

## Linear Machinery for Morphological Distortion

FRED L. BOOKSTEIN\*

*Center for Human Growth and Departments of Statistics and Biostatistics, University of Michigan,  
Ann Arbor, Michigan 48109*

Received October 13, 1977

In 1917 D'Arcy Thompson reduced the problem of comparing two homologous shapes to the construction and depiction of a mathematical distortion in the plane. Attempts at algorithms for this computation, found mostly in the biological literature, ignore the primacy of the boundary correspondence within the data. One can define roughness of a map as the extent to which the image of the centroid of a square deviates from the centroid of the images of its corners; analytically, this is the sum of the squared Laplacians of its real and imaginary parts. When data are supplied geometrically in the form of a boundary correspondence and homologous point pairs, one can compute by wholly linear methods the function (splined over a mesh) which accords with geometric homologies and has least integral roughness. The necessary high-order matrix operations are available in the numerical-analysis literature under the rubric of "fast Poisson solvers." The resulting explicit smooth functions lend themselves naturally to diagrammatic display in terms of the eigenstructure of the symmetrized geometric strain and integral curves of its principal directions.

### INTRODUCTION

The classic literature of interpolation (cf. Davis (10)) treats mainly of maps of  $R^1$ , the real line, into  $R^1$ . Within this field, the mapping into  $R^1$  of  $R^m$ ,  $m$  greater than 1, is a subspeciality, recently reviewed in Schumaker (23). On problems associated with the mapping of  $R^m$  into  $R^n$  for  $n$  greater than 1 there is no general literature, but only the records of specific fields, notably photogrammetry and remote sensing, grappling with their data-specific problems: see, for instance, Doyle (11). Such fields, by happenstance, operate at a precision approaching that of celestial mechanics; their technology is very well suited for functions of a very few parameters explicitly measurable. The proprietor of softer data, notably measures of biological growth and form, cannot make much use of algorithms so demanding.

Morphometrics, the subject of the application in this article, has been seriously retarded by this methodological lacuna. Through all its history it has found no technical aids suited to its quantitative needs except for the relatively unspecialized maneuvers of multivariate statistical analysis. Morphometrics' central problem, the

\* Send correspondence to: Fred L. Bookstein, 311 Victor Vaughan, Center for Human Growth, University of Michigan, Ann Arbor, MI 48109.

characterization of shape contrast and shape change, is neither subtle nor arcane; yet it has not been able to share a practical lore with any neighboring disciplines that might provide fresh insights. It has but one indigenous method, whose demands have perplexed a small band of workers for 60 years.

D'Arcy Thompson's classic *On Growth and Form* (25) proposed the method of "Cartesian transformations" for the inspection of specific distortions. It was Thompson's happy idea to compare two related shapes by contemplating a single geometrical object, the diffeomorphism between them that accorded with biological

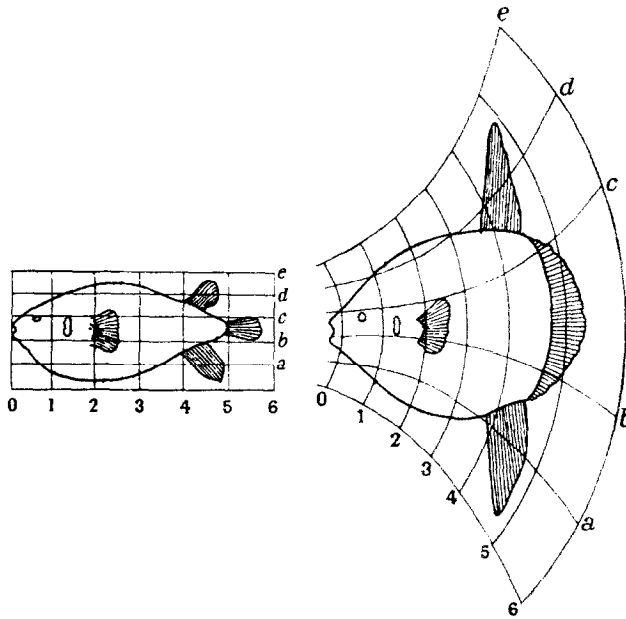


FIG. 1. Transformation grid from *Diodon* to *Orthogoriscus*, from Thompson (25).

homologies. This map is depicted by its effect upon a square ("Cartesian") grid superimposed over one of the shapes. The apparent distortion of the biologically homologous grid in the other shape embodies the shape transformation in a diagram beholden not to any aspects of the original shapes themselves, but only to the distortion between them. Even should the shapes individually be intricate and indescribable, the resulting grids might demonstrate the distortion to be "homogeneous," the expression of a simple system of forces. Figures 1 and 2 display two of Thompson's best-known examples.

Several generations of biometricians have essayed quantifications of Thompson's suggestion. I reviewed the vicissitudes of their ideas in Bookstein (4). Quantification has hitherto been considered a matter of reduction of dimensionality: to describe usefully a particular distortion function using a small number of parametric

“features” which permit the summary and comparison of populations. This quest has proved quite frustrating in practice. When formalisms capture the shifting details of empirical distortions, they are statistically intractable; when they are drawn smooth and homogeneous, they are not subtle enough to capture the manifold local aspects of form change. In Thompson’s original exposition, this dilemma was avoided. The master drew the grids himself, selecting pairs of shapes for which the distortion was already quite homogeneous and simplifying further by selective inaccuracies of

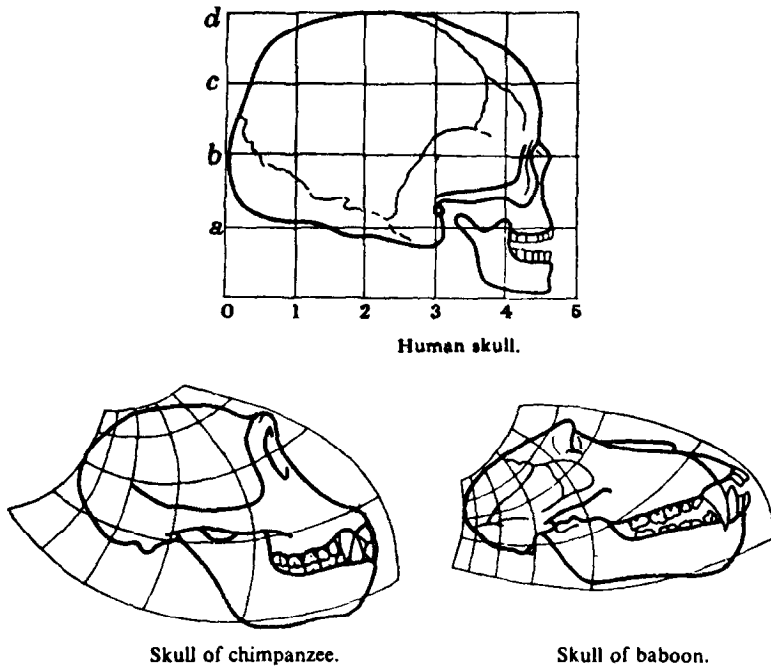


FIG. 2. Transformation grids from *Homo sapiens* to *Pan* and to *Papio*, from Thompson (25).

penmanship. He left no algorithm, no notes by which we might fit such smooth grids to latter-day data. Since his time there have been several rigorous attempts to place the engridments on an algorithmic basis.

The botanist G. S. Avery (3) realized that these distortions can be visualized in a noninvasive manner during the ordinary course of growth. On an immature tobacco leaf, he inked a square grid of dots and photographed their configuration as they moved apart during the leaf’s maturation. The series of photographs encapsulated the ongoing distortion directly. Different regions of the original leaf form displayed spatial variations in rate and principal orientation of growth. In the elegant growth regulation of plant meristems, as reviewed by Schüepp (22), there is a similar aid to visualization. The pattern of distortion is nearly invariant as a spatial field defined in

relation to the moving meristematic apex. Any snapshot of such a tissue contains images of its own transforms in the visually apparent cell lineages gotten by tracing patterns of cell walls.

