*Statistical Science*
1986, Vol. 1, No. 2, 181–242

# Size and Shape Spaces for Landmark Data in Two Dimensions

**Fred L. Bookstein**

*Abstract.* Biometric studies of the forms of organisms usually consider size and shape variations in the geometric configuration of landmarks, points that correspond biologically from form to form. The size variables may be usefully considered the linear vector space spanned by the set of all distances between pairs of landmarks. The shape of a single triangle $\Delta ABC$ of landmarks may be reduced to a single pair of shape coordinates locating the vertex $C$ in the coordinate system with landmark $A$ sent to $(0,0)$ and landmark $B$ to $(1,0)$. A useful space of shape variables is the span of all such shape coordinate pairs for various triples of landmarks. On a convenient null model of identical circular normal perturbations at each landmark independently, one size variable $S$, which may be taken as the mean square of all the interlandmark distances, has covariance zero with every shape variable. Then associations between shape and size may be tested by the $F$ ratio for multiple regression of $S$ on any basis for shape space. For a single triangle of landmarks, the existence of any mean difference or mean change in shape may be tested by Hotelling's $T^2$ applied to any pair of shape coordinates for that triangle. When such a difference is statistically significant, it may be interpreted as the ratio of a pair of size variables measured along directions at an angle averaging 90° in the samples of forms. One size variable will bear the greatest mean rate or ratio of change between the forms, the other the least. Analysis of configurations of more than three landmarks reduces to consideration of size variables involving at most three landmarks. These techniques are demonstrated in a study of the growth of the head in 62 normal Ann Arbor youth. Each comparison of interest is summarized in its own orthogonal coordinate system, the biorthogonal grid pair.

*Key words and phrases:* Morphometrics, size variables, shape variables, allometry, landmark data, tensor biometrics, biorthogonal grids, craniofacial growth, statistics of deformations.

## SUMMARY

Morphometrics, the study of the geometrical form of organisms, combines themes from biology, geometry, and statistics. Data for morphometric study usually include geometric locations of *landmarks*, points that correspond biologically from form to form. In some applications, one measures configurations of landmark points by variables that express aspects of *size* or *shape* of single specimens: for instance, distances or ratios of distances. In other applications, the morphometrician directly measures the relation between one form and another as a *deformation*, a smoothly varying rearrangement of the configuration of landmarks considered as a whole. Either approach may be turned to the investigation of group differences in size or shape and of associations, called *allometry*, between size and shape or between size change and shape change.

In this essay I attempt to unify these two approaches to morphometrics. From a set of landmark points located on each of a sample of specimens, one can construct two linear spaces, one for size variables and the other for shape variables, which together exhaust all the information about landmark locations. These variables have a spatial as well as a statistical

*Fred L. Bookstein is a Research Scientist with the Center for Human Growth and Development, the Consortium for Research in Developmental and Reproductive Biology, and the Department of Biostatistics of the University of Michigan, Ann Arbor, Michigan 48109.*

structure. It is possible to pass back and forth between the deformational and the multivariate forms of description as follows: A group difference or growth trend detected by looking at variables from the size and shape spaces can be interpreted as a deformation of the whole configuration of landmarks; but also, every deformation prescribes specific size and shape variables in terms of which its effect is most clearly presented.

For purposes of statistical inference, there is an interesting null model in which every landmark is distributed about an (unobservable) centroid by independent, identically distributed normal perturbations in each Cartesian coordinate separately. On this model, one can use conventional multivariate statistical methods, such as $T^2$ ratios and $F$ ratios, to test unambiguously for the significance of apparent group differences in shape or apparent associations of shape with size. One does not need to have decided in advance upon particular size and shape variables for statistical analysis. Rather, using the tie between these spaces and the deformation model, after an effect has been detected one can determine the two size variables per triangle of landmarks which show greatest and least fractions of increase over the change. In most forms of the sample, the segments these distances measure lie at nearly 90°. Along with their ratio, they are the most suitable variables for reporting the findings and displaying them in optimally legible form.

This essay presents the theorems and algebraic and geometric derivations underlying this mixed spatial/statistical method, and demonstrates it in examples from paleontology and craniofacial biology.

## 1. INTRODUCTION

Within biometrics there is a subfield, morphometrics, which analyzes the geometric forms of organisms. Throughout experimental and descriptive biology it is useful to know when two populations of organs or organisms have the same typical form; to indicate *size allometry*, the typical shape changes accompanying size increase over the lifespan of a single organism or along a graded series of specimens or species; and to characterize the typical difference between the sexes, the typical response of form to therapeutic intervention or environmental variation, the typical facies of an anomaly, or the most reliable dimensions of individual difference.

In practical applications, these themes tend to be interwoven in many ways. For example, many human birth defects present gross malformations of the bones of the face. One syndrome, Apert's Syndrome, involves an upper jaw and nasal region which appear to have been "caved in" toward the back of the head. There are associated serious dental and ophthalmic problems and many other features clearly visible on x-rays. To correct the facial deformity, the upper jaw is cut apart from the rest of the head, moved forward and downward, and wired back in. One set of questions about the syndrome (Bookstein, 1984b) involves the detailed loci and origin of the deformity. From a sample of cases we can learn which quantitative features are of normal size but incorrectly positioned and which are always too small. The closer these latter are to the base of the brain, the more likely they are to have been initiating sites for the craniofacial manifestations of the syndrome.

Another set of questions about the same sample deals with the design of individual surgical treatment. In planning these operations, one knows the extent of malformation, but not the extent of feasible surgical correction. We need statistical summaries of treatment success—maintenance of the altered form—as a function of both the need for and the extent of repositioning. In younger patients, we also wish to determine whether growth following surgery is normal; if not, the operation must "overcorrect" to anticipate failure of the jaw to grow normally after it is repositioned. Populations submitted to different surgical protocols may average different success rates; more crucially, different procedures may succeed upon patients of different starting forms. At every turn in the clinical management of a single patient, we face statistical questions about the relations among starting form, change during surgery, and change after surgery. A similarly complicated set of questions about the interaction of condition with treatment arises in diagnosing the location and extent of a heart attack, and then assessing the success of a treatment based on that diagnosis, by observing changes in the shape change which is cardiac contraction.

Morphometric study of questions like these requires that information be drawn from two sources: biological homology and geometric location. A biological *homology* is a spatial or developmental correspondence between individuals, a correspondence among definable structures or "parts"—separate bones, nerves, muscles, and the like. In the context of morphometrics it becomes a homology *mapping*, a correspondence not of parts to parts but of points to points. For any choice of point or curve upon or inside any particular form, the homology map associates well-defined and biologically acceptable counterparts, the *homologues* of the point or curve, on all the other geometric forms in the data set. Morphometrics studies the empirical geometry of homology—variation in the relative locations of sets of homologous points over a sample of forms.

One normally samples this map at a small number of discrete points, called *landmarks*, defined intrinsically in terms of the anatomy in their vicinity. Landmarks are points pointed out by biologists when they

talk about form whether or not it is quantified: "the tip of the axis," "the bridge of the nose." There are two general modes of characterizing landmarks. Some are located by juxtaposition of different identifiable structures. In fishes, "anterior fin base" and "posterior fin base" delimit the fin upon the body outline; in the human skull, Nasion, the bridge of the nose, is at the intersection of three bony sutures visible upon x-rays. (In craniometrics, the quantitative analysis of skull form, landmarks have proper names, usually in a sort of mock Greek or Latin: for instance, Sella, Nasion, Menton, and Gonion in Figure 12.) Other landmarks are located by geometric properties. For instance, the tip of a tooth may be taken at the point where the curvature of the edge is greatest. Such a feature, we must assume, is the relic of a highly positional developmental process we cannot observe directly. Homology maps smoothly extend the spatial correspondence between forms from landmarks like these into the regions of anonymous tissue, blood, or air in between them. The morphometrician and the biologist must collaborate especially closely at this stage of a project.

Figure 1 exemplifies the decisions involved in working out these notions in practice. Figure 1a presents typical raw data: detailed drawings of three primate skulls, lower jaws removed, in midsagittal section (that is, cut flat up the middle of the head in a front-to-back vertical plane). The parts and regions labeled in the figure are conventionally accepted as homologous from species to species. In Figure 1b, the same three forms are abstracted into line in order to characterize nine landmark points in whose homology we have some confidence (not only in this little data set but in most other higher primates as well). There seem to be no landmarks over the vault of the skull, nor any upon most of the teeth (as they do not lie in this plane). Figure 1c shows how seven of the landmarks may be taken to delimit a region, the splanchnocranium (nonneural skull), that we may consider as related by homology mappings among these forms. After the fashion of D'Arcy Thompson (1961, Figures 178–180), the homology map may be displayed as a deformation of *Homo* into each of the other forms. These maps were computed by an algorithm described later in this essay, in conjunction with Figure 15. More on the analysis of these and other primates may be found in Chapter VII of Bookstein (1978a).

