Empirical Bayes hierarchical models for regularizing maximum likelihood estimation in the matrix Gaussian Procrustes problem

Douglas L. Theobald* and Deborah S. Wuttke

Department of Chemistry and Biochemistry, UCB 215, University of Colorado, Boulder, CO 80309

Edited by Richard V. Kadison, University of Pennsylvania, Philadelphia, PA, and approved September 26, 2006 (received for review September 30, 2005)

Procrustes analysis involves finding the optimal superposition of two or more "forms" via rotations, translations, and scalings. Procrustes problems arise in a wide range of scientific disciplines, especially when the geometrical shapes of objects are compared, contrasted, and analyzed. Classically, the optimal transformations are found by minimizing the sum of the squared distances between corresponding points in the forms. Despite its widespread use, the ordinary unweighted least-squares (LS) criterion can give erroneous solutions when the errors have heterogeneous variances (heteroscedasticity) or the errors are correlated, both common occurrences with real data. In contrast, maximum likelihood (ML) estimation can provide accurate and consistent statistical estimates in the presence of both heteroscedasticity and correlation. Here we provide a complete solution to the nonisotropic ML Procrustes problem assuming a matrix Gaussian distribution with factored covariances. Our analysis generalizes, simplifies, and extends results from previous discussions of the ML Procrustes problem. An iterative algorithm is presented for the simultaneous, numerical determination of the ML solutions.

heteroscedasticity | morphometrics | Procrustes analysis | superpositions | least-squares

The goal of Procrustes analysis is to superpose non-identical shapes in an optimal manner via scalings, translations, and rotations (1–3). Classical Procrustes analysis uses the unweighted least-squares (LS) criterion for finding the optimal transformations. LS implicitly assumes that the landmarks describing the forms' shapes are uncorrelated and that they have identical variances (i.e., that they are homoscedastic). However, in many practical applications these assumptions are known to be violated. For instance, when superpositioning macromolecular proteins, individual atoms (the landmarks) are connected via covalent chemical bonds, and thus the variance of a given atom can be correlated with the variance of atoms to which it is connected. Furthermore, certain atoms may have larger mobilities than others, or they may have spatial positions with relatively greater uncertainty due to experimental error. Similarly, when comparing the skulls from different members of a species, some homologous features may be highly variable relative to others. Hence, different landmarks can have widely different variances. Under these conditions, ordinary LS can give misleading results, even with large samples of data (4).

The method of maximum likelihood (ML) is a common alternative to LS and is widely considered to be fundamental for statistical modeling and parameter estimation (5). Given the proper model, ML can provide accurate and robust estimates of parameters in the presence of both heteroscedasticity and correlation. Incorporating heterogeneous variances and non-zero correlations into the ML Procrustes problem involves weighting by two covariance matrices (a landmark and a dimensional covariance matrix). However, estimation of these covariance matrices has been a major stumbling block prohibiting a viable ML Procrustes analysis (1, 3, 6–10). Without special assumptions, simultaneous estimation of the landmark covariance matrix and the translations is generally impossible. By parametrically constraining the landmark covariance matrix using a hierarchical, empirical Bayes likelihood model, we provide a relatively simple solution that permits valid covariance estimation.

In addition to allowing simultaneous estimation of the landmark covariance matrix and the translations, we address several additional challenges that have thus far impeded successful ML inference in the Procrustes problem. (*i*) As usually described, Procrustes analysis gives degenerate estimates of the covariance matrices even when the translations are known (7). We suggest a method for uniquely identifying the covariance matrices, using the fact that the absolute reference frame in Procrustes analysis is arbitrary. (*ii*) The dimensional covariance matrix introduces further difficulties for optimal rotation estimation, and current methods involve a circuitous and computationally intensive numerical procedure. We provide a simplified, analytic solution to estimating the optimal rotations when using an arbitrary dimensional covariance matrix. (*iii*) We give a simplified ML estimator for the scaling factors and also present a hierarchical model for estimating the scalings that can improve estimation in common biological cases. (*iv*) The ML estimator of the mean form is known to be inconsistent (i.e., it converges to the wrong value as the number of forms increases). Based on the approximate χ^2 distribution of the Procrustes sum of squared distances, we derive an easily calculable bias-corrected ML estimator for the mean. Given that the errors are not large, this estimator is consistent for landmarks with uncorrelated, nonisotropic ellipsoidal variability. (*v*) Finally, we present a relatively simple and stable iterative algorithm for the simultaneous solution of these unknown parameters.

Formulation of the ML Procrustes Problem

In statistical shape theory, the shape of a geometrical configuration is commonly defined as the geometrical information that is unchanged after translation, rotation, and scaling of the object. Particular instances of an object, which are described by a set of points summarizing the object's shape, we refer to as a "form." We also define "size-and-shape" as those geometrical aspects of a form that are unchanged after translation and rotation. In its most general formulation, Procrustes analysis involves the optimal matching of two or more form matrices. Here we consider only sets of forms where each form matrix has the same dimensions.

Consider N different forms (X_i, $i = 1 \ldots N$), each with K corresponding points or "landmarks" that depict the form's shape. Each form is defined as a $K \times D$ matrix holding K rows

Author contributions: D.L.T. and D.S.W. designed research; D.L.T. performed research; D.L.T. contributed new reagents/analytic tools; D.L.T. and D.S.W. analyzed data; and D.L.T. and D.S.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Freely available online through the PNAS open access option.

Abbreviations: LS, least squares; ML, maximum likelihood.

*To whom correspondence should be sent at the present address: Brandeis University, MS013, Waltham, MA 02454. E-mail: dtheobald@brandeis.edu.

of landmarks, where each landmark is a D-vector in $\Re^D$. For example, when superposing protein structures, the forms are independent protein models and the landmarks are the three-dimensional atomic coordinates of each atom in the protein structures. Furthermore, each form is observed in a different, arbitrary, and unknown coordinate system (1, 10, 11). The absolute reference frame for the forms is arbitrary; only relative rotational and translational transformations are of interest. We wish to estimate the optimal orthogonal rotations, translations, and scaling factors that, when applied to this set of form matrices, predict the observed data with the highest likelihood.