These techniques both take ungeneralizable advantage of the manner of leaf and shoot growth, which consists in multiplication of biomass between all pairs of loci without morphogenetic movement or tissue attrition. Most applications of Thompson's technique, however, involve sibling forms, or a form and an archetype, rather than a succession of material modifications. One does not have the luxury of a coordinate system fixed in the body of the tissue, but the correspondence of grid points must be constructed indirectly. C. F. A. Moorrees, for instance, diagnoses abnormalities of orthodontic form via Thompson grids which he constructs by bending lines around the irregularly strewn points that customarily provide the orthodontist his data. On any sagittal cephalogram (xray of the head from the side), there can be identified a considerable number of biological landmarks, loci of specific form or function: the ear hole, the frontmost incisor, the bridge of the nose, etc. These are fairly evenly distributed over the whole image of the face and jaw. A "standard" form can be constructed for any population by the statistical averaging of these landmarks' locations according to some conventional coordinate system. Should a Cartesian grid be superimposed over the configuration of landmarks for the standard image, each landmark is fixed at a particular locus inside some square of the standard mesh. Moorrees (16, 17) reverses this determination—he imagines the intersections of the grid as located by relative coordinate inside polygons whose vertices are landmarks surrounding them. For any cephalogram encountered in clinical practice, grid intersections can be constructed from the empirically observed positions of the landmarks just as the standard nodes were reconstructed from the standard positions of the landmarks. When grid nodes are connected by suitably smooth-looking curves, there results an empirical distortion grid which is quite legible by the trained orthodontist.

Instead of constructing a grid directly, one might construct a mapping function explicitly, then infer the grid by holding constant each Cartesian variable in turn. P. H. A. Sneath (24) suggested constructing the mapping function by least-squares bivariate polynomial fit (trend-surface analysis) to the separate Cartesian coordinates of landmark displacement from one image to the other. The resulting function summarizes the broad features of a grid in a small set of coefficients. But the function is insensitive to local fluctuations: it depicts only an approximation to the mapping. The error of the approximation is a further distortion awaiting further display.

#### THE IMPORTANCE OF BOUNDARY

None of the preceding methods recognize that the images analyzed in these computations are biological outlines: simple closed curves, together with their interiors, that are homologous all around their boundaries. The literatures of shape

measurement have not taken this twofold composition into account. In certain methods, each boundary is measured as a separate point-set in the plane, with no pointwise correspondence from instance to instance. Shape contrast, then, becomes a matter of subtracting pairs of parameter vectors, a process quite distant from the biological processes by which they might have been related. Good reviews of the techniques of this approach include Zusne (26, Chap. 5) and Duda and Hart (12, Chap. 9). In another tradition, commonly found in cephalometrics (cf. Merow (15)), the positions of corresponding points are scrutinized closely, but the phenomenon of the outline itself, which is mainly responsible for the perceptual components of shape, bulges and bumps and vaults, is ignored. The tangents and curvatures of the outlines are disregarded, and with them a massive amount of information not associated with landmark positions at all. I explained the complementary failures of these two standard approaches in my dissertation (5, Chaps. 2–3).

Acknowledging the primacy of the extended boundary correspondence alters the structure of the interpolation problem in all essentials. The data by which a pair of shapes is to be compared present not a collection of methodologically equivalent isolated points but a sample from a continuous function that maps boundary onto boundary, together with some correspondences of interior pairs. The homology is *extended inward* from the boundary: the computational problem is, indeed, one of interpolation. The bulk of the information for computing a Thompson transform is implicit in the pairing of boundary forms, a crucial locus hitherto overlooked.

Proceeding from the primacy of the boundary, I construct in this article a new framework for quantification of distortion which combines attractive features from all the pioneering efforts I have reviewed. From Avery I preserve the notion of representing this function by explicit grid points in a distorted image; from Moorrees the systematic adjustment of these grid points to preserve a proper relation to a small number of movable interior landmarks; from Sneath the convenience of estimating real and imaginary components of the mapping function separately in linear analyses. Unlike all the earlier efforts, however, my algorithm divides the computation into two steps explicitly reflecting the dualism of the data: an interpolation from boundary values only, followed by an adjustment for such scattered interior data as may be present. At each step the interpolant may be computed by applying nearly standard linear methods.

#### THE MEASURE OF ROUGHNESS

Algorithms for the “computation” of an interpolating function usually *select*, instead, from within a prescribed vector space of functions. That space may have dimension much higher than the dimension of the data supplied, so that an entire subflat of the function space will be found to fit the data exactly, and selection of a particular interpolant must proceed via some ancillary condition. It accords with the biotheoretical foundations of this problem that the computed interpolant should in all

cases be as “smooth” as possible. I shall set forth a measure of “roughness” instead, which shall be minimized.

Let there be given a pair of shapes  $C, C'$  and a biological homology in the form of a distortion function mapping between them. A suitable operational interpretation of “smoothness” is as follows. Consider a small dot square in  $C$  and its image in  $C'$ , as in Fig. 3a. I suggest a map be considered smooth if the center of the Cartesian square maps into the centroid of the distorted square. As in Fig. 3b, roughness will be characterized for this little cell by the *squared distance* between the image of the centroid and the centroid of the image configuration.

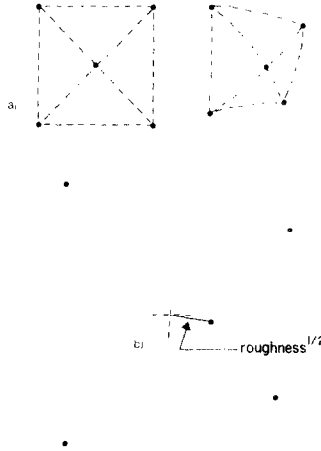


FIG. 3. (a) A small square and its image. (b) Roughness is the squared distance between the image of the centroid and the centroid of the image.

Roughness has a simple analytic approximation. Let the input square have corners of coordinate  $(x_0 \pm h, y_0 \pm h)$  and let the mapping function  $(x, y) \rightarrow (u, v)$  be expressible around  $(x_0, y_0)$  as a Taylor series of second order. Then the four corners of the image configuration are, to second order in  $h$ ,

$$\left[ \begin{aligned} &u(x_0, y_0) + h \left( \pm \frac{\partial u}{\partial x} \pm \frac{\partial u}{\partial y} \right) + \frac{h^2}{2} \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \pm 2 \frac{\partial^2 u}{\partial x \partial y} \right), \\ &v(x_0, y_0) + h \left( \pm \frac{\partial v}{\partial x} \pm \frac{\partial v}{\partial y} \right) + \frac{h^2}{2} \left( \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} \pm 2 \frac{\partial^2 v}{\partial x \partial y} \right) \end{aligned} \right]$$

of centroid

$$(u(x_0, y_0), v(x_0, y_0)) + 2h^2 (\nabla^2 u, \nabla^2 v)$$

where  $\nabla^2$  is the Laplacian operator  $\partial^2/\partial x^2 + \partial^2/\partial y^2$  and all derivatives are taken at  $(x_0, y_0)$ . The value of the roughness is then just  $4h^4((\nabla^2 u)^2 + (\nabla^2 v)^2)$ , to fourth order in  $h$ . A suitable quantity independent of the scale  $h$  of the square is just  $(\nabla^2 u)^2 + (\nabla^2 v)^2$ .

Conformal mappings are smooth and look smooth. The elegance of illustrations of the classic conformal maps, such as Fig. 4, lies in the even seamlessness of the image. The distortion seems to pass from one point to any other with utter smoothness. In

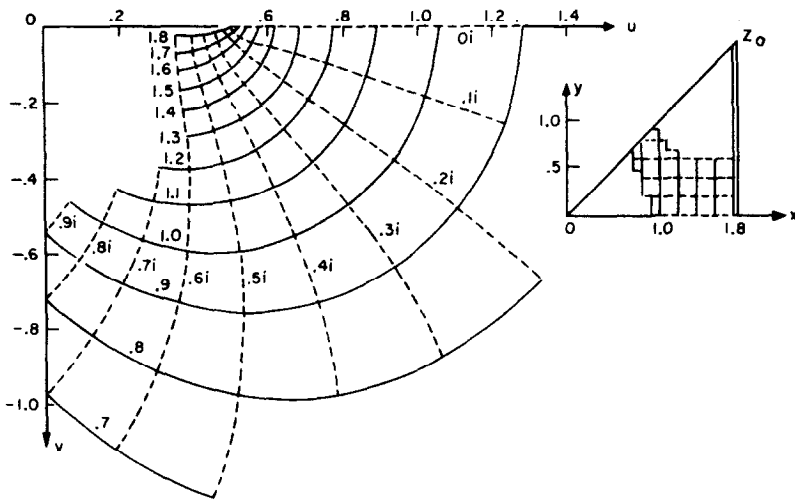


FIG. 4. Conventional diagram of a conformal mapping.

fact, for any conformal map  $(x, y) \rightarrow (u, v)$ ,  $u$  and  $v$  are each harmonic functions, which is to say,  $\nabla^2 u = \nabla^2 v = 0$ , so that the roughness is everywhere identically zero. But it is not necessary for  $u$  and  $v$  to be conjugate harmonic functions for a map to have zero roughness. Any two harmonic functions  $u, v$  define a map  $(x, y) \rightarrow (u, v)$  of zero roughness. For instance, the bilinear map of Fig. 5, which is by no means conformal, is perfectly smooth. As  $u$  and  $v$  are separately bilinear in the Cartesian coordinates, all derivatives of order greater than one vanish.