The statistical analyses put forward in this essay will implicitly assume that correspondences of landmark positions are a discrete sample of paired points from an extended correspondence of regions such as those in the figure. In one aspect, however, these examples are atypical. They are at the limit of "small shape variation" for which the algebraic methods of this essay are adapted. For smaller changes the depiction by explicit homology maps, as in Figure 15a, is peculiarly bland. The viewing eye cannot detect the patterns of spatial variation embodied in these maps. Nevertheless patterns are present; they may be made patent, as in Figure 15c, by appropriate morphometric tactics.

Topological variations in the assembly of organisms out of organs often lead to fascinating ambiguities of biological homology, and biologists may reasonably argue the facts of the homology map relating sets of observed landmarks—the very mapping we wish to study. In particular, points with a perfectly good operational definition, like "the tip of the nose," may not correspond from specimen to specimen in any biologically meaningful or useful way. Such disagreements are beyond the scope of this essay (but see Bookstein, 1978a, pp. 120–123, or Bookstein et al., 1985, Section 5.1).

Landmark locations may be augmented by information about the curving of external or internal boundaries between landmarks, but this extension will not be demonstrated here. All the geometric data analyzed in this essay will be discrete landmark locations in the plane of a photograph, diagram, or digitizing tablet: two-dimensional Cartesian coordinate pairs of homologous points over a sample of forms.

### The Two Main Strategies of Morphometrics

Data bases of landmark locations are usually analyzed by one of two general strategies: multivariate morphometrics or deformation analysis. In the practice of multivariate morphometrics (Reyment et al., 1984), configurations of landmarks are measured one at a time in collections of "morphometric variables." Some of these measures are size variables which increase exactly as some power, preferably the first, of geometric scale: for instance, distances between landmarks in pairs. Other morphometric variables are shape variables having a value algebraically independent of geometric scale. Shape variables include ratios of distances or other pairs of size variables, together with various functional transforms of these ratios, such as angles.

Multivariate morphometrics does not insist that size or shape measures be based in the locations of homologous landmarks. The methods are often applied without modification to distances extracted without regard for homology, such as maximal or locally minimal diameters of a form, and to coefficients from the expansion of a boundary curve as a series of orthogonal functions. Collections of these morphometric variables, whatever their provenance, are submitted to conventional multivariate statistical analysis—linear modeling, discriminatory analysis, component extraction—and any findings are interpreted coefficient by coefficient. The geometric origin

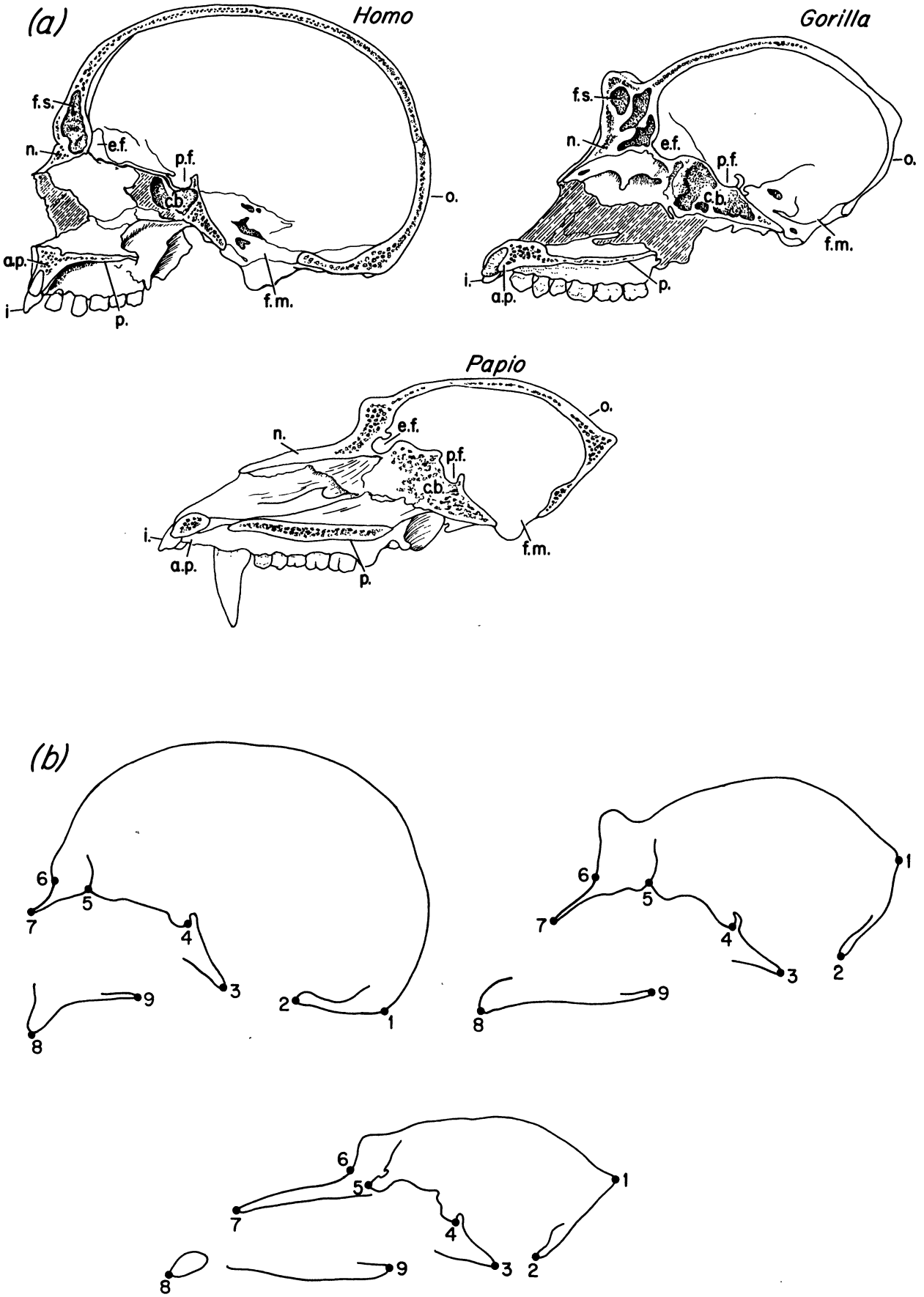
(a)
Homo
f.s.
n.
e.f.
p.f.
c.b.
o.
a.p.
i
p.
f.m.
Gorilla
f.s.
n.
e.f.
p.f.
c.b.
o.
f.m.
i.
a.p.
p.
Papio
o.
n.
e.f.
p.f.
c.b.
i.
f.m.
a.p.
p.
(b)
6
5
7
4
3
9
2
1
8
6
5
7
4
3
2
1
9
8
6
5
7
4
3
2
1
8
9

FIG. 1a–b.

