Higher-Order Features of Shape Change for Landmark Data

Fred L. Bookstein

Center for Human Growth and Development
University of Michigan
Ann Arbor, Michigan 48109

Abstract

This chapter reviews a decomposition of shape change or shape variation into geometrical and statistical components which, together, often support useful interpretations. The techniques apply to data in the form of homologous landmark locations. All the shape features discussed here are independent of the decision as to whether to standardize position, orientation, and scale in any way and, if so, in what way that standardization is carried out. In particular, the features that describe the shape relations of a pair of landmark configurations are the same as those that describe the residuals from the fit of either to the other by an isotropic Procrustes transformation, as discussed elsewhere in this volume. The nonuniform features are the same whether or not a uniform part is estimated and corrected first. For drawing shape changes, I will use my two-point registration. Shapes of triangles are expressed as shape coordinate pairs of one landmark in a coordinate system defined by fixing two others along a baseline at Cartesian locations (0,0) and (1,0). More general landmark reconfigurations are treated as vectors of displacement of all but two of the landmarks after reduction to sets of shape coordinate pairs one by one. Multivariate analyses of such representations, properly interpreted, are very nearly independent of baseline.

The present chapter is, intended mainly to sketch the assumptions, computations, and dogmata underlying the two examples to follow in this Section. It is no substitute for a detailed study of the full statistical feature space for landmark-based shape. These matters were the subject of several lectures at the Michigan workshop and also of Chapter 7 in the preliminary draft of my Morphometric Tools for Landmark Data (Bookstein, 1991). Portions of the same material have appeared in Bookstein (1985, 1986, 1987, 1989a, 1989b) as well.

A Summary of the Basic Ideas

- Any change of shape for a configuration of landmarks has a uniform part and a non-uniform part, and any observed sample of landmark configurations incorporates variation of both the uniform part and the non-uniform part about a mean configuration.
- The uniform and the non-uniform parts of any change or scatter represent complementary subspaces of the full vector space of shape variation.
- The uniform part may be imagined as the change or variation of a "typical" triangle,
rigorously interpolated or extrapolated so as to apply to every landmark triangle in the same way.

- A purely uniform transformation leaves parallel lines parallel. In the two-point registration, all landmarks are displaced by multiples of a single vector; each multiplier is proportional to the landmark's distance from the baseline.

- There are various ways of estimating the uniform part of a shape change that is not exactly uniform. The most convenient is as a factor score, an average of all landmark shifts weighted by their distance from the baseline.

- To any sample of shapes corresponds a two-dimensional distribution of this uniform factor score. It may have up to two uniform statistical components, which are eigenvectors of uniform shape variation with respect to the anisotropy metric (log ratio of principal strains).

- To any transformation of landmarks there is a bending energy, which may be thought of as the net energy required to bend an infinite, infinitely thin metal plate over one set of landmarks so that its height over each landmark is equal first to the $x$-, then the $y$-coordinate of the corresponding landmark in another set. Uniform transformations involve tilting and re-rolling this plate, not bending it, and so require zero bending energy.

- Any single non-uniform transformation may be expressed as a finite sum of principal warps, eigenfunctions of the bending energy corresponding to Procrustes-orthogonal displacements of the "metal plate" at the landmarks. These warps emerge in descending order of an eigenvalue, bending energy per unit summed squared Procrustes displacement, which can be identified with the inverse geometrical scale or information localizability.

- Because the uniform part of a mixed transformation can be defined in many different reasonable ways, the non-uniform part is likewise not unique. However, all of its variants have the same bending energy.

- A sample of shape changes, or their residuals after subtraction of an estimated uniform part, may be usefully decomposed into a series of relative warps, which are eigenvectors of the variance-covariance matrix of landmark coordinates with respect to bending energy. These are analogous to ordinary principal components, in that they emerge in order of their power to account for transformations of landmark locations distributed as widely as possible over the form. They are calibrated by eigenvalues which represent variance per unit bending energy.

- The uniform transformations can be thought of as the zeroth of these relative warps, with eigenvalue infinite and zero bending energy per unit variance.

- The metrics for the uniform and non-uniform parts of shape change or variation are wholly incommensurate: for the uniform part, an anisotropy; for the non-uniform part, an energy. Describing the "magnitude" of a change of landmark configuration requires at least three "distances": anisotropy, bending energy, and also a size difference score. There is no good way to combine these into one single metric; the hope of a meaningful unitary matrix of distances between shapes is vain.

