Syst. Biol. 52(1):66–69, 2003
DOI: 10.1080/10635150390132759

Computing the Uniform Component of Shape Variation

F. JAMES ROHLF[1] AND FRED L. BOOKSTEIN[2]

[1]*Department of Ecology and Evolution, State University of New York, Stony Brook, New York 11794-5245, USA; E-mail: rohlf@life.bio.sunysb.edu*
[2]*Institute of Anthropology, University of Vienna, Austria; and Institute of Gerontology, 300 North Ingalls, University of Michigan, Ann Arbor, Michigan 48109, USA; E-mail: fred@brainmap.med.umich.edu*

Abstract.—Any change in shape of a configuration of landmark points in two or three dimensions includes a uniform component, a component that is a wholly linear (affine) transformation. The formulas for estimating this component have been standardized for two-dimensional data but not for three-dimensional data. We suggest estimating the component by way of the complementarity between the uniform component and the space of partial warps. The component can be estimated by regression in either one space or the other: regression on the partial warps, followed by their removal, or regression on a basis for the uniform component itself. Either of the new methods can be used for both two- and three-dimensional landmark data and thus generalize Bookstein's (1996, pages 153–168 *in* Advances in morphometrics [L. F. Marcus et al., eds.], Plenum, New York) linearized Procrustes formula for estimating the uniform component in two dimensions. [Affine; geometric morphometrics; Procrustes; regression; shape statistics.]

Informally, the shape of a set of p landmarks in k-dimensional space is the information about the points that is not changed when the configuration of all p points is moved or rescaled as a whole. This information includes familiar shape variables such as angles and distance ratios but also goes beyond them with the aid of a fairly rigorous biometric method that has emerged over the last approximately 15 years. The new approach has greatest power for statistical testing and biological understanding when variation in shape is small (i.e., when shapes are sufficiently close to their mean). Thus, it is most useful when faced with the amounts of morphological variation systematists often find when describing adults in species of the more complex animal phyla.

When the variation in shape is sufficiently small, statistical analyses of shape variation can be performed using standard multivariate statistical methods. These analyses are performed by making a linear approximation to Kendall's shape space (a multidimensional non-Euclidean space of all possible shapes of configurations of landmark points; Kendall, 1984, 1985). The approximation is called a tangent space approximation because it can be visualized as a Euclidean space tangent to shape space (Bookstein, 1991; Rohlf, 1999; Slice, 2001). The point corresponding to the average shape is usually used as the point of tangency (and is often called the reference). The Procrustes average shape for a sample can be computed using a generalized Procrustes analysis (GPA; see Rohlf and Slice, 1990, where it was called generalized orthogonal least-squares fitting). The specimens are then superimposed on this average shape to remove variation of location, orientation, and scale, specimen by specimen. It is often convenient to treat the average, $\mathbf{X}_c$ (a p by k matrix, the subscript c stands for consensus), as a row vector, $\bar{\mathbf{x}}$, of kp elements. As a set of Cartesian coordinates, $\bar{\mathbf{x}}$ defines a plane S that is orthogonal to it, the hyperplane of shape coordinates with respect to $\bar{\mathbf{x}}$. Even though $\bar{\mathbf{x}}$ has kp elements, the space is of just $2p-4$ dimensions for two-dimensional (2D) data or $3p-7$ dimensions for three-dimensional (3D) data because of the constraints on variation imposed by GPA. This space is called Kendall's tangent space because it is tangent to Kendall's shape space (see Rohlf, 1999 and Slice, 2001, for discussion of the tangent space and its relation to Kendall's shape space, especially for the case where $p=3$ and $k=2$, triangles in the plane). We assign coordinates on S with the point $\bar{\mathbf{x}}$ as the zero vector.