Roughness is invariant under rotation of the coordinate system for either shape. Rotation through an angle  $\theta$ , in either coordinate plane, multiplies the roughness by  $(\cos^2\theta + \sin^2\theta)^2$ , which is, of course, unity. Then the roughness computed for infinitesimals aligned with the square  $\diamond$  is the same as for the square  $\square$ , or any other. In the definition of roughness as  $(\nabla^2 u)^2 + (\nabla^2 v)^2$ , therefore we did not have to fix the Cartesian axes in the biological form; the computation is invariant with respect to their orientation.

In the algorithms that follow, all distortion functions will be required to have minimum net roughness in the set of all functions affording the requisite homologies. (By "net" roughness I refer to the integral of  $|\nabla^2 f|^2$ , in continuous analyses, but to the unweighted sum of discrete approximations therefor in the discrete-mesh version.) This is an explicit generalization of certain characterizations of splines. In one dimension, the familiar cubic spline has the minimum of  $\int (y'')^2 dx$  subject to

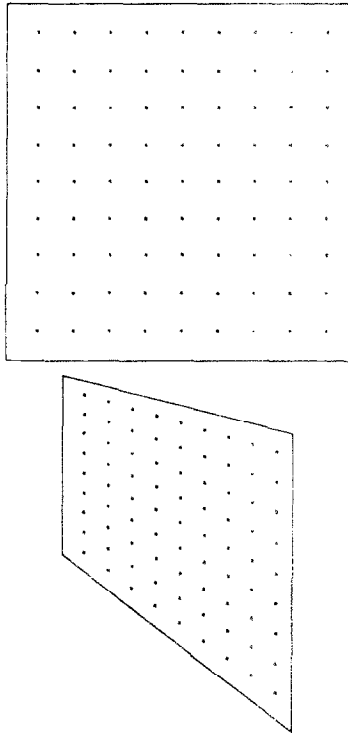


FIG. 5. A bilinear map which is not conformal but which has roughness identically zero.

constraints at knots and at the boundary (cf. Ahlberg *et al.* (1)). The integrand here is the square of the numerator of the simple curvature of the function  $y(x)$  as a locus in the plane. The minimum is attained for a piecewise cubic, which perforce has  $y^{iv} = 0$  except at the knots. For the two-dimensional, complex analysis,  $(y'')^2$  is generalized to  $|\nabla^2 f|^2$ . It can be shown by use of Green's formula (cf. Briggs (7) or Collatz (9, Sect. V.5.6)) that any function which minimizes  $\int |\nabla^2 f|^2$  over a region satisfies  $\nabla^2 \nabla^2 f = 0$  there. This neatly generalizes the finding that  $y^{iv} = 0$  for minimum-curvature one-dimensional splines. Now  $\nabla^2 \nabla^2 f = 0$  is the biharmonic equation, the equation satisfied by thin elastic metal sheets subject to point displacements, just as the spline equation was originally derived to model one-dimensional elastica pinned



to a draftsman's board. That my interpolation function satisfies the elastic equation is not a postulate of the model, however, but a deeper consequence of the more intuitive, data-oriented mathematical formulation in terms of smoothness of interpolation.

THE VECTOR SPACE AND ITS ASSOCIATED FUNCTIONS

Let  $C$  be some simple smooth closed curve in the plane. Let  $h$  be some small spacing, perhaps 5% of the diameter of  $C$ , and let  $(x_0, y_0)$  be some fixed point.

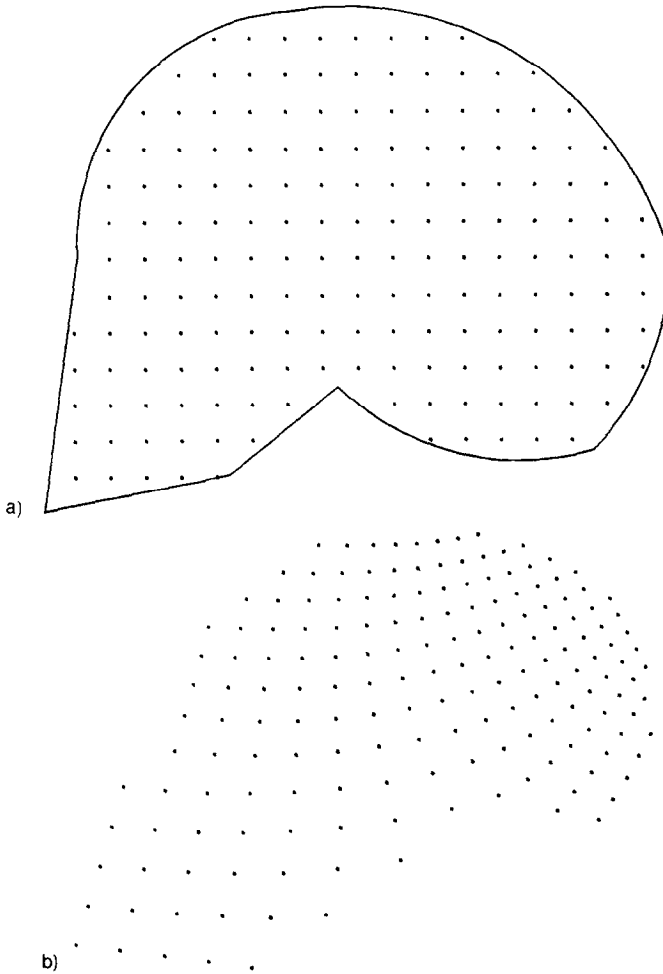


FIG. 6. (a) The set  $S$  for preassigned boundary  $C$ , arbitrary origin  $(x_0, y_0)$ , and spacing  $h$ . (b) The set  $S'$  which depicts a particular vector defined "over"  $S$ .

Consider the set  $S$  of points  $(x, y)$  properly inside  $C$  which satisfy the equations  $(x - x_0 = hn, y - y_0 = hm$  for some integers  $n, m$ .  $S$  is in fact the restriction of a square lattice to the inside of  $C$ , as in Fig. 6a. The number of points of  $S$  will be denoted by  $N$ . The elements of  $S$  need to be assigned some fixed order, perhaps by columns within rows from the top.

For later use I shall need the set  $S_i \subset S$  of "interior points" of  $S$ , points all four of whose nearest lattice neighbors lie in  $S$ . Let  $N_i$  be the count of points in  $S_i$ . I will assume throughout that for no point in  $S$  are both its lattice neighbors above and below, or both its lattice neighbors left and right, outside  $S$ . In other words,  $S$  must be everywhere "wider than 1."

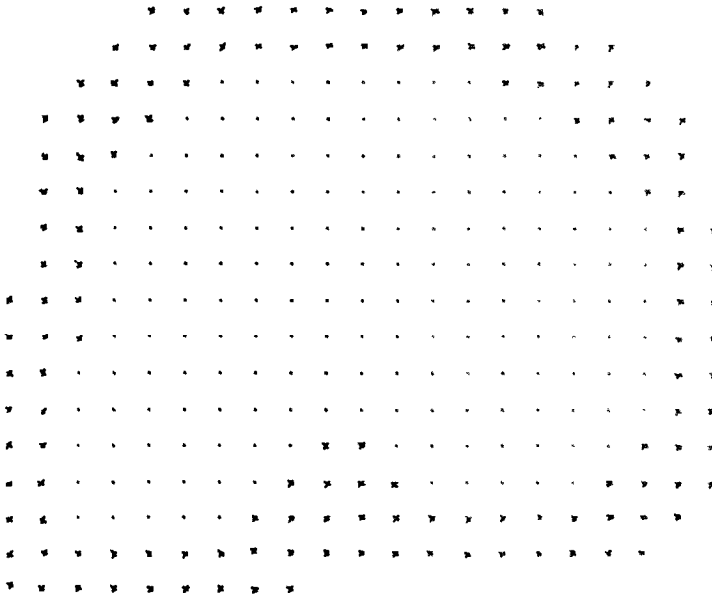


FIG. 7. The gnomon  $G$  around the boundary of  $S$ .