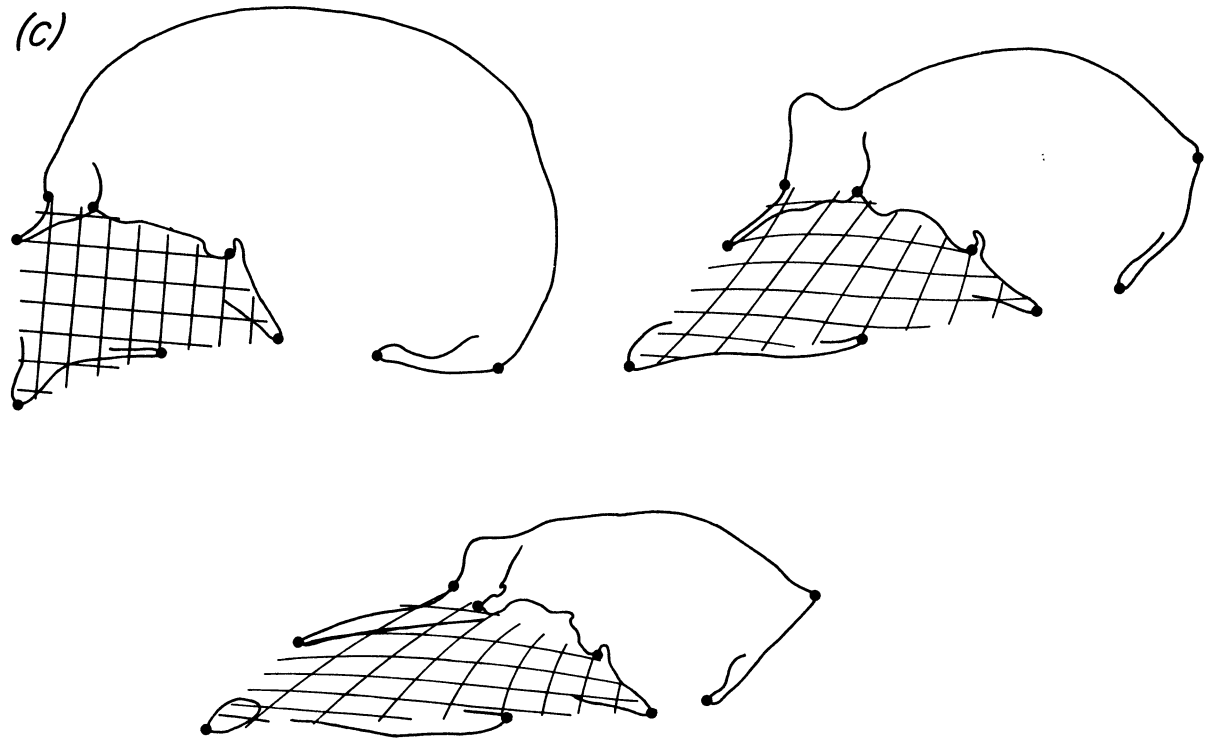


FIG. 1. *Landmarks and the homology map.* (a) *Careful drawings of three primate skulls in midsagittal section. Homo (modern European) and Gorilla (female) after Hofer* (1954); *Papio sphinx (male mandrill) after Hofer* (1965). *The forms are not to the same scale. Parts and regions with the same names in different figures are homologous in the usual biological senses. Abbreviations: o., occipital bone; f.m., foramen magnum (opening to the spinal column); c.b., cranial base; p.f., pituitary fossa; e.f., ethmoid fossa; f.s., frontal sinus; n., nasal bones; a.p., alveolar process or ridge; i., front incisors; p., hard palate.* (b) *Abstractions of these figures, with indications of nine landmarks.* 1, *external occipital protuberance ("Inion"); 2, posteriormost point of the foramen magnum ("Opisthion"); 3, anteriormost point of the foramen magnum ("Basion"); 4, deepest point of pituitary fossa; 5, deepest point of ethmoid fossa; 6, deepest point of curvature of nasal bones ("Nasion"); 7, terminus of internasal suture ("Rhinion"); 8, tip of alveolar bone between the upper central incisors ("Prosthion"); 9, posterior end of the hard palate ("Staphylion").* (c) *Computed homology maps corresponding to the region delimited by landmarks 3–9. The computation is by the algorithm of Bookstein* (1978a); *cf. Fig.* 15.

of the measured variables is generally not exploited further. (Along with myself, Charles Oxnard is an exception to this generalization; see, for example, Oxnard, 1984.)

In multivariate morphometrics, the relation of analysis by size variables to analysis by shape variables or by a combined list of size *and* shape variables is presently obscure (Sprent, 1972). The one solid result in the area is a theorem of Mosimann (1970). It deals with the case of a fixed list of size variables $X_1, \cdots, X_J$ and a list of shape variables generated by dividing all the $X$'s by a size variable that is an algebraic combination of the $X$'s: for instance, the lengths of subdivisions of a vertebral column as proportions of the length of the whole. Mosimann's theorem states that at most one size variable can be stochastically independent from the space of shape variables spanned by a list constructed in this way. In other words, although "size" and "shape" are verbally orthogonal, computationally and conceptually they are inextricably entangled.

The other main morphometric tradition concentrates on the theme of *deformation* introduced into descriptive biology by D'Arcy Thompson (1961) under the label of "Cartesian transformation." A deformation is a mapping which takes neighboring points to neighboring points and which alters lengths of little segments by factors which never get too large or too small. The notion is an informal version of what the mathematician calls a *diffeomorphism*, a one-to-one transformation which, along with its inverse, has a derivative at every point of a region and its image.

Thompson argued that a comparison of biological forms ought to be observed directly, as a geometric object of measurement in its own right, rather than as the mere numerical difference of measures made upon forms separately. He suggested that the form change be construed as a deformation of the picture plane corresponding closely to what biologists already knew as homology: the smooth mapping of one form onto the other sending landmarks onto their homologues and interpolated suitably in between. The object of

study thereby becomes the mapping relating the pair of forms, rather than the configuration of either form separately.

I have reviewed the history of Thompson's suggestion in prior publications (Bookstein, 1978a, 1982a). In spite of the visual attractiveness of Thompson's own transformation diagrams and those following later, they have proved surprisingly resistant to statistical consideration in groups. I think this is principally because statistical techniques suited for dealing with the requisite richness of parameters, nearly twice the number of landmarks, have hitherto had difficulty incorporating the spatial context in which the measurements originated. Instead, most techniques for the geometric study of mappings choose to model the reconfiguration of landmarks by a map from some algebraically simple family, and then interpret either the few coefficients of the fitted map or else its distributed "error" of fit.

A typical approach of this sort is that of Sneath (1967). He expresses each Cartesian coordinate of the landmarks of one form by a cubic bivariate polynomial (estimated by multiple regression) in the $x$ and $y$ coordinates of the same landmarks in another form. The resulting coefficients are apparently impossible to interpret directly. Sneath's purpose is instead to summarize them in a single net measure of dissimilarity between forms, a measure which might be processed afterward by a clustering algorithm.

A more specialized model in the same spirit (Siegel and Benson, 1982; Benson et al., 1982) postulates a conserved subset of at least half the landmarks: a region of the form in which the mapping is nearly a pure change of scale or isometry. "Errors" in the fit of this simple model are represented as little vectors of displacement tying "expected" location to observed location for each landmark separately. But the fitted mapping cannot recognize differences in change of scale, however slight, as a function of either position or direction. Hence the shape changes observed, which are the features of greatest interest to Benson and his co-authors, are estimated as a sort of mis-specification term for the isometric model actually fitted. An appropriate analysis of these "residuals" requires instead the regional quantification of shape differences by the methods of this essay.

Another technique (Tobler, 1978) computes a shape change that is constant throughout the form. His model admits directional differences in rates of change of length, but cannot detect regional gradients in those rates.

In applications to biological data these models all bear, in my view, too few parameters to properly measure most of the phenomena of shape change to which they are likely to be applied. In contrast, Goodall (1983) shares with me a willingness to deal with the spatial complexity inherent in describing reconfigurations of many individual points. From a botanical data base detailing the tessellation of a growing shoot's surface by its cells, Goodall computes parameters of shape change separately cell by cell. The dependence of growth rate upon position and direction throughout the form is displayed directly. Goodall also keeps track of relative rotations between the different regions of the shoot. As of this writing, his machinery has not yet been extended to the summary of many specimens at once by way of reliably located landmarks.

## Deformations and Factors

Between the two styles of morphometric analysis, the multivariate and the deformational, there is a profound tie. A deformation moves landmarks around. In so doing, it changes all distances that one might consider measuring upon a configuration of landmarks; it changes them whether or not they are measured. In this action the deformation behaves very much like a *factor* in the classic psychometric sense: an "underlying" construct which accounts for simultaneous changes in a large number of conceivable variables, only some of which are actually observed (Bookstein, 1986). When a number of variables appear to change simultaneously with each other, we attempt to model each one as varying according to a regression upon the *factor score* which estimates their common trend. The regression coefficients are called *loadings*. When two variables both load upon the same factor, their observed covariance includes a term for the product of their separate loadings; in a good factor model, these products of loadings will closely match the covariances actually observed (Bookstein et al., 1985, Section 4.2). The systematic patterns of mean deformation commonly encountered in biology have exactly this sort of effect on distance measures: they account for most of the observed covariation.

Prediction of a variable by a factor upon which it loads is not by multiple regression upon the factor's indicators, but instead by simple regression of the dependent variable upon a suitable estimate of the factor score. Factor scores are usually estimated as appropriately weighted sums of the variables which are supposed to load on them, the variables whose observed correlations are sufficiently high. But factors need not be estimated so. Deformations have a great advantage over the sort of factors one normally encounters in multivariate analysis: they are explicitly observable in the same planes in which the data lie, so that their loadings have a spatial as well as a statistical structure. We can compute the loading of any distance upon any deformation, to great accuracy, without measuring the distance at all. Up to effects of

curvature, the loading is just the net increase in length the deformation engenders over the straight line path followed by the length measure. To control for starting length, we usually measure this change upon a logarithmic scale—change per original unit length, in dimensionless units such as mm/mm—whereupon the loading of any distance is identified with the derivative of the deformation in its direction.