Using shape coordinates in place of conventional shape variables such as angles or ratios, the new statistical shape analysis treats the shape of the landmark point configuration as a whole so as to permit the systematist to ask new kinds of questions that could not be formulated before. One of the simplest of these questions turns out to be surprisingly fundamental: the issue of how to describe overall changes of proportion that are "the same everywhere." The geometric metaphor here, already used by D'Arcy Thompson, is the uniform transformation or affine transformation from classic analytic geometry. Affine transformations leave all sets of parallel lines parallel; these transformations take square graph paper into graph paper made up of parallelograms, or the equivalent in three dimensions, taking a system of identically oriented cubes into a system of identical and identically oriented parallelepipeds. Corresponding to this insight is a statistical question. Any landmark rearrangement has some component of change of this sort, and some component that is on the contrary local, essentially different in the vicinity of different landmarks. This local component is usually called the subspace of pure bending. One can ask whether either of these components is informative about differences between groups of specimens or whether one or the other is correlated with causes or effects of form, etc. In particular, systematic effects on a uniform component often arise as responses to uniform causes, such as biomechanical aspects of ontogeny or changes in typical biomechanical requirements over phylogeny. Causes such as these are typically interpreted in the language of stresses and strains interpreted likewise as uniform tensors. This interpretation is often used in analyzing specific parts of a form that in life are concerned with the delivery of force; in mammals, these parts might include jaws, burrowing apparatus, the scapula, and the pelvis. When explanations via some

biomathematical tensor are among those entertained, it is advantageous to be able to couch any claimed data signal in a fully quantitative form, whereby hypotheses of group difference, ecophenotypy and the like, may be challenged by statistically phrased null hypotheses; principal strain given standard errors of estimate, etc. In contrast, local shape changes are often more difficult to rationalize as expressions of tensor explanations. Explanations of local shape changes are typically expressed in terms of local morphogenetic processes, and the corresponding morphometric tests will not involve principal strain terms.

The essential first step in answering these various questions is to formalize the division into uniform versus nonuniform components in an objective way. We used a mathematical model of the Kendall tangent space that divides this space into two subspaces, one just for the uniform shape changes and one that permits the systematist to reconstruct all the others. In the 1990s, most of the work by morphometric toolbuilders went into the second part, the description of pure bending, even though that description is mathematically more difficult. Here, we present a new approach for partitioning shape variation into these two components and provide formulas for the uniform shape component that can be used on 3D landmark data and on 2D data. This approach clarifies some of the relationships between shape spaces, bending energy, and the thin-plate spline.

The Kendall tangent space S decomposes into a vector sum of two subspaces that are each of interest in morphometric studies (cf. Bookstein, 1991; Rohlf, 1999):

$$S = U \oplus B, \tag{1}$$

where U is the subspace of uniform transformations and B is the subspace of those transformations that are "pure bending," combinations of the eigenvectors of the bending energy matrix that correspond to the nonzero eigenvalues. The symbol $\oplus$ indicates the direct sum of two vector spaces. Equation 1 means that any set of shape coordinates, a vector s of kp elements, can be written uniquely as a sum $s = u + b$ of one vector from U and one vector from B.

S is a Euclidean space and so are its subspaces U and B. The Euclidean geometry of S is the Procrustes geometry of the original coordinate space: The length of any vector in S is its Procrustes distance from the sample mean shape $\bar{\mathbf{x}}$. Moreover, U and B are perpendicular in this Euclidean metric, meaning that if $s = u + b$ then $|s|^2 = |u|^2 + |b|^2$, the usual Pythagorean decomposition of squared distance. This perpendicularity is important because it means that we can use standard multivariate statistical methods within these subspaces separately without confounding the uniform part by the bending part or vice versa. The Procrustes metric is not equal to bending energy even though the Procrustes subspace B is spanned by the eigenvectors of nonzero bending (this seems to be a common misunderstanding). The vector space S can be described usefully by many different metrics (Euclidean distance, generalized distance, etc.), but not by bending energy.

In the geometric morphometric literature, two different bases are used to describe the subspace B of pure bending: principal warps (eigenvectors of the bending energy matrix arising out of the thin-plate spline) that are orthonormal in the Procrustes geometry, and relative warps (principal components) that are orthogonal in the covariance structure of the data. That is, the principal warps are geometrically orthogonal but correlated in most data sets, whereas relative warps are uncorrelated and not geometrically orthogonal. Each set is just a rotation of the other, although the principal warps are usually computed first.