Gaussian Perturbation Model with Factored Covariances

To analyze the Procrustes problem from a likelihood perspective, one must choose a statistical model that is assumed to generate the observed data. We assume a perturbation model in which each form X_i is distributed according to a Gaussian (normal) probability density (7, 11). In this Gaussian perturbation model, each X_i can be considered to be an arbitrarily scaled, rotated, and translated Gaussian perturbation of a mean form M

$$\mathrm{X}_i = \frac{1}{\beta_i}(\mathrm{M} + \mathrm{E}_i)\mathrm{R}_i' - 1_K\mathrm{t}_i, \qquad [1]$$

where R_i is a $D \times D$ rotation matrix, β_i is a scaling parameter, t_i is a $1 \times D$ row vector for the translational offset, and 1_K denotes the $K \times 1$ column vector of ones. The matrix E_i is a zero-mean matrix of Gaussian random errors $\mathrm{E}_i \approx N_{K,D}(0, \Sigma, \Xi)$ with the same dimensions as the mean form M, where Σ and Ξ are covariance matrices described in detail below. In other words, to generate each of the observed forms in D-dimensional Euclidean space, each landmark in an idealized mean form M is randomly perturbed according to a Gaussian distribution. The perturbations for each landmark may be unequal on average and may be correlated with each other, as specified by the covariance matrices. Each perturbed form is then arbitrarily scaled, rotated, and translated, the results of which constitute the observed data.

Covariance matrices describe the variability and correlations among the landmarks in the forms. In the above model, we use factored covariances by assuming that the overall $KD \times KD$ covariance matrix Ω can be decomposed adequately as the Kronecker product of Σ and Ξ (i.e., $\Omega = \Sigma \otimes \Xi$). Here, Σ is a $K \times K$ covariance matrix for the rows of the forms (the landmarks), and Ξ is a constrained $D \times D$ covariance matrix for the columns of the forms (the dimensions). In the protein structure example, the landmark covariance matrix Σ describes the variances and correlations among the atoms. Additionally, the dimensional covariance matrix Ξ describes the variances and correlations between the three coordinate axes, for example, if atomic positions are on average more variable in one dimension than another. We note that, in many cases, the factored covariance model is physically unrealistic, because it assumes that each of the landmarks have the same ellipsoidal principal axes of variability. However, we accept this assumption as a simplifying approximation, and note that it is more general than the assumptions underlying classical LS Procrustes analysis.

Although the product $\Sigma \otimes \Xi$ is uniquely identified, Σ and Ξ independently are not, because for all $c > 0$, $\Omega = \Sigma \otimes \Xi = c\Sigma \otimes 1/c\ \Xi$. Thus, to jointly identify both Σ and Ξ, we impose the constraint that $\mathrm{tr}\Xi = D$. Furthermore, as a generality in Procrustes analysis, the absolute rotational frame of reference is arbitrary, and only the relative rotations among the forms are of interest. As a result, Ξ is only identifiable up to an arbitrary orthogonal similarity transformation (7). However, any given Ξ can be spectrally decomposed such that $\Xi = \mathrm{V}\Delta\mathrm{V}'$. Then, the diagonal matrix of eigenvalues Δ is equivalent to the Ξ that would result if the mean form M were rotated with the orthogonal matrix V' of left eigenvectors. Thus, for unique identifiability, we require that the axes of the mean form's reference frame be aligned with the principle axes of the dimensional covariance matrix Ξ. This requirement also guarantees that Ξ is diagonal.

Procrustes Matrix Gaussian Likelihood Equation

A likelihood analysis requires an explicit statement of the likelihood function derived from the probability density function that describes the assumed statistical model. The full joint likelihood equation for the Gaussian Procrustes model given in Eq. **1** is obtained from a multivariate matrix normal distribution (12, 13). To simplify the appearance of the likelihood equation, we define the squared Frobenius Mahalanobis matrix norm as $\|\mathrm{U}\|^2_{\mathrm{V,W}} = \mathrm{tr}\{\mathrm{VU}'\mathrm{WU}\}$, which can be regarded as the sum of the squared elements of U with columns weighted by V and rows weighted by W. Letting $|\mathrm{U}|$ be the determinant of matrix U, the full Procrustes log-likelihood $\ell(\beta, \mathrm{R}, \mathrm{t}, \mathrm{M}, \Xi, \Sigma; \mathrm{X}) = \ell_P$ is given by

$$\ell_P = -\frac{1}{2}\sum_i^N \|\beta_i(\mathrm{X}_i + 1_K\mathrm{t}_i)\mathrm{R}_i - \mathrm{M}\|^2_{\Xi^{-1},\,\Sigma^{-1}} - \frac{NKD}{2}\ln(2\pi) - \frac{NK}{2}\ln|\Xi| - \frac{ND}{2}\ln|\Sigma|. \qquad [2]$$

The Procrustes log-likelihood equation above (Eq. **2**) is the key conceptual contribution underlying the ML analysis that follows. In order to solve the ML Procrustes problem, ML estimators must be derived for each of the unknown parameters in the Procrustes log-likelihood equation. We provide the individual ML estimates for each of these unknowns: the Ξ, Σ, M, and the N β_i, t_i, and R_i (for derivations see *Appendices*).

Regularization of the Covariance Matrices Allows Simultaneous Estimation

Estimation of the factored covariance matrices has been the primary impediment for a successful, practical implementation of a joint non-isotropic ML treatment of the Procrustes problem (1, 3, 6–10). We first present the unrestricted ML estimators of the covariance matrices and then discuss the difficulties that our approach remedies.

The unrestricted ML estimators of the covariance matrices, $\hat{\Xi}_U$ and $\hat{\Sigma}_U$, are given by straightforward variants of the usual covariance estimators for the matrix normal distribution (*Appendix I*) (4, 13).

$$\hat{\Sigma}_U = \frac{1}{ND}\sum_i^N (\beta_i\check{\mathrm{X}}_i\mathrm{R}_i - \hat{\mathrm{M}})\Xi^{-1}(\beta_i\check{\mathrm{X}}_i\mathrm{R}_i - \hat{\mathrm{M}})' \qquad [3]$$

$$\hat{\Xi}_U = \frac{1}{NK}\sum_i^N (\beta_i\check{\mathrm{X}}_i\mathrm{R}_i - \hat{\mathrm{M}})'\Sigma^{-1}(\beta_i\check{\mathrm{X}}_i\mathrm{R}_i - \hat{\mathrm{M}}), \qquad [4]$$

where $\check{\mathrm{X}}$ is a row-weighted centered form.[†] Because one of the covariance matrices must be normalized for identifiability, we (arbitrarily) choose to normalize $\hat{\Xi}_U$ by

$$\tilde{\Xi} = \frac{D\hat{\Xi}_U}{\mathrm{tr}\hat{\Xi}_U}, \qquad [5]$$

and use $\tilde{\Xi}^{-1}$ in place of Ξ^{-1} in calculating $\hat{\Sigma}_U$ given in Eq. **3**.