I represent the  $N$ -dimensional vector space  $W$  over the complex numbers  $\mathbb{C}$  by assigning a complex number to each point of  $S$ . A basis for this space may be taken as the set of vectors  $b_s$  equal to  $(1, 0)$  at the point  $s$  and  $(0, 0)$  at all other points of  $S$ . For any vector  $v$  of  $W$ , on another picture plane we place the points whose complex coordinates are the very numbers  $v_s$  assigned to points  $s$  of  $S$  by the vector  $v$ , that is, its components in the basis  $\{b_s\}$ . These points form a set  $S'$ , shown in Fig. 6b. In fact,  $N!$  elements of  $W$  have this same picture. By connecting each point of  $S'$  to the images of those points which were the nearest neighbors of its carrier in  $S$ , one achieves a visual suggestion of distortion. (For all but at most one of the  $N!$  vectors in  $W$  corresponding to any set  $S'$ , the connections will cross somewhere, unless  $S$  has improbable symmetries. The crossing-free representation suggests the distortion I mean.)

A gnomon  $G_1$  of lattice points may be placed about  $S$  such that the set  $S \cup G_1$  includes all lattice points which are within one lattice step of  $S$  in any of the eight cardinal directions (north, northeast, east, southeast, south, southwest, west, northwest). A gnomon  $G_2$  of lattice points is placed about  $S \cup G_1$  similarly. The set  $G = G_1 \cup G_2$  is now a gnomon two thick around the boundary of  $C$ , as in Fig. 7.

To each vector  $v$  in  $W$ , I associate a smooth function  $f_v$  defined throughout  $C$ , which has continuous derivatives of first order within  $C$ , as follows. Each point of  $C$  lies inside some square of side  $h$  whose corners are all points of  $S \cup G_1$ . Suppose that for every point of  $S \cup G_1$  there is supplied a (complex) value of the function  $f_v$ , its derivatives  $\partial f_v/\partial x$  and  $\partial f_v/\partial y$  in the  $x$  and  $y$  directions, and its mixed derivative  $\partial^2 f_v/\partial x \partial y$ . Within each square  $\{(x, y) | x_0 \leq x \leq x_0 + h, y_0 \leq y \leq y_0 + h\}$  the function  $f_v$  will be set to the "Coons patch":

$$f_v(x, y) = [u^3 \ u^2 \ u \ 1] \mathbf{M} \mathbf{B} \mathbf{M}^t [w^3 \ w^2 \ w \ 1]^t. \tag{1}$$

Here  $u = (x - x_0)/h$ ,  $w = (y - y_0)/h$ ,  $\mathbf{M}$  is the matrix

$$\begin{pmatrix} 2 & -2 & 1 & 1 \\ -3 & 3 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix},$$

and  $\mathbf{B}$  is a matrix of complex numbers selected from those presumed supplied:

$$\mathbf{B} = \begin{pmatrix} f_v(x_0, y_0) & f_v(x_0, y_0 + h) & \left. \frac{\partial f_v}{\partial y} \right|_{(x_0, y_0)} & \left. \frac{\partial f_v}{\partial y} \right|_{(x_0, y_0 + h)} \\ f_v(x_0 + h, y_0) & f_v(x_0 + h, y_0 + h) & \left. \frac{\partial f_v}{\partial y} \right|_{(x_0 + h, y_0)} & \left. \frac{\partial f_v}{\partial y} \right|_{(x_0 + h, y_0 + h)} \\ \left. \frac{\partial f_v}{\partial x} \right|_{(x_0, y_0)} & \left. \frac{\partial f_v}{\partial x} \right|_{(x_0, y_0 + h)} & \left. \frac{\partial^2 f_v}{\partial x \partial y} \right|_{(x_0, y_0)} & \left. \frac{\partial^2 f_v}{\partial x \partial y} \right|_{(x_0, y_0 + h)} \\ \left. \frac{\partial f_v}{\partial x} \right|_{(x_0 + h, y_0)} & \left. \frac{\partial f_v}{\partial x} \right|_{(x_0 + h, y_0 + h)} & \left. \frac{\partial^2 f_v}{\partial x \partial y} \right|_{(x_0 + h, y_0)} & \left. \frac{\partial^2 f_v}{\partial x \partial y} \right|_{(x_0 + h, y_0 + h)} \end{pmatrix}.$$

This function  $f_v$  is the unique bicubic polynomial in the Cartesian coordinates which satisfies 16 corner conditions, namely, assigned values of  $f_v$ ,  $\partial f_v/\partial x$ ,  $\partial f_v/\partial y$ , and  $\partial^2 f_v/\partial x \partial y$  at each of the four corners of each little square (cf. Forrest (13) and Rogers and Adams (21, Chap. 6)). It can be shown that this function is  $C^1$  at all lattice points and on all vertical or horizontal segments through lattice points. Inside each square, of course, it is  $C^\infty$  by virtue of its polynomial form.

The values of  $f_v$  on  $S$  are available—they are the components  $v_s$  of  $v$ , the points of  $S'$ . To compute the partial derivatives for the preceding formula, it is convenient to

estimate values of  $f_v$  upon the gnomon  $G$  as well. The extension is by extrapolation along the eight cardinal directions. For any point  $g$  of  $G$ , let  $s_0$  be the closest point of  $S$  such that  $s_0 = g + l(e_x, e_y)$ , where  $l$  is equal to 1 or 2 and each of  $e_x, e_y$  is 0 or  $\pm 1$ . Let  $n_g$  be the minimum  $n$  greater than zero for which  $g + (l + n)(e_x, e_y)$  is outside  $S$ —see Fig. 8. If  $n_g$  exceeds 4, set it equal to 4, to avoid instabilities engendered by distant values. The set of points  $\{g + (l + n)(e_x, e_y) | n = 0, 1, \dots, n_g - 1\}$  is a one-dimensional lattice upon which we can pass a unique  $(n_g - 1)$ th-degree polynomial through the values of  $f_v (= \text{complex numbers } v_s)$  assigned there. The value of  $f_v$  at  $g$  is set equal to the value of this polynomial at  $g$ . (A special adjustment to this routine for the case  $n_g = 1$  searches for a parallel line of lattice points upon which to fit a polynomial which is then shifted. In practice it is invoked rarely.)

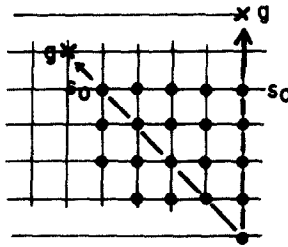


FIG. 8. Domain for extrapolation into  $G$ .

With values for  $f_v$  now assigned at every point of  $S \cup G$ , values of  $\partial f_v / \partial x$ ,  $\partial f_v / \partial y$ , and  $\partial^2 f_v / \partial x \partial y$  are assigned at any point  $s$  of  $S \cup G$  as follows:

$\partial f_v / \partial x$  is set to the derivative at  $s$  of the unique fourth-order polynomial through the values of  $f_v$  assigned at the points  $\{s + l(h, 0) | l = 0, \pm 1, \pm 2\}$ , should these points all be in  $S \cup G$ ; otherwise to the derivative at  $s$  of the unique fourth-order polynomial through the values assigned at  $\{s + l(h, 0) | l = -1, 0, 1, 2, 3\}$  or  $\{s + l(h, 0) | l = -3, -2, -1, 0, 1\}$ , whichever of these sets is entirely within  $S \cup G$ ;

$\partial f_v / \partial y$  is set to the output of a similar numerical differentiation along the ordinate, that is, replacing  $(h, 0)$  by  $(0, h)$  in the previous paragraph;

$\partial^2 f_v / \partial x \partial y$  is set to  $(4h^2)^{-1} \{f_v(s + (h, h)) + f_v(s - (h, h)) - f_v(s + (h, -h)) - f_v(s + (-h, h))\}$ , a simple four-point approximation.

These formulas may be compared with those of Akima (2), which are based on nonlinear estimators. Neither my technique nor his should be confused with ordinary bicubic spline interpolators (cf. Schumaker (23)), whose values are determined globally by all the values  $f_v$  together with certain boundary conditions. These bicubics themselves minimize certain integrals in  $|\nabla^2 f|^2$ ; in the algorithm I present it is the vector, not the function  $f$ , which bears optimality properties.

Values of  $f_v$  and its derivatives now having been set throughout  $S \cup G$ , the values of  $f_v$  anywhere inside  $C$  may be computed square by square according to the

interpolation formula [1] preceding. Since  $f_v$  maps each point of  $S$  into the corresponding point of  $S'$ —the point  $v_s$  carried by the  $s$ th component of  $v$ —it may be considered a  $C^1$  interpolant of the mapping  $S \rightarrow S'$  as a function defined throughout  $C$ .