- Instead, the combination of two uniform components with some number of relative warps provides a useful feature space in which to search for evidence of diverse morphogenetic processes at multiple geometric scales.

The Simplest Example: Transformations of a Square

I shall assume that the reader agrees with my much-published view, beginning with Bookstein (1984), about how best to carry out the multivariate statistical analysis of a triangle of landmarks. The analysis of shape variability for a triangle reduces to
a scatter of single pairs of shape coordinates. Distance in this shape coordinate plane is proportional to log-anisotropy, log of the ratio of diameters of the ellipse into which any circle is taken by the simple shear (cf. Figure 3) consistent with the landmark locations. All the conventional sorts of biometric hypotheses dealing with covariates of shape or shape change, such as size or group, may be rigorously tested in this space, and any effects found as displacements or trends in the shape coordinate plane may be interpreted immediately as specific scalar shape variables: ratios of homologously measured distances aligned with the principal strains of the effect construed as a deformation.

The question naturally arises as to whether the general change of shape for more than three landmarks can be described in equally simple language. The problem is clear in the context of even the very simplest configuration of four landmarks, the square in Figure 1. Consider two transformations of the square, one to a parallelogram and one to a kite-shaped object. Let us inspect the effects of these changes upon the two triangles into which we can divide the square. Figure 1a shows this analysis for the transformation to a parallelogram. The analyses of shape change for the two triangles agree regarding the principal directions and principal strains of the shape change. But in the change of square to kite (Figure 1b), the analyses of the shape change for the two triangles are somewhat in disagreement. The direction of greater strain for each is the direction of lesser

![Diagrams](image_url)

Figure 1. Uniform and non-uniform transformations of a square. a) Square to parallelogram: the strain tensors agree between the triangles. b) Square to kite: the strain tensors reverse greater and lesser principal strains. c) The same for a different starting triangulation. d) When the two strains are taken to refer to the identical starting triangle, the implied landmark displacements are equal for the change to the parallelogram, but opposite for the change to the kite.
strain for the other. When we switch to the other triangulation (Figure 1c), the agreement or disagreement continues.

It seems that the transformation of square to parallelogram can be described by a single triangle in some sense in which that of square to kite cannot. We can attempt to quantify this, and, in fact, we arrive at the actual uniform and non-uniform spaces of shape change—for squares only!—if in each of Figures 1a and b we rotate one of the triangles by $180^\circ$ around the baseline so that the starting positions of the third landmark are now superimposed in the same location. Then in the transformation of a square to a parallelogram (Figure 1d) the resulting displacements of the "same" point are identical, while in the transformation of square to kite, they are opposite. This agreement or disagreement of features, which is the same regardless of original triangulation (cf. Figure 1e), suggests a decomposition of any observed change of shape of a square of landmarks into two parts (Figure 1f). One part, representing the difference of the displacements of the two movable landmarks (which becomes their average after one is "flipped"), is the square-to-parallelogram part of the transformation, the same no matter how one triangulates the form. The other part, incorporating the average of the moving landmarks before flipping (e.g., their difference after flipping), seems to represent the pure contradiction between the alternate triangulations, likewise in a way that here seems independent of the triangulation.

A suggestive visualization of this distinction treats differences in the landmark locations between the square and the outcome form as if displacements were perpendicular to the picture rather than within the plane of the paper. Then the uniform transformation (Figure 2a) appears to involve a distorted square which is tipped with respect to the
original picture—it looks like a projected image of the original form—whereas the non-uniform transformation (Figure 2b) appears to bend the square. In neither case can the transformation be "localized" to any single landmark or subset of landmarks. Both appear to be distributed evenly over the whole set of four. We will return to this rather potent metaphor shortly.

Figure 2. Metaphor for landmark displacements perpendicular to the plane of the starting square. a) Uniform transformation, square is tipped and "foreshortened." b) Non-uniform transformation, square is irrevocably bent.

The apparent agreement of these analyses between the triangulations corresponds to invariance of one's biometric analysis under such changes (Bookstein, 1987). If the square and the kite represented mean forms in two populations, conventional statistical tests of the difference between the groups would yield exactly the same significance levels whether based on optimal distance-ratios from the first triangulation or on those from the second. Different variables would be involved, of course, in discriminating the kite from those involved in discriminating the parallelogram.