[†]Row-wise weighted centering can be applied to an uncentered structure X_i by $\check{\mathrm{X}}_i = \mathrm{X}_i + 1_K\hat{\mathrm{t}}_i$, where $\hat{\mathrm{t}}_i$ is the ML estimate of the translation. The ML translation is independent of the rotations R_i, the scaling factors β_i, and the dimensional covariance matrix Ξ, and it is given by $-(1_K\Sigma^{-1}\mathrm{X}_i/1_K\Sigma^{-1}1_K)$, where X_i is an uncentered structure (11).

Both of these covariance matrix estimators involve an inverse of a covariance matrix, as do several of the other ML solutions, a fact that presents three important and distinct difficulties for ML estimation in the Procrustes problem. First, in general, the ML estimators of the matrix normal covariance matrices only exist when the estimators are invertible, a requirement that obtains when $N \geq (K/D) + 1$ and $N \geq (D/K) + 1$ (13). In practice, these conditions are usually unsatisfied due to a paucity of available data. Second, Procrustes analysis requires centering the form matrices (e.g., Eq. **6**), which imparts a common linear constraint on the columns of the forms. Even with ample data, then, the sample landmark covariance matrix is collinear and rank deficient, with at least one zero eigenvalue (3, 6), and consequently it is noninvertible. Third, and most fundamentally, the displacements due to Σ and the translations t_i are linearly entangled (see Eq. **1**), and therefore unrestricted simultaneous estimation of both is impossible (6, 10). This latter problem is especially grave, because the unrestricted Procrustes estimators of the translations t_i and of the covariance matrix Σ are strongly interdependent; translations that closely superpose a given landmark decrease the estimated variance, and a small variance in turn preferentially weights the translations to center on that landmark. For example, it is always possible to find translation vectors that will perfectly superpose any selected landmark, thereby rendering its sample variance exactly zero, $\hat{\Sigma}$ singular, and the likelihood infinite.

To overcome the first two of these problems it is sufficient to guarantee the invertibility of the estimated landmark covariance matrix, i.e., all eigenvalues must be positive. In the third difficulty, the possibility of an infinite likelihood immediately points to a need for regularization, either of the translations (6) or of the covariance matrix, or of both.

In real data sets, the variances often are unable to take arbitrary values, but rather they describe coupled components of a larger, integrated system. For instance, the atoms in a macromolecule are linked by chemical bonds, and the bones in a skull are physically joined by ligaments and at sutures. In such cases, the variances for each landmark are similar in magnitude, and extremely small or large variances are improbable and physically unrealistic. Thus, the problem of simultaneous estimation of the covariance matrices and the translations can be circumvented by realistically restricting the variances of the landmarks.

An alternative, then, is to view each of the eigenvalues (λ_j) of the landmark covariance matrix as a draw from a positive distribution (where $\lambda_j > 0$). The eigenvalues of the covariance matrix can be considered as variances with all correlations eliminated, e.g., the eigenvalues of a diagonal covariance matrix are the variances themselves. The inverse gamma ($I\gamma$) distribution is used conventionally in Bayesian statistics as a natural proper prior for the variances in a multivariate Gaussian model (14). The flexibility of the $I\gamma$ distribution is convenient as it can adopt a wide variety of shapes approximating other distributions. Therefore, we regularize our matrix Gaussian Procrustes model using a hierarchical empirical Bayes approach (an extended likelihood) (5, 15) by assuming that the eigenvalues of the landmark covariance matrix belong to a common $I\gamma$ distribution. The full, regularized Procrustes log-likelihood (ℓ_r) is then the sum of the Procrustes log-likelihood from Eq. **2** and the log-likelihood of an $I\gamma$ distribution of the eigenvalues (see Eq. **15** in *Appendix II*): $\ell_r = \ell_P + \ell(\lambda)_{I\gamma}$.

Assuming an $I\gamma$ distribution for the eigenvalues of the landmark covariance matrix (with scale parameter α and shape parameter γ), the extended ML estimator $\hat{\Sigma}_{I\gamma}$ of Σ is a simple linear function of the unrestricted ML estimate $\hat{\Sigma}_U$ from Eq. **3** (see *Appendix II*)

$$\hat{\Sigma}_{I\gamma} = \frac{ND}{ND + 2(\gamma + 1)} \left(\frac{2\alpha}{ND} \mathrm{I} + \hat{\Sigma}_U \right). \qquad [6]$$

In this hierarchical model, the $I\gamma$ parameters α and γ are point estimates determined by the data, unlike when using a bona fide Bayesian prior. Eq. **6** can be interpreted as a shrinkage estimator that contracts the eigenvalues of the covariance matrix towards the mode of the inverse gamma distribution. Given positive α and γ parameters, this regularized model constrains all eigenvalues (and variances) of the landmark covariance matrix to be positive, thereby guaranteeing that $\hat{\Sigma}_{I\gamma}$ is invertible. Furthermore, this treatment enables simultaneous estimation of the landmark covariance matrix Σ and the translations by decoupling the strong dependence of their estimators. Thus, the inverse gamma regularization is sufficient to overcome all three of the difficulties listed above. In our experience, the $\hat{\Sigma}_{I\gamma}$ estimator is also well conditioned. We note that our regularized hierarchical estimator of the multivariate Gaussian covariance matrix also has potential for wide applicability beyond Procrustes analysis.

Simplified ML Estimates of the Rotations

Maximizing the Procrustes likelihood with respect to the rotations is equivalent to maximizing the first term of Eq. **2**, which requires weighting by the inverse of the dimensional covariance matrix Ξ. Unfortunately, there is apparently no general algebraic solution to this maximization problem using an arbitrary, symmetric, dimensional weight matrix (2, 11, 16). Koschat and Swayne have proposed a solution for an arbitrary diagonal $D \times D$ weight matrix using a computationally intensive, iterative numerical algorithm assuming that $\Sigma = \mathrm{I}$ (16). However, the Koschat–Swayne algorithm is reported to have poor convergence properties and is sensitive to the seed values (16).