By using factor notions rather than notions of multiple regression, we are freed from having to measure the deformation on any particular scale or by any particular set of variables, and we need not be concerned that the distance being accounted for is involved in specifying the precise nature of the deformation under study. We simply *observe* a deformation as a single construct, the relative reconfiguration of a set of landmarks. Ultimately, the reason for recourse to this indirect language of factors is the infinite variety of shape variables available for describing any configuration of landmarks. A deformation represents a simultaneous change in all these variables, each with its own loading computable from the mean geometry.

Because the derivatives of deformations may be studied analytically, we may use their geometry to select particular size and shape variables customized for the description of particular comparisons of form. The size variables that most interest us are those with the largest or smallest loadings on the deformation. Any such pair generates a shape variable, namely their ratio, which changes fastest (on the logarithmic scale) of all shape variables across a comparison or along a trend. We will see in Section 5 that the pairs of distances so selected have the happy property of remaining at 90° over the course of the deformation; this invariance considerably simplifies the reports and diagrams of findings.

Thus deformations may be usefully interpreted as factors. Can we manage the reverse of this tactic, representing multivariate factors of size and shape as deformations? Like the computation of loadings from the derivatives of a deformation, this step is made possible by the two-dimensional geometry of the plane in which our data lie and by the finite dimension of the landmark configuration itself. A deformation of a triangle of landmarks, for instance, is determined by the loadings it induces upon any three distance measures. These effects, in turn, may be represented by the deformation's effect upon any two shape measures together with just one size measure shared by all the triangles of a configuration. In this way, we may read statistical analyses of suitable spaces of size and shape variables as pertaining directly to deformations, and thereby we generate novel size and shape variables that best report the deformation.

It is the purpose of this essay to exploit this combination of spatial and statistical analysis for the comparison of samples of forms and the description of samples of form change.

## Symbols and Assumptions

The algebra of the models introduced here interprets size and shape as functions of many complex variables. Therefore, the raised horizontal rule $\overline{\phantom{x}}$ (e.g., "$\overline{dz}$") is used to indicate not a sample mean but the operation of complex conjugation, reflection in the real axis. The absolute value operator (e.g., "$|Z_i - Z_j|$") represents the *modulus* of its argument, its distance from (0, 0) in the complex plane. The operator Re returns the real part of a complex number, Im the imaginary part. When we need an expected value, we will write it using a boldface **E**; thus, $\mathbf{E}\,|Z_i - Z_j|$ denotes the expected value of the distance between landmarks $Z_i$ and $Z_j$. As $i$ (and also the engineers' $j$) are ubiquitous as statistical subscripts, the quantity $\sqrt{-1}$, whenever it needs to be explicitly mentioned, will be written out in full. The symbol $\Delta$, when followed by a list of three landmarks, should be read as "triangle" (e.g., "$\Delta ABC$").

Sets of $K$ landmarks observed upon a single form will be modeled as lists of $K$ complex numbers $Z_1, \cdots, Z_K$. Each $Z_i$ corresponds to one landmark; in the sample, it is assumed to vary around a centroid $W_i$ by displacements $dz_i$ the algebraic treatment of which is like that of the differentials of classic complex analysis. The variation of any of these locations $Z_i$ about its own centroid $W_i$ is assumed considerably smaller than the distance between any pair of centroids:

$$|dz_i| = |Z_i - W_i| \ll |W_j - W_i|,$$
$$i, j = 1, \cdots, K, i \neq j.$$

This variation includes both measurement error (from tracing and from digitizing) and true differences between individual configurations.

The symbol $d$ for differential, already used in referring to small displacements $dz_i$ of the landmark locations $Z_i$, will also serve for the changes in value introduced in functions of the $Z$'s and for summaries of differences in those functions observed between groups or over the extent of a trend. Thus $d\,|Z_i - Z_j|$, subject of the first Lemma, is the differential of the function $|Z_i - Z_j|$, the difference $|(Z_i + dz_i) - (Z_j + dz_j)| - |Z_i - Z_j|$ through terms of first order in the $dz$'s. We rely upon the differential notation primarily for the linear algebra of deviations of the $Z$'s, and functions of them, from their mean values.

We postulate, formally, a *null model* for analysis of these configurations. In this model, the $Z_i$ are independently and identically distributed about their centroids $W_i$ as circular normal deviates of a common variance $\sigma^2$, as indicated schematically in Figure 2.

(By "circular normal" is meant "bivariate normal with covariance matrix a multiple of the identity." The sense related to distributions upon a circle is not used in this essay.) This variance subsumes all measurement error, biological noise, and the like. By the assumption about $|Z_i - W_i|$ preceding, we may conclude that the expected value of the distance between landmarks $i$ and $j$, $\mathbf{E}|Z_i - Z_j|$, is almost exactly $|W_i - W_j|$. For expressions in which the coefficients of linear forms in the differentials $dz$ are replaced by their expected values in terms of the $W$'s, and which are therefore accurate through terms of first order in $\sigma$, we reserve the symbol $\simeq$, thus: $\mathbf{E}|Z_i - Z_j| \simeq |W_i - W_j|$.

The assumptions of this model are most directly verifiable by considering the circularity of certain derived scatters representing perturbations of landmark positions taken three at a time, in triangles. There will be a comment about this matter in Section 3.

Other models to be tested against this null model would have certain aspects of the joint distribution of the $Z$'s depend on other circumstances: upon a biological grouping variable, perhaps, or a size measure (likewise defined in terms of the $Z$'s).

### The Variables of the Null Model Are Not Directly Observable

The coordinate pairs $Z_i$ of our model are not defined absolutely, only with respect to each other, and so cannot be observed. (This is analogous to the notion of the hypothetical "open" or "parent" array of geomathematical variables from which covariances observed among empirical proportions of components in mineral ores are presumed to derive. See Chayes, 1971.) We cannot detect net translations or rotations that replace each $Z_i$ by $Z_i + Z$ or by $\alpha Z_i$, where $Z$ is any complex number and $\alpha$ is a complex number of unit modulus; for there is no reference coordinate frame with respect to which to detect these changes. We can observe only an equivalence class of sets of these coordinates.

Likewise, the centroids $W_i$ are unobservable; they have only relative locations. Furthermore, they are "fuzzy points" that cannot, in general, be located properly with respect to one another in a Cartesian plane. In the course of proving Lemma 2 in the Appendix, for instance, we will note that lines such as $W_1W_2$, $W_3W_4$ are usually skew: no linear combination of one pair of landmarks is ever at mean square

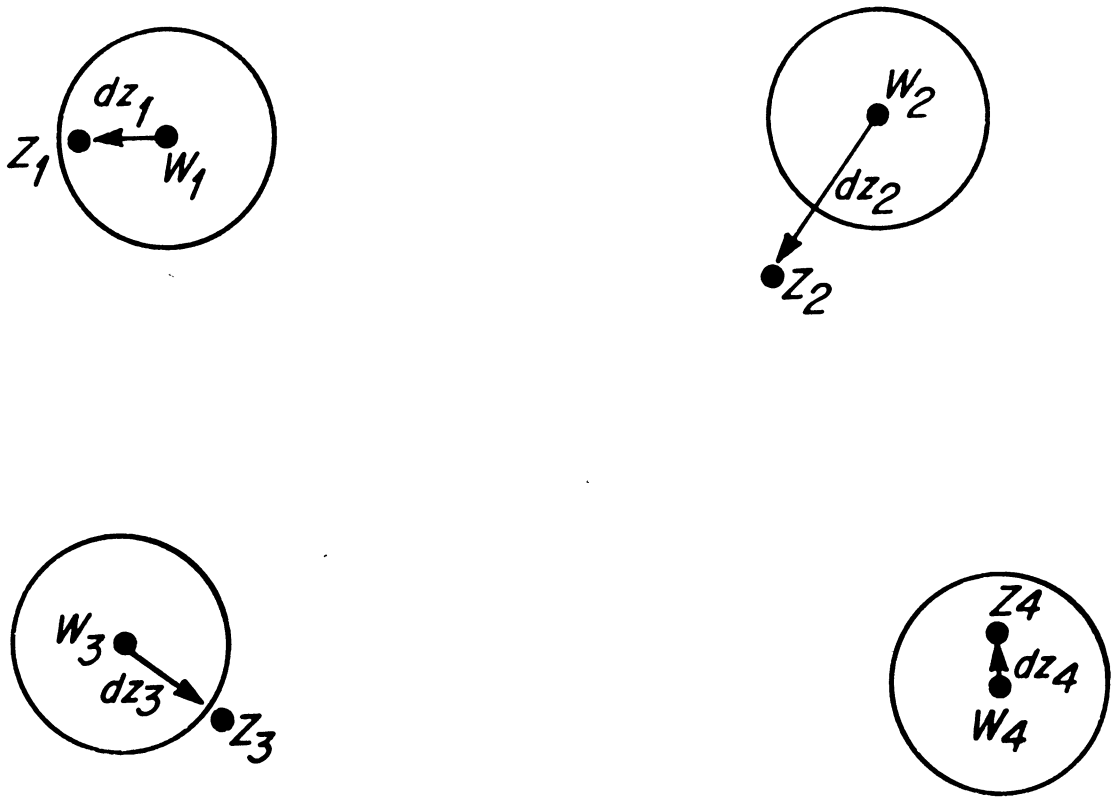


FIG. 2. *The unit of morphometric analysis is a configuration of K landmarks. These are modelled as sets of K complex numbers* $Z_1, \cdots, Z_K$ *varying by displacements* $dz_1, \cdots, dz_K$ *around centroids* $W_1, \cdots, W_K$. *The radius of the circles in the figure represents the standard deviation* $\sigma$ *assumed for each component, real or imaginary, of each independent and identically distributed displacement* $dz_i$.

distance zero from any linear combination of another pair.