These functions  $f_v$  are themselves clearly a vector space over  $\mathbb{C}$  isomorphic to  $W$ . For each  $s$ , the formulas for the derivatives are linear on  $W$ , as is the formula for interpolation within lattice squares. Then for fixed  $(x, y) \in C$ , lattice point or no, the value  $f_v(x, y)$  is a linear functional on  $W$ . We will use this fact presently.

It is useful to consider an alternative basis for the vector space  $W$ , one naturally associated with the roughness measure. Let  $s$  be any point of  $S_i$ . Then the  $s$ th element of the alternate basis is the vector

$$r_s = b_s - \frac{1}{4} \sum_{\substack{s' \text{ neighbors of } s \\ \text{north, east, south, west}}} b_{s'}$$

Let  $M_s$  be the vector of coefficients of  $r_s$  in the ordering of  $S$ ; in general it looks like

$$(0, \dots, 0, -\frac{1}{4}, 0, \dots, 0, -\frac{1}{4}, 1, -\frac{1}{4}, 0, \dots, 0, -\frac{1}{4}, 0, \dots, 0). \quad [2]$$

A well-known approximation in numerical analysis states that  $h^2 \nabla^2 f(x_0, y_0) \sim f(x_0, y_0) - \frac{1}{4}(f(x_0 + h, y_0) + f(x_0 - h, y_0) + f(x_0, y_0 + h) + f(x_0, y_0 - h))$ . In the current notation, this means that the roughness of  $f_v$  at  $s$  is approximately  $|r_s v'|^2$ , where  $|z|$  is the Gaussian norm  $(z\bar{z})^{1/2}$ . I shall hereinafter identify roughness with this value.

For points of  $s$  not in  $S_i$ , the formula for roughness at  $s$  cannot be that given above. The usual approximation is instead the Shortley-Weller (19, p. 442). If  $s$  is not in  $S_i$ , it has one or two neighbors outside  $C$ . The segments from  $s$  to those neighbors cut the boundary of  $S$  as shown in Fig. 9. At such cuts there will eventually be

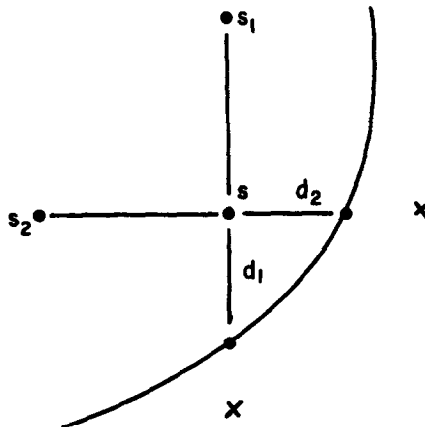


FIG. 9. Geometry of the roughness computation near the boundary of  $S$ .

assigned scalar values  $b_1$  and  $b_2$ , "boundary values," derived from the one-to-one correspondence between boundaries. On either arm of the cross through  $s$ , through the three values unevenly spaced there passes a unique quadratic polynomial, which, extrapolated, yields a value to assign to the missing neighbor outside  $S$ . From the reconstructed full-size cross the roughness may be computed as usual; it is a Hermitian form in the components and the scalars. For the case illustrated in Fig. 9, the approximate roughness is the square of the norm of

$$-2v_{s_1}/(1+d_1) - 2v_{s_2}/(1+d_2) + (2/d_1 + 2/d_2)v_s - 2b_1/d_1(1+d_1) - 2b_2/d_2(1+d_2), \quad [3]$$

where  $d_1, d_2$  are the distances to boundary shown. Should only one neighbor of  $s$  be outside  $S$ , the formula is modified by setting the other  $d$  to unity and replacing the boundary value (a scalar) by the appropriate  $v_s$  (a component). The possibility of other forms of the cross has been eliminated by the assumption that  $S$  is everywhere wider than 1.

The homogeneous part of Eq. [3] supplies a vector  $M_s$  for the points  $s$  of  $S - S_i$ , just as (2) supplies  $M_s$  for  $s \in S_i$ . Construct the matrix  $M$ , which is  $N \times N$ , by  $M = (M_1^t | M_2^t | \dots | M_N^t)$ . Most monographs on the numerical solution of partial differential equations prove that this matrix is nonsingular. (See, for instance, Collatz (9, Sect. I.5.5, Theorem 2 and p. 348).) Then this change of basis is proper, i.e., of full rank. A special algorithm for computing its inverse will be described in due course.

#### INTERPOLATION FROM BOUNDARY VALUES

Let  $B$  be the vector in  $W$ , of components indexed by  $s$  in  $S$ , which has entries 0 for  $s$  in  $S_i$  and the appropriate nonhomogeneous term of an equation like [3] for  $s$  not in  $S_i$ . Consider the system of equations  $v_0 M = B$ , where  $M$  is the matrix defined just above. Since  $M$  nonsingular, we can solve this system of equations for  $v_0$ :  $v_0 = BM^{-1}$ . By the manner in which  $M$  and  $B$  were constructed, the vector  $v_0$  can be characterized as the unique vector which, in conjunction with the boundary values which entered into the components of  $B$ , has approximate roughness exactly zero at every point of  $S$ .

The function  $f_{v_0}$  corresponding to  $v_0$  will not take on exactly the values used in the computation of  $B$  at points such as those drawn in Fig. 9, for two reasons:

- (a) for points  $s$  with two neighbors outside  $S$ , there are two free parameters  $b_1, b_2$  in the correspondence, but only one degree of freedom in the system of equations;
- (b) the simple roughness formula presumes a quadratic approximation to  $f$  between grid points, while  $f$  is cubic by construction.

In spite of these difficulties, in all applications for which the boundary values are smooth functions of position, the fit between  $f_{v_0}$  and the components of  $B$  will be reasonably close.

Suppose, now, that the boundary values entering into the components of  $B$  are assigned as boundary *points*, themselves complex numbers, of a second shape  $C'$ . The assigned values are to be the exact homologs, under some conventional interpolation, of the boundary points of  $C$  which served as surrogate neighbors in the Shortley-Weller formula [3]. Then the set of components of  $v_0$ , plotted as points making up a set  $S'$ , will depict a distortion of  $S$  which interpolates into the interior of

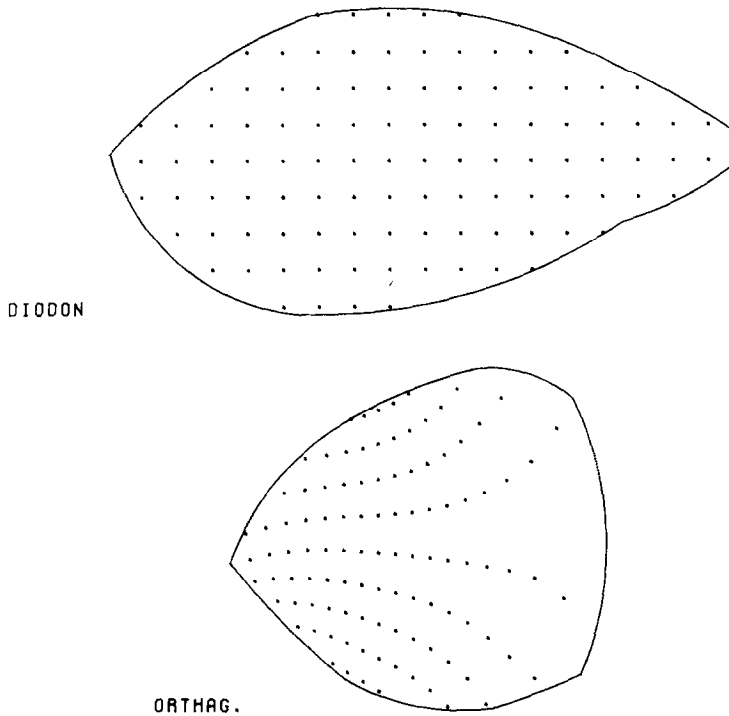


FIG. 10. Boundary-driven correspondence between *Diodon* and *Orthogoriscus*, driven by a seven-landmark digitization of the data in Fig. 1.

$C'$  with roughness exactly zero, and the function  $f_{v_0}$  will be just the distortion function needed in the D'Arcy Thompson formulation for this mesh.