In our ML Procrustes treatment, the dimensional weight matrix is not arbitrary; rather, we weight specifically by the inverse of the dimensional covariance matrix (i.e., $\hat{\Xi}_U^{-1}$). In this case, there is a much simpler, analytic solution to maximizing the estimated ML target for the optimal rotation. Here the dimensionally weighted Procrustes problem reduces conveniently (see *Appendix III*) to maximizing

$$\sum_i^N \mathrm{tr}(\hat{\Xi}^{-1}\hat{\mathrm{M}}'\hat{\Sigma}^{-1}\check{\mathrm{X}}_i\mathrm{R}_i). \qquad [7]$$

The rotations that maximize Eq. **7** can be found with the usual Procrustes singular value decomposition (SVD) solution (11). Let the SVD of an arbitrary matrix D be $\mathrm{U}\Lambda\mathrm{V}'$. Then the maximizing rotations $\hat{\mathrm{R}}_i$ are given by

$$\hat{\Xi}^{-1}\hat{\mathrm{M}}'\hat{\Sigma}^{-1}\check{\mathrm{X}}_i = \mathrm{U}\Lambda\mathrm{V}'$$
$$\hat{\mathrm{R}}_i = \mathrm{VPU}', \qquad [8]$$

where rotoinversions can be avoided by constraining the determinant of $\hat{\mathrm{R}}_i$ to be 1 by using $\mathrm{P} = \mathrm{I}$ if $|\mathrm{V}||\mathrm{U}| = 1$ or $\mathrm{P} = \mathrm{diag}(1, \ldots, 1, -1)$ if $|\mathrm{V}||\mathrm{U}| = -1$.

ML Estimation of the Scaling Factors

When superposing and comparing the shapes of different sized objects, scaling factors must be estimated to remove differences due to size alone. The unconstrained ML estimates of the scaling parameters β_i are given by

$$\beta_i = \frac{\mathrm{tr}(\Xi^{-1}\mathrm{M}'\Sigma^{-1}\check{\mathrm{X}}_i\mathrm{R}_i)}{\mathrm{tr}(\Xi^{-1}\mathrm{R}_i'\check{\mathrm{X}}_i'\Sigma^{-1}\check{\mathrm{X}}_i\mathrm{R}_i)}. \qquad [9]$$

However, unlike the rotations which are generally arbitrary with real data, it may be reasonable to model the scaling factors themselves as random variables drawn from a relevant probability distribution. For instance, the log-normal distribution arises as the limiting distribution of many small, multiplicative

random effects, as may be expected for random deviations in size (17). Many biological entities subject to random variations in growth naturally fit a log-normal distribution. Assuming that $\beta_i \approx LN(\theta, \theta)$, where

$$P(\beta_i) = (2\pi\theta)^{-\frac{1}{2}}\beta_i^{-1}e^{\frac{-(\ln \beta_i - \theta)^2}{2\theta}}, \quad [10]$$

then the probabilities $P(\beta_i) = P(1/\beta_i)$, and the mode of the distribution is 1 as required. The ML estimate $\hat{\beta}_i$ for the lognormally distributed scaling factors is the positive root of (*Appendix IV*)

$$\beta_i^2 \mathrm{tr}(\Xi^{-1}\mathrm{R}_i'\check{\mathrm{X}}_i'\Sigma^{-1}\check{\mathrm{X}}_i\mathrm{R}_i) - \beta_i \mathrm{tr}(\Xi^{-1}\mathrm{M}'\Sigma^{-1}\check{\mathrm{X}}_i\mathrm{R}_i) + \frac{\ln \beta_i}{\theta} = 0. \quad [11]$$

The $(\ln \beta_i)/\theta$ term in Eq. **11** can be thought of as a "penalty term" that concentrates the $\hat{\beta}$ estimates around unity; otherwise it is identical to the nonparametric solution (Eq. **9**).

ML Estimate of the Mean Form

Maximizing the likelihood with respect to the mean form M gives the familiar average (11)

$$\hat{\mathrm{M}} = \bar{\mathrm{X}} = \frac{1}{N}\sum_i^N \beta_i\check{\mathrm{X}}_i\mathrm{R}_i. \quad [12]$$

However, the average $\bar{\mathrm{X}}$ is an asymptotically biased estimate of the mean form M (it is too large) for constant K when superpositioning without scaling (i.e., setting $\beta_i = 1$ for all i) (1, 7, 18). For the case when Ξ and Σ are diagonal (i.e., no correlations), we offer the following easily computable, bias-corrected ML Procrustes estimator $\tilde{\mathrm{M}}$ of the mean (*Appendix V*)

$$\tilde{\mathrm{M}} = \bar{\mathrm{X}}\sqrt{1 - \frac{D^2 + D}{2\mathrm{tr}(\Xi^{-1}\bar{\mathrm{X}}'\Sigma^{-1}\bar{\mathrm{X}})}}. \quad [13]$$

This bias-corrected estimate is consistent (for constant K as $N \to \infty$) provided that the errors are not very large, specifically when $\mathrm{tr}\,(\Xi^{-1}\bar{\mathrm{X}}'\Sigma^{-1}\bar{\mathrm{X}}) \geq (D^2 + D)/2$. For three-dimensional forms, for example, this estimator is valid when the landmark variances are less than $K/6$ times the squared radius of gyration of the average form.

An Iterative Algorithm for ML Procrustes Analysis

In practice, the ML solutions given above must be solved simultaneously via numerical methods, because each of the unknown parameters is a function of the others. We have developed the following iterative algorithm (11, 13, 19) for determining solutions to each of the Procrustes ML estimators.