Six edge lengths among four points determine a tetrahedron. Whenever these lengths are averaged separately over a sample of forms, whether as root mean square expectations or otherwise, for any four centroids $W$ the resulting tetrahedron is usually not quite planar. The centroids $W_i$ of the null model will not lie exactly in a plane unless these sample statistics are relaxed very slightly. For an algorithm suited to this purpose, see Strauss and Bookstein (1982).

Nevertheless, we must draw the points $W_i$ *somewhere* upon the printed page. For this purpose and this purpose only, it is reasonable to average Cartesian coordinates in a system registered upon (that is, placed and oriented with respect to) certain landmarks. For instance, one might assign landmark $Z_1$ to the Cartesian origin $(0, 0)$ and constrain the $x$ axis to pass through another landmark $Z_2$, so that the line $Z_1Z_2$ is the $x$ axis of the Cartesian system and the perpendicular to this line through $Z_1$ is the $y$ axis. But the statistical properties of these coordinates separately are highly misleading. On the null model, the two Cartesian coordinates of any landmark so registered no longer have a joint distribution that is circular normal; more seriously, measurement variance in the points of registration is propagated to all other landmarks as covariance of coordinates between all landmarks in pairs. Although we will use particular coordinate pairs from conventional registrations to display mean configurations, we cannot execute morphometric statistical analysis upon them directly.

What we may observe of our set of points $Z_i$ is in fact the rigid triangulation of their configuration, a specification, case by case, of every distance $|Z_i - Z_j|$ and every inner product $(Z_i - Z_j)(\bar{Z}_k - \bar{Z}_m)$. The algebraic development of the next three sections will deduce aspects of the distributions of these observables from the assumptions of the null model. What we observe of the $W_i$ are statistics upon observables regarding the $Z_i$. The distances $|W_i - W_j|$ are taken as $\sqrt{\mathbf{E}(|Z_i - Z_j|^2)}$, root mean squares of the corresponding distances as observed in the sample, and the inner products $(W_j - W_i)(\bar{W}_k - \bar{W}_m)$ are the corresponding expected values $\mathbf{E}[(Z_j - Z_i)(\bar{Z}_k - \bar{Z}_m)]$.

*Prospectus.* This essay explores spaces of size and shape descriptors for comparisons of such configurations, relating them to the ordinary algebraic machinery of statistical hypothesis testing and linear modeling and also to the geometrical machinery of plane mappings. Relying on the algebra of the $W$'s and $dz$'s, Section 2 introduces a linear space of size variables spanned by the set of interlandmark distances. Section 3 introduces a different linear space of shape variables, a space that can be analyzed without any reference to size variables, and identifies a convenient basis in certain complex ratios embodying the shapes of single triangles.

These two spaces, defined on the same configurations of landmarks, must be closely linked. Sections 4 and 5 explore the ties between them, which permit the geometric interpretation of familiar multivariate maneuvers and the multivariate interpretation and testing of geometrical extrema. Theorem 1 demonstrates that, under the null model, one particular size variable is independent of all shape variables. This independence makes possible a single unambiguous $F$ test for size allometry, demonstrated in an application from micropaleontology. It is then shown how Hotelling's $T^2$ may be used to test for the presence of any shape difference between two populations of triangles. Any difference found will be measured afterward by the ratio of two distances along directions at 90° to one another, one distance bearing the greatest loading (ratio of change) over the change and the other one the least. Theorem 2 indicates how the search for distances bearing extrema of loading in configurations of more than three landmarks reduces to the consideration of triangles, for which the necessary extrema can be computed in closed form.

Section 6 is devoted to an extended example of all these procedures using a data set of four cephalometric landmarks observed in two groups at two ages. The example exploits the method of biorthogonal grids for geometric visualization of shape differences and shape factors that vary smoothly over a polygon of landmarks. Several shape differences are displayed, tested for significance, and reported by optimal computed geometric proportions. The example goes on to depict the nature of size allometry in this data set, and then, passing from consideration of difference to the consideration of invariance, uses the grid transformation diagram to interpret a factor for shape stability computed by ordinary canonical analysis of the shape space at two ages. Section 7 suggests some further methodological developments that would extend the efficient exploitation of geometric data throughout the biological and biomedical sciences.

## 2. A LINEAR SPACE FOR SIZE VARIATION

### Summary

The simplest size variable is the distance between a pair of landmarks. In addition to the landmarks present in the data, we consider "constructed landmarks," arbitrarily weighted centroids of sets of landmarks, and the set of all distances measured between pairs of these: the real numbers $|\sum c_i Z_i|$ with $\sum c_i = 0$. For

shapes constrained to a sufficiently modest range, this set of distances generates a linear vector space, here called the space of size variables. The set of differentials of all distances between landmarks in pairs forms a (redundant) basis for this space. Of special interest is the size variable $S$ which is the sum of all squared distances between the landmarks in pairs. $S$ is statistically equivalent to the sum of distances from each landmark to their joint centroid, each distance weighted by its own sample mean.

### Distances between Landmarks and the Space They Span

Suppose we have a data set of configurations of $K$ landmarks $Z_1, \cdots, Z_K$. The simplest sort of size variable is the distance $|Z_i - Z_j|$ between two of these landmarks, as measured in each landmark configuration of the sample. So that we may rely upon linear machinery for statistical purposes, we will make use of the quantity $|Z_i - Z_j|$ by way of its differential $d|Z_i - Z_j|$, the deviation from its mean $|W_i - W_j|$, under small displacements $dz_i$, $dz_j$ of the landmarks from their (unobservable) mean positions.

LEMMA 1. *If $f(Z_i, Z_j) = |Z_i - Z_j|$, a real-valued function of two complex arguments, then*

$$df = \mathrm{Re}\left[\frac{Z_i - Z_j}{|Z_i - Z_j|}(\overline{dz_i} - \overline{dz_j})\right].$$

PROOF. The change in the distance $|Z_i - Z_j|$ is the sum of the projections of $dz_i$ and $-dz_j$ separately upon the segment $(Z_i - Z_j)$. Rotate that segment to lie along the real axis—a multiplication by the unit complex number $(\overline{Z_i} - \overline{Z_j})/|Z_i - Z_j|$. The projection we seek is effected by taking the real part of the product by $(dz_i - dz_j)$:

$$df = \mathrm{Re}\left[\frac{\overline{Z_i} - \overline{Z_j}}{|Z_i - Z_j|}(dz_i - dz_j)\right].$$

The assertion of the Lemma follows from the invariance of the real part of a complex number under complex conjugation.

It follows that if $f(Z_i, Z_j) = |Z_i - Z_j|^2$, then

$$df = 2|Z_i - Z_j|\,d|Z_i - Z_j| = 2\,\mathrm{Re}[(Z_i - Z_j)(\overline{dz_i} - \overline{dz_j})].$$

To convert one of these differentials $d|Z_i - Z_j|$ or $d|Z_i - Z_j|^2$ into a variable suitable for linear statistical analysis, we evaluate it in the vicinity of the sample mean form. This involves replacing the data-dependent coefficient $Z_i - Z_j$ by its expected value $W_i - W_j$, the same function of the $W$'s. Once this is done, to variation in either the distance $|Z_i - Z_j|$ or its square there corresponds, up to a scale factor irrelevant for the linear statistical manipulations to follow, the linear form

$$\mathrm{Re}[(W_i - W_j)(\overline{dz_i} - \overline{dz_j})].$$

Consider now a real-weighted average of landmark locations: a new point $\sum_{i=1}^K a_i Z_i$ for which all $a_i$ are real and $\sum a_i = 1$. Because

$$\sum_{i=1}^K a_i(Z_i + Z) = \sum_{i=1}^K a_i Z_i + \sum_{i=1}^K a_i Z = \sum_{i=1}^K a_i Z_i + Z,$$

and

$$\sum_{i=1}^K a_i(\alpha Z_i) = \alpha \sum_{i=1}^K a_i Z_i,$$

the quantity $\sum a_i Z_i$ moves precisely in accordance with rigid motions of the landmarks used to compute it, so that it may fairly be called a *constructed landmark.* (For instance, the midpoint of the segment between two landmarks $Z_j$ and $Z_k$ is $\sum_{i=1}^K a_i Z_i$ with $a_j = a_k = \frac{1}{2}$ and all other $a_i$ equal to zero.) The distance between two constructed landmarks, itself a valid measure of size, is $|\sum a_i Z_i - \sum b_i Z_i| = |\sum c_i Z_i|$ with $\sum c_i = \sum a_i - \sum b_i = 0$.

The differential of this variable, its difference from its mean for small deviations $dz_i$ of the $Z_i$ from *their* centroids $W_i$, is

$$d|\sum c_i Z_i| \simeq \mathrm{Re}\left[\frac{\sum c_i W_i}{|\sum c_i W_i|}(\sum c_i\, \overline{dz_i})\right].$$

Is this a linear combination of the differentials $\mathrm{Re}[(W_i - W_j)(\overline{dz_i} - \overline{dz_j})]$ corresponding to the simple interlandmark distances $|Z_i - Z_j|$? In view of the constraint $\sum c_i = 0$, we may rewrite $\sum c_i Z_i$ as $\sum c_i(Z_i - Z_1)$. Then $d|\sum c_i Z_i|$ is proportional to

$$\mathrm{Re}\{[\sum c_i(W_i - W_1)][\sum c_i(\overline{dz_i} - \overline{dz_1})]\}.$$

This expression may be expanded into $(K-1)^2$ terms:

$$d|\sum c_i Z_i| \propto \sum_{i=2}^K \sum_{j=2}^K \mathrm{Re}[c_i c_j (W_i - W_1)(\overline{dz_j} - \overline{dz_1})],$$

where all the $c$'s are real. The terms for $i = j$ are elements of the basis already. All the remaining terms may be collected in pairs

$$\mathrm{Re}\{[c_i c_j[(W_i - W_1)(\overline{dz_j} - \overline{dz_1}) + (W_j - W_1)(\overline{dz_i} - \overline{dz_1})]\}.$$

But the real part of the term in inner brackets may be rewritten as the sum of *three* elements of our basis: $\mathrm{Re}[(W_i - W_1)(\overline{dz_i} - \overline{dz_1})]$, $\mathrm{Re}[(W_j - W_1)(\overline{dz_j} - \overline{dz_1})]$, and $\mathrm{Re}[(W_i - W_j)(dz_i - dz_j)]$. Hence the set of differentials $d|Z_i - Z_j|$ spans the set of all homologously defined size measures $d|\sum c_i Z_i|$ and $\sum c_i = 0$—all the homologous distance measurements that might be made upon our configuration of landmarks.

This, then, will be our space of size variables: the differentials of distances between landmarks, and all of their linear combinations, as shown in Figure 3a.

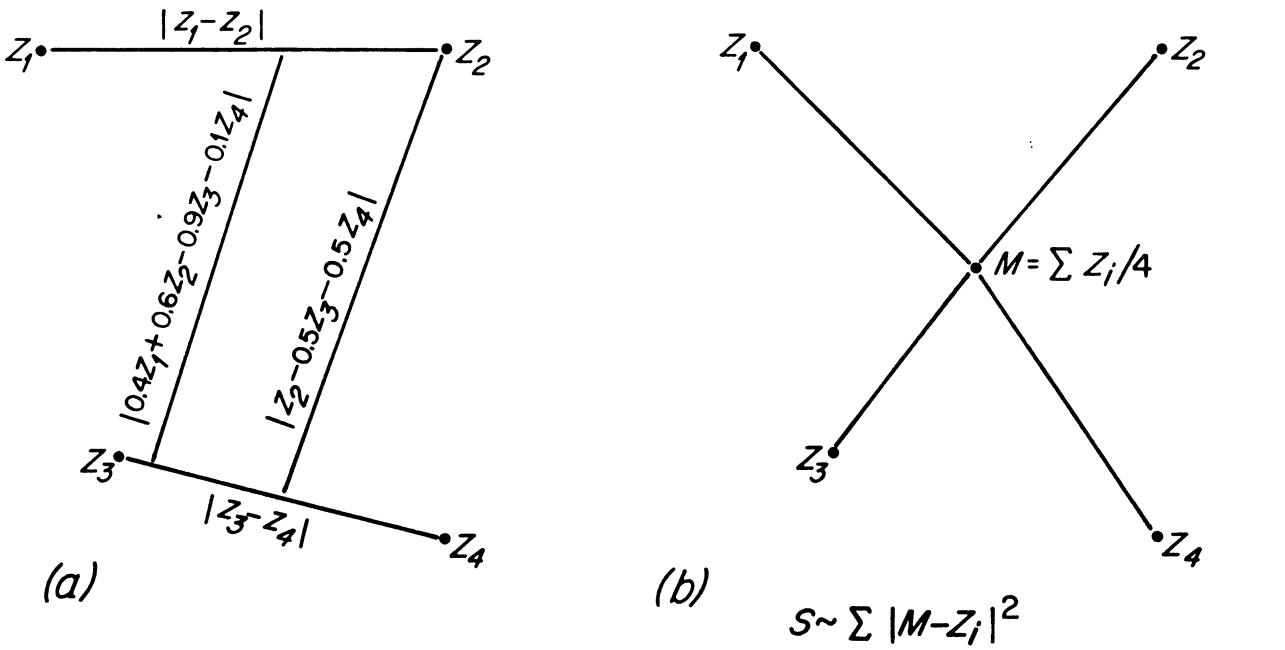


FIG. 3. *Size variables.* (a) *In this essay, size variables are distances between weighted averages of landmark locations. Their linearized space is spanned by the set of all distances between landmarks.* (b) *The special size variable S is the sum of all squared distances between pairs of landmarks. It is geometrically equivalent to the sum of squared distances of each landmark from their centroid.*

For configurations of more than three landmarks, this basis is redundant: it contains $K(K-1)/2$ distance measures, whereas the set of equivalence classes of landmarks $Z_1, \cdots, Z_K$ has only $2K-3$ geometric degrees of freedom.

Most familiar size measures may be located in this space. For instance, the area of a triangle of landmarks is spanned by this basis. By the formula of Heron, the area of a triangle with edge lengths $a, b, c$ is

$$A = \sqrt{s(s-a)(s-b)(s-c)},$$

where $s = (a+b+c)/2$. The differential $dA$ of area may thereby be expanded as a sum of products of $da, db, dc$ by real coefficients that are functions of $\mathbf{E}(a), \mathbf{E}(b), \mathbf{E}(c)$; but $da, db, dc$ are themselves basis elements of this size variable space.

Every size variable $|\sum c_iZ_i|$ is a multiple of an actual observable distance between two constructed landmarks. For by renumbering the landmarks we may reorder the sum into a run of the positive terms followed by a run of its negative terms

$$\sum c_iZ_i = \sum_{c_i>0} c_iZ_i - \sum_{c_i<0} -c_iZ_i.$$

The coefficients of the two partial sums must have the same total (as their difference is the sum of all the original $c$'s, namely, zero). Rescaling through by this constant, we may arrange that the coefficients of each sum separately total 1. Then each term is a constructed landmark, the weighted average location of the landmarks in its run.

This class of size variables, although algebraically very tractable, includes a subset which proves to be of no morphometric use. The two components $\sum_{c_i>0} c_iZ_i$, $\sum_{c_i<0} -c_iZ_i$ may have the same expected location in the complex plane specimen by specimen: for instance, they may be the midpoints of the two diagonals for four landmarks arranged approximately in a square. In this case, the squared distance between the constructed landmarks has a positive expected value $\sum c_i^2\sigma^2$, but the distance between the comparable expressions replacing the $Z$'s by their centroids is zero. So large a relative error is unacceptable. We are more than sufficiently protected against this possibility if we restrict our attention in practice (specifically, in Theorem 2) to those size measures $|\sum c_iZ_i|$ for which the landmarks involved in the two partial sums have nonoverlapping convex hulls. We call these size variables *admissible.* In the case of the square, this class includes (but is not limited to) distances between constructed landmarks defined on opposite edges, and excludes those which are distances between points on different internal diagonals.

Cartesian coordinates locating one landmark in a system registered upon some other landmarks are suited for computed averaged forms as mentioned in Section 1. But they are not members of the size variable space specified here. One Cartesian coordinate of $Z_3$, for instance, might be the distance of $Z_3$ from the line connecting $Z_1$ and $Z_2$ (or from the line through $Z_1$ perpendicular to the segment $Z_1Z_2$). Being the distance from a point to a line, the coordinate is

properly the distance from $Z_3$ to the foot of the perpendicular from $Z_3$ dropped onto the line $Z_1Z_2$. But the location of the foot of this perpendicular is not homologous from configuration to configuration in a sample—it will divide the segment between $Z_1$ and $Z_2$ in varying proportion. Thus the $x$ coordinate of $Z_3$ is not expressible as the distance between $Z_3$ and any constructed landmark $a_1Z_1 + a_2Z_2$. That is, it is not a member of the space of size variables.

### A Preferred Size Variable

One particular linear combination of these size measures will be the subject of a theorem in Section 4. Consider the function $S$ which is the sum of all $\binom{K}{2}$ *squared* interlandmark distances upon the configuration:

$$(1) \qquad S(Z_1, \cdots, Z_K) = \sum\sum_{i<j} |Z_i - Z_j|^2.$$

(The letter $S$ stands for "size.") For small variations of the $Z_i$ about their mean positions $W_i$ we have

$$dS = 2 \sum\sum_{i<j} |Z_i - Z_j| \, d|Z_i - Z_j|$$
$$\simeq \sum\sum_{i<j} |W_i - W_j| \, d|Z_i - Z_j|,$$

to first order. Thus $S$ behaves like a weighted sum of edge length measures, each edge weighted by its own population mean length.

We may expand the expression (1) for $S$ term by term according to the comment following the preceding Lemma:

$$dS \simeq 2 \operatorname{Re}\left[\sum\sum_{i<j} (W_i - W_j)(\overline{dz_i} - \overline{dz_j})\right].$$

Because the terms for $i = j$ are identically zero, twice the sum for $i < j$ is equivalent to a single repeated summation $\sum_{i=1}^{K} \sum_{j=1}^{K}$. Hence

$$dS \simeq \operatorname{Re}[\sum\sum (W_i - W_j)(\overline{dz_i} - \overline{dz_j})].$$

Collecting the $2K$ terms in each differential separately, this becomes

$$(2) \qquad dS \simeq 2 \operatorname{Re}\left[\sum_{i=1}^{K} \overline{dz_i}\left(KW_i - \sum_{m=1}^{K} W_m\right)\right]$$
$$= 2K \operatorname{Re}\left[\sum_{i=1}^{K} \overline{dz_i}\left(W_i - \frac{1}{K}\sum_{m=1}^{K} W_m\right)\right].$$

To the right-hand side of this equation we may add the term

$$\operatorname{Re}\left[\sum_{m=1}^{K} \overline{dz_m}\left(\sum_{m=1}^{K} W_m - \frac{K}{K}\sum_{m=1}^{K} W_m\right)\right],$$

which is identically zero. There results, finally, the expression

$$dS \simeq 2K \operatorname{Re}\left[\sum_{i=1}^{K}\left(W_i - \frac{1}{K}\sum_{m=1}^{K} W_m\right)\left(\overline{dz_i} - \frac{1}{K}\sum_{m=1}^{K} \overline{dz_m}\right)\right]$$
$$= K\, d\left[\sum_{i=1}^{K}\left|Z_i - \frac{1}{K}\sum_{m=1}^{K} Z_m\right|^2\right].$$

Hence $S$ is statistically equivalent to a simpler size variable, the sum of squared distances of each landmark from their centroid case by case, as shown in Figure 3b.

For triangles of landmarks, the net size measures one most commonly finds in the biometric literature are perimeter and area. The variable $S$ is statistically equivalent to either of these only when the mean triangular shape is equilateral; in that case, perimeter, area, and $S$ are multiples of each other up to terms of first order in the $dz$'s.

## 3. A LINEAR SPACE FOR SHAPE VARIATION

### Summary

The shape of a single triangle of landmarks may be reduced to a single pair of shape coordinates, the complex number $Q = (C - A)/(B - A)$ locating any vertex $C$ in the coordinate system with landmark $A$ at $(0,0)$ and landmark $B$ at $(1,0)$. Any conventional shape variable is linearly equivalent to the projected coordinate of this point in an appropriate direction. Permutation of the landmarks $A, B, C$ in this construction results, to first order, in mere rotation and rescaling of empirical shape scatters. If $A$, $B$, and $C$ are multivariate normal about their centroids, so is $Q$. The set of all such shape coordinate pairs $Q$ for various triples of landmarks spans the (linearized) space of all ratios of size variables as defined in the preceding section. This will be called the space of shape variables.

### A Geometric Construction

Generally one arrives at shape data by scaling observed forms inversely to some distance, or other net size measure, explicitly measured upon the forms (Mosimann, 1970). The morphometric analysis of shape variables generated in this way is sound only insofar as it is ultimately independent of this arbitrary decision taken at the outset. In studies based on landmark locations, surely the most convenient size measures are the distances between landmarks. Let us temporarily select one pair of landmarks—say, $A$ and $B$—as a standard of length, and consider the shape that they make up in combination with a third landmark $C$: that is, the shape of the triangle $\Delta ABC$. We will scale the triangle so that the length of the edge $AB$ is constant at 1 unit.

After scaling so, we may "register on edge $AB$," placing landmark $A$ at the Cartesian point $(0,0)$ and landmark $B$ at the point $(1,0)$ as in Figure 4. If landmark $A$ were originally at $(x_A, y_A)$, $B$ at $(x_B, y_B)$, and $C$ at $(x_C, y_C)$, then this registration assigns landmark $C$ the Cartesian coordinates $(\eta_1, \eta_2)$ where

$$\eta_1 = \frac{(x_B - x_A)(x_C - x_A) + (y_B - y_A)(y_C - y_A)}{(x_B - x_A)^2 + (y_B - y_A)^2},$$

$$\eta_2 = \frac{(x_B - x_A)(y_C - y_A) - (y_B - y_A)(x_C - x_A)}{(x_B - x_A)^2 + (y_B - y_A)^2}.$$

All the shape information there is to be had about $\Delta ABC$ must be encoded in these two coordinates of the normalized point $C$.

Consider any familiar sort of shape variable that can be computed from a triangle of landmarks. Figure 5a, for instance, illustrates the angle $\angle ACB$. Considering $A$ and $B$ now to be fixed in position at $(0,0)$ and $(1,0)$, and shape variable is constant on some curve through $C$; the variable $\angle ACB$ happens to be constant as $C$ varies along circles through $A$ and $B$. Neighboring, equally spaced curves, in this case adjacent circles through $A$ and $B$, correspond to neighboring, nearly equally spaced values of the shape measure.

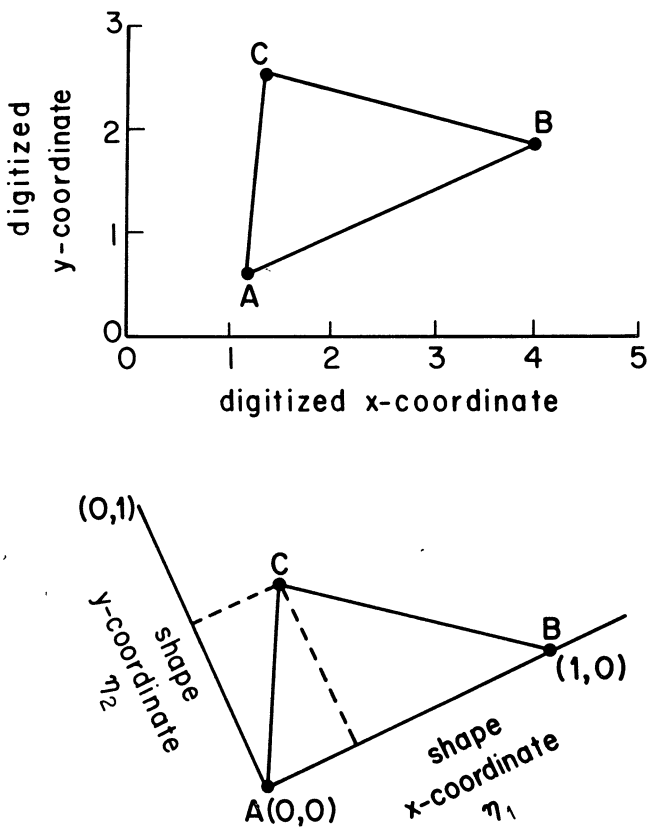


FIG. 4. *One set of shape coordinates of triangle $\Delta ABC$: the coordinates of point $C$ in a Cartesian system with $A$ at $(0,0)$ and $B$ at $(1,0)$.*

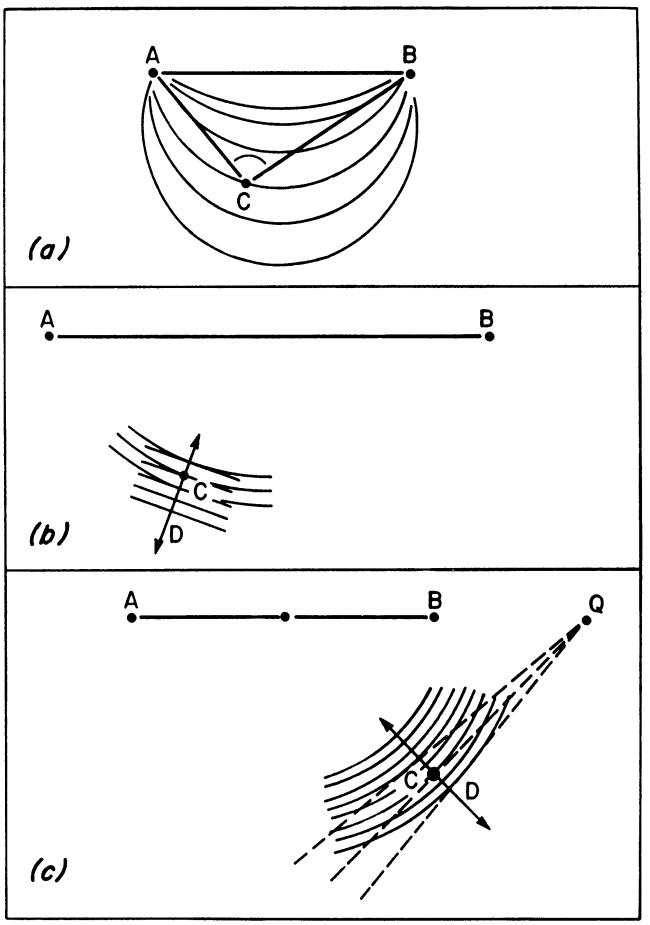


FIG. 5. *The geometry of shape variables in small regions of the shape coordinate plot.* (a) *Example: isopleths of the shape variable $\angle ACB$ are circles through $A$ and $B$.* (b) *Approximating a shape variable by a Cartesian coordinate. The variable is represented by $D$, the direction of its gradient.* (c) *All shape variables with the same gradient near a mean form are statistically equivalent. Dashed lines, isopleths of the angle $\angle AQC$; solid curves, isopleths of distance to the midpoint of $AB$.*

In a small region of this plot, such as corresponds to observed cranial shape variation of human beings of school age and beyond, this set of curves can be approximated by a family of parallel, equally spaced straight lines (Figure 5b). The shape variable serves as a coordinate indexing the lines of this family and varying along the gradient $D$ perpendicular to them. In a population of shapes resembling $\Delta ABC$, up to this first order approximation any shape variable may be identified with a bundle of Cartesian coordinate lines perpendicular to some gradient direction through $C$. For linear statistical purposes, their spacing is immaterial.

Then there is one family of linearly equivalent shape variables for every direction through the point $C$. For instance, the direction $D$ in Figure 5c is the gradient of the angle $\angle AQC$, where $Q$ is the point at 1.5 on the $x$ axis (in the coordinate system with $A$ at 0 and $B$ at

1). This is a perfectly reasonable shape measure, although unfamiliar. Another measure with the same gradient is the ratio to $AB$ of the distance from $C$ to the midpoint of $AB$. These two shape measures are linearly equivalent at $C$, as are all others bearing the same gradient there.

Since there is a semicircle's worth of directions around any point $C$, there is only a semicircle's worth of linearly different shape variables in the vicinity of a typical shape $\Delta ABC$. For any two distinct shape gradients $D_1$, $D_2$ at $C$, every other shape variable $D$ is linearly equivalent to (i.e., has the same gradient near $C$ as) some linear combination $a_1D_1 + a_2D_2$ of $D_1$ and $D_2$, or any multiple of this linear combination. Any shape variable would be perfectly predicted as a linear combination of any other two shape variables of distinct gradient were it not for the nonlinearities ignored in this shape space formalism.

### Analytical Representation of the Shape of a Triangle

For a single triangle of landmarks, the construction just introduced has a simple expression in terms of the complex variables $Z_i$, $Z_j$, $Z_k$: the point $(\eta_1, \eta_2)$ is just the *complex affine ratio*

$$Q(Z_i, Z_j, Z_k) = \frac{Z_k - Z_i}{Z_j - Z_i},$$

as can be verified directly from the formulas for $(\eta_1, \eta_2)$ preceding. The ratio $Q$ is real if $Z_i$, $Z_j$, $Z_k$ are in a straight line, and it lies between 0 and 1 if $Z_k$ lies between $Z_i$ and $Z_j$ on that line. If the angle $\angle Z_kZ_iZ_j$ is $\pm 90°$, $Q$ is purely imaginary. The effect of the assumption $|Z_i - W_i| \ll |W_j - W_i|$, part of the null model, is to ensure that the scatter of these points $Q$ is bounded away from the origin.

This shape function is most useful in differential form. We have

$$\begin{aligned} dQ &= d\left[\frac{Z_k - Z_i}{Z_j - Z_i}\right] \\ &= \frac{(Z_j - Z_i)(dz_k - dz_i) - (Z_k - Z_i)(dz_j - dz_i)}{(Z_j - Z_i)^2} \\ &= -\frac{(Z_j - Z_k)\,dz_i + (Z_k - Z_i)\,dz_j + (Z_i - Z_j)\,dz_k}{(Z_j - Z_i)^2} \\ &\simeq -\frac{(W_j - W_k)\,dz_i + (W_k - W_i)\,dz_j + (W_i - W_j)\,dz_k}{(W_j - W_i)^2} \end{aligned} \tag{3}$$

to first order terms.

There is considerable symmetry to this form, owing to the cycling of subscripts in the numerator. When the arguments $Z_i$, $Z_j$, $Z_k$ are permuted, all that changes of $dQ$ is its denominator and (for odd permutations) its sign. Under change of baseline for the $(\eta_1, \eta_2)$ construction, the space of shape variables merely counter-rotates twice as fast as the baseline, possibly with an additional half-turn, and changes its scale. This invariance is demonstrated in Figure 6 for some data to be analyzed later. The invariance remains when the baseline involves a point dividing one side in fixed fraction—for instance, a baseline chosen from $Z_i$ to $(Z_j + Z_k)/2$.

This Euclidean invariance of $dQ$ space is very important for the multivariate statistical analysis to follow. Whenever a statistical method is invariant under Euclidean transformations of variable space, its application to distributions of these $dQ$ will result in nearly the same findings regardless of the pair of landmarks $Z_i$ and $Z_j$ chosen for the baseline, the denominator in the definition of $Q$. Standard multivariate statistics like $F$ ratios and $T^2$ are invariant under rotation and rescaling in the appropriate way. Thus, findings of group difference or growth allometry based on any particular form for $dQ$ will apply to the whole triangle of landmarks. In this way we execute a multivariate analysis for which the choice of baseline, although required for computations to proceed, is irrelevant to the findings, just as multivariate analysis of size measures is the same whether one measures in centimers or millimeters.

Note from equation (3) that whenever the individual displacements $dz$ are independent circular normals, $dQ$ is likewise circular normal. If the variance of each component of each $dz$ is $\sigma^2$, the variance of either component of $dQ$ is $\sigma^2 p^2/|W_i - W_j|^4$, where $p^2$ is the sum of squares of the sides of the mean triangle $\Delta W_iW_jW_k$. This is as closely as one can approach to a direct visual verification of the null distribution assumptions. We cannot observe any single $dz_i$, and the observable combinations $d|Z_i - Z_j|$ of two $dz$'s are only scalars; but the combination of three $dz$'s into $dQ$ gives rise to an observable scatter in the complex plane, as exemplified in Figure 6. If these scatters are not approximately circular (cf. Figure 7b), the null model is untenable. More detailed verifications of the model will not be considered here.

We are free to use as an ordinary shape variable the projection of $dQ$ on any direction in its plane: that is, any linear combination of the real and imaginary parts of $dQ$. One useful projection is along the direction through $(W_k - W_i)/(W_j - W_i)$ from $(0,0)$. Because $(0,0)$ is the image of $W_i$ under the mapping represented by this formula for $Q$, the projection in question is along the gradient of the length $|Z_k - Z_i|$ in the registration holding $|Z_j - Z_i|$ constant—that is, it is along the gradient of $|Z_k - Z_i|/|Z_j - Z_i|$, ratio of two sides of the triangle $\Delta Z_iZ_jZ_k$. Similarly, the direction from $(1,0)$ through $(W_k - W_i)/(W_j - W_i)$ is along the gradient of $|Z_k - Z_j|/|Z_i - Z_j|$. Ratios with other denominators may be obtained by rotating the construction to a different baseline, which, as we noted