This entire approximating procedure will be well defined once a rule is at hand for the provenance of the boundary values. In my implementation, the boundaries of shapes are constrained to be segments or arcs of conic sections between a small number ( $\ll 100$ ) of landmarks. Between corresponding segments the homology between boundaries is linear in distance; between corresponding arcs it is linear in arc length. Examples of this computation are displayed in Figs. 10 and 11. Corresponding to their exactly zero roughness, they appear smooth throughout.

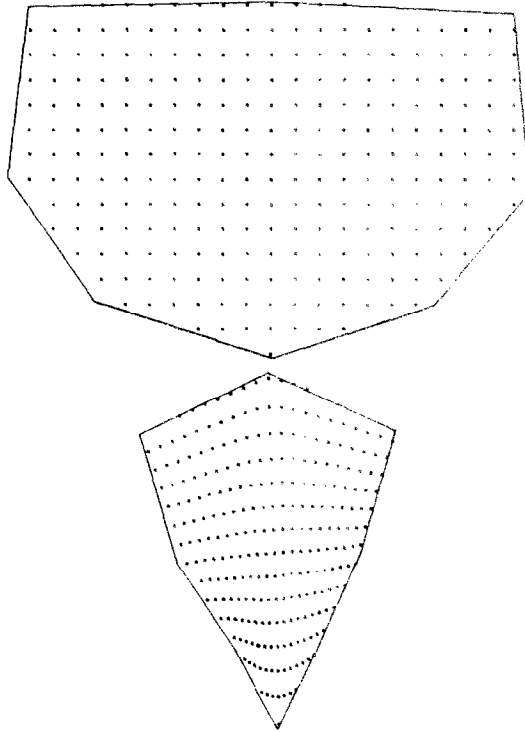


FIG. 11. Shape change from "stage 13" to "stage 15" of the protonotochord of the newt, boundary driven interpolation only. Data from A. G. Jacobson and R. Gordon. *J. Exp. Zool.* **197**, 191 (1976).

#### INTERPOLATION FROM BOUNDARY VALUES AND INTERIOR POINTS

The boundary-driven interpolation function  $f_{v_0}$  has roughness zero at every point of  $S$ . This function maps those boundary points on grid lines very nearly onto their homologs; it maps the boundary *landmarks* onto their homologs to the extent that they are represented by the nearby boundary points on grid lines. But the function  $f_{v_0}$  takes no cognizance at all of interior points—as indeed it cannot, for its vector  $v_0$  is the solution of an exactly determined linear system, with no degrees of freedom to spare.

Insisting that a set of interior points of  $C$  be mapped exactly onto homologs in  $C'$  will require a modification of  $f_{v_0}$ . To this end, consider the computation of  $v_0$  not as the solution of a linear system but as the minimization of the total roughness  $|vM - B|^2$ , where  $|\cdot|$  is now the Hermitian vector norm,  $|(v_1, \dots, v_N)| = (\sum v_i \bar{v}_i)^{1/2}$ . The unconstrained minimum of this form is zero, attained at the value  $v_0$  just computed.

To homologies on interior pairs correspond constraints on the vector  $v$ . Each constraint will be of the form  $f_v(x_1^i, y_1^i) = (x_2^i, y_2^i)$ . By the construction of the



function  $f_v$ , its value at  $(x_1^i, y_1^i)$  is a linear combination in the components of  $v$ , the values of  $f_v$  at the grid points around  $(x_1^i, y_1^i)$ . The coefficients of this linear combination are functions of the relative location of  $(x_1^i, y_1^i)$  in the grid square surrounding it and the proximity of those grid points to  $G$ .

Let us write  $f_v = f_{v_0} + f_{v_1} = f_{v_0+v_1}$ , where  $f_{v_0}$  is the interpolant corresponding to zero roughness, boundary determination only, and  $f_{v_1}$  is an adjustment for interior points. We must have  $f_{v_1}(x_1^i, y_1^i) = (x_2^i, y_2^i) - f_{v_0}(x_1^i, y_1^i)$  for every pair of points constrained to correspond. Since  $f_{v_0}$  fits the distortion very accurately on the boundary of  $C$ , that is, at the intersections of the boundary of  $C$  with grid lines, we also must have  $f_{v_1}$  identically zero there. We wish to compute the vector  $v_1$  subject to these constraints for which the roughness of  $f_{v_1}$ , which is the same as the roughness of  $f_v$ , is minimal.

For the estimation of  $v_1$ , the appropriate vector  $B$  is identically zero, for components corresponding to  $s$  in  $S_i$  and in  $S - S_i$  alike. In light of this all-zero boundary condition, the roughness of  $f_{v_1}$ , summed over all the points of  $S$ , is just equal to  $(v_1 M)(\bar{v}_1 M)^t$ . We wish the vector for which this quantity is minimal, subject to a set of constraints regarding interior pairs  $(x_1^i, y_1^i)$ ,  $(x_2^i, y_2^i)$ . Let the constraints be assembled in the single matrix form  $v_1 Q = R$ , where  $Q$  is the sandwich of coefficient column vectors expressing each  $f_v(x_1^i, y_1^i)$  as a weighted sum of the components of  $v$ , and  $R$  is a vector with  $i$ th entry  $(x_2^i, y_2^i) - f_{v_0}(x_1^i, y_1^i)$ .

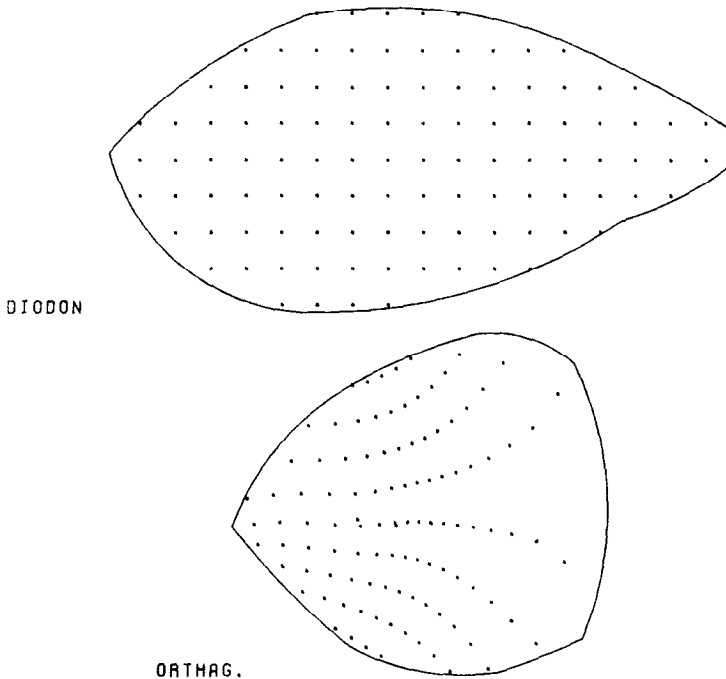


FIG. 12. Boundaries as in Fig. 10, with gill and center of pectoral fin as interior homologs.

Let us change to the alternate basis by writing  $v_2 = v_1 M$ . The problem is now to calculate  $v_2$  such that  $v_2 \bar{v}_2^t$  is minimal subject to  $v_2 M^{-1} Q = R$ . This is a standard problem in quadratic forms (cf. Rao (20, Sect. 1f.1)); the solution is at the orthogonal projection of the origin onto the flat  $v_2 M^{-1} Q = R$ , which is

$$v_2 = R((M^{-1}Q)^t(M^{-1}Q))^{-1}(M^{-1}Q)^t.$$

The vector  $v_2$  supplies an imputed roughness at every point of  $s$ ; the actual complex loci for  $v_1$  are gotten by reversing the change-of-basis— $v_1 = v_2 M^{-1}$ . For this  $v_1$ , the

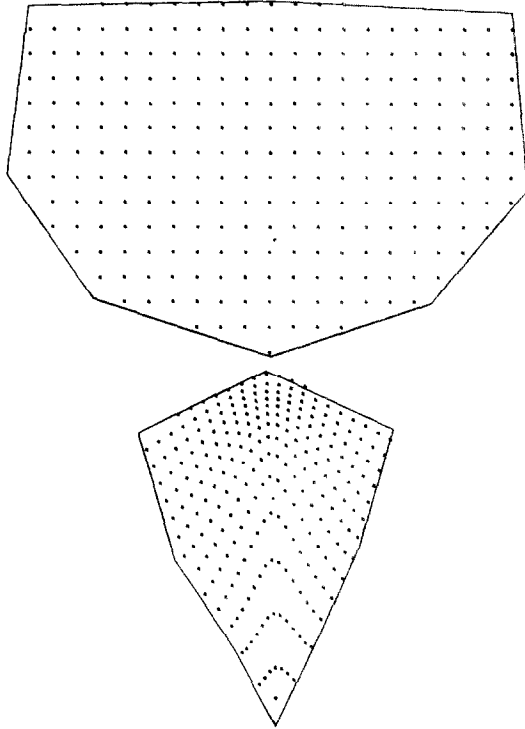


FIG. 13. Boundaries as in Fig. 11, with tip of protonotochord as interior homolog.

function  $f_{v_0 + v_1}$  maps  $C$  onto  $C'$ , approximately preserves the boundary homology and the homology of scattered interior pairs, and minimizes roughness over the set of all  $v$  whose  $f_v$  can be so described. It is the smoothest distortion function possible for applications of the Thompson method to data which include interior correspondences.