1. **Initialize.** Set $\hat{\Sigma} = \mathrm{I}$, $\hat{\Xi} = \mathrm{I}$, and $\beta_i = 1$ for all i; choose one of the X_i to serve as $\hat{\mathrm{M}}$.
2. **Translate.** Obtain each $\check{\mathrm{X}}_i$ by centering X_i according to footnote †.
3. **Rotate.** Calculate each $\hat{\mathrm{R}}_i$ according to Eq. **8**, and set each $\mathrm{X}_i = \check{\mathrm{X}}_i\hat{\mathrm{R}}_i$.
4. **Scale.** For lognormally distributed scale factors, first find the lognormal shape/location parameter θ with Eq. **19.** Newton–Raphson can then be used to find the lognormal scale estimates with an initial guess equal to the arbitrary scale factor estimate Eq. **9** or to the scale factor from the previous iteration.
5. **Estimate the mean.** Find the new average $\hat{\mathrm{M}}$ according to Eq. **12** or **13.** Return to step 3 and loop to convergence.
6. **Estimate the inverse gamma distributed eigenvalues.** Calculate $\hat{\Sigma}_U$ from Eq. **3**. The $\hat{\Sigma}_U$ covariance matrix must be decomposed, and the positive eigenvalues fit iteratively to an inverse gamma distribution. The rank of the sample covariance matrix is $\min(ND - D - 3, K - 3)$. At convergence, the rank is reduced by two due to the collinearity induced by the linear constraints from the scale and shape parameters of the inverse gamma regularization (see Eq. **6**). The rank is further reduced (by one) due to additional collinearity from centering the form matrices (6). Thus, the maximally obtainable number of positive eigenvalues in our hierarchical treatment is $K - 3$; this rank deficiency is inherent in the Procrustes sample covariance matrix and cannot be eliminated by increasing the number of forms. Thus, we treat all zero eigenvalues as missing data using an Expectation-Maximization algorithm and include only the full-rank, positive eigenvalues of the sample covariance matrix when estimating the parameters of the inverse gamma distribution. If the covariance matrix is assumed to be diagonal, one must include only the $K - 3$ largest variances in the inverse gamma fit, as the smallest three are known *a priori* to be a zero-valued eigenvalues. Estimation of the regularized eigenvalues thus cycles until convergence between two steps: (*i*) fit the current estimates of the full-rank eigenvalues to an inverse gamma distribution, and (*ii*) obtain updated estimates of the eigenvalues by modification of the sample eigenvalues analogous to Eq. **6**. $\lambda_{I\gamma,j} = (ND\lambda_{U,j} + 2\alpha)/(2(1 + \gamma) + ND)$.
7. **Estimate the landmark covariance matrix.** Calculate $\hat{\Sigma}_{I\gamma}$ from Eq. **6** using the $\hat{\alpha}$ and $\hat{\gamma}$ inverse gamma parameters determined in the previous step.
8. **Estimate the dimensional covariance matrix.** Calculate $\hat{\Xi}$ from Eq. **5.** Return to step 6 and loop to convergence.
9. **Align the superposition with $\hat{\Xi}$.** Let the spectral decomposition of $\hat{\Xi}$ be $\mathrm{V}\Delta\mathrm{V}'$. Rotate the entire superposition, including the average form $\hat{\mathrm{M}}$, with the matrix V' of left eigenvectors (i.e. set $\hat{\mathrm{M}} = \hat{\mathrm{M}}\mathrm{V}'$) so that $\hat{\Xi}$ is equivalent to the diagonal matrix Δ.
10. **Loop.** Return to step 2 until convergence (e.g., stop if the relative difference between rotation matrices between iterations fall below a specified tolerance, such as 10^{-7}).

We have implemented this algorithm in THESEUS, a UNIX C program intended for likelihood analysis of biological macromolecules (25). Because of the complexity of the problem, the efficiency of the algorithm is difficult to determine generally. If Σ is assumed to be diagonal, then the algorithm is very efficient ($O(NK)$). In our experience, even with large three-dimensional problems, when assuming a diagonal landmark covariance matrix (e.g., $K = 200$ and $N = 500$), the algorithm converges rapidly after a few dozen iterations (from milliseconds to a few seconds on a 500 MHz G4 laptop). When using full, arbitrary covariance matrices on a low-dimensional problem, the algorithm is $O(K^3)$, as the limiting step is the inversion of the landmark covariance matrix $\hat{\Sigma}$; in this case, the calculation may take a few minutes.

Comparison with Previous Analyses

Our likelihood analysis offers several ML Procrustes estimators, including the covariance matrices ($\hat{\Xi}$ and $\hat{\Sigma}$), a bias-corrected average form ($\tilde{\mathrm{M}}$), lognormally distributed scaling factors ($\hat{\beta}_i$), and simplified estimates of the rotations ($\hat{\mathrm{R}}_i$) when weighting the dimensions. Several of our solutions differ significantly from corresponding solutions to similar problems proposed previously in the Procrustes literature (1, 7, 8, 10, 11, 16). Unlike the unconstrained covariance estimators reported in Goodall and recounted by others (1, 7, 8, 11), each of our covariance estimators is itself a function of the other covariance matrix. Equations 10.2 and 10.3 of Goodall (11) are only valid when either Ξ or Σ is constrained to be equal to the identity matrix. Additionally, to find the optimal rotations when weighting with a factored dimensional covariance matrix, Goodall proposes a method using the Koschat–Swayne algorithm (16). Goodall's

method aims to maximize numerically the same estimated likelihood target as our method, corresponding to the first term of Eq. **2.** In contrast to the Goodall and Koschat–Swayne treatments, our analytic solution (Eq. **8**) is shown to be valid for a nondiagonal dimensional weight matrix Ξ^{-1} and for $\Sigma \neq I$. Finally, we present estimates of the scaling factors that differ from that reported in Goodall (10). When $\bar{X}_i R_i$ is identical to the mean form M, the scaling factor must necessarily equal unity, since no scaling is required when the mean form is superposed on itself. Unlike the scale factor estimator given in Goodall (equation 2.4), both of our ML estimators (Eqs. **9** and **11**) conform to this logically necessary condition.

Several authors have reported that the Procrustes estimate of the mean form (corresponding to the estimated ML estimator, Eq. **12**) is inconsistent (1, 7), being too large as the number of forms approaches infinity. The inconsistency of the average form is a result of ignoring the uncertainty introduced by using estimates of the rotations and translations in Eq. **12**, rather than using their true values, which are unknown. However, we note that Eq. **23** implies that the average $\bar{X}$ is in fact a consistent estimator of the mean M (assuming nonisotropic and uncorrelated errors) in the limit as both the numbers of landmarks (K) and forms (N) increase together, since then the bias factor approaches 1 in probability.

Performance of ML Procrustes Analysis with Simulated Data

We conducted simulations to investigate the ability of our ML Procrustes method to correctly infer known parameters of a family of forms. One common application of Procrustes methods in structural biology is superposing the atomic coordinates of multiple protein structures, each adopting a similar yet variable molecular conformation. To create an artificial data set, 400 perturbed protein structures were generated randomly, assuming a matrix Gaussian error distribution with a known mean form and known covariance matrices. At this stage of the simulation, the resulting family of perturbed structures consists of a known superposition, since the structures have not yet been subjected to arbitrary rotations and translations. As the final step in creating our test set of forms, we randomly translated and rotated each of the perturbed structures.

We then applied our ML Procrustes procedure to this simulated data set, obtaining estimates of the coordinates of the mean structure, the covariance matrices, and the original superposition. To illustrate the robustness of the ML method to violations of the assumptions, in this superposition analysis we assumed a diagonal landmark covariance matrix, even though the true covariance matrix contains large correlations. For comparison, we also performed a standard LS Procrustes analysis on the same data set. Results of a representative simulation are shown in Fig. 1.