Bookstein Linearized Procrustes Approach

Bookstein (1996b) presented, in effect, an analog of the principal warps for the subspace U_2 for uniform shape changes in 2D data. For this case, one can write down a fixed basis for U_2 that is guaranteed orthonormal in the Procrustes geometry (i.e., the two vectors are geometrically orthogonal in the tangent space although the statistical scatter of samples along these vectors may be correlated in a particular data set), no matter what the mean form may be. Using this basis, the uniform shape scores for a specimen are

$$\begin{aligned} u_1 &= \left(\alpha \sum y_i \Delta x_i + \gamma \sum x_i \Delta y_i\right) \Big/ \sqrt{\alpha\gamma} \\ u_2 &= \left(-\gamma \sum x_i \Delta x_i + \alpha \sum y_i \Delta y_i\right) \Big/ \sqrt{\alpha\gamma}, \end{aligned} \tag{2}$$

where $\alpha = \sum x^2$, $\gamma = \sum y^2$, x_i, y_i are coordinates of the ith landmark in the average shape (to simplify the equations, the average shape has been rotated to align it to its principal axes), and Δx_i and Δy_i are the differences in coordinates between a specimen (Procrustes aligned to the average) and the average shape. Note that $\alpha + \gamma = 1$ (the average shape is taken at unit centroid size). Rohlf (1996) gave a slightly different version of these equations (based on an early draft of Bookstein's, 1996b, manuscript), and Rohlf (1999) expressed Equation 2 using matrices.

This basis, a function only of the average, is easy to visualize in an intuitive way: u_1 is a horizontal shear, and u_2 is a vertical dilation. The statistically uncorrelated basis, the relative warps of the uniform component (Bookstein, 1996a), is not really necessary for 2D data. One could just look at the 2D scatter of u_1 and u_2, note the directions along which specimens differ most, and construct the principal strains of the corresponding deformations by ruler and protractor (Bookstein, 1991).

The space B_3 of pure bending energy for 3D data has a simple basis that arises by an eigenanalysis of bending energy just as it did for 2D data, but we were not able to produce the analogous set of principal warps for the space U_3 spanning the five uniform dimensions. Bookstein (1996b) suggested several possible approaches rather than a single explicit formula. Here, we provide

explicit formulations for the relative warps in the 3D case. These basis vectors for U_3 are guaranteed orthonormal in the Procrustes geometry of the tangent space and also will be uncorrelated in the data being analyzed. Their disadvantage is that they are not principal warps; they are functions of the data set as a whole, not just of the mean form. Their advantage is they have simple formulations nevertheless and can be used for both 2D and 3D data.

Two New Approaches

The first approach is based on the observation that if $S = U \oplus B$, then for any vector s of S, if $s = u + b$ then $u = s - b$. Given a shape S, if we have b we can get u and vice versa. The second approach involves regression and comes from yet another set of basis vectors of U_3 that span the set of nine affine transformations of each Procrustes component (x, y, or z) that are shears with respect to each of the same three components (x, y, or z) of the reference form. A singular-value decomposition (SVD; Eckart and Young, 1936; Jackson, 1991) is used to reduce the set of nine different one-dimensional projections to the common space of five dimensions they all lie within.

For both approaches, a generalized Procrustes analysis is performed first to obtain the average configuration of landmarks that will be used to construct the tangent space. The landmark configurations for each specimen are then Procrustes aligned to the reference shape and projected onto the tangent space that is orthogonal to the $\bar{\mathbf{x}}$ vector.

The computational steps are given below. Both algorithms assume that the p by k matrices of landmark coordinates for each specimen (where k is the number of coordinates for each landmark) are strung out as row vectors with kp elements and collected in a single matrix $\mathbf{X}$.

Complement of the Space of Pure Bending Shape Variation

The Procrustes alignment of each specimen to the mean shape removes any nonshape variation among the specimens. Nonuniform shape variation can be removed by the multiplication by a matrix that projects onto the space orthogonal to the partial warps. Any variation that remains in the coordinates must represent uniform shape variation because the uniform plus the nonuniform shape components account for all shape variation. An SVD can then be used to extract the dimensions corresponding to the uniform shape component because they are the only ones that still contain variation. The algorithm is as follows.

First, construct a p by p matrix:

$$\mathbf{N} = \mathbf{I}_p - \mathbf{E}(\mathbf{E}^t\mathbf{E})^{-1}\mathbf{E}^t, \tag{3}$$

where $\mathbf{E}$ is a p by $p-k-1$ matrix of principal warps of the bending energy matrix (see Bookstein, 1991; Rohlf, 1999) based on the reference and $\mathbf{I}_p$ is a p by p identity matrix. Multiplication by $\mathbf{N}$ projects points onto a subspace orthogonal to the subspace spanned by the column vectors in the $\mathbf{E}$ matrix. This operation is the same type as used by Burnaby (1966) to remove variation parallel to a specified size vector. It is also the matrix used in regression analysis to compute residuals.