Examples of this interpolation are shown in Figs. 12 and 13. These involve the same boundary correspondences as Figs. 10 and 11, respectively, but require in addition the correspondence of the one or two pairs of inside points located at those periods (·) which do not appear to be part of the mesh.

NOTES ON COMPUTATION

The matrix  $M$  is of considerable size. For Fig. 11, for instance, it is 162 by 162, and in practical problems where greater accuracy is required it could easily be as large as  $1000^2$ . To construct its inverse would be a thankless task. We do not ever need  $M^{-1}$  written out explicitly, but only the result of its operating on certain individual vectors:  $B$  and the rows of  $Q$ .

The fast computation of the product of  $M^{-1}$  by arbitrary vectors  $T$  has occupied numerical analysts for some 50 years. Before 1970 it was the accepted practice to compute  $TM^{-1}$  by solving the equation  $vM = T$  according to a variety of successively more sophisticated iterative methods. Shortly after the publication of the fast Fourier transform in 1965, several analysts realized more or less simultaneously that in the special case when  $C$  is a rectangle of sides parallel to the grid lines, the vector  $TM^{-1}$  can be computed "directly" with remarkable speed by taking advantage of the highly regular structure of  $M$ . The eigenstructure of  $M$  can be written out in closed form in these cases. When the grid count is a power of 2, in particular, the

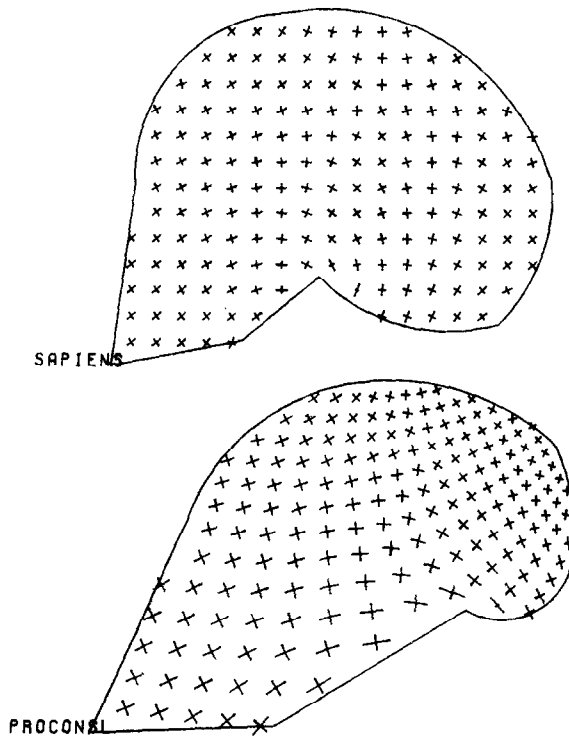


FIG. 14. Principal directions, displayed at points of  $S$  and  $S'$ , for the data in Fig. 6. The upper form is the cranium of *Homo sapiens*, omitting the jaw; the lower form is a reconstruction of the form of *Proconsul*, a 15-million-yr-old fossil. Data from Bookstein (6).

eigenvectors are so patterned that the components of an arbitrary vector with respect to them as a basis can be computed by the FFT in a small fraction of the time necessary for a general orthogonal rotation. In the early 1970s it was noted that for a general geometry, such as concerns us here, the matrix  $M$  is modified by a matrix of rank merely  $N - N_i$  from the patterned case. The modification of  $M^{-1}$  is then a procedure of rank  $N - N_i$  also, rather than rank  $N$ . This brief, brilliant software development may be reviewed in Refs. (8), (14), and (19).

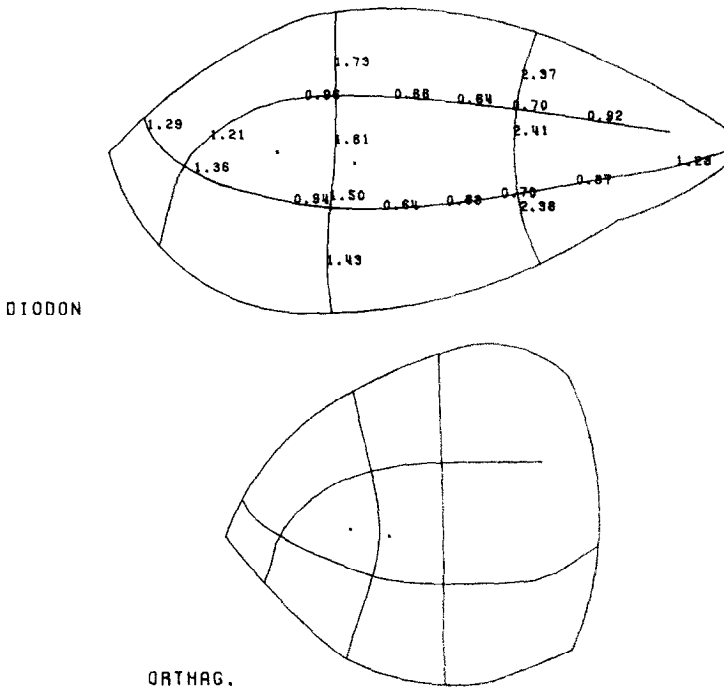


FIG. 15. A sample of integral curves, with selected dilatations, from the biorthogonal grid for the map of Fig. 12.

In Proskurowski and Widlund (18) there is listed a computer program which solves this equation  $vM = T$  for arbitrary real vectors  $T$  and a quite general class of object geometries (sets  $S$ ) by an easy series of calls from a Fortran driver. By executing the computations twice, once with the real part of the boundary vectors and interior homologs, again with the imaginary parts, the complex problem is exactly solved; for the whole estimation of  $v$  is separable, roughness and all, into its real and imaginary components. It is not necessary to understand the mathematical physics of the Helmholtz equation, and the equivalence of the low-rank modification of  $M^{-1}$  to an electrostatic problem of capacitances, to invoke this algorithm effectively. The subsystem published in (18) computed the vectors  $v_0$  and  $v_1$  for all the grids of this article for about forty cents apiece.

Once the vector  $v$  and the function  $f_v$  are in hand, it is straightforward to construct a legible diagrammatic summary of the transformation in accordance with the principles set forth in Bookstein (4, 6). The set  $S'$  of components of  $v$  depicts distortion as Thompson did in 1917: it does not lead the eye directly toward features implicit in the differentials of  $f_v$ , but suggests, irrelevantly, shears imposed from outside the forms. Summary is much more convenient by way of the *principal*

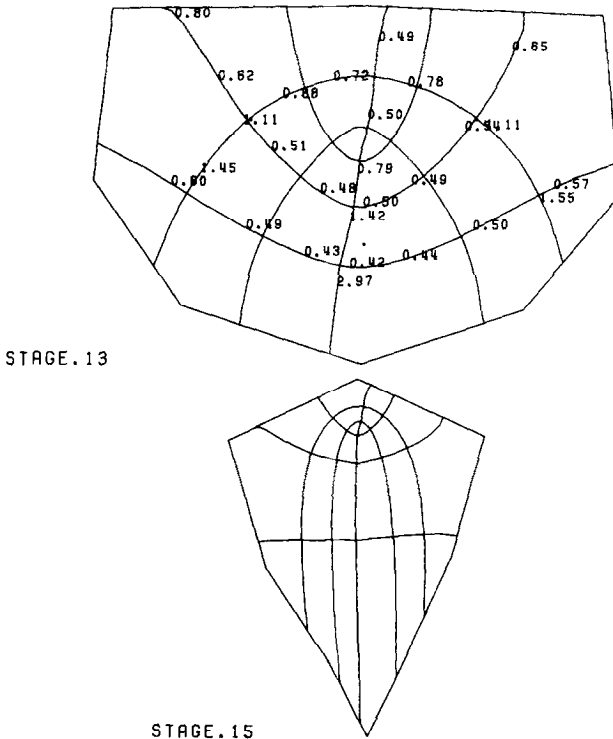


FIG. 16. A sample of integral curves, with selected dilatations, from the biorthogonal grid for the map of Fig. 13.