Visual inspection clearly shows the remarkably close correspondence between the ML superposition and the known "true" superposition (Fig. 1). In contrast, the LS superposition is largely inaccurate, with a marked excess of positional variability in the regions of the structure with the least true variance. A residuals plot of the LS superpositions shows the tell-tale signs of extremely heteroscedastic data, indicating that the LS assumption of homoscedastic data has been violated (Fig. 2, which is published as supporting information on the PNAS web site). In contrast, a residuals plot of the ML superposition shows a very even spread for each landmark, as would be expected if the covariance matrices were estimated accurately. Similarly, probability plots of the residuals indicate large deviations from normality for the LS method, whereas ML recovers the true Gaussian distribution very well (Fig. 3, which is published as supporting information on the PNAS web site). A χ^2 analysis comparing the inferred covariance matrices to the known covariance matrices also indicates the superior accuracy of ML vs LS superpositioning (reduced $\chi^2 = 0.83$, $P = 1.0$ vs. $\chi^2 = 14{,}293$, $P = 0.0$, respectively). Finally, both methods recovered the mean structure with good accuracy, although the ML method was slightly better than LS (0.074 vs 0.11 Å root mean squared error, respectively). Simulations with a large set of forms (400 here) should reflect the asymptotic behavior of the methods. However, qualitatively similar results are also obtained when superposing much smaller numbers of simulated structures (even as few as $N = 2$, see *Data Sets 1* and *2*, which are published as supporting information on the PNAS web site).

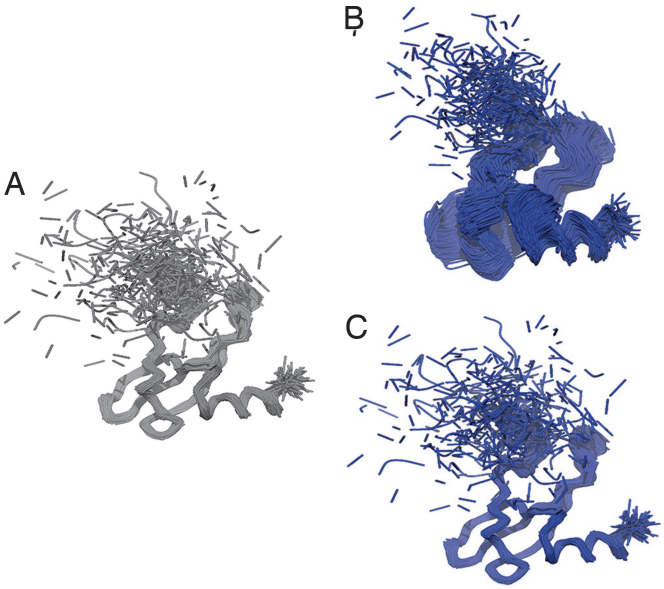

Fig. 1. ML and LS superposition of simulated protein structure data. (*A*) The "true" superposition, generated from a known mean structure with known covariance matrices, before arbitrary translations and rotations have been applied. In generating the simulated data, the set of alpha carbon atoms from model 1 of Protein Data Bank entry 2SDF (www.pdb.org) was used as the mean form (67 atoms/landmarks, squared radius of gyration = 152 Å^2). The nondiagonal 67 × 67 landmark covariance matrix was based on values calculated from the superposition given in 2SDF, with variances ranging from 0.01 to 80 Å^2 and correlations ranging from 0 to 0.99 (see *Data Sets 1* and *2*). The known dimensional covariance matrix had eigenvalues of 0.16667, 0.33333, and 0.5 corresponding to the *x*, *y*, and *z* axes of 2SDF model 1, respectively. (*B*) An ordinary LS superposition of the simulated data. (*C*) A ML superposition of the simulated data, assuming a diagonal landmark covariance matrix Σ (i.e., no correlations), $\Xi = 1$, and inverse gamma distributed variances.

The Procrustes likelihood analysis described here should be of widespread applicability for quantitative comparison of shape. Unlike LS, our ML analysis results in approximately Gaussian distributed residuals, opening the possibility for the application of standard statistical analyses that rely on Gaussian errors for their validity. Additionally, because of the greatly improved estimation of the covariance matrices, it should be possible to accurately analyze major modes of covariation in a family of forms via principal components analysis (18).

Appendices

In the following derivations, all log-likelihoods with respect to a given parameter are stated up to an arbitrary constant. Likelihoods are maximized by standard application of matrix derivatives (20), by finding the first derivative of the log-likelihood function with respect to the parameter of interest, setting it to zero, and solving for the parameter.

Appendix I: ML Estimates of the Factored Covariance Matrices. We first define the error displacement matrix E_i as the matrix of differences from the mean M for each translated, scaled, and rotated form X_i

$$E_i = \beta_i(X_i + 1_K t_i)R_i - M.$$

The log-likelihood for the row-wise covariance matrix Σ is

$$\ell(\Sigma) = -\frac{1}{2}\sum_i^N \mathrm{tr}(\mathrm{E}_i\Xi^{-1}\mathrm{E}_i'\Sigma^{-1}) - \frac{ND}{2}\ln|\Sigma| + C.$$

Taking the derivative with respect to Σ and setting it equal to zero gives

$$\frac{\partial\ell(\Sigma)}{\partial\Sigma} = \sum_i^N \Sigma^{-1}\mathrm{E}_i\Xi^{-1}\mathrm{E}_i'\Sigma^{-1} - \frac{1}{2}\mathrm{I} \odot \left(\sum_i^N \Sigma^{-1}\mathrm{E}_i\Xi^{-1}\mathrm{E}_i'\Sigma^{-1}\right) - ND\Sigma^{-1} + \frac{ND}{2}\,\mathrm{I} \odot \Sigma^{-1} = 0$$

$$= \sum_i^N \Sigma^{-1}\mathrm{E}_i\Xi^{-1}\mathrm{E}_i'\Sigma^{-1} - ND\Sigma^{-1} = 0,$$

where $\mathrm{A} \odot \mathrm{B}$ denotes the Hadamard (element-wise) product of two matrices A and B. Thus, after simplification and multiplication by Σ twice, we have $\hat{\Sigma} = (1/ND)\,\Sigma_i^N \mathrm{E}_i\Xi^{-1}\mathrm{E}_i'$, which is equivalent to Eq. **3**. Derivation of the ML estimator of Ξ proceeds similarly.