Second, perform an SVD on the matrix $\mathbf{V}(\mathbf{N} \oplus \mathbf{I}_k)$, where $\mathbf{V} = \mathbf{X} - \mathbf{1}_n\bar{\mathbf{x}}$ is an n by kp matrix of differences between the aligned specimens and the reference configuration (arranged as a row of kp values for each specimen), $\mathbf{1}_n$ is a column vector of n 1s, $\mathbf{I}_k$ is a k by k identity matrix, and $\otimes$ indicates the matrix direct product operation. The result is

$$\mathbf{LSR}^t = \mathbf{V}(\mathbf{N} \oplus \mathbf{I}_k), \tag{4}$$

where $\mathbf{L}$ is an n by kp matrix of left singular vectors, $\mathbf{S}$ is a kp by kp diagonal matrix of singular values (of which at most $k + \frac{1}{2}k(k-1) - 1$ are nonzero), and $\mathbf{R}$ is a kp by kp matrix of right singular vectors.

The n rows of the first $k + \frac{1}{2}k(k-1) - 1$ columns (two for 2D data and five for 3D data) of the product $\mathbf{LS}$ give scores for the uniform component of shape differences for the n specimens. The kp rows of the corresponding columns of the $\mathbf{R}$ matrix give the coefficients that define the uniform components as linear combinations of the kp coordinates.

Construction Using Regression

The same results can be obtained by regressing each specimen's Procrustes aligned coordinates onto the coordinates of the reference shape. The resulting k by k matrix of regression coefficients for the ith specimen is stored as the ith row of a matrix with k^2 columns. Because these k^2 vectors are redundant, we use an SVD of this matrix to yield the columns that span the space of the uniform shape component (these are the composite vectors, combinations of the columns, with nonzero singular values). The computational steps are as follows.

First, the regression coefficients for the x, y, and z coordinates are computed separately as

$$\mathbf{B}_x = \left(\mathbf{X}_c^t\mathbf{X}_c\right)^{-1}\mathbf{X}_c^t\mathbf{X}_x^t, \tag{5}$$

where $\mathbf{X}_x$ is the n by p matrix of x-coordinates of the aligned specimens, and $\mathbf{B}_y$ and $\mathbf{B}_z$ are defined similarly as k by n matrices for the y and z coordinates. This regression is the same as used in equation 23 of Rohlf and Slice (1990) to compute the affine fit of one configuration of landmarks onto another. The regression coefficients are then combined into a single n by k^2 matrix $\mathbf{B} = [\mathbf{B}_x^t|\mathbf{B}_y^t|\mathbf{B}_z^t]$. As shown by Rohlf and Slice (1990), no intercept term is needed for the regression because the mean and the aligned specimens are centered on the origin.

Second, perform an SVD of $\mathbf{B}$ to yield

$$\mathbf{LSR}^t = \mathbf{B}\,(\mathbf{X}_c^t \otimes \mathbf{I}_k). \tag{6}$$

Where **L** is an n by n matrix of left singular vectors, **S** is an n by kp diagonal matrix of singular values (at most $k + \frac{1}{2}k(k-1) - 1$ on the diagonal are nonzero), and **R** is a kp by kp matrix of right singular vectors.

As above, the first $k + \frac{1}{2}k(k-1) - 1$ columns (two for 2D data and five for 3D) of the product **LS** give scores for the uniform component of shape differences for the n specimens.

This method is valid because for "small" shape variations the two least-squares steps involved in generating the Procrustes uniform estimate, the Procrustes superposition and the projection down to the 2D or 5D linear subspace, pertain to orthogonal linear subspaces of the full space of Cartesian coordinate variation and thus can be carried out in any order. Subject to the constraints that have already been applied by projection into Procrustes shape space, the linear term that minimizes the overall sum of squares must minimize it coordinate by coordinate. The role of the SVD is simply to pull this 2D or 5D subspace out of the redundant 4D or 9D representation. (The other dimensions pertain to the rotational and isotropic terms of the original Procrustes fit, which have already been removed by the Procrustes alignment of each specimen to the average configuration of landmarks.)

DISCUSSION

Both approaches have the advantage that they do not require the reference configuration to be aligned to its principal axes. Thus, it is easier to display the results of an analysis with the reference in its original alignment, which may correspond to a more natural or standard orientation for the organisms being studied.

