

Invariance and Morphometrics

Subhash Lele
Department of Biostatistics
The Johns Hopkins University
Baltimore, MD 21205

Charles E. McCulloch
Biometrics Unit,
Department of Statistical Science
Cornell University
Ithaca, NY 14850

BU – 1440 - M

February, 1998

Keywords: Shape, size, multivariate, rotation, translation

Abstract

The form of an object is that characteristic which remains invariant under a group of transformations consisting of translation, rotation and possibly reflection. Group invariance thus naturally plays an important role in the statistical analysis of forms. We examine the existing methods for the statistical analysis of form from the invariance perspective. We begin with a review of the important basic ideas behind invariance, derive a maximal invariant under the group of transformations consisting of rotation, reflection and translation and its distribution under the Gaussian perturbation model. We discuss the issues of identifiability of various parameters in the single population case. We show that method of moments estimators based on the distances between landmarks and method of maximum likelihood based on the size and shape coordinates are invariant. However, a number of commonly used methods are not invariant. We compare the statistical and computational efficiencies of the method of moments and maximum likelihood. It is shown that the method of moments substantially simplifies the estimation procedure computationally at a small loss of statistical efficiency. We then extend the discussion of invariance to the comparison of forms and show that many commonly used methods for the study of form difference are not invariant.

1. Introduction

The analysis of the form (or shape and size) of an object has a long history in the biological sciences (e.g., Thomson, 1917, Huxley, 1932). One of the reasons for the importance of the study of form is the belief that the form is directly related to the function of the organism (Thompson, 1917). Using such relationships is not only useful in the study of evolution of organisms but also in medical statistics, e.g. in plastic surgery where prediction of the effect of the surgical procedure on the growth of a child is necessary to decide on the best course of action (Dufresne and Richtsmeier, 1995, Richtsmeier et al., 1997). Morphometrics, or the quantitative analysis of form, is thus an important methodology in a number of fields, including evolutionary biology (Gould, 1977), surgery, brain mapping (Richtsmeier, 1993), genetics and protein science (Godzik, 1996).

The analysis of form has also attracted substantial attention in the statistical literature since the publication of review papers by Bookstein (1986) and Kendall (1989). Intuitively, the definition of the "form" of a geometrical object should be invariant under the group of transformations consisting of translation, rotation and possibly reflection. Bookstein, et al (1985) argue that whenever an arbitrary choice is made to resolve a lack of invariance that "its impact upon subsequent statistical analysis (be) explicitly investigated." The main purpose of the present paper is to take a careful look at the relationship between this innate invariance and various methods developed for the

statistical analysis of forms. Unfortunately, the conclusion is that many commonly used and/or intuitively appealing methods are *not* invariant.

The outline of the paper is as follows. We start with a brief discussion of the nature of landmark coordinate data and specify the basic stochastic model we use. In Section 3, we study the single population case and discuss identifiability and estimation of the parameters. In Section 4, we study the problem of comparison of two forms using deformation and superimposition of mean forms as well as comparison in terms of distances between landmarks. We conclude by summarizing the implications of invariance to the statistical analysis of landmark coordinate data.

2. A brief introduction to landmark coordinate data and the perturbation model

Practical analysis of biological form is hindered by the complexity of quantifying an entire form. One approach to reducing this complexity is to consider a few biologically important points on the form under consideration and assume that the configuration of these points approximates the underlying form adequately for the problem at hand. Such points are called "landmarks". These landmarks not only summarize the form of the object in the geometrical sense but also define a correspondence of biological structures between the objects under study. Hence they are called "homologous landmarks". There have been a number of papers discussing the biological basis of "homology and

landmarks". (See for example Roth, 1988, Van Valen, 1982, Hall, 1994 and Valeri et al. 1997)

Suppose that the choice of the landmarks to be collected is decided upon. They can be easily quantified using current imaging technology. For example, we could fix a coordinate system arbitrarily and locate the object in this system. Next, we calculate the coordinates of each landmark with respect to this arbitrarily chosen coordinate system using a digitizer or other device (Corner et al., 1992, Richtsmeier et al., 1995). If there are K landmarks on a D dimensional object then we can construct a $K \times D$ matrix whose j -th row consists of the D coordinates of the j -th landmark. In typical applications, D is usually either 2 or 3 and K is assumed to be larger than D to avoid degeneracy. We will use X_i to denote the $K \times D$ matrix of coordinates for the i -th individual.

Perturbation model: We model the inter-individual variability by the Gaussian perturbation model described in, e.g., Goodall (1991) or Lele (1993). The perturbation model may be thought of as representing the following process. To generate a random geometrical object or equivalently a K point configuration in D dimensional Euclidean space, nature first chooses a mean form (represented by matrix M) and perturbs the elements of this matrix by adding noise to this mean form according to a matrix valued Gaussian distribution. The K point configuration so obtained is then rotated and/or reflected by an unknown angle and translated by an unknown amount. Such perturbed, translated, rotated and/or reflected K point configurations constitute our data. The above description can be put in a mathematical form as follows.

Let M be a $K \times D$ matrix corresponding to the mean form. Let E_i be a $K \times D$ matrix valued Gaussian random variable (Arnold, 1981, pages 309-323) representing the error for the i -th individual. We assume E_i is Gaussian with mean matrix 0 and covariance matrices given by Σ_K and Σ_D . The matrix Σ_K describes the covariances between elements within the same column of E and Σ_D describes the covariances within the rows of E_i . Notationally we have $E_i \sim N(0, \Sigma_K, \Sigma_D)$. More precisely, if we stack the matrix, E_i , into a vector, $\text{vec}(E_i)$, then we have $\text{var}(\text{vec}(E)) = \Sigma_K \otimes \Sigma_D$. Let R_i be an orthogonal matrix corresponding to the rotation of the i -th individual and let t_i be a $1 \times D$ matrix corresponding to the translation of the i -th individual. Then the landmark coordinate matrix corresponding to the i -th individual may be represented as:

$$X_i = (M + E_i)R_i + \underline{1}_i^T$$

where $\underline{1}$ is a $K \times 1$ matrix of 1's. The random matrices X_i thus follow:

$$X_i \sim N(MR_i + \underline{1}_i^T, \Sigma_K, R_i \Sigma_D R_i^T)$$

An important thing to notice is that the matrices R_i and t_i are unknown and unknowable. These are nuisance parameters while the parameters of interest are (M, Σ_K, Σ_D) . In fact, there are more unknown parameters than there are number of observations and the number of parameters grows with the sample size. This problem thus falls in the class of

problems described by Neyman and Scott (1948).

Invariance and elimination of nuisance parameters

One way to eliminate the nuisance parameters is by considering a maximal invariant under a group of transformations. We briefly review the definition of a maximal invariant and an important consequence of using a maximal invariant for the statistical inference to the identifiability of the underlying parameters. For an excellent discussion of invariance, we refer the reader to Lehmann(1959, Chapter 6), Berger (1982, Chapter 6) or Arnold (1981,Chapter 1).

Definition 1: Let G be a group of invertible functions from a set C to itself. A function $T(c)$ is called a maximal invariant if it satisfies the following two conditions:

- a) $T(g(c))=T(c)$ for all $g \in G$ and $c \in C$.
- b) If $T(c_1) = T(c_2)$, then there exists $g \in G$ such that $c_2 = g(c_1)$.

For any set C and any group G of invertible functions from C into itself, there exists a maximal invariant. Any one to one function of a maximal invariant is itself a maximal invariant. Hence maximal invariants are not unique. However, all maximal invariants are equivalent in the sense that their sets of constancy coincide.

Let X be distributed according to a probability distribution $P_\theta, \theta \in \Theta$. Denote by gX the

random variable that takes on the value gx when $X=x$, and suppose that when the distribution of X is P_θ , the distribution of gX is $P_{\theta'}$, with θ' also in Θ . The element θ' associated with θ in this manner will be denoted by $\bar{g}\theta$. The transformation \bar{g} of Θ onto itself, defined in this manner, is one to one provided the distributions P_θ corresponding to different values of θ are distinct. Invariance, by reducing the data to a maximal invariant statistic T , typically also shrinks the parameter space. The following theorem (Lehmann, 1959, Chapter 6, Theorem 3) is a precise version of this statement.

Theorem 1: If $T(x)$ is invariant under G , and if $\nu(\theta)$ is maximal invariant under the induced group \bar{G} , then the distribution of $T(X)$ depends only on $\nu(\theta)$.

In other words, the principle of invariance identifies all parameter points that are equivalent with respect to \bar{G} .

Now let us consider the landmark coordinate data problem. The landmark coordinate matrix is denoted by X . Recall that the group of transformations involved in the analysis of landmark coordinates consists of translation, rotation and possibly reflection. Thus

$gX = XR + \underline{1}l^T$. It is easy to see that if $X \stackrel{d}{=} N(M, \Sigma_K, \Sigma_D)$, then

$gX \stackrel{d}{=} N(MR + \underline{1}l^T, \Sigma_K, R^T \Sigma_D R)$, where $\stackrel{d}{=}$ means equal in distribution. From the

above theorem, the parameters are identifiable only up to the membership on the orbits defined by this group of transformations. Which particular member is obtained depends

on the choice of the particular maximal invariant.

We now consider one particular maximal invariant under this group and derive the parameters that are obtained under it, $T(\cdot)$ and $\nu(\theta)$ in the above notation. Since all the maximal invariants are equivalent, the identifiability issues remain the same for all of them.

Let L be a $(K-1) \times K$ matrix whose first column consists of -1's and the rest of the matrix is an identity matrix of dimension $(K-1) \times (K-1)$. Now define $T(X) = LXX^T L^T$. Since $L\underline{1} = \underline{0}$ and because R is an orthogonal matrix i.e. $RR^T = I$, it is easy to see that $T(X) = T(XR + \underline{1}^T)$ and hence is invariant. To show that it is a maximal invariant, we need to show that, given $T(X)$, one can map it back to a unique orbit in the original space. This can be proved by using the fact that $T(X)$ is a centered inner product matrix and that there exists a unique (up to rotation, translation and reflection) mapping from the centered inner product matrix to a coordinate matrix, see Lele (1991, 1993). Next we derive the distribution of this maximal invariant in order to determine which parameters are identifiable. We introduce some additional notation.

Let $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_D\}$ denote the diagonal matrix of the eigenvalues of Σ_D and let

$\Sigma_K^* = L\Sigma_K L^T$. Consider a spectral decomposition of the matrix

$$(LM)\Sigma_D^{-1}(LM)^T = \sum_{j=1}^D d_j d_j^T \text{ where } d_j \text{'s are the scaled eigenvectors of}$$

$(LM)\Sigma_D^{-1}(LM)^T$, that is, $\|d_j\| = \eta_j$, the j -th eigenvalue of $(LM)\Sigma_D^{-1}(LM)^T$. Let

$d_j d_j^T = \Delta_j$ and note that Δ_j is of rank one.

Theorem 2: The maximal invariant $T(X) = LXX^T L^T$ is distributed as a linear combination of non-central Wishart matrices. More precisely stated,

$T(X) = \sum_{j=1}^D \lambda_j Z_j$ where $Z_j \stackrel{d}{=} W_K(1, \Delta_j, \Sigma_K^*)$, a non-central Wishart matrix of dimension

$K \times K$ with non-centrality parameter Δ_j and scale parameter Σ_K^* .

Proof: Notice that $LX \sim N(LM, L\Sigma_K L^T, \Sigma_D)$. The above result follows by taking A to be an identity matrix in Theorem 2 on page 249 and remark on page 248-249 of deGunst (1983).

The maximal invariant induced on the parameter space by $T(X)$ is thus given by

$(\lambda_1, \lambda_2, \dots, \lambda_D, \Sigma_K^*, \Delta_1, \Delta_2, \dots, \Delta_D)$. Only these parameters or functions thereof are

identifiable. Thus the eigenvalues of the covariance matrix Σ_D are identifiable but not the

eigenvectors. Moreover Σ_K itself is not identifiable but only $L\Sigma_K L^T$ is identifiable. It

will be shown in the next section, that the Δ_i 's can be used to construct the mean form M

up to rotation, reflection and translation.

Another maximal invariant suggested in the literature is form coordinates (also known as

size and shape coordinates; Bookstein, 1986; Kendall, 1989). This was studied in detail by Dryden and Mardia (1991). This maximal invariant does not include reflection as part of the group transformation. Dryden and Mardia (1991) also study the maximal invariant under an additional component of scaling in order to consider only 'shape coordinates'. The exact distributions of 'form' as well as 'shape' coordinates are obtained in Dryden and Mardia (1991). Identifiability issues were not dealt with explicitly in any of these papers.

3. Estimation of parameters in the single sample case

Having established the identifiability of certain parameters, the next natural question is whether these parameters can be estimated in a practically suitable and statistically desirable fashion. There have been various methods suggested in the literature. The most commonly used method is based on Generalized Procrustes Analysis. The review article by Goodall (1991) provides a description of this method. Alternatively, maximum likelihood estimation can be conducted using the exact shape distributions derived by Mardia and colleagues or estimators can be constructed via the method of moments (Stoyan, 1990; Lele, 1993)

Method of moments estimators are based on the moments of the distribution of the maximal invariant described in the previous section. Lele(1993) showed that under the assumption that $\Sigma_{pp} = I$, the estimating functions based on the method of moments have a unique, analytical solution. Moreover, Lele (1993) also shows that the estimator of the

mean form M can be estimated up to translation, rotation and reflection consistently and the covariance parameter $\Sigma_K^* = L\Sigma_K L^T$ is also estimable consistently. However, assuming $\Sigma_D = I$ imposes restrictions on the applicability of this model so we next extend the method to the situation where $\Sigma_D \neq I$.

Method of moments estimators

For notational simplicity, let $Y=T(X)$. Let $Y = [Y_{lm}]$ where $l = 1, 2, \dots, K; m = 1, 2, \dots, K$ denoting the individual elements of the matrix Y . Similarly let $\Sigma_K^* = [\sigma_{lm}]$ and

$\Delta_j = [\delta_{lm}^j]$. Notice that the matrices Δ_j 's are symmetric and that $\delta_{lm}^j = \sqrt{\delta_{ll}^j \delta_{mm}^j}$.

This follows easily from the definition of these matrices.

From the previous section, we know that Y is distributed as a linear combination of non-central Wishart random variables. The following expressions provide the first two moments of the random matrix Y (Alam and Mitra, 1990).

$$E(Y_{lm}) = \sum_{j=1}^D \lambda_j \{ \delta_{lm}^j + \sigma_{lm} \}$$

$$ar(Y_{lm}) = \sum_{j=1}^D \lambda_j^2 \{ \sigma_{lm}^2 + \sigma_{ll} \sigma_{mm} + \sigma_{ll} \delta_{ll}^j + \sigma_{mm} \delta_{mm}^j + 2\sigma_{lm} \delta_{lm}^j \}$$

$$Cov(Y_{lm}, Y_{np}) = \sum_{j=1}^D \lambda_j^2 \{ \sigma_{ln} \sigma_{mp} + \sigma_{lp} \sigma_{mn} + \sigma_{ln} \delta_{np}^j + \sigma_{lp} \delta_{mm}^j + \sigma_{mn} \delta_{lp}^j + \sigma_{mp} \delta_{ln}^j \}$$

Equating the sample moments with the population moments, one can numerically obtain the estimates of the parameters $(\lambda_1, \lambda_2, \dots, \lambda_D, \Sigma_K^*, \Delta_1, \Delta_2, \dots, \Delta_D)$. Given these estimates, how do we obtain the mean form of the object (up to translation, rotation and reflection)? From the definition of Δ_j 's, it is clear that they are symmetric matrices with rank 1. One can write $\Delta_j = d_j d_j^T$, where d_j is a scaled eigenvector as described earlier. Thus, for $D=3$, the estimate of the mean form up to translation, rotation and reflection is given by

$$\hat{M} = \begin{bmatrix} 0 & 0 & 0 \\ \sqrt{\hat{\lambda}_1} \hat{d}_1 & \sqrt{\hat{\lambda}_2} \hat{d}_2 & \sqrt{\hat{\lambda}_3} \hat{d}_3 \end{bmatrix}. \text{ These estimators are computationally simple and are}$$

especially convenient with a large number of landmarks and easily extend from two- to three-dimensional data.

Maximum likelihood estimators

Dryden and Mardia (1991) derived the exact distributions of the form coordinates as well as of the shape coordinates. One can maximize the likelihood function to obtain maximum likelihood estimators of the mean form as well as the covariance parameters. It should be noted again that these parameters are identifiable only to within the orbit defined by the group structure. A significant drawback of this approach is that the exact distributions of form and shape coordinates are mathematically complex for large numbers of landmarks and for three-dimensional objects. It is not uncommon in medical problems to need at least 10, 15 or more landmarks to represent an object reasonably. In

such a situation, the exact shape distribution is extremely complicated (see equation 2.4 of Dryden and Mardia, 1991). It involves telescoping sums increasing geometrically with the number of landmarks.

All the maximal invariants are equivalent in the sense that maximum likelihood estimators based on any of them give estimates on the same orbit. The use of maximal invariance to eliminate nuisance parameters usually leads to some loss of efficiency as compared to the situation where the nuisance parameters are known. As shown earlier, the method of moments estimators based on the maximal invariant $T(X)$ are easy to obtain. We next study the loss of efficiency of method of moments as compared to maximum likelihood based estimators.

Table 1 gives the result of a small simulation study comparing the method of moments method based on the inter-landmark differences (EDMA), the MLE based on the Dryden and Mardia distribution and the MLE based on the (unobserved) data before translation or rotation. Thus the “unobservable MLE” represents an idealized situation for comparison since the nuisance parameters, R_i and t_i , are assumed known.

In estimating the inter-landmark distances or the variance covariance parameters, none of the methods exhibited any bias, hence Table 1 concentrates on the relative root mean square error, which primarily reflects variability. Since performance of the methods was similar within the mean parameters, within the variance parameters and within the covariance parameters, the results are averaged across the (three in each case) parameters.

We can see that, for estimating the mean form, both EDMA and the MLE do as well as the unobservable MLE. For estimating the variance parameters EDMA and the MLE do worse than the unobservable MLE and EDMA does slightly worse than the MLE. For the covariance parameters there is some loss of efficiency in using EDMA compared to the MLE, but not a dramatic amount.

Another point is worth mentioning in comparing EDMA with the MLE: even for such a simple case of three landmarks in two dimensions, we found the likelihood very difficult to maximize numerically using the Dryden and Mardia (1991) distribution. Despite estimating the variance-covariance matrix using a Cholesky decomposition to guarantee positive semi-definiteness, many data sets required new starting values and/or needed a preliminary sizing step for the variance-covariance parameters in order to achieve convergence. Despite repeated attempts there were still 2 data sets (out of 100) in the first parameter configuration for which we could not obtain convergence. More details of the simulation are given in the Appendix.

Superimposition-based methods of estimation

As a practical alternative to maximum likelihood estimation, Goodall and Bose (1987) and Goodall (1991) discuss superimposition-based estimators. See also Bookstein (1986) for a similar estimator based on shape coordinates. The basic idea is to first superimpose the objects as closely as possible and then use least squares to estimate the parameters. Intuitively, this attempts to remove the unknown translations and rotations so that the

data can be treated as multivariate normal with common mean and variance. If the unknowns could be removed, estimation of the parameters would be straightforward. But this can not be done. As a consequence, these methods lead to inconsistent estimates in the general case where $\Sigma_K \neq I$ and $\Sigma_D \neq I$ (Lele, 1993; Kent and Mardia, 1997). A further and an important implication of this result is that the covariance estimator is also inconsistent. Lele (1993) illustrates this fact by simulations and provides an intuitive explanation for it. Estimation of the variability is central to any statistical analysis. The lack of consistency of the covariance parameters estimators based on Generalized Procrustes Analysis is thus all the more problematic.

4. Comparison of form: The two-sample case

We now proceed to apply the principle of invariance to the important case of comparison between two or more populations. Along the lines developed in the first part of this paper, it can be shown that the mean forms are identifiable only up to translation, rotation and reflection. Identifiability of the covariance parameters is also similar to the one-sample case. Now suppose, using two samples, we have estimated the mean forms for each of the population using either the method of moments or the maximum likelihood based on the maximal invariants. The question of scientific interest is that of studying the difference in the two mean forms.

As shown earlier, mean forms are identifiable only up to the orbit to which they belong

and not the precise location on the orbit. Hence we can only estimate differences in mean form which depend on the orbits. Conversely, we cannot estimate differences that require information as to a specific location on the orbit. Hence we make the following definition.

Definition 2: Let $Diff(M_1, M_2)$ be a measure of the difference between two mean forms M_1 and M_2 . Then $Diff(M_1, M_2)$ is an invariant measure of form difference if and only if $Diff(M_1, M_2) = Diff(M_1 R_1 + \underline{1}t_1^T, M_2 R_2 + \underline{1}t_2^T)$ for all orthogonal matrices R_i and translation vectors t_i .

In other words, the notion of form difference should not depend on which members of the two orbits, corresponding to M_1 and M_2 , are chosen to calculate the form difference.

One should be able to take any member of the first orbit, any member of the second orbit, and calculate a form difference that is invariant to the choice of the particular members.

There are several different measures of form difference that are used in practice. We examine some of the prominent ones in the light of the above definition.

The deformation approach to form difference

Let M_1 and M_2 be two mean forms under consideration. A typical way to represent the relationship between M_1 and M_2 is to consider M_1 being deformed into M_2 by a

function h . Typically a parametric family of functions represents this deformation. Thus we have $M_2 = h(M_1, \phi)$ and the values of the parameters ϕ help us understand the underlying biological mechanisms that are associated with this deformation.

To make the basic ideas understandable, let us consider the simplest case of two-dimensional objects with three landmarks each. Suppose we consider the simplest form of deformation, the affine deformation. Under this model: $M_2 = M_1 A + \underline{t}^T$ where A is any 2×2 matrix and \underline{t} is a translation vector. Again the parameter of interest is A and \underline{t} is the nuisance parameter. It is easy to see that $A = (LM_1)^{-1}(LM_2)$ where L is the translation matrix defined earlier. A physical interpretation can be given to this matrix A by writing its singular value decomposition in the following manner (Goodall and Green, 1986)

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} p & 0 \\ 0 & q \end{pmatrix} \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}$$

The angle θ corresponds to the initial rotation to the principal axes of M_1 , the singular values p and q correspond to the stretching along the principal axes and the angle φ corresponds to the rotation of the deformed M_1 to match exactly with M_2 . From the scientific point of view the parameters of interest are (θ, p, q) because they contribute to the deformation, whereas the parameter φ does not. Let us examine if this measure of

form difference is invariant in the sense of definition 2 above. In the notation used in the definition 2, $Diff(M_1, M_2) = (LM_1)^{-1}(LM_2)$, whereas,

$Diff(M_1 R_1 + \underline{1}r_1^T, M_2 R_2 + \underline{1}r_2^T) = R_1^{-1} Diff(M_1, M_2) R_2$. Thus this measure of form difference is not invariant. It depends on which members of the orbits are chosen to evaluate the affine deformation. However, if one defines the form difference to be the singular values of the affine deformation matrix A, then it is clear that singular values are an invariant measure of form difference.

The above result can also be proved by considering the distribution of the maximal invariant and an appeal to Theorem 1. Consider the case when $\Sigma_D = I$ and for notational

simplicity we write $R(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$. The maximal invariants $T(X)$ and $T(Y)$

follow non-central Wishart distribution with non-centrality parameters $LM_1(LM_1)^T$ and $LM_2(LM_2)^T$. Recall the fact that $M_2 = M_1 A + t$. Consider the parameter

combinations $(M_1, A, \Sigma_{K,1}, \Sigma_{K,2}, I, I)$ and $(M_1 R(\theta), \begin{pmatrix} p & 0 \\ 0 & q \end{pmatrix}, \Sigma_{K,1}, \Sigma_{K,2}, I, I)$. For

these two parameter combinations, the inner product matrices $LM_1(LM_1)^T$ and

$LM_2(LM_2)^T$ are identical and hence the distributions of the maximal invariant are

identical. Hence the singular values of the affine deformation matrix A are identifiable

but not its eigenvectors. The more general situation follows in a similar fashion.

The affine deformation is the simplest type of deformation function. In the literature on

the statistical analysis of landmark data, more complex forms of deformation functions are suggested. For example, Bookstein (1989, 1991, and 1996) suggests the use of thin plate splines as a deformation function. A touted advantage of thin plate splines is that it depicts the deformation geometrically and that the total deformation can be decomposed into several orthogonal components which, in turn, can be studied in order to localize the form differences. Such localization, presumably, helps understand the biological processes that might be responsible for the form change. We refer the reader to Bookstein's work (for example, see the figure on page 347 in Bookstein, 1996) for more details and various situations where thin plate splines are applied to real world problems.

The idea behind thin plate splines is very similar to the kriging methodology (Cressie, 1990). We have two sets of coordinates (x,y) and (z,w) 's. We find a continuous function that maps (x,y) coordinates onto (z,w) coordinates in such a fashion that the integral of the second derivative (called the bending energy) is minimized. The solution to this optimization problem is the well known cubic splines interpolant (Wahba, 1990). The function $h(\cdot)$ that maps one object onto the other in this fashion can be decomposed into orthogonal components (each component corresponding to the basis functions), the first component corresponds to the affine component described earlier and the rest are the non-affine components. The plots depicting non-affine components are claimed to help localization of form differences in the two forms. Is such a decomposition invariant to the group action of translation, rotation and reflection. Instead of describing the algebraic details of this result, we show the implications of such lack of invariance graphically in Figure 1. Two, two-dimensional mean forms with five landmarks each were chosen. The

landmark coordinates for the two figures are reported in Table 3. We used the TPSPLINE (Rohlf, 1994) program that provides graphical presentation of these decompositions. The left-hand column of the figure depicts the orthogonal decompositions obtained for the original figures. The right hand column depicts the orthogonal decompositions obtained after rotating one of the figures by 90 degrees. It is obvious that the two sets of figures are different from each other that is, the decompositions are not invariant to rotation. This greatly complicates the use of thin plate spline methodology since it is difficult to separate the invariant features from those that are not invariant when looking at a graphical representation of the results.

Superimposition methods for comparison of forms

We now examine the superimposition method for comparing two forms. For details on this method, we refer the reader to Siegel and Benson (1982), Rohlf and Slice (1990) and Lele (1991).

The basic idea is to first “superimpose” the two estimated forms M_1 and M_2 by translation and/or rotation and centering and then directly calculate the difference. There are a number of different ways to achieve the superimposition, but we need not be concerned with the details here. Suppose M_1 is rotated using R_1 and translated using t_1 , where the rotation and translation can depend on either or both of the estimated mean forms. The superimposition estimate of mean form difference is then: $LM_1R_1 - LM_2R_2$. Is this an invariant measure of form difference?

No - for the simple reason that, since the mean forms are only determined up to their orbit this measure cannot distinguish between different rotations of either of the true mean forms. Stated another way, if we describe mean form difference using

$$\text{Diff}(M_1, M_2) = LM_1 - LM_2, \text{ this is not invariant, since}$$

$$\text{Diff}(M_1 R_1, M_2 R_2) = LM_1 R_1 - LM_2 R_2.$$

Let us look at superimposition based measures of form difference using the distribution of the maximal invariant and Theorem 1. Let us define $S(\delta) = LM_2 R(\delta) - LM_1$ where the rotation matrix may depend on one or both the mean forms. Following the same arguments as in the case of non-identifiability of the eigenvectors of the affine deformation, it is clear that the distribution of the maximal invariant for the parameters $(M_1, S(\delta), \Sigma_{K,1}, \Sigma_{K,2}, \Sigma_{D,1}, \Sigma_{D,2})$ is identical for all values of $\delta \in (0, 2\pi]$. Thus the form difference $S(\delta) = LM_2 R(\delta) - LM_1$, defined in terms of coordinatewise differences is non-identifiable.

We would like to point out that the quantity that is estimated by the superimposition method of form difference depends on two arbitrary choices, the centering matrix L and the superimposition metric or the fitting criterion. These choices are external to the group structure. Fixing of these choices is analogous to the fixing of a linear constraint among the α_i in the model for a one-way ANOVA: $E[y_{ij}] = \mu + \alpha_i$. For the sake of estimation, an arbitrary constraint may be necessary, however when it comes to scientific and statistical inference, such a constraint should have no effect. For example, in the

analysis of variance, inferences are based only on the contrasts between the α_i and not on the α_i themselves.

Comparisons based on distances

In the classical approaches to the problem of size and shape (Huxley, 1932; Mosimann, 1970), the forms and shapes are compared in terms of the distances between landmarks. This approach is extended by Lele and his colleagues (Lele, 1991, 1993; Lele and Richtsmeier, 1991, 1992, 1995, Richtsmeier and Lele, 1992, Lele and Cole, 1996). Do these methods satisfy the invariance principle?

Consider a mean form coordinate matrix M . Define a new matrix, denoted by $F(M)$, where $F_{ij}(M)$ = Euclidean distance between landmarks i and j . This is a symmetric, $K \times K$ matrix with diagonal elements 0. This is called a Euclidean Distance Matrix corresponding to the form M . In Lele (1991, 1993) it is proved that this is also a maximal invariant under the group of transformations consisting of translation, rotation and reflection. Notice that the maximal invariant $T(X)$, the inner product matrix, defined in the first part of the paper as well as the form coordinates, depend on the choice of L , the translation matrix. However, the Euclidean distance matrix, which is also a maximal invariant, does not depend on the choice of the translation matrix L as well. Euclidean Distance Matrix Analysis (EDMA) consists of the statistical procedures based on the analysis of these matrices.

Because $F_{ij}(M)$'s are continuous functions, it is clear that $F(\hat{M}_1)$ and $F(\hat{M}_2)$ are consistent estimators of the corresponding Euclidean distance matrices for the mean form of the two populations. These quantities are thus estimable. Moreover, any continuous function of $F(M_1)$ and $F(M_2)$ is also estimable. Thus, any procedure that is based only on the entries of these matrices, is estimable and is invariant to the choice of the translation matrix L and rotation and reflection. For example, the testing procedures and confidence intervals proposed in Lele and Richtsmeier (1991, 1995) satisfy the principle of invariance. Similarly procedures suggested by Mosimann (1970) and by Rao and Suryawanshi (1996) are also invariant because they are based on distances between the landmarks.

Comparison of growth patterns and s-sample comparisons

One of the major problems confronting modern biology is to understand how complex morphological structures arise during development and how they are altered during evolution (Atchley and Hall, 1991). (SL: this is not in the references) An important component of these processes is the way in which component structures rearrange relative to one another with overall increase in size. A growth pattern is defined as the composite of geometrical changes in structure occurring through time (Richtsmeier and Lele, 1993). Gould (1977) discusses the relationships between growth patterns and their implications in the evolutionary context. Richtsmeier and her colleagues study the

implications of growth patterns in clinical situations involving genetic disorders (Dufresne and Richtsmeier, 1995). A theoretical framework for studying growth using the EDMA was developed in Richtsmeier and Lele (1992).

In the following, we provide the definition of an invariant growth comparison measure, parallel to form difference measure. Let M_1 and M_2 denote the mean forms at age category 1 and 2 respectively for one of the species. Let N_1 and N_2 denote the mean forms at age category 1 and 2 respectively for the other species. We want to know how the change from M_1 to M_2 compares with the change from N_1 to N_2 .

Definition 3: Let $GDiff(M_1, M_2, N_1, N_2)$ be a measure of growth difference between the two classes. We say that this measure is invariant if and only if

$$GDiff(M_1, M_2, N_1, N_2) = GDiff(M_1 R_1 + \underline{v}_1^T, M_2 R_2 + \underline{v}_2^T, N_1 R_3 + \underline{v}_3^T, N_2 R_4 + \underline{v}_4^T).$$

In other words, the growth comparison measure should not depend on which members of the four orbits are chosen to calculate it. Again appealing to the one sample case, it is clear that only the orbits are identifiable and not the individual members and hence growth comparison measure should depend only on the orbit and not on the choice of particular members. The growth difference matrix used in Richtsmeier and Lele (1992) for growth comparison can be shown to satisfy the invariance criterion. We are not aware of measures of growth comparison based on deformation or superimposition. However, any such measures, should they be proposed, should satisfy the invariance principle

before any scientific or statistical inferences based on them are drawn. Invariance of multi-species comparison measures can be defined in a like manner.

Remark: There are two types of data that may be available for studying growth, longitudinal and cross-sectional. The above definition is clearly relevant when the data are cross-sectional. However there might be situations, for example in the study of plant growth, where the individual is naturally sessile and the coordinate system becomes fixed at the first time point. In such a situation the growth of a particular individual can be studied with respect to this fixed coordinate system. We would then define *GDiff* with $R_1 = R_2$, $R_3 = R_4$, $t_1 = t_2$, and $t_3 = t_4$. Thus the comparison of growth patterns between two individuals (or two populations) should be invariant to the choices of R_1, t_1, R_2 , and t_2 .

5. Discussion

This paper studied implications of the presence of nuisance parameters and group invariance to the statistical analysis of landmark coordinate data. We illustrated the use of maximal invariance in the estimation of the parameters of the Gaussian perturbation model. A method based on the method of moments is shown to be computationally simple and almost as efficient as the one based on the method of maximum likelihood. Another advantage of the method of moments is its relative ease of application in the presence of missing data.

The principle of invariance was then applied to the important situation of comparison of two populations. It was shown that the popular methods of superimposition and deformation do not satisfy this principle in general. It was also shown that the method of comparison based on distances between landmarks does satisfy the principle of invariance. The goal of morphometrics is not limited to classification of individuals (numerical taxonomy) or simple null hypothesis testing of the equality of shapes or forms. It goes beyond that. When the null hypothesis of equality of forms or shapes is rejected, the scientifically important issue is to localize the differences between the two forms. Such localization helps scientists to gain insight into the biological processes that may be driving the form changes. This has important implications in the fields of corrective plastic surgery, brain mapping, and genetics among others. Invariance is an important issue when such localization is attempted. It is our contention that the methods for the analysis of biological shapes (morphometrics) be scrutinized carefully with the principle of invariance in mind. Our conclusion from such a scrutiny of some of the popular methods is that the methods based on deformation and superimposition do not satisfy this principle in general, whereas the methods based on the distances between landmarks do.

We contend that whenever a new technique for the statistical analysis of landmark coordinates is proposed, invariance or non-invariance of such a technique be checked by relating the quantities utilized in the technique to the underlying parameters and showing that the parameters that are being estimated are in fact identifiable and estimable under the group structure.

References

- Alam, K. and Mitra, A. (1990) On estimating the scale and noncentrality matrices of a Wishart distribution. *Sankhya, Ser. B* 52:133-143.
- Arnold, S.F. (1981) *Theory of Linear Models and Multivariate Analysis*. John Wiley and Sons, New York, U.S.A.
- Atchley, W.R. and Hall, B.K. (1991) A model for development and evolution of complex morphological structures. *Biological Reviews* 66, 101-157.
- Berger, J.O. (1980) *Statistical Decision Theory*. Springer-Verlag, NY.
- Bookstein, F. (1986) Size and Shape Spaces for Landmark Data in Two Dimensions. *Stat. Science* 1:181-242.
- Bookstein, F.L. (1989) Principal Warps: Thin plate splines and the decomposition of deformations *IEEE Trans. Pattern Anal. Machine Intelligence* 11:567-585.
- Bookstein, F.L. (1991) *Morphometric Tools for Landmark Data: Geometry and*

Biology. Cambridge University Press, Cambridge, UK.

Bookstein, F.L. (1996) Biometrics, Biomathematics and the morphometric synthesis. *Bulletin of Mathematical Biology* 58:313-365.

Bookstein, F.L., B. Chernoff, R. Elder, J. Humphries, G. Smith and R. Strauss (1985) *Morphometrics in Evolutionary Biology*. Philadelphia: Academy of Natural Sciences of Philadelphia.

Cressie, N.A.C. (1990) Reply to Wahba's letter. *American Statistician*, 44, 256-258.

Corner, B.D., S. Lele and J.T. Richtsmeier (1992) Measuring Measuring precision of three dimensional landmark data. *Quantitative Anthropology* 3(4):347-359.

deGunst, M.C.M. (1982) On the distribution of General Quadratic Forms in Normal Variables. *Statistica Nierlandica*. 245-251.

Dufresne, C. and Richtsmeier, J.T. (1995) The interaction of craniofacial dysmorphology, growth and the prediction of surgical outcome. *J. of Craniofacial Surgery* 4(4):234-244.

Dryden, I.L. and Mardia, K.V. (1991) General shape distributions in a plane. *Adv. Appl. Probab.* 23:259-276.

Godzik, A. (1996) The structural alignment between two proteins: Is there a unique answer? *Protein Science* 5:1325-1338.

Goodall, C. (1991) Procrustes Methods in the Statistical Analysis of Shape. *J. Royal Stat. Soc. Ser. B* 53:285-339.

Goodall C. and Bose, A. (1987) Models and Procrustes Methods for the Analysis of Shape Difference. *Proc. 19th Symp. Interface between Computer Science and Statistics*, Philadelphia, PA. 86-92.

Goodall, C.R. and P.B. Green (1986) Quantitative Analysis of Surface Growth. *Botanical Gazette*, 147:1-15.

Gould, S. J. (1977) *Ontogeny and Phylogeny*. Belknap Press, Cambridge, Massachusetts, USA.

Hall, B.K. (1994) *Homology: The hierarchical basis of comparative biology*. Academic Press, NY.

Huxley, J. (1932) *Problems of Relative Growth*. MacVeagh, London, UK.

Kendall, D.G. (1989) The Statistical Theory of Shape (with discussion) *Statist. Science* 4:87-120.

Kent, J.T. and Mardia, K.V. (1997) Consistency of Procrustes Estimators. *J. R. Statist. Soc. Ser. B.* 59:281-290.

Lehmann, E. L. (1959) *Testing Statistical Hypothesis*. John Wiley and Sons, NY.

Lele, S. (1991) Some comments on Coordinate Free and Scale Invariant Methods in Morphometrics. *Am. J. Phys. Anthropol.* 85:407-418.

Lele, S. (1993) Euclidean Distance Matrix Analysis (EDMA): Estimation of mean form and mean form difference. *Mathe. Geol.* 25: 573-602.

Lele, S. and Cole, T. M. (1996) A new test for shape difference when variance-covariance matrices are unequal. *J. Human Evolution.* 31:193-212.

Lele, S. and Richtsmeier, J.T. (1991) Euclidean Distance Matrix Analysis: A Coordinate Free Approach for comparing biological shapes. *Am. J. Phys. Anthropol.* 86:415-427.

Lele, S. and Richtsmeier, J.T. (1992) On comparing biological shapes: Detection of influential landmarks. *Am. J. Phys. Anthropol.* 87:49-65.

Lele, S. and Richtsmeier, J.T. (1995) Euclidean Distance Matrix Analysis: Confidence Intervals for Form and Growth Differences. *Am. J. Phys. Anthropol.* 98:73-86.

Mosimann, J.E. (1970) Size Allometry: Size and Shape variables with characterizations of the lognormal and generalized gamma distributions. *J. Am. Statist. Assoc.* 63:930-978.

Neyman, J. and Scott, E. J. (1948) Consistent Estimates Based on Paritally Consistent Observations. *Econometrika* 16:1-32.

Rao, C.R. and Suryawanshi, S. (1996) Statistical analysis of shape of objects based on landmark data. *Proc. Natl. Acad. Sci.* 93:12132-12136.

Richtsmeier, J.T. (1993) Beyond morphing: visualization to predict a child's skull growth. *Advance Imaging* 8(7):24-27.

Richtsmeier, J.T., Fang, R. and Raghavan, R. (1997) Volumetric morphing, the study of craniofacial growth and the prediction of surgical outcome. (submitted manuscript)

Richtsmeier, J.T. and Lele, S. (1993) A coordinate free approach to the analysis of growth patterns: models and theoretical considerations. *Biol. Rev.* 68:381-411.

Rohlf, F.J. and Slice, D. (1990) Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks. *Syst. Zool.* 39:40-59.

Rohlf, F.J. (1994) TPSPLINE, version 9|14|94 DPMI Stony Brook, NY: Department of

Ecology and Evolution, State University of New York, U.S.A.

Roth, V.L. (1988) The Biological Basis of Homology. In *Ontogeny and Systematics* (C.J. Humphries, ed.)

Siegel, A.F. and Benson, R.H. (1982) A robust comparison of biological shapes.
Biometrics 38:341-350.

Stoyan, D. (1990) Estimation of distances and variances in Bookstein's landmark model.
Biometr. J. 32:843-849.

Thompson, D. W. (1917) *On growth and form*. Cambridge University Press, Cambridge, UK.

Valeri, C, Cole, T. C., Lele, S. and Richtsmeier, J.T (1997) Capturing data from three dimensional surfaces using fuzzy landmarks (submitted to *American Journal of Physical Anthropology*)

Van Valen, L.M. (1982) Homology and causes. *J. Morphol.* 173:305-312.

Wahba, G. (1990) *Spline Models for Observational Data*. SIAM, Philadelphia.

Table 1: Efficiency comparisons between the method of moments and the method of maximum likelihood. The following are the percent relative root mean squared errors for the two methods based on 100 simulations. Samples of size 30 were generated under two different mean forms and two different covariance structures (see Table 2). The column 'Unobserved MLE' corresponds to the relative root mean squared error for the maximum likelihood estimators based on assuming the nuisance parameters are known. Hence these values represent best achievable results.

Scenarios	Parameters	EDMA	MLE	Unobserved MLE
$(M_1, \Sigma_{K,1}, I_2)$	Mean distances*	1.0	1.0	1.0
	Variances*	23.2	19.5	16.3
	Covariances*	45.6	37.7	28.4
$(M_1, \Sigma_{K,2}, I_2)$	Mean distances	1.5	1.5	1.5
	Variances	21.4	18.7	14.6
	Covariances	36.4	31.0	24.7
$(M_2, \Sigma_{K,1}, I_2)$	Mean distances	2.2	2.2	2.2
	Variances	24.2	19.4	14.9

Covariances		48.4	41.3	30.1
-------------	--	------	------	------

* Because of the non-convergence of the maximum likelihood routine, these are based on only 98 out of 100 simulations.

Table 2: The mean form coordinates (centered) and the covariance matrices used in the simulation study.

$$\Sigma_{K,1} = \begin{bmatrix} 0.78 & 0.44 & 0.28 \\ 0.44 & 0.66 & 0.40 \\ 0.28 & 0.40 & 0.53 \end{bmatrix}$$

$$\Sigma_{K,2} = \begin{bmatrix} 0.78 & 0 & 0 \\ 0 & 0.66 & 0 \\ 0 & 0 & 0.53 \end{bmatrix}$$

$$M_1 = \begin{pmatrix} 0 & 4.72 \\ 7.07 & -2.36 \\ -7.07 & 2.36 \end{pmatrix}$$

$$M_2 = \begin{pmatrix} -1 & 2.33 \\ 4 & -0.67 \\ -3 & -1.66 \end{pmatrix}$$

Table 3: The landmark coordinates for the two objects whose orthogonal decompositions based on thin plate splines are depicted in Figure 1.

Object 1 =	$\begin{pmatrix} 3.6929 & 10.3819 \\ 6.5827 & 8.8386 \\ 6.7756 & 12.0866 \\ 4.8189 & 11.2047 \\ 5.6969 & 10.0748 \end{pmatrix}$	Object 2 =	$\begin{pmatrix} 3.9724 & 6.5354 \\ 6.6969 & 4.1181 \\ 6.5394 & 7.2362 \\ 5.4016 & 6.4528 \\ 5.7756 & 5.1142 \end{pmatrix}$
------------	---------------------------------------------------------------------------------------------------------------------------------	------------	-----------------------------------------------------------------------------------------------------------------------------

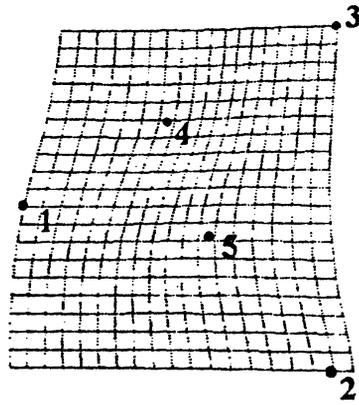
APPENDIX: Computational Details for the Simulation

Data were simulated from the Gaussian perturbation model using MATLAB's built-in normal random number generator. Method of moments estimators were calculated using the formulas from Lele (1993). Maximum likelihood estimators were calculated by numerical maximization of the likelihood using GAUSS and its quasi-Newton optimization routine, OPTMUM. To maintain positive definiteness of the estimated Σ_K the Cholesky square root of the matrix was estimated. Estimation of the parameters in Σ_K caused problems for a large percentage of data sets. For those where preliminary non-convergence occurred, the parameters were scaled in order to improve the numerical aspects of the problem. Despite these precautions and the choice of different options governing the type of quasi-Newton algorithm used, there were two data sets (out of the 100 simulated) in the first scenario (Table 1) for which convergence could not be achieved.

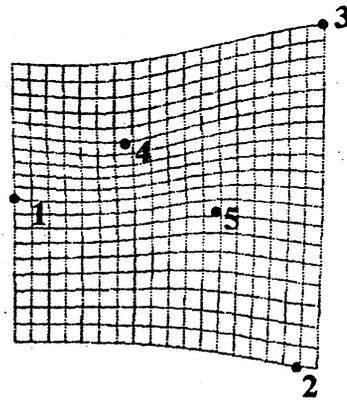
Figure 1 Caption:

Graphical depiction of orthogonal decomposition based on Thin plate splines. First column corresponds to the decomposition based on the original orientation of the two objects whereas the second column corresponds to the decomposition obtained after rotating the second object by 90 degrees. It is clear that the orthogonal decompositions based on Thin plate splines are not invariant to rotation.

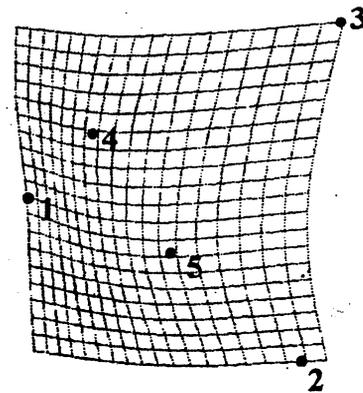
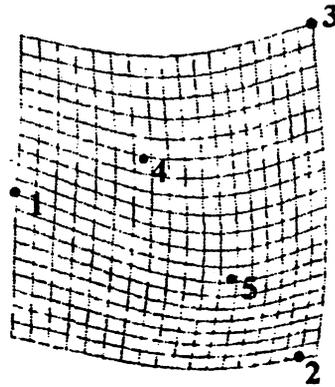
**Object 2 in
Original
Orientation**



**Object 2
Rotated
90 Degrees
Clockwise**



Partial Warp 1



Partial Warp 2