Appendix II: Inverse Gamma Distributed Variances. The probability distribution function for an inverse gamma distribution of the eigenvalues of the row-wise covariance matrix is

$$P(\lambda_j) = \frac{\alpha^\gamma}{\Gamma(\gamma)}\lambda_j^{-(1+\gamma)}e^{-\frac{\alpha}{\lambda_j}}.$$

The corresponding log-likelihood of the eigenvalues is

$$\ell(\lambda)_{I\gamma} = -(1+\gamma)\sum_j^K \ln\lambda_j - \alpha\sum_j^K \frac{1}{\lambda_j} + K\gamma\ln\alpha - K\ln\Gamma(\gamma)$$

$$= -(1+\gamma)\ln|\Sigma| - \alpha\,\mathrm{tr}\,\Sigma^{-1} + K\gamma\ln\alpha - K\ln\Gamma(\gamma). \quad [14]$$

The unrestricted log-likelihood for Σ can be written as (compare *Appendix I*)

$$\ell(\Sigma) = -\frac{ND}{2}\sum_i^N \mathrm{tr}(\hat{\Sigma}_U\Sigma^{-1}) - \frac{ND}{2}\ln|\Sigma| + C.$$

The Procrustes extended log-likelihood (5, 15) assuming inverse gamma distributed eigenvalues $\ell(\Sigma)_{I\gamma}$ is simply the sum $\ell(\lambda)_{I\gamma} + \ell(\Sigma)$. Taking the derivative of $\ell(\Sigma)_{I\gamma}$ with respect to Σ and setting it equal to zero gives

$$\frac{\partial\ell(\Sigma)_{I\gamma}}{\partial\Sigma} = -(1+\gamma)\Sigma^{-1} + \alpha\Sigma^{-2} + \frac{ND}{2}\Sigma^{-1}\hat{\Sigma}_U\Sigma^{-1} - \frac{ND}{2}\Sigma^{-1} = 0,$$

which, after pre- and postmultiplication by Σ^{-1} and rearrangement, gives Eq. **6**.

Appendix III: ML Rotations with Dimensional Weighting by Ξ^{-1}. If all structures have been centered, the likelihood equations simplify. For centered form matrices, the log-likelihood with respect to the rotations is

$$\ell(\mathrm{R}) = -\frac{1}{2}\sum_i^N \mathrm{tr}\{\Xi^{-1}(\beta_i\check{\mathrm{X}}_i\mathrm{R}_i - \mathrm{M})'\Sigma^{-1}(\beta_i\check{\mathrm{X}}_i\mathrm{R}_i - \mathrm{M})\} + C. \quad [15]$$

The nuisance parameters Ξ, Σ, and M are in general unknown, and as a consequence the likelihood given in Eq. **15** cannot be maximized directly. Therefore, as in Goodall (11), we use the method of estimated likelihood (5, 21, 22) to construct a modified target likelihood function by replacing these nuisance parameters with their ML estimates ($\hat{\Xi}$, $\hat{\Sigma}$, and $\hat{\mathrm{M}}$, respectively). To simplify the derivation here, we choose Σ to be the normalized member of the factored covariance matrices. After substitution with ML estimates, Eq. **15** can be expanded as the estimated likelihood $\ell_e(\mathrm{R}) = \ell(\mathrm{R}, \hat{\Xi}, \hat{\Sigma}, \hat{\mathrm{M}})$

$$\ell_e(\mathrm{R}) = \beta_i\sum_i^N \mathrm{tr}(\hat{\Xi}_U^{-1}\hat{\mathrm{M}}'\hat{\Sigma}^{-1}\check{\mathrm{X}}_i\mathrm{R}_i) - N\,\mathrm{tr}\,(\hat{\Xi}_U^{-1}\hat{\mathrm{M}}'\hat{\Sigma}^{-1}\hat{\mathrm{M}}) - \frac{1}{2}\,\mathrm{tr}\cdot\left\{\hat{\Xi}_U^{-1}\sum_i^N(\beta_i^2\mathrm{R}_i'\check{\mathrm{X}}_i'\hat{\Sigma}^{-1}\check{\mathrm{X}}_i\mathrm{R}_i - \hat{\mathrm{M}}'\hat{\Sigma}^{-1}\hat{\mathrm{M}})\right\} + C. \quad [16]$$

After substitution with Eq. **12** for the estimate of the mean form $\hat{\mathrm{M}}$, the ML estimate of the dimensional covariance matrix $\hat{\Xi}_U$ can be expressed as

$$\hat{\Xi}_U = \frac{1}{NK}\sum_i^N(\beta_i^2\mathrm{R}_i'\check{\mathrm{X}}_i'\hat{\Sigma}^{-1}\check{\mathrm{X}}_i\mathrm{R}_i - \hat{\mathrm{M}}'\hat{\Sigma}^{-1}\hat{\mathrm{M}}). \quad [17]$$

Substitution of Eq. **17** into Eq. **16** gives

$$\ell_e(\mathrm{R}) = \beta_i\sum_i^N \mathrm{tr}(\hat{\Xi}_U^{-1}\hat{\mathrm{M}}'\hat{\Sigma}^{-1}\check{\mathrm{X}}_i\mathrm{R}_i) - N\,\mathrm{tr}(\hat{\Xi}_U^{-1}\hat{\mathrm{M}}'\hat{\Sigma}^{-1}\hat{\mathrm{M}}) - \frac{NK}{2}\,\mathrm{tr}(\hat{\Xi}_U^{-1}\hat{\Xi}_U) + C. \quad [18]$$

In the final term of Eq. **18** the two matrices cancel reducing to $NKD/2$. Hence, when maximizing the estimated likelihood $\ell_e(\mathrm{R})$ with respect to the rotations (holding other nuisance parameters fixed), both the last two terms of Eq. **18** are constant. Maximization of the first term alone (as given in Eq. **8**) provides a ML estimate of the rotation R_i.

Appendix IV: ML Procrustes Estimate of Lognormally Distributed Scale Factors. In the lognormal distribution given in Eq. **11**, the shape and location parameter both equal θ to constrain the mode of the distribution to 1. Given a sample of N scale factors β_i, the ML estimate of θ is

$$\hat{\theta} = \frac{1}{2}\left(\sqrt{1 + \frac{4}{N}\sum_i^N(\ln\beta_i)^2} - 1\right). \quad [19]$$

The log-likelihood for the scale factors β given a lognormal LN:(θ,θ) distribution is

$$\ell(\beta_i)_{LN} = -\frac{1}{2}\ln\,\theta - \frac{\theta}{2} - \frac{(\ln\,\beta_i)^2}{2\theta}.$$

The Procrustes extended log-likelihood $\ell_{P,LN}$ (5, 15) assuming lognormally distributed scale factors is then given by the sum of $\ell(\beta_i)_{LN}$ and ℓ_P (from Eq. **2**). With respect to β_i the joint log-likelihood is

$$\ell(\beta_i)_{P,LN} = -\ln\,\beta_i - \frac{1}{2}\ln\,\theta - \frac{\theta}{2} - \frac{(\ln\,\beta_i)^2}{2\theta}$$
$$-\frac{1}{2}\,\mathrm{tr}(\beta_i^2\Xi^{-1}\mathrm{R}_i'\check{\mathrm{X}}_i'\Sigma^{-1}\check{\mathrm{X}}_i\mathrm{R}_i)$$
$$+\frac{1}{2}\,\mathrm{tr}(2\beta_i\Xi^{-1}\mathrm{M}'\Sigma^{-1}\check{\mathrm{X}}_i\mathrm{R}_i) + C.$$

Taking the derivative of $\ell(\beta)_{P,LN}$ with respect to β_i, setting it equal to zero, and multiplying through by $-\beta_i$ yields Eq. **11**.