Although the new methods will be more convenient to use in most applications, the Bookstein (1996b) method for 2D landmarks can still be useful. That method has the advantage that the meaning of the u_1, u_2 axes do not depend on the data and thus are useful as a fixed coordinate system for comparing different data sets; these axes were used in this way by Rohlf (1999, 2000b). The method for 2D data also involves much less computation. The more general approaches are needed for 3D landmark data. They are also useful for 2D data because they do not require the reference configuration to be rotated to its principal axes. Of the two new approaches, the regression method is the more practical because it is direct and involves much less computation.

The method using regression does not require the computation of the bending energy matrix. Thus, shape variables spanning the space of pure bending shape variation, B, can be constructed without reference to the bending energy matrix or its eigenvectors (the principal warps) or to thin-plate splines. These shape variables can be constructed as the complement of the space of the uniform component by using the regression of the uniform component rather than the principal warps in Equation 3 and keeping the $2p - 6$ (for 2D data) or $3p - 12$ (for 3D data) columns of **LS** in Equation 4 that have nonzero singular values.

This new procedure is now used in NTSYSpc Rohlf (2000a) and will be implemented in the tps software (Rohlf, 1998a, 1998b, 2000c).

ACKNOWLEDGMENTS

This work was supported in part by a grant (IBN-0090445) from the Ecological and Evolutionary Physiology program of the National Science Foundation. Helpful comments by Dean C. Adams and Dennis E. Slice are appreciated. This is contribution 1104 from the Graduate Studies in Ecology and Evolution, State University of New York, Stony Brook. Fred Bookstein was supported in part by USPHS grant GM-37251 to the University of Michigan.

REFERENCES

BOOKSTEIN, F. L. 1991. Morphometric tools for landmark data: Geometry and biology. Cambridge Univ. Press, New York.

BOOKSTEIN, F. L. 1996a. Combining the tools of geometric morphometrics. Pages 131–151 *in* Advances in morphometrics (L. F. Marcus, M. Corti, A. Loy, G. J. P. Naylor, and D. E. Slice, eds.). Plenum, New York.

BOOKSTEIN, F. L. 1996b. A standard formula for the uniform shape component in landmark data. Pages 153–168 *in* Advances in morphometrics (L. F. Marcus, M. Corti, A. Loy, G. J. P. Naylor, and D. E. Slice, eds.). Plenum, New York.

BURNABY, T. P. 1966. Growth-invariant discriminant functions and generalized distances. Biometrics 22:96–110.

ECKART, C., AND G. YOUNG. 1936. The approximation of one matrix by another of lower rank. Psychometrika 1:211–218.

JACKSON, J. E. 1991. A user's guide to principal components. John Wiley & Sons, New York.

KENDALL, D. G. 1984. Shape-manifolds, Procrustean metrics and complex projective spaces. Bull. Lond. Math. Soc. 16:81–121.

KENDALL, D. G. 1985. Exact distributions for shapes of random triangles in convex sets. Adv. Appl. Probab. 17:308–329.

ROHLF, F. J. 1996. Morphometric spaces, shape components and the effects of linear transformations. Pages 117–129 *in* Advances in morphometrics (L. F. Marcus, M. Corti, A. Loy, G. J. P. Naylor, and D. E. Slice, eds.). Plenum, New York.

ROHLF, F. J. 1998a. tpsRegr: Shape regression, version 1.13. Department of Ecology and Evolution, State Univ. New York, Stony Brook.

ROHLF, F. J. 1998b. tpsRelw: Analysis of relative warps, version 1.15. Department of Ecology and Evolution, State Univ. New York, Stony Brook.

ROHLF, F. J. 1999. Shape statistics: Procrustes superimpositions and tangent spaces. J. Classif. 16:197–223.

ROHLF, F. J. 2000a. NTSYS-pc: Numerical taxonomy and multivariate analysis system, version 2.1. Exeter Software. Setauket, New York.

ROHLF, F. J. 2000b. Statistical power comparisons among alternative morphometric methods. Am. J. Phys. Anthropol. 111:463–478.

ROHLF, F. J. 2000c. tpsTree: Fitting shapes to trees, version 1.12. Department of Ecology and Evolution, State Univ. New York, Stony Brook.

ROHLF, F. J., AND D. E. SLICE. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst. Zool. 39:40–59.

SLICE, D. 2001. Landmarks aligned by Procrustes analysis do not lie in Kendall's shape space. Syst. Biol. 50:141–149.

First submitted 1 May 2002; reviews returned 20 June 2002; final acceptance 19 August 2002
Associate Editor: Junhyong Kim