The General Picture of a Uniform Transformation

We need to generalize the preceding discussion so as to refer to any number of landmarks located anywhere, not just four changing from the form of a square. A very useful model for this generalization is that shown in Figure 3: the class of transformations for which the shape change tensors computed from all triangles of landmarks are the same in their lengths and orientation upon "tissue". It can be shown that such transformations are simply the uniform or affine transformation that keep parallel lines parallel and preserve ratios of lengths measured in the same direction. These transformations take circles to ellipses whose axes are the principal strains of the transformation. For a review, see Bookstein et al., 1985.

In an arbitrary superposition, such as one resulting from a best-fitting isotropic Procrustes transformation, it is not at all clear when a transformation is uniform in this sense (Figure 4a). Things are much clearer when superposition is by means of shape coordinates to any baseline pair of landmarks (Figure 4b). In the shape coordinate plane, a uniform transformation displaces all landmarks in the same direction, by multiples of a single vector. Landmarks at the same height are displaced by the same vector regardless of their location along the baseline. The distance by which each landmark is displaced is proportional to the distance that landmark began above the baseline.
Landmarks below the baseline are displaced by multiples of the opposite of that vector, corresponding to their negative distance "above." All this corresponds perfectly to what we already noted about the square in Figure 1.

When a transformation is in fact uniform, there is no disagreement about what uniform transformation it is. When a transformation differs from the uniform, either by mere digitizing noise or by additional biologically real features, it is no longer obvious what we should consider to be the uniform "part" of the shape change. Several of us in morphometrics are working on this problem from different points of view.

The best solution, in my opinion, is that which I have proposed in the course of my papers on the thin-plate spline (e.g., Bookstein, 1989a). Any reconfiguration of landmarks in two dimensions can be uniquely expressed as the sum of a uniform transformation together with vector multiples of the function \( r_i^2 \log r_i^2 \), where \( r_i \) is the ordinary distance to the \( i \)th landmark of one form. The origin of these strange functions will be revealed presently. Because this decomposition is exact, it requires no decision about the direction in shape space along which to measure the "error of fit" to the uniform transformation which is to be minimized. The transformation is linear in the coordinates of any set of landmarks—indeed, it is expressed by the last three rows of the matrix \( L^{-1} \) in the next section. The resulting uniform part of each transformation may be expressed as a shape scatter in the usual way, by reference to its effect on a standard triangle, and biometrics proceeds from there.

In this approach, the non-uniform part of the interpolation function is confounded with the uniform part. If it were known a priori that the non-

---

**Figure 4.** Visualizing and estimating uniform transformations as factors. a) It is very difficult to detect evidence of uniform transformations in the residuals left by fitted models of scale change. b) In the shape coordinate plane, a uniform transformation displaces all landmarks in a single direction by amounts proportional to (signed) distance from the baseline. c) One may thereby estimate the uniform part of any observed shape change using the usual formula for estimating factor scores (see text).
uniform part is meaningless, pure noise, then one might wish to estimate a uniform part which is the "closest" to the actual transformation for some reasonable measure of nearness, averaging away the apparent non-uniformity as best one can instead of compensating for it. Rohlf (Chapter 10) and Goodall (1989) both present least-squares methods for this computation under slightly different assumptions about the structure of error. Bookstein and Sampson (1987) present another, lifting the assumption that error is independent at the several landmarks, and Mardia (1989) suggests a maximum-likelihood technique.

The papers in this section use a much simpler method than any of these others. The scheme of Figure 4b, in which each vector of displacement is proportional to one single vector according to a known multiplier, is exactly analogous to the usual scheme of estimation of factor scores (Figure 4c). In the usual factor model, if a factor score is postulated to predict each of a family of indicators with known regression slopes and independent regression errors, then the best estimate of the factor score itself, given only its sequelae, is proportional to the average of the observed outcomes, each weighted by the inverse of its own error of prediction. (No factor-analyst would let me fail to remind the reader that this is true only if those errors are, in fact, uncorrelated—only if there is no secondary factor structure.)

We can apply this model to our landmark data by treating the uniform part as a vector-valued factor score, the uniform factor, which predicts each observed displacement of shape coordinates via a regression coefficient that, for each landmark, equals its distance from the baseline. The "error of prediction" of each displacement by this common score is technically unobservable, but can be guessed as approximately the same for each landmark (perhaps on the assumption that they are expressing the same digitizing noise); then the relative precision of each landmark-specific "prediction" is directly proportional to its distance from the baseline. The factor score is the sum of all these inverse predictions, divided by a suitable constant. (The exactly analogous formula for factors not referring to landmarks can be found on page 89 of Bookstein et al., 1985.)

There results the following formula for the estimated uniform transformation underlying any observed change of shape coordinates \((x_i, y_i) \rightarrow (x_i, y_i) + (\Delta x_i, \Delta y_i)\):

\[
(\bar{u}_x, \bar{u}_y) = \frac{\sum y_i (\Delta x_i, \Delta y_i)}{\sum y_i^2}.
\]

Here the sums are taken over all the landmarks; each weight is the mean vertical shape coordinate (relative distance from baseline) for its landmark. If the context is that of the description of shape variation rather than shape change, then the \(\Delta\)'s here should be the deviations of the observed shape coordinates from their sample means. If the transformation is indeed uniform—that is, if each \((\Delta x_i, \Delta y_i)\) equals \(y_i \alpha\) for some common vector \(\alpha\)—then these formulas (without the bars over the \(y\)'s) recover the vector \(\alpha\) exactly whatever the distribution of ordinates \(y_i\). If the transformation is not exactly uniform, then this estimate will disagree slightly with estimates to other baselines and with estimates provided by Procrustes algorithms, my own projection algorithm, or the exact spline fit. In my view the simplicity of the formula above more than compensates for its not being an embodiment of any least-squares optimum.

The uniform component arrived at by this or any other estimation rule may be considered as if it were indeed the observed effect of the shape change or variation in question on one big, fuzzy triangle. For a single shape change, it may be re-expressed in terms of its principal strains by the construction in Bookstein et al., 1985. For a sample of forms, there results a sample of these estimated uniform "factors," which may be scattered for the cases of a sample, scanned for outliers, regressed into exogenous variables to find shape trends, or referred to a conventional component analysis of their own—which turns out to be with respect to
anisotropy, a sensible choice (see below). There result up to two uniform statistical components of this uniform factor. Each may be inspected, just as any other principal component may be inspected, to see if it suggests some underlying biological process.

**Pictures of Bending Energy**

One might imagine "the" non-uniform part of a transformation to be the residual reconfiguration of landmarks left after one has undone the effects of a uniform part fitted to the data by my factor approximation, Rohlf’s least-squares algorithm, or any other. It is not a trivial task to unearth a descriptor of such a residual that is independent of the algorithm used for producing the uniform part whose residual it is. Instead, one needs a method which extracts non-uniform parts of observed shape changes or variations directly, without requiring the (arbitrary) projection onto a fitted uniform part as an intermediate step. Such a procedure is available, I noted above, in the course of the decomposition of any observed change of landmark configuration as a thin-plate spline.

The general theory of thin-plate splines is somewhat technical (cf. Bookstein, 1989a, 1991), and it is inappropriate to review it here in any detail. Briefly, let $P_1 = (x_1, y_1), P_2 = (x_2, y_2), \ldots, P_n = (x_n, y_n)$ be $n$ points in the ordinary Euclidean plane according to any convenient Cartesian coordinate system. Write $r_{ij} = |P_i - P_j|$ for the distance between points $i$ and $j$, and $U(r)$ for the function $r^2 \log r^2$. Define matrices

$$K = \begin{bmatrix} 0 & U(r_{12}) & \cdots & U(r_{1n}) \\ U(r_{21}) & 0 & \cdots & U(r_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ U(r_{n1}) & U(r_{n2}) & \cdots & 0 \end{bmatrix}, \quad n \times n;$$

$$P = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & y_n \end{bmatrix}, \quad 3 \times n;$$

and

$$L = \begin{bmatrix} K \\ P^T \end{bmatrix} \begin{bmatrix} P \end{bmatrix}, \quad (n+3) \times (n+3),$$

where $^T$ is the matrix transpose operator and $0$ is a $3 \times 3$ matrix of 0’s.

Let $V^T = (v_1, \ldots, v_n)$ be any $n$-vector, and write $Y^T = (V^T | 0 \ 0 \ 0)$. Define the vector $W^T = (w_1, \ldots, w_n)$ and the coefficients $a_1, a_x, a_y$ by the equation

$$Y^T L^{-1} = (W^T | a_1, a_x, a_y).$$

Use the elements of $Y^T L^{-1}$ to define a function $f(x,y)$ everywhere in the plane:

$$f(x,y) = a_1 + a_x x + a_y y + \sum_{i=1}^n w_i U(|P_i^{-1}(x,y)|).$$

Then the following three propositions hold:

1. $f(x_i,y_i) = v_i$, for all $i$.
2. The function $f$ minimizes the nonnegative quantity

$$I_f = \int \int \left( \left( \frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left( \frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left( \frac{\partial^2 f}{\partial y^2} \right)^2 \right)$$

over the class of such interpolants. This is a constant multiple of the physical bending energy of an infinite, uniform, thin metal plate originally flat and level and now constrained to pass through all the points $(x_i, y_i, v_i)$. The function $f$ in fact gives the actual form of that plate, as it takes a position which minimizes precisely this energy.

3. The value of $I_f$ is proportional to

$$W^T K W = V^T (L_n^{-1} K L_n^{-1}) V,$$

where $L_n^{-1}$ is the upper left $n \times n$ subblock of $L^{-1}$.

This form is zero only when all the components of $W$ are zero: in this case, the computed interpolant is $f(x,y) = a_1 + a_x x + a_y y$, a linear function.

In the present application we take $V$ to be the $2 \times n$ matrix
\[
V = \begin{pmatrix}
  x'_1 & y'_1 \\
  x'_2 & y'_2 \\
  \vdots & \vdots \\
  x'_n & y'_n
\end{pmatrix}
\]

where each \((x'_i, y'_i)\) is a point "homologous to" \((x_i, y_i)\) in another copy of \(\mathbb{R}^2\). The resulting function \(f\) now maps each point to its homologue \((x'_i, y'_i)\) and is least bent (according to the measure \(I_f\), integral quadratic variation over all \(\mathbb{R}^2\), computed separately for real and imaginary parts of \(f\) and summed) over all such functions. In effect, our metric is the bending energy of a four-dimensional thin plate: two dimensions of plate, displaced in two "other" perpendicular directions.

It is instructive to view the form of the plate for the transformation of square into kite. Figure 5a shows the interpolation, Figure 5b the plate. It is the energy of this bending that is proportional to the quadratic form in point (3) above. In this figure one can finally see how it is that the non-uniform transformation is indeed localized: while the tilt of Figure 2a extrapolates out to infinity unchanged, the spline of Figure 5b goes asymptotically flat a short distance from the landmarks involved.

This particular transformation involves changes in one shape coordinate only (the ordinate). In general, bending energy derives from both shape coordinates. Figure 6a shows five landmarks, Figure 6b the two-dimensional spline interpolant \(f\) just introduced. Figures 6c and 6d, separately, show the \(x\)- and \(y\)-components of this interpolant after the uniform part is graphically suppressed. The energies of these two are 0.0205 and 0.0225 (in arbitrary units). This example is discussed at much greater length in Bookstein, 1989a.

The bending energy may be imagined as a metric (a distance measure) on shape space (cf. Bookstein, 1991, Appendix 2). Landmark configurations that differ by a uniform transformation are at distance zero from one another in this metric: bending energy zeroes out all transformations that exactly fit any combination of those simple models. We already know how to describe those changes of configurations, however: by a combination of changes of position, orientation, and scale, together with a uniform "distance" measured as the logarithm of the ratio of the principal strains. Just as change of position is incommensurate with change of orientation (centimeters and degrees don't mix), and just as both are incommensurate with log anisotropy, so all of these natural metrics for uniform changes are incommensurate with the bending energy introduced here.

In most multivariate statistical applications, the appropriate picture of a distance measure is a generalized ellipsoid. "Statistical distance" in all directions is variously proportional to the Euclidean distances of the "natural" descriptor space. (Such metrics arise, for instance, when one refers to the difference between population mean forms in units of within-sample covariance, the underpinnings of Hotelling's \(T^2\).) By contrast, the bending energy is a deficient metric. Its picture is a cylinder in landmark configuration space, not an ellipsoid. The generators of the cylinder—"the straight lines" on it—are in fact the sets of all transformations derived from a given non-uniform warping by application of any additional uniform transformation. All such additional transforms are at the same "distance" (bending energy) from the starting form.

The axes of this cylinder are the principal warps of the landmark configuration (Bookstein, 1990). These are eigenvectors of the bending energy with respect to summed squared landmark displacement in their original coordinate system. Each principal warp specifies the displacement of each landmark by a particular distance (positive or negative, summing to zero) in an unspecified direction that is the same for all landmarks (Bookstein, 1989c). The cylinder pairs these axes into circles of equivalent bending in any direction of the plane. The first principal warp represents the pattern of displacements having largest bending energy per unit root-mean-square landmark displacement; usually, it is the relative displacement of the two
landmarks closest together with only small contributions from the others, which are effectively "at infinity." At the other extreme, the last principal warp is the largest-scale nonlinearity that can be considered to leave landmarks at infinity fixed: it usually looks just like the square-to-kite transformation in Figure 5a.

**Relative Warps as Eigenfunctions with Respect to Bending Energy**

In ordinary principal components analysis, the principal components are the directions in feature space which have the successively greatest variances as ratios of their "lengths" in a geometry for which direction cosines in all directions are weighted equally. This is almost always unreasonable in practice, as everybody fails to consider whether lengths in different directions of feature space really ought to be considered as calibrated by Euclidean distance regardless of direction. Nevertheless, in the analysis of the uniform factor I introduced above, the uniform components are taken as principal components in just that way: directions in the space of the uniform factor which have the greatest and least sample variance per unit length. This is justified here because in that space we know the meaning of length: it is exactly proportional to anisotropy, our preferred measure of the extent of a shape change, and as such really is independent of direction.

In the method of relative warps, the eigenanalysis of the observed variance-covariance matrix of shape coordinates is taken with respect to the bending-energy matrix $L_n^{-1}KL_n^{-1}$ described above. We extract components of purely non-uniform shape variation, the relative warps, as directions in shape space of successively greatest variances in relation to bending energy. The computation of these directions by conventional matrix operations is a bit delicate (see Figure 5. The thin-plate spline for the square-to-kite transformation. a) The interpolant, treated as a vertical displacement over corners of a square: up at the ends of one diagonal, down at the ends of the other. Note that the bending is localized in the region of the four landmarks, though it is evenly distributed over the four. b) The equivalent three-dimensional "plate."
Bookstein, 1991)—and will not be reviewed here. A program which computes them has been included on the set of software accompanying this publication.

In the relative warps, it was my intention to provide the strongest possible analogue to the notion of principal component for the very highly structured data of landmark locations. Intuitively, a principal component is attempting to find the dimensions of variability that the whole list of variables hold "most in common." The ordinary first principal component conflates covariances among all pairs of the original measures into that axis of the covariance ellipsoid which has the greatest length. In the translation into landmark-based morphometrics, "length" remains sample
variance of coordinates, but we must be careful about what is meant by "most in common." In the metaphor of bending energy, reducing the scale of a set of landmarks by half multiplies the bending energy of a given set of displacements fourfold. Recall that the dimensions of shape space may be ranked in terms of intrinsic bending energy, the "size" of the square-to-kite transformation that best suits them. For a single set of landmark displacements, large regions have the lowest bending energy, and small ones the greatest.

The analysis by relative warps, which I have proposed, weights sampling variance inversely to this apparent "scale" of geometric nonlinearity before searching for the series of successive extrema which are the eigenvectors, the relative warps. A two-up-two-down transformation (Figure 5a) on the farthest corners of the form, for instance, will need only one-fourth the sampling variance of the same features restricted to a quadrant to emerge as the first relative warp. In this way, the relative warps pull out geometrically orthogonal dimensions of nonlinear shape variability in order of variance scaled inversely by feature size. This ensures at least one aspect in which the relative warps do not generalize the conventional components. There, analysis of the covariance matrix is independent of the sample means. But the matrix for bending energy is a function of the mean landmark configuration—the same landmark variance-covariance matrix leads to different relative warps as the variance is taken about different mean forms.

The uniform transformations have bending energy zero, and thus have "infinite" shape variance per unit bending energy regardless of direction. In that sense, they may be considered as the "zeroth" eigenvectors of this system. (They cannot be computed in that fashion, however, as we could not identify which direction held the largest variance of the uniform factor without reference to anisotropy, which is a different metric entirely.)

Extended examples of the complete analysis of a system of landmarks by its decomposition into the uniform statistical components, the relative warps, and their correlations are the concern of the paper by Tabachnick and myself later in this volume, and will not be repeated here. It is useful, however, to attempt a certain clarification of nomenclature:

The higher-order features of shape to which the title of this chapter refers apply to samples of shapes and, equally, to residuals from their analysis by any combination of partial fits to changes of size, orientation, and position. Any single shape change, and likewise any sample of deviations of shapes from a sample mean, may be usefully decomposed into two parts: uniform and non-uniform. The uniform part for each change or deviation of shape is representable as a single vector of length 2, the uniform factor. There are several ways of estimating it which differ only in statistical details. A sample of these factors may be usefully expanded in terms of its (first and second) uniform statistical components, each of which is an eigenvector of the observed sampling variability of the uniform factor with respect to its anisotropy. Complementary to the subspace of uniform transformations, those described exhaustively by their uniform parts in this sense, is the subspace of non-uniform transformations, those having no uniform part at all, according to whatever estimation routine one prefers. Sample variability of this non-uniform part is usefully considered in terms of the (first, second, etc.) relative warps of the sample, which are the first few eigenvectors of the observed sample covariance matrix with respect to the deficient metric that is bending energy. The uniform factor itself may be thought of as the first pair of eigenvectors of this computation, having "infinite" relative eigenvalue, owing to the fact that uniform transformations have no bending energy. These two metrics, bending and anisotropy, are incommensurate; it requires values for both of these distances to describe the "magnitudes" of any shape transformation. (And therefore it requires at least three distances to describe the general reconfiguration of landmarks: not only these two shape metrics, but a third, for size change, as well.)
Concluding Comment

Nothing here is meant to imply that other features of shape space might not be interesting in particular applications, only that one needs a reason to look at them. I have published several examples of the reporting of shape changes by specific features of change associated with reasonable biological processes. Of the conventional models for shape change, the uniform model sometimes fits real data (Bookstein, 1987), and the model of growth-gradients is very compatible with the highest-order relative warps as described here. In comparison, a shape change that is truly limited to one single landmark moving upon a background of all the others in an unchanging configuration, such as is postulated by the robust Procrustes methods, is smeared out into a series of relative warps by this method—so much for nonlinearity of the smallest quadrilateral around it, so much for the second-smallest, and so on—and thereby becomes unrecognizable. In one study of the rat cranium (Bookstein, 1989b) the first two relative warps are explicitly identified with the two dimensions of an obvious candidate for explanation of form change, the rigid motion of the vault of the skull with respect to the cranial base. Abe et al. (1988) find that a cubic growth-gradient accounts cleanly for some changes in a lineage of ostracodes. In another cranial data set (Grayson et al., 1985), the appropriate explanation of an observed difference between typical syndromal and normal forms of the human cranial base hinges upon one single landmark that is being pushed in two directions by two separate sequelae of the abnormality. In Tabachnick's chapter below, the first uniform component and first relative warp correlate 0.96, and together embody Raup's parameter $\theta$ for a spiral form in one view; the same uniform component in another view is interpreted quite differently. In all these papers, shape coordinates were computed first, and then configurations of their differences were inspected to see what simple features were suggested.

The method of shape coordinates permits both the inspection of single residuals and the construction of large-scale gradient patterns. I recommend the formal consideration of all of these components and their correlations as a necessary step in the morphometric analysis of any set of landmark data, regardless of whether the intended end-point of the investigation is an understanding of process. Any fitting of landmark data to restricted models, with or without a uniform part, by Procrustes methods or any other, is useless without a close inspection of the geometric and statistical covariances of the residuals it induces. When that latter analysis is done according to the decomposition I am recommending here, the appropriate fits are optional, a-posteriori approaches to summaries of effects already noted. They are properly taken as confirmatory, not exploratory, techniques.

Acknowledgment

The preparation of this chapter and the development of new methods it puts forward were supported in part by United States National Institutes of Health grant GM-37251 to the University of Michigan (Fred L. Bookstein, Principal Investigator).

References