Appendix V: Bias-Corrected ML Procrustes Estimate of the Mean Form. In the following, we assume that Ξ and Σ are both diagonal, and that $\beta_i = 1$ for all i. The sum of squared distances from the mean for each landmark (G_j) in a set of optimally superimposed forms is given by

$$G_j = \sum_i^N (x_{i,j}\Xi^{-1}x_{i,j}') - N\bar{x}_j\Xi^{-1}\bar{x}_j', \qquad [20]$$

where $x_{i,j}$ is the jth $1 \times D$ row vector in the optimally translated and rotated form matrix $\check{\mathrm{X}}_i\mathrm{R}_i$, such that $\check{\mathrm{X}}_i\mathrm{R}_i = [x_{i,1}' \ldots x_{i,j}']'$. It can be shown from standard multivariate analysis that (12)

$$\frac{1}{N}\sum_i^N x_{i,j}\Xi^{-1}x_{i,j}' \to \mu_j\Xi^{-1}\mu_j' + D\sigma_j^2 \text{ as } N \to \infty, \qquad [21]$$

where similarly μ_j is the jth row vector in the mean form matrix M. Assuming that G_j is approximately χ^2 distributed (1, 23, 24), then $G_j \approx \sigma_j^2\chi_{F_j}^2$ with F_j degrees of freedom. For each of the K landmark row vectors in the N forms there are D degrees of freedom. However, for each of the N forms, D degrees of freedom are lost due to translational estimation, and ½ $D(D-1)$ degrees are lost due to rotational estimation (1, 24). Thus, by distributing the lost degrees of freedom over each of the K row vectors in proportion to the contribution of each landmark to the minimization criterion ($w_j = \mu_j\Xi^{-1}\mu_j'/[\sigma_j^2\,\mathrm{tr}(\Xi^{-1}\mathrm{M}'\Sigma^{-1}\mathrm{M})]$), we have

$$F_j = N\left(D - \frac{\mu_j\Xi^{-1}\mu_j'(D^2+D)}{2\sigma_j^2\,\mathrm{tr}(\Xi^{-1}\mathrm{M}'\Sigma^{-1}\mathrm{M})}\right).$$

It then follows trivially that (17)

$$\frac{G_j}{N} \to \sigma_j^2\left(D - \frac{\mu_j\Xi^{-1}\mu_j'[D^2+D]}{2\sigma_j^2\,\mathrm{tr}(\Xi^{-1}\mathrm{M}'\Sigma^{-1}\mathrm{M})}\right) \text{ as } N \to \infty \qquad [22]$$

Combining Eqs. **20**–**22** yields the asymptotic result as $N \to \infty$

$$\bar{x}_j\Xi^{-1}\bar{x}_j' \to \mu_j\Xi^{-1}\mu_j'\left(1 + \frac{D^2+D}{2\,\mathrm{tr}(\Xi^{-1}\mathrm{M}'\Sigma^{-1}\mathrm{M})}\right). \qquad [23]$$

The final factor on the right hand side of Eq. **23** is an estimate of the asymptotic bias of the average. Setting $\mu_j = B\bar{x}_j$ (implying $\mathrm{M} = B\bar{\mathrm{X}}$) and solving for B gives Eq. **13**.

We thank Meredith Betterton, Ian Dryden, Colin Goodall, Subhash Lele, and Norm Pace for helpful comments and suggestions. This work was supported by the Arnold and Mable Beckman Foundation and National Institutes of Health Grant GM59414. D.L.T. was supported by Postdoctoral Fellowship Grant PF-04-118-01-GMC from the American Cancer Society.

1. Dryden IL, Mardia KV (1998) *Statistical Shape Analysis* (Wiley, New York).
2. Gower JC, Dijksterhuis GB (2004) *Procrustes Problems* (Oxford Univ Press, Oxford), Vol 30.
3. Lele S, Richtsmeier JT (2001) *An Invariant Approach to Statistical Analysis of Shapes* (Chapman and Hall/CRC, Boca Raton, FL).
4. Mardia K, Goodall C (1993) in *Multivariate Environmental Statistics,* eds Patil G, Rao C (North-Holland, New York), Vol 6, pp 347–385.
5. Pawitan Y (2001) *In All Likelihood: Statistical Modeling and Inference Using Likelihood*, Oxford Science Publications (Clarendon, Oxford).
6. Lele S, Richtsmeier JT (1990) *Syst Zool* 39:60–69.
7. Lele S (1993) *Math Geol* 25:573–602.
8. Glasbey C, Horgan G, Gibson G, Hitchcock D (1995) *Biometrical J* 37:481–495.
9. Goodall C (1991) *J Roy Stat Soc B Met* 53:334–339.
10. Goodall C (1995) in *Proceedings in Current Issues in Statistical Shape Analysis*, eds Mardia KV, Gill CA (Leeds Univ Press, Leeds, UK), pp 18–33.
11. Goodall C (1991) *J Roy Stat Soc B Met* 53:285–321.
12. Arnold SF (1981) *The Theory of Linear Models and Multivariate Analysis* (Wiley, New York).
13. Dutilleul P (1999) *J Stat Comput Sim* 64:105–123.
14. Broemeling L (1985) *Bayesian Analysis of Linear Models* (Dekker, New York).
15. Bjornstad JF (1996) *J Am Stat Assoc* 91:791–806.
16. Koschat MA, Swayne DF (1991) *Psychometrika* 56:229–239.
17. Evans M, Hastings N, Peacock JB (2000) *Statistical Distributions* (Wiley, New York), 3rd Ed.
18. Kent JT, Mardia KV (1997) *J Roy Stat Soc B Met* 59:281–290.
19. Dempster AP, Laird NM, Rubin DB (1977) *J Roy Stat Soc B Met* 39:1–38.
20. Dwyer PS (1967) *J Am Stat Assoc* 62:607–625.
21. Royall R (1991) *Statistical Evidence: A Likelihood Paradigm* (Chapman and Hall, New York), Vol 71.
22. Royall R, Tsou T-S (1995) *J Am Stat Assoc* 90:316–320.
23. Langron SP, Collins AJ (1985) *J Roy Stat Soc B Met* 47:277–284.
24. Sibson R (1979) *J Roy Stat Soc B Met* 41:217–229.
25. Theobald DL, Wuttke DS (2006) *Bioinformatics* 22:2171–2172.