrm{U}(1)$, in the notation of Kent (1994). Note the important restriction of invariance under scalar rotations. We propose here to carry out a basic investigation of the problem of point and regional estimation of shape change on shape space, rather than on a Euclidean approximation to shape space. Goodall (1991) addresses this problem using Euclidean approximations, but here we propose to explore the consequences of investigating shape change by representing it as a complex rotation and developing estimation and test procedures on $\mathbb{C}P^{k-1}$.

Received November 1993; revised March 1995.
AMS 1991 *subject classifications*. Primary 62H10; secondary 62H11.
Key words and phrases. Shape, unitary matrices, spherical regression, configuration.

Although it is now conventional to use the coordinate system of Kendall (1984) on shape space, using complex contrasts generated by premultiplication with the Helmert matrix, there are some advantages in using a Fourier alternative, described in detail below. After centring and scaling, each X_j and Y_j is a unit-length complex $(k+1)$-vector with entries summing to zero. Premultiplication by the Helmert matrix constructs k linearly independent contrasts, but there is no particular reason to choose a coordinate system on the complex sphere in this way. Indeed, one criticism levelled at the use of the complex sphere in shape analysis is that the coordinates do not measure anything biologically meaningful. The following alternative does, however, give complex spherical coordinates with a geometric meaning. We propose to replace the Helmert matrix by

$$F = (f_{rs}) = \left((k+1)^{-1/2}\exp(2\pi irs/(k+1))\right), \qquad r = 1,\dots,k;\ s = 1,\dots,k+1,$$

so that the shape vectors u_j and v_j are unit norm discrete Fourier transforms of the centred and scaled X_j and Y_j, respectively. Note that $FF^* = I_k$ and $F^*F = I_{k+1} - (k+1)^{-1}11^T$, as for the Helmert matrix. Thus, for example, if X_j represents a configuration of landmarks in standard numerical order at the vertices of a regular n-sided polygon in the plane, then $u_j = (1,0,\dots,0)$ and, in general, its first element u_{j1} is a measure of the extent to which X_j is similar to such a polygon. Similarly u_{jt} measures the extent to which X_j is similar to a configuration with vertices at $\exp(-2\pi ist/(k+1))$, $s = 1,\dots,k+1$.

For the case of triangle, $k = 2$, suppose that the correspondence quoted by Kent [(1994), page 292] between $\mathbb{C}S^1$ and S^2 is modified by relabelling as follows. Given $z = (z_1, z_2)^T \in \mathbb{C}S^1$, define $t = (t_1, t_2, t_3)^T \in S^2$ by

$$t_1 = -2\,\mathrm{Re}(z_1^* z_2), \qquad t_2 = 2\,\mathrm{Im}(z_1^* z_2), \qquad t_3 = z_1 z_1^* - z_2 z_2^*.$$

Then the Fourier choice of coordinate system above maps into the conventional spherical blackboard of Kendall (1984).

In this paper we propose to begin exploration of the statistical analysis of shape change on Kendall's shape space viewed as a normed frequency domain. The parallels with spectral analysis of time series are obvious, but there is the important difference that we usually have a relatively small number of landmarks (length of time series) with a reasonably large number of replicated measurements, as compared with only one realisation of a time series.

A complex spherical regression model is suggested in Section 2 as a basic model for shape change in the frequency domain. The main theoretical results needed are stated in Section 3, and estimation and test procedures are exemplified in Section 4. In Section 5, these ideas are applied to the data analysed by Mardia and Walder (1994a, b)—a set of landmark coordinates from lateral cephalograms of male rats from a close bred European strain, for which x-ray results are available at a number of different times for the same subjects. It is, of course, necessary to assume that the shape of the outline of each rat skull is adequately summarised by a small set of points, known as biological landmarks. At ages 90 and 150 days and 7 and 14 days, analyses of shape change in the frequency domain are carried out. The discussion given in Section 6 highlights some advantages of this method. For a general background to directional statistics, we refer to Mardia (1972), Watson (1983) and Fisher, Lewis and Embleton (1987).

2. A complex spherical regression model. Interest will centre here on a nested sequence of three hypotheses $H_0 \subset H_1 \subset H_2$ concerning a complex rotation A. The simplest possible statement about any shape change is H_0: $A = I_k$. If H_0 is true, then the shapes v_j are assumed to be the same as the shapes u_j except for composition with a small isotropic error. The most general model considered here is H_2: $A \in \mathrm{SU}(k)$, where $\mathrm{SU}(k)$ denotes the group of special unitary matrices, that is, complex rotations, $k \times k$ complex matrices A such that $AA^* = A^*A = I$ and $\det(A) = +1$. Hence if H_2 is true, then, ignoring scalar rotations, the shapes v_j are assumed to be the shapes Au_j, except for composition with a small isotropic error, where here A does not depend on j, but could be any fixed complex rotation in $\mathrm{SU}(k)$. This is the most general form of a rigid motion on the complex sphere. We shall also reserve the standard notation $\mathrm{U}(k)$ for general unitary matrices, having determinant with modulus 1. The model described below assumes that all inferences are made conditional on $u_1, \dots, u_n$. Note that we assign no particular biological or geometric meaning to the components of the $(k^2 - 1)$-dimensional parameter A. Our motivation in this exploratory paper is a preliminary investigation of global shape-space alternatives to conventional linear multivariate analyses on convenient tangent spaces.

Let u_0 and v_0 denote the unit length dominant eigenvectors of the population complex moment of inertia matrices of the u_j and v_j, respectively, that is, the "population mean shapes" of the u_j and v_j in the sense of Kent [(1992), page 121]. The intermediate hypothesis we consider here is that u_0 is unaffected by the complex rotation A. Since all eigenvalues of $A \in \mathrm{SU}(k)$ have unit modulus the hypothesis H_1: $Au_0 = \exp(i\theta)u_0$, for some unspecified real θ, is merely a statement that u_0 is an eigenvector of A. The hypothesis H_1 asserts that the "population mean after-shape" is the same as the "population mean before-shape," since any shape is unaffected by scalar rotations, although it allows that the distribution of the individual v_j need not be centred at their corresponding u_j. We shall investigate the problem of estimating A on the general alternative hypothesis H_2 and also subject to the restrictions imposed by H_1.

We consider in this paper a "complex axial regression" model: $u_1, \dots, u_n$ are fixed points on $\mathbb{C}P^{k-1}$, $v_1, \dots, v_n$ are independent random points on $\mathbb{C}P^{k-1}$ such that the probability density of v_j, with respect to uniform measure on $\mathbb{C}S^{k-1}$, is of the form

$$g\left(\|v_j^* A u_j\|^2\right)$$

for some unknown A in $\mathrm{SU}(k)$. Large sample statistical procedures for estimating and testing the unknown parameter A are described here. Since complex rotations are the only maps on $\mathbb{C}S^{k-1}$ which preserve inner products, the parameter A is a fundamental measure of shape change on Kendall's shape space.

The development of these procedures is an extension of the work of Chang (1987) on spherical regression, which is the corresponding problem for signed directions on the real hypersphere S^{k-1}, with the v_j having a probability density dependent only on $v_j^T A u_j$. Here, in the complex axial case, A and $A\exp(i\theta)$ are distinguishable in $\mathrm{U}(k)$, but induce the same transformation in complex projective space, so we shall say that A is "unique" (in quotes) if

$$A_2 = A \otimes \bar{A} \tag{2.1}$$

is unique (without quotes). Note that $A_2 \in \mathrm{SU}(k^2)$.

It is convenient to work with k^2-vectors

$$u_j^{(2)} = u_j \otimes \bar{u}_j - e \tag{2.2}$$

and $v_j^{(2)} = v_j \otimes \bar{v}_j - e$, where

$$e = k^{-1} \operatorname{vec}(I_k), \tag{2.3}$$

so that $u_j^{(2)}$ is unchanged by scalar rotations of u_j and has zero expectation if u_j is uniformly distributed. It is reasonable to estimate A by the matrix $\hat{A}$ which maximises

$$r(A) = n^{-1} \sum \|v_j^* A u_j\|^2, \tag{2.4}$$

a complex axial analogue of the "vector correlation" of Stephens (1979); see also Jupp and Mardia (1980). Since

$$\|v_j^* A u_j\|^2 - k^{-1} = v_j^{(2)*} A_2 u_j^{(2)} \tag{2.5}$$

with A_2 as in (2.1) and $v_j^{(2)}, u_j^{(2)}$ as in (2.2), we may equivalently choose to estimate A by the matrix A which maximises

$$r_2(A) = n^{-1} \sum v_j^{(2)*} A_2 u_j^{(2)}, \tag{2.6}$$

which is in the form of a vector correlation on a subspace of a k^2-sphere of radius $(1 - k^{-1})^{1/2}$, provided

$$\mathscr{Y}_0 = E\left(v_j^{(2)*} A_2 u_j^{(2)}\right) > 0, \tag{2.7}$$

or, equivalently, provided $E(\|v_j^* A u_j\|^2) > k^{-1}$. Note that $\mathscr{Y}_0 = 0$ if the v_j are uniformly distributed and that $r_2 = r - k^{-1}$, $0 \le r \le 1$.

One iterative scheme to estimate the complex rotation matrix A is to start at some suitable $A^{(0)}$ and choose $A^{(1)}, \dots, A^{(t)}, \dots$ so $A^{(t)}$ maximises $\operatorname{tr}(A^{(t)} n^{-1} \sum u_j u_j^* A^{(t-1)*} v_j v_j^*)$, using (complex) singular value decompositions at each iteration as in Prentice (1989). We call this $\hat{A}$ the least squares estimate of A because $\hat{A}$ minimises

$$\sum \left(v_j^{(2)} - A_2 u_j^{(2)}\right)^* \left(v_j^{(2)} - A_2 u_j^{(2)}\right) = 2n - 2 \sum \|v_j^* A u_j\|^2.$$

In Section 3, we find the asymptotic distribution of $\hat{A}$ under the assumption that $\lim_{n\to\infty} n^{-1}\sum u_j u_j^* = M$, a positive definite symmetric matrix (Theorem 1), and asymptotic confidence regions for A will be based on Theorem 1. For closed subgroups $G' \subseteq G$ of $\mathrm{U}(k)$, we also find the asymptotic distribution of $r_2(G') = \sup r_2(A)$ and of

$$A \in G,$$

$r_2(G) - r_2(G')$, when $A \in G'$ (Theorem 2).

Note that if the underlying distribution is complex Dimroth–Watson with probability density

$$b(K)\exp(K\|v^*Au\|^2), \tag{2.8}$$

then the procedures of this paper are just maximum likelihood estimation and likelihood ratio testing. Note also that we assume that u_j are fixed, or else make inferences conditional on the u_j. Complex axial regression with errors in variables is not discussed here [see Chang (1989) for spherical regression with errors in variables], but complex axial regression with highly concentrated errors is certainly of interest; see Rivest (1989) for the spherical case and our remarks on F tests in Section 4.

3. Statement of the main asymptotic results. The tangent space at the identity I_p of $\mathrm{SU}(k)$ is the collection of skew-Hermitian $k \times k$ matrices with zero trace, that is, the matrices H such that $H + H^* = 0$ and $\mathrm{tr}(H) = 0$. We denote this tangent space by $L(\mathrm{SU}(k))$ and define the exponential map ϕ: $L(\mathrm{SU}(k)) \to \mathrm{SU}(k)$ by

$$\phi(H) = \sum_{s=0}^{\infty} (s!)^{-1} H^s. \tag{3.1}$$

Note that if G is a closed subgroup of $\mathrm{SU}(k)$ and $L(G)$ is the tangent space at I_p of G, then $L(G)$ is a vector subspace of $L(\mathrm{SU}(k))$ and $L(G)$ is the set of H in $L(\mathrm{SU}(k))$ such that $\phi(tH)$ is in G for all real t. Note that $\dim G = \dim(L(G))$ and $\dim \mathrm{U}(k) = k^2$, where $\dim \mathrm{SU}(k) = k^2 - 1$.

The statement and proof of Theorem 1 given below is made easier by introducing the following notation. If $A \in G \subseteq \mathrm{U}(k)$, then define A_2 as in (2.1) and let G_2 be the collection of such A_2, as A varies over G. Then G_2 is a subgroup of $\mathrm{SU}(k) \otimes \mathrm{SU}(k)$, which is itself a subgroup of $\mathrm{SU}(k^2)$, as A_2 has determinant $+1$. The tangent space at I_p of $\mathrm{SU}(k) \otimes \mathrm{SU}(k)$ is the collection of $k^2 \times k^2$ matrices [Prentice (1989)]

$$H^{(2)} = (I \otimes H) + (H \otimes I),$$

where $H \in L(\mathrm{SU}(k))$. We denote this tangent space by $L(\mathrm{SU}(k) \otimes \mathrm{SU}(k))$ and define ϕ_2: $L(\mathrm{SU}(k)) \to L(\mathrm{SU}(k) \otimes \mathrm{SU}(k))$ by

$$\phi_2(H) = \phi(H^{(2)}) \tag{3.2}$$

with ϕ as in (3.1), acting on $k^2 \times k^2$ matrices. Note that $\dim L(G_2) = \dim L(G)$ for all $G \subseteq \mathrm{SU}(k)$, so, in particular, $\dim \mathrm{SU}(k) \otimes \mathrm{SU}(k) = k^2 - 1$.

If v on $\mathbb{C}S^{k-1}$ has a complex axially symmetric probability density of the form $g(\|v^*u\|^2)$, then it is useful to define constants $\mathscr{Y}_0$ and $\mathscr{Y}_2$ by $E(v^{(2)}) = \mathscr{Y}_0 u^{(2)}$, that is, $E(vv^*) = \mathscr{Y}_0 uu^* + k^{-1}(1-\mathscr{Y}_0)I_k$, with $\mathscr{Y}_0$ as in (2.7), and

$$\mathscr{Y}_2 = (k-1)^{-1}E\big(\|v^*u\|^2(1-\|v^*u\|^2)\big). \tag{3.3}$$

Thus if Θ is a random variable representing the great circle distance between v and u, then

$$\mathscr{Y}_0 = (k-1)^{-1}kE(\cos^2\Theta - k^{-1}) = (2k-2)^{-1}kE(\cos 2\Theta + 1 - 2k^{-1}),$$

$$\mathscr{Y}_2 = (k-1)^{-1}E(\sin^2\Theta\cos^2\Theta) = \tfrac{1}{4}(k-1)^{-1}E(\sin^2 2\Theta).$$

By contrast, in the case of spherical regression, Chang (1986, 1987) found it useful to work with $C_0 = E(\cos\Theta)$ and $C_2 = (k-1)^{-1}E(\sin^2\Theta)$. Note the equivalence, after angle doubling, of $\mathscr{Y}_0, 4\mathscr{Y}_2$ and C_0, C_2 in the case $k=2$. Complex axial regression reduces to spherical regression for the case of triangles: SU(2) is isomorphic to S^3, the unit quaternions, and SO(3) is obtained from S^3 by identifying antipodal pairs. A statistical study of such identifications has been carried out by Prentice (1986).

It is also of some importance to characterise the variance matrix of $v^{(2)}$. Let $z = (u,U)^*v$, where $(u,U)\in \mathrm{SU}(k)$, so

$$\begin{aligned} &u^*U = 0, \qquad U^*U = I_{k-1},\\ &UU^* = I_k - uu^*, \qquad (u,U)^*u = e_1 = (1,0,\dots,0)^T. \end{aligned} \tag{3.4}$$

In this canonical form $E(z^{(2)}) = \mathscr{Y}_0 e_1^{(2)}$ and $E(zz^*) = \mathscr{Y}_0 e_1 e_1^T + k^{-1}(1-\mathscr{Y}_0)I_k$. We assume $z^{(2)}$ has a $k^2\times k^2$ variance matrix, with the symmetries of the complex normal distribution, of the form $W = (w_{qr,st}) = E(z^{(2)}z^{(2)*}) - \mathscr{Y}_0^2 e_1^{(2)}e_1^{(2)T}$, with obvious symmetries, and where $(w_{jq,jq}) = (v_{jq})$ displays complex analogues of the special structure described in Lemma 5.1 of Prentice (1984). However, it turns out that only $w_{12,12}$ $(= w_{13,13} = \cdots = w_{1k,1k})$ is of any real concern in the proofs of the main results below, and of course $w_{12,12} = E(z_1^*z_1z_2^*z_2) = (k-1)^{-1}E(z_1z_1^*(1-z_1z_1^*)) = \mathscr{Y}_2$, as in (3.3).

If $Y = (u,U)^*$, then $E((v^{(2)} - \mathscr{Y}_0u^{(2)})(v^{(2)} - \mathscr{Y}_0u^{(2)})^*) = Y_2^*WY_2$. The most obvious parametric model fitting into this general description is the Dimroth–Watson density (2.8).

Our proof of the complex axial version of Theorem 1 of Chang (1986) is modelled very closely on that of Chang, but some of the calculations apparently necessary are considerably more intricate. Further, note that we apparently cannot state Theorem 1 as any special case of Theorem 1 of Chang (1986), and we make no claims about the asymptotic distribution of $\mathrm{tr}(H_n^{(2)2}M_2) = 2(\mathrm{tr}(H_n^2M))^2$, where

$$M_2 = \lim_{n\to\infty} n^{-1}\sum u_j^{(2)}u_j^{(2)*}. \tag{3.5}$$

As M_2 is singular and $v^{(2)}$ does not have a complex axially symmetric distribution on the entire complex k^2-sphere, it is not possible to construct any useful complex axial proofs from special cases of his Theorem 1.

THEOREM 1. *Let G be a closed subgroup of* $\mathrm{U}(k)$, *so G_2 is a closed subgroup of $T \subseteq \mathrm{SU}(k^2)$, and suppose each v_j has a density $g(\|v_j^* A_0 u_j\|^2)$, where $A_0 \in G$. Suppose furthermore that $\mathscr{Y}_0 > 0$, with $\mathscr{Y}_0$ as above* (3.3), *and that $n^{-1}\sum u_j u_j^*$ converges to a Hermitian positive definite matrix M, where $M = N_1 + iN_2$ with N_1 real symmetric positive definite and N_2 real skew-symmetric. Then*:

(a) *The least squares estimate of A_2 in G_2, denoted $\hat{A}_{n2}(G_2)$, is consistent for $(A_0)_2$. We may write $\hat{A}_{n2}(G_2) = (\hat{A}_n(G_2))_2$ "uniquely," so $\hat{A}_n(G_2)$ is consistent for A_0.*

(b) *Write $(A_0)_2^* \hat{A}_{n2}(G_2) = \phi_2(H_n)$ for $H_n \in L(G)$, with ϕ_2 as in* (3.2), *so $A_0^* \hat{A}_n(G_2) = \phi(H_n)$, with ϕ as in* (3.1). *Then H_n is asymptotically complex multivariate normal with mean* 0 *and probability density* [*with respect to Lebesgue measure on $L(G)$*] *proportional to* $\exp[(2\mathscr{Y}_2)^{-1}\mathscr{Y}_0^2 n\, \mathrm{tr}(H_n^2 N_1)]$ *so that $(-n\mathscr{Y}_0^2/\mathscr{Y}_2)\mathrm{tr}(H_n^2 N_1)$ is asymptotically $\mathscr{X}^2(\dim G)$.*

Note the similarity of parts (a) and (b) of our Theorem 1 and Theorem 1 of Chang (1986): N_1 is the real part of M and $\mathrm{tr}(H_n^2 N_1) \equiv \mathrm{tr}(H_n^2 M)$.

An outline of the proof of Theorem 1 is given in the Appendix.

THEOREM 2. (a) *If $A_0 \in G$, then $r_2(G)$, defined in* (2.6), *has a limiting normal distribution with mean $\mathscr{Y}_0$ and variance*

$$n^{-1}V(G) = n^{-1}(1 - k^{-1})\big[(1 - \mathscr{Y}_0)\big(\mathscr{Y}_0 + k^{-1}(1 - \mathscr{Y}_0)\big) - k^{-1}\mathscr{Y}_2\big].$$

(b) *If $A_0 \in H \subseteq G$, then*

$$(n\mathscr{Y}_0/\mathscr{Y}_2)(r_2(G) - r_2(H)) = (n\mathscr{Y}_0/\mathscr{Y}_2)(r(G) - r(H))$$

has a limiting $\chi^2(\dim G - \dim H)$ distribution.

(c) *If $A_0 \in K \subseteq H \subseteq G$, then*

$$\begin{aligned}&[(\dim G - \dim H)/(\dim H - \dim K)]\\&\quad\times[(r_2(H) - r_2(K))/(r_2(G) - r_2(H))]\end{aligned}$$

is asymptotically $F(\dim H - \dim K, \dim G - \dim H)$.

Again, note some similarity with Theorem 2 of Chang (1986). In (a), his variance was $n^{-1}(C_1 + C_2) = n^{-1}(E(\cos^2 \Theta) - C_0^2)$, whereas here it is $n^{-1}(E(\cos^4 \Theta) - (\mathscr{Y}_0(1 - k^{-1}) + k^{-1})^2)$, and in (b) and (c) the results are identical except for a missing factor of 2 in (b).

THEOREM 3. *Let $A_0^T \hat{A}_n(G_2) = \phi(H_n(G_2))$ for $H_n(G_2) \in L(G)$. Then $n^{1/2}(r_2(G) - \mathscr{Y}_0)$ and $n^{1/2}H_n(G_2)$ are asymptotically independent.*

The proofs of Theorems 2 and 3 are available from the authors on request.

If the density g is unknown, then to use Theorem 1 and 2 we need consistent estimators of $\mathscr{Y}_0$ and $\mathscr{Y}_2$. Using Theorem 2(a), we can estimate $\mathscr{Y}_0$ consistently by $g_0 = r_2(G)$ if $A_0 \in G$, and a consistent estimator of $\mathscr{Y}_2$ is

$g_2 = [n(k-1)]^{-1}\sum e_j(1-e_j)$, where $e_j = \|v_j^* A_G u_j\|^2$ and A_G maximises $r_2(G)$.

4. Some simple test procedures. Suppose H is a closed subgroup of $\mathrm{SU}(k)$, where $k > 2$. For now we exclude the case of triangles $k+1=3$. If $\mathscr{Y}_0$ and $\mathscr{Y}_2$, defined in (3.3), are known, Theorem 2(b) can be used to test, in large samples, if the true complex rotation matrix A is in H. If a test of the simple H_0: $A = A_0$ against the general alternative H_2: $A \in \mathrm{SU}(k) - H_0$ is required, then from Theorem 2(b), $n\mathscr{Y}_0\mathscr{Y}_2^{-1}(r_2(\mathrm{SU}(k)) - r_2(A_0))$ is asymptotically distributed as $\chi^2(k^2-1)$ if H_0 is true. Here

$$r_2(\mathrm{SU}(k)) = \left(n^{-1}\sum\|v_j^*\hat{A}\,u_j\|^2\right) - k^{-1} \tag{4.1}$$

and

$$r_2(A_0) = \left(n^{-1}\sum\|v_j^*A_0u_j\|^2\right) - k^{-1}. \tag{4.2}$$

An alternative procedure, involving more computational effort, is provided by Theorem 1, but not pursued further here. Given a sample estimate A_n of the unknown complex rotation, suppose $A_n^*A_0$ has spectral decomposition $X^*\exp(i\Lambda)X$, where $X \in \mathrm{U}(k)$ and Λ is a real diagonal matrix. Then if $H_n = iX^* \wedge X$, a test of H_0: $A = A_0$ is provided by referring $-(n\mathscr{Y}_0^2/\mathscr{Y}_2)\mathrm{tr}(H_n^2N_1)$ to $\chi^2(k^2-1)$.

It is also possible to investigate the hypothesis H_1: $Au_0 = \lambda u_0$, where here u_0 is the population mean before-shape and λ is a complex number of unit modulus. The hypothesis H_1 asserts that u_0 is an eigenvector of A and hence that the mean before-shape u_0 is the same as the mean after-shape v_0, since all shapes are unchanged by scalar rotations. There may be systematic differences between each before-shape u_j and its corresponding after-shape v_j, but $u_0 = v_0$. Since the subgroup of matrices A in $\mathrm{SU}(k)$, such that u_0 is an eigenvector of A, is isomorphic to $\mathrm{SU}(k-1)$, it follows from Theorem 2(b) that a test of H_1 against H_2: $A \in \mathrm{SU}(k) - H_1$ is provided by referring $n\mathscr{Y}_0\mathscr{Y}_2^{-1}(r_2(\mathrm{SU}(k)) - r_2(\mathrm{SU}(k-1)))$ to $\chi^2(2k-1)$, as $(k^2-1) - ((k-1)^2 - 1) = (2k-1)$. Here $r_2(\mathrm{SU}(k))$ is as in (4.1) and $r_2(\mathrm{SU}(k-1))$ is found most easily by first rotating all u data and all v data so that $M = n^{-1}\sum u_ju_j^*$ is diagonal, $M = \mathrm{diag}(s_1, \dots, s_k)$, where $s_1, \dots, s_{k-1}$ and $1 - s_k$ are typically small. After this change of coordinate system, $u_j^T = (z_j^T, a_j)$, $v_j^T = (w_j^T, b_j)$, where the z_j and w_j are $(k-1)$ vectors and the a_j and b_j are real scalars. Then

$$r_2(\mathrm{SU}(k-1)) = \left(n^{-1}\sum\|v_j^*\hat{A}_cu_j\|^2\right) - k^{-1} \tag{4.3}$$

and where $\hat{A}_c$ is constrained to be of the form block $\mathrm{diag}(\hat{A}_1, 1)$. The iterative search procedure for the constrained least squares estimate $\hat{A}_1$ is identical to that for $\hat{A}$, except that the dimensionality of the problem is reduced by 1.

Of course in practice $\mathscr{Y}_0$ and $\mathscr{Y}_2$ must be estimated by g_0 and g_2, say, and the validity of the test procedures is then even more approximate. For a practical large sample test of H_0: $A = I$ (no systematic shape change) against

$$H_2: A \in SU(k) - H_0,$$

refer

$$T_2 = [ng_0(\mathrm{SU}(k))/g_2(\mathrm{SU}(k))][r_2(\mathrm{SU}(k)) - r_2(I)], \tag{4.4}$$

to $\chi^2(k^2 - 1)$, where $g_0(\mathrm{SU}(k)) = r_2(\mathrm{SU}(k))$ as in (4.1), and $g_2(\mathrm{SU}(k)) = (n(k-1))^{-1}\Sigma c_j(1-c_j)$, $c_j = \|v_j^* \hat{A} u_j\|^2$, $0 \le c_j \le 1$. Similarly, to test only whether the mean shape has changed, that is, test H_1: $u_0 = v_0$ against H_2: $A \in \mathrm{SU}(k) - H_1$, refer

$$\begin{aligned} T_1 = {} & [ng_0(\mathrm{SU}(k-1))/g_2(\mathrm{SU}(k-1))] \\ & \times [r_2(\mathrm{SU}(k)) - r_2(\mathrm{SU}(k-1))] \end{aligned} \tag{4.5}$$

to $\chi^2(2k-1)$, where $g_0(\mathrm{SU}(k-1)) = r_2(\mathrm{SU}(k-1))$ as in (4.3) and $g_2(\mathrm{SU}(k-1)) = (n(k-1))^{-1}\Sigma d_j(1-d_j)$, where $d_j = \|w_j^* \hat{A}_1 z_j + b_j^* a_j\|^2$, $0 \le d_j \le 1$.

Note incidentally that a test of H_0: $A = I$ against H_1: $u_0 = v_0$ is provided by referring $T_0 = T_2 - T_1$ to $\chi^2(k(k-2))$. Thus it is possible to demonstrate systematic shape differences ($A \neq I$) even when the mean before-shape is the same as the mean after-shape.

These χ^2 tests based on large sample size asymptotics can be replaced, where appropriate, by F tests derived from high concentration asymptotics, as in Rivest (1989) for the real case. Following the same approach, we note that the complex analogue of Rivest [(1989), formula 5] is simply

$$\begin{aligned} F_{\mathrm{obs}} = {} & [n(2k-1) - (k^2-1)][r_2(\mathrm{SU}(k)) - r_2(G)] \\ & \times \left((k^2 - 1 - g)[1 - k^{-1} - r_2(\mathrm{SU}(k))]\right)^{-1}, \end{aligned} \tag{4.6}$$

with null distribution $F(k^2 - 1 - g, n(2k-1) - (k^2-1))$. This is the general form of the F test of the null hypothesis that $A \in G$, a closed proper subgroup of $\mathrm{SU}(k)$.

5. An application. The procedures of Section 4 are applied here to a standard data set analysed most recently by Mardia and Walder (1994a, b). The data consists of landmark coordinates from lateral cephalograms of 18 male rats from a close bred European strain as summarised graphically by Bookstein [(1991), Figure 7.6.7; reproduced here as Figure 1]. We consider first the change between ages 90 and 150 days in the shape of the lateral cephalograms of the 18 male rats, as measured by the $k+1 = 8$ landmarks described there. Various analyses have been carried out on this data set, notably by Bookstein [(1991), Section 7.6]. Here we explore some statistical inferences which can be made about the shape change on $\mathbb{C}P^6$, using the Fourier coordinate system described in the Introduction. The mean shape at age 90 days is

$$\begin{aligned} u_0 = 10^{-3}(&939, -33 + 104i, -80 - 73i, 65 - 20i, 5 + 39i, \\ & 26i - 103i, 54 + 271i)^T \end{aligned}$$

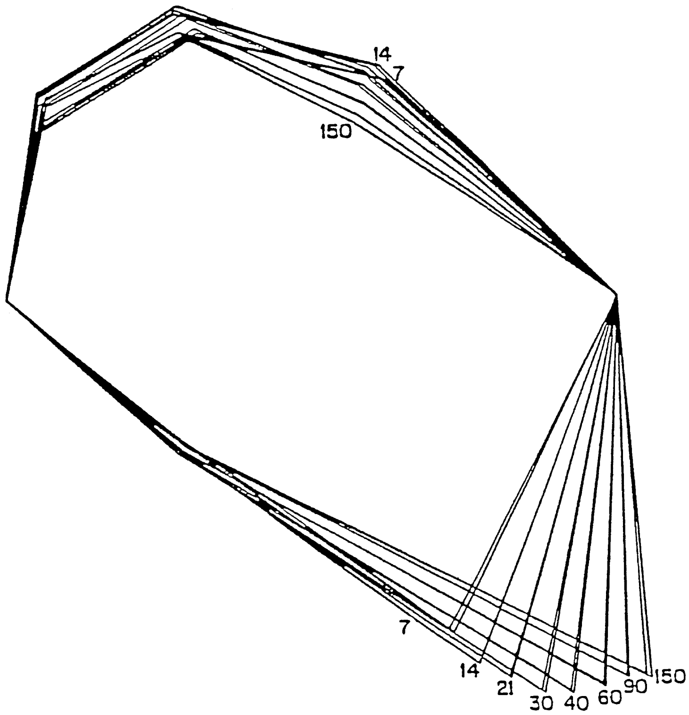

FIG. 1. *Evolution of rat calvarial shape to Basion–Bregma baseline. Reproduced from Bookstein* [(1991), *page* 356].

to three decimal places, and the mean shape at age 150 days is

$$v_0 = 10^{-3}(935, -40 + 98i, -71 - 74i, 68 - 20i, 5 + 41i, \\ 31 - 106i, 54 + 286i)^T,$$

where both shapes have been made unique by requiring that their first components are real and positive. The unconstrained least squares estimate of the complex rotation A is such that v_0 and Au_0 coincide to three decimal places and $\hat{A}$ has an eigenvector

$$w = 10^{-3}(820, 22 + 182i, -120 + 132i, -165 - 98i, \\ 108 - 1i, 7 - 65i, -240 + 391i)^T,$$

quite close to both u_0 and v_0. In passing we note that u_0, v_0 and w are all reasonably close to $e_1 = (1, 0, 0, 0, 0, 0, 0)^T$. The fact that the first components of u_0 and v_0 are so large confirms what is evident from Figure 2 [reproduced from Bookstein (1991), Figure 3.4.1], that these cephalograms have landmarks labelled 1 to 8 in anticlockwise order, and are approximately evenly spaced. If they were regular octagons, then the shape vectors would have been e_1, but of course there is no biological reason to expect this.

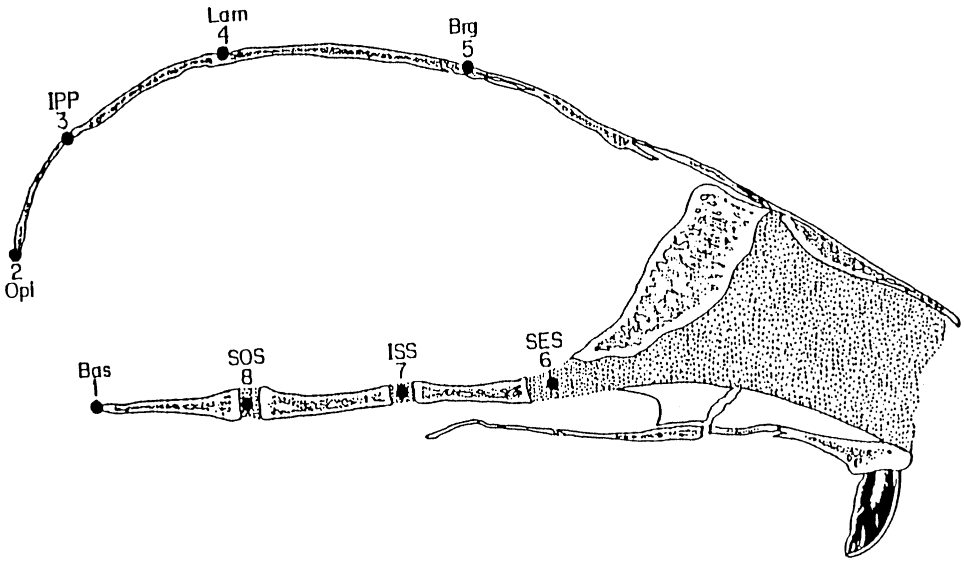

FIG. 2. *Representation of rat calvaria mid-sagittal section. Reproduced from Bookstein* [(1991), *page* 68].

For these data, $(7/6)r_2(\mathrm{SU}(7)) = 0.9996271$ (note that $0 \leq r_2 \leq 6/7$), with convergence to seven significant figures in only three iterations, and $g_2(\mathrm{SU}(7)) = 0.00005325$. For a test of H_1: $u_0 = v_0$, $(7/6)r_2(\mathrm{SU}(6)) = 0.9995555$ with convergence to seven significant figures achieved after only two iterations. The statistic (4.5) for testing H_1 against the general alternative H_2 is $T_1 = 17.79$, to be compared with $\chi^2(13)$, significant at about the 0.16 level if the asymptotic distributional assumptions are valid. With these distributional assumptions there is no definite evidence that the population mean shape at age 150 days is different from the population mean shape at 90 days. Note also that the statistic T_2, as in (4.4), for testing H_0: $A = I$ against the completely general H_2, is 156.5. This gives incontrovertible evidence that H_0 is false when compared with $\chi^2(48)$. Equally, for a test of H_0 against H_1, the statistic $T_0 = T_2 - T_1$ is 138.7, also highly significant when compared with $\chi^2(35)$. We have strong evidence of systematic shape change, $A \neq I$, but no real evidence that the mean after-shape differs from the mean before-shape.

A similar analysis of the shape changes between ages 7 and 14 days leads to different conclusions. The mean shape at 7 days is

$$u_0 = 10^{-3}(958, 45 + 161i, -125 - 54i, 30 - 34i,$$
$$19 + 66i, 0 - 43i, 26 + 165i)^T$$

and the mean shape at age 14 days is

$$v_0 = 10^{-3}(954, 33 + 159i, -113 - 57i, 36 - 31i,$$
$$13 + 59i, -60i, 65 + 181i)^T,$$

where again both have been made unique by requiring first components real and positive. In this case, $(7/6)r_2(\mathrm{SU}(7)) = 0.9988397$ and $g_2(\mathrm{SU}(6)) = 0.0001656$, rather bigger than in the first data set. Also $(7/6)r_2(\mathrm{SU}(6)) = 0.9985120$ and the statistic (4.5) for testing H_1 against H_2 is $T_1 = 41.51$, highly significant when compared with $\chi^2(13)$. There is definite evidence that the mean shape at 7 days is different from that at age 14 days.

The use of high concentration asymptotics, as in Rivest (1989), is also of interest here. The use of the F test in (4.6) for testing H_1 against H_2 gives $F_{\mathrm{obs}} = 4.04$, to be referred to the distribution $F(13, 186)$ and leading to the same conclusion. Clearly these procedures based on high concentration asymptotics are of particular importance when both sample size n and number of landmarks $k + 1$ are quite small.

6. Discussion. One standard multivariate approach to testing whether there has been any change of shape of a configuration of landmarks in the plane is to standardise all sample configurations to a common baseline and then carry out a Hotelling's T^2 test on the differences [e.g., Bookstein (1991), Section 5.4]. A criticism that may be levelled at this by a theoretician is that the use of such a procedure can give results which depend on which two landmarks were used as the baseline, but, as is well known, this criticism has little strength since most shape data sets are very highly concentrated so that the Hotelling's T^2 values obtained from different choices of baseline are in practice virtually identical [Bookstein (1991), Section 5.2]. In this paper we have developed an alternative approach to the fundamental location-shift problem on Kendall's shape space, the complex projective hyperplane. The test procedures developed in Section 4 will be completely unaffected by baseline choice, even for shape data sets which are more diffuse than usual, and there is the added advantage that systematic shape changes not detectable as location shifts by standard Euclidean analysis can be found, as in the numerical example of Section 5: the fact that calculations are performed in the Fourier domain gives a different insight into the changes taking place. Presumably the shape changes detected here show up in a Euclidean analysis as a change in variance matrix rather than a location shift. For the case of triangles with $k + 1 = 3$ landmarks in the plane (the simplest possible case), the procedures described here reduce to conventional spherical regression on the sphere S^2, using Kendall's spherical blackboard, for precisely the reason that a complex Bingham distribution on $\mathbb{C}S^1$ is equivalent to a Fisher distribution on S^2 [Kent (1994), page 292]. The test procedure for investigating H_1: $u_0 = v_0$ is quite simply that described in Chang [(1986), Section 2, second example], with the specified axis being the population mean before-shape, a unit direction on S^2.

It is instructive to consider an artificial example as a demonstration on S^2 of the kind of shape change detected in the first numerical example of Section 4 in more dimensions. Such changes are not evident as location shifts using standard Euclidean methods, but can be found on Kendall's shape space. Consider a "data set" of $4m$ matched pairs of triangles with shapes on S^2 all close to the equilateral shape $(0, 0, 1)^T$, using the "spherical blackboard" coordinate system of Kendall [(1984), Figure 2, page 101]. Specifically, consider the artificial data set consisting of matched pairs (u_j, v_j), where $u_j = (a\cos j\Theta,\ a\sin j\Theta,\ (1-a^2)^{1/2})$ for small positive a, $\Theta = \pi/2m$, and $v_j = u_{j+m}$. Note that $u_j = u_{j+4m}$. All shapes u_j and v_j are slight perturbations of an equilateral triangle.

Clearly, if these shapes on S^2 are converted back into triangles standardised to any baseline whatsoever, then the Hotelling's T^2 statistic calculated from the changes in location of the remaining landmark will be negligibly small. Hence a standard matched pairs Euclidean analysis to detect a change in mean shape will conclude that there is no change. However, a spherical regression analysis on Kendall's blackboard leads to a different conclusion. The sample mean before-shape and sample mean after-shape are both $(0,0,1)^T$, the equilateral triangle. The hypothesis H_0: $A = I_3$ of no systematic shape change is rejected in favour of H_1: $u_0 = v_0$, but there is no evidence in favour of the general alternative H_2: $u_0 \neq v_0$. The conclusion to be drawn is that there is a systematic shape change, even though there is no evidence that the mean before-shape is any different from the mean after-shape. Indeed, let A be a rotation through $\pi/2$ about $(0,0,1)^T$, that is,

$$A = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then the systematic change in shape is given by $v_j = Au_j$.

APPENDIX

Proof of Theorem 1. The major modification of the proofs of Chang (1986) requires the use of vectors that are here quadratic in the original shape measurements, leading to more complicated terms in Taylor series expansions. We refer to the corresponding sections in Chang (1986) for ease of comparison. Let $X_n = n^{-1}\Sigma u_j^{(2)} v_j^{(2)*}$ and $Z_n = n^{-1}\Sigma(u_j v_j^*) \otimes (u_j v_j^*)^{*T}$. The existence of M implies the existence of M_2, as in (3.5), and of

$$M^{+} = \lim n^{-1}\Sigma(u_j u_j^*) \otimes (u_j u_j^*)^{*T}. \tag{A.1}$$

Since $\hat{A}_n(G_2) \to A_0$, for large enough n we can write $A_0^* \hat{A}_n(G_2) = \phi(H_n)$, where $H_n \in L(G)$ is chosen to have smallest magnitude. We can assume $A_0 = I$ by replacing v_j with $A_0^* v_j$. By analogy with Chang [(1986), page 911] pick a specific $B \in L(G)$ and define a real-valued function on $L(G)$,

$$g_n^B(H) = \frac{d}{dt}\bigg|_{t=0} \operatorname{tr}\big(\phi_2(H + tB)Z_n\big),$$

so that $g_n^B(H_n) = 0$. We expand g_n^B in a Taylor series about 0:

$$g_n^B(0) = \frac{d}{dt}\bigg|_{t=0} \operatorname{tr}\big(\phi_2(tB)Z_n\big) = \operatorname{tr}\big(B^{(2)}Z_n\big).$$

However, $\operatorname{tr}(B^{(2)}X_n) = \operatorname{tr}(B^{(2)}Z_n)$ since $\operatorname{tr}(B^{(2)}e(v^* \otimes v^T)) = 0$. If $H \in L(G)$,

$$(g_n^B)'(0) \cdot H = \frac{d}{ds}\bigg|_{s=0} \frac{d}{dt}\bigg|_{t=0} \operatorname{tr}\big(\phi_2(sH + tB)Z_n\big) = \operatorname{tr}\big(C(B,H)Z_n\big),$$

where $C(B,H) = [\frac{1}{2}(HB + BH)]^{(2)} + H \otimes (B^{*T}) + B \otimes (H^{*T})$. Thus $0 = g_n^B(H_n) = \operatorname{tr}(B^{(2)}X_n) + \operatorname{tr}(C(B,H_n)Z_n) + R$, where R is suitably bounded and so, using Lemma 1, it follows that if $\alpha_n(B) = -\operatorname{tr}(B^{(2)}\sqrt{n}\,X_n)$, then $\alpha_n(B) = \mathscr{Y}_0 \operatorname{tr}(\sqrt{n}\,C(B,H_n)M^+) + (1 - \mathscr{Y}_0)\operatorname{tr}(\sqrt{n}\,C(B,H_n)\operatorname{vec}(M)e^T) + R_n$, where R_n is suitably bounded. It remains to simplify $\operatorname{tr}(C(B,H)M^+)$ and $\operatorname{tr}(C(B,H)\operatorname{vec}(M)e^T)$. With M^+ as in (A.1),

$$\begin{aligned}\operatorname{tr}(C(B,H)M^+) &= \tfrac{1}{2}\big\{\operatorname{tr}((BH \otimes I)M^+) + \operatorname{tr}((HB \otimes I)M^+)\\ &\qquad + \operatorname{tr}\big((I \otimes (BH)^{*T})M^+\big) + \operatorname{tr}\big((I \otimes (HB)^{*T})M^+\big)\big\}\\ &= \operatorname{tr}(HMB) + \operatorname{tr}(BMH) = 2\operatorname{tr}(HN_1B).\end{aligned}$$

Similarly, $2\operatorname{tr}(C(B,H)\operatorname{vec}(M)e^T)$ simplifies eventually to 0.

Thus

$$\alpha_n(B) = -\operatorname{tr}\big(B^{(2)}\sqrt{n}\,X_n\big) = 2\mathscr{Y}_0 \operatorname{tr}\big(\sqrt{n}\,(H_nN_1B) + R_n\big),$$

where R_n is suitably bounded.

As in Chang [(1986), Lemma 4], we have the following lemma:

LEMMA 1. *α_n has covariance quadratic form*

$$\mathscr{Y}_2 Q(B_1, B_2) = -4\mathscr{Y}_2 \operatorname{tr}(B_1N_1B_2) \quad \text{for } B_1, B_2 \in L(G).$$

PROOF.

$$\begin{aligned}&E\Big(\big(B^{(2)}u_j^{(2)}\big)^*\big(v_j^{(2)} - \mathscr{Y}_0u_j^{(2)}\big)\big(v_j^{(2)} - \mathscr{Y}_0u_j^{(2)}\big)^*B^{(2)}u_j^{(2)}\Big)\\ &\qquad = u_j^{(2)*}B^{(2)*}(Y_j)_2^*W(Y_j)_2B^{(2)}u_j^{(2)},\end{aligned} \tag{A.2}$$

where $Y_j = (u_j, U_j)^*$ and W is the variance matrix of the canonical random variable $z^{(2)}$.

Now $Y_ju_j = e_1$ and $Y_jBu_j = f$ (say), where $f = (0, f_2, \dots, f_p)^T$ and f may be assumed real, as arbitrary scalar rotations may be applied to the columns of U_j. Thus (A.2) reduces to g^TWg, where $g = f \otimes e_1 + e_1 \otimes f$ and the entries in g are nonzero only if its (double) suffix includes the value 1 exactly once. Hence (A.2) reduces to

$$4\sum_{q=2}^{k}\sum_{m=2}^{k}\omega_{1q,1m}f_qf_m = -4\mathscr{Y}_2 \operatorname{tr}(Bu_ju_j^*B)$$

since $\omega_{1q,1m} = \delta_{qm}\mathscr{Y}_2$ with $\mathscr{Y}_2$ as in (3.3), and $f^Tf = \operatorname{tr}(u_j^*B^*Bu_j)$. We obtain the covariance quadratic form of $\alpha_n(B)$, which is real, by averaging over j. Thus $\mathscr{Y}_2Q(B_1, B_2) = -4\mathscr{Y}_2 \operatorname{Re}(\operatorname{tr}(B_1N_1B_2))$. Since $\alpha_n(B) = 2\mathscr{Y}_0 \operatorname{Re}(\operatorname{tr}(H_nN_1B)) + o_p(1)$, it follows by the same argument as that of Chang [(1986), Lemma 4] that H_n is asymptotically complex normal with probability density proportional to $\exp(n(2\mathscr{Y}_0)^2/(8\mathscr{Y}_2)\mathrm{tr}(H_n N_1 H_n))$. Hence $-(n\mathscr{Y}_0^2/\mathscr{Y}_2)\mathrm{tr}(H_n^2 N_1)$ is asymptotically distributed as $\chi^2(\dim G)$. This proves Theorem 1. □

Acknowledgments. We are grateful to the Associate Editor and the two referees for their helpful comments. In particular, one of the referees provided meticulous details for the revised version. We are also grateful to Dr. Alastair Walder for his assistance.

REFERENCES

BOOKSTEIN, F. (1991). *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge Univ. Press.

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

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

CHANG, T. (1989). Spherical regression with errors in variables. *Ann. Statist.* **17** 293–306.

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

GOODALL, C. (1991). Procrustes methods in the statistical analysis of shape (with discussion). *J. Roy. Statist. Soc. Ser. B* **53** 285–340.

JUPP, P. E. and MARDIA, K. V. (1980). A general correlation coefficient for directional data and related regression problems. *Biometrika* **67** 163–173.

KENDALL, D. G. (1984). Shape manifolds, procrustean metrics and complex projective spaces. *Bull. London Math. Soc.* **16** 81–121.

KENT, J. T. (1992). New directions in shape analysis. In *The Art of Statistical Science* (K. V. Mardia, ed.) Wiley, New York.

KENT. J. T. (1994). The complex Bingham distribution and shape analysis. *J. Roy. Statist. Soc. Ser. B* **56** 285–300.

MARDIA, K. V. (1972). *Statistics of Directional Data*. Academic Press, New York.

MARDIA, K. V. and DRYDEN, I. L. (1989). Shape distributions for landmark data. *Adv. in Appl. Probab.* **21** 742–755.

MARDIA, K. V. and WALDER, A. N. (1994a). Shape analysis of paired landmark data. *Biometrika* **81** 185–196.

MARDIA, K. V. and WALDER, A. N. (1994b). Size-and-shape distributions for paired landmark data. *Adv. in Appl. Probab.* **26** 893–905.

PRENTICE, M. J. (1984). A distribution-free method of interval estimation for unsigned directional data. *Biometrika* **71** 147–154.

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

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

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

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

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

DEPARTMENT OF MATHEMATICS
AND STATISTICS
EDINBURGH UNIVERSITY
EDINBURGH
SCOTLAND

DEPARTMENT OF STATISTICS
UNIVERSITY OF LEEDS
LEEDS LS2 9JT
ENGLAND