*directions* of the strain field, those directions at any point which are at  $90^\circ$  in *both* forms. In terms of the function  $f_v$ , these principal directions are the eigenvectors of the matrix  $A^t A$ , where  $A$  is the  $2 \times 2$  affine derivative of  $f_v$ , considered as a map from  $R^2$  into  $R^2$ . Along each of the principal directions at a point, the map  $f_v$  acts by dilatation (stretch or shrink) and upon each  $90^\circ$  cross of directions by rotation. The mapping may be expressed, then, by the field of these crosses, diagrammed with arms of length proportional to the dilatation thereupon. Figure 14 shows some of the crosses for the map whose  $S, S'$  were shown in Fig. 6.

Even better would be the integration of this field of principal directions into a whole network of curves intersecting at  $90^\circ$  throughout the image. One could then extract typical extended loci which summarize whole areas of the map, after the fashion of (4, 6). The formulation of  $f_v$  in the manner I have set forth makes it possible to carry out such an integration directly, since the principal directions, functions of the affine derivative matrix, can be computed at any point inside  $C$  from

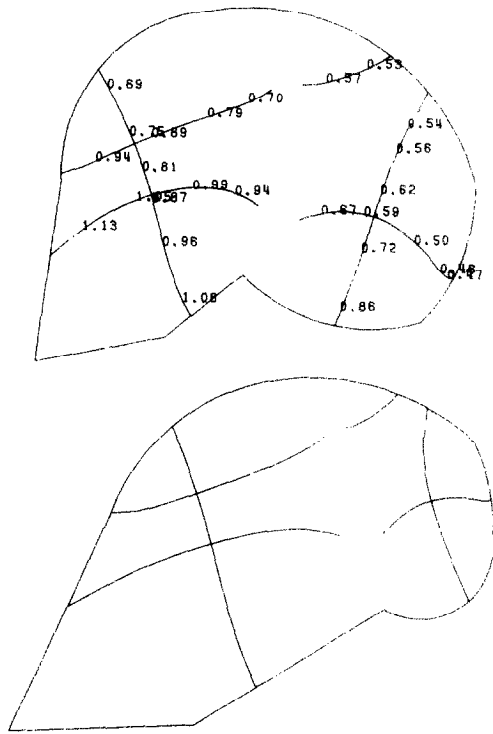


FIG. 17. A sample of integral curves, with selected dilatations, from the biorthogonal grid for the map of Figs. 6, 14. This display is strictly analogous to Fig. 4 of Bookstein (6) but embodies smoothly curving boundary lines and a superior algorithm for integration.

the explicit bicubic formula for  $f_v$  there. As  $f_v$  is continuously differentiable throughout  $C$ , its principal directions are themselves continuous functions of position except possibly at isolated singularities (cf. Bookstein (4, pp. 206–210)). Samples of such integral curves, and some local dilatations along them, are drawn out in Figs. 15, 16, and 17 for the data of Figs. 12, 13, and 14.

For a fixed region  $C$ , maps to multiple forms  $C'$ , or maps with various sets of internal homologies, can all be described by the same data structure, the vector  $v$  and associated discrete geometric information. The exact distribution of homologous points is not encoded in this structure in any fashion, any more than the input data of

a scatter are encoded in the correlation coefficient summarizing it. The representation  $f_v$  of a two-dimensional correspondence is thus highly transportable. Sets of these  $f_v$  can be visualized and compared and statistics of conveniently low dimension extracted by the tensor methods outlined in Bookstein (5, Chap. 8). These computations make practical the display and summary of the maps Thompson invoked in 1917, which have proved hitherto refractory.

All the computations leading to the figures of this article have been combined in a single Fortran program, BIORTHOG, which runs under MTS, the University of Michigan's highly terminal-oriented operating system. BIORTHOG acquires homologous points and piecewise-conic outlines and computes  $S$ ,  $v_0$ ,  $v_1$ , and the various diagrammatic representations of  $f_v$  in a fully interactive manner. At present writing (October 1977), I am engaged in (a) extending these techniques to three dimensions, and (b) designing analogues of multivariate statistics to extract scalar information from the tensor fields underlying the display.

#### ACKNOWLEDGMENTS

Preparation of this article was supported by grant SOC77-21102 from the National Science Foundation. I wish to acknowledge the patient advice of D. P. O'Leary of the University of Maryland and W. R. Tobler of the University of California at Santa Barbara.

#### REFERENCES

1. AHLBERG, J. H., NILSON, E. N., AND WALSH, J. L. "Theory of Splines and their Applications." Academic Press, New York, 1967.
2. AKIMA, H. A method of bivariate interpolation and smooth surface fitting based on local procedures. *C.A.C.M.* **17**, 18 (1974).
3. AVERY, G. S., JR. Structure and function of the tobacco leaf. *Amer. J. Bot.* **20**, 565 (1933).
4. BOOKSTEIN, F. L. The study of shape transformation after D'Arcy Thompson. *Math. Biosci.* **34**, 177 (1977).
5. BOOKSTEIN, F. L. "The Measurement of Biological Shape and Shape Change," Springer-Verlag, New York, 1978.
6. BOOKSTEIN, F. L. Orthogenesis of the hominids: an exploration using biorthogonal grids. *Science* **197**, 901 (1977).
7. BRIGGS, I. A. Machine contouring using minimum curvature. *Geophysics* **39**, 48 (1974).
8. BUZBEE, B. L., GOLUB, G. H., AND NIELSON, C. W. On direct methods for solving Poisson's equation. *SIAM J. Numer. Anal.* **7**, 627 (1970).
9. COLLATZ, L. "The Numerical Treatment of Differential Equations", 3rd ed. Springer-Verlag, Berlin, 1960.
10. DAVIS, P. J. "Interpolation and Approximation." Blaisdell, Waltham, Mass. 1963.
11. DOYLE, F. J. Cartographic presentation of remote sense data. In "Manual of Remote Sensing" (R. G. Reeves, Ed.). American Society of Photogrammetry, Falls Church, Va. 1975.
12. DUDA, R. O AND HART, P. E. "Pattern Classification and Scene Analysis." Wiley, New York, 1973.
13. FORREST, A. R. On Coons and other methods for the representation of curved surfaces. *Comput. Graphics Image Proc.* **1**, 341 (1972).

14. HOCKNEY, R. W. The potential calculation and some applications. *Meth. Comp. Phys.* **9**, 136 (1970).
15. MEROW, W. W. Cephalometrics. In "Handbook of Facial Growth" (D. M. Enlow, Ed.). Saunders, Philadelphia, 1975.
16. MOORREES, C. F. A. AND LEBRET, L. The mesh diagram and cephalometrics. *Angle Orthod.* **32**, 214 (1962).
17. MOORREES, C. F. A. *et al.* The computerized mesh diagram analysis. In "Transactions of the Third International Orthodontic Congress." Mosby, St. Louis, 1975.
18. PROSKUROWSKI, W. AND WIDLUND, O. "On the Numerical Solution of Helmholtz's Equation by the Capacitance Matrix Method." ERDA Research and Development Report C00-30077-99, Courant Institute, New York University, 1975.
19. PROSKUROWSKI, W. AND WIDLUND, O. On the numerical solution of Helmholtz's equation by the capacitance matrix method. *Math. Comput.* **30**, 433 (1976).
20. RAO, C. R. "Linear Statistical Inference and its Applications." 2nd ed. Wiley, New York, 1973.
21. ROGERS, D. F. AND ADAMS, J. A. "Mathematical Models for Computer Graphics." McGraw-Hill, New York, 1976.
22. SCHÜEPP, O. Wachstum und Zellanordnung im Sprossgipfel. *Ber. Schweiz. Bot. Ges.* **62**, 592 (1952).
23. SCHUMAKER, L. L. Fitting surfaces to scattered data. In "Approximation Theory, II" (G. G. Lorentz *et al.*, Eds.). Academic Press, New York, 1976.
24. SNEATH, P. H. A. Trend-surface analysis of transformation grids. *J. Zool. Lond.* **151**, 65 (1967).
25. THOMPSON, D'A. W. "On Growth and Form." abr. ed. The University Press, Cambridge, England, 1961 (1917).
26. ZUSNE, L. "Visual Perception of Form." Academic Press, New York, 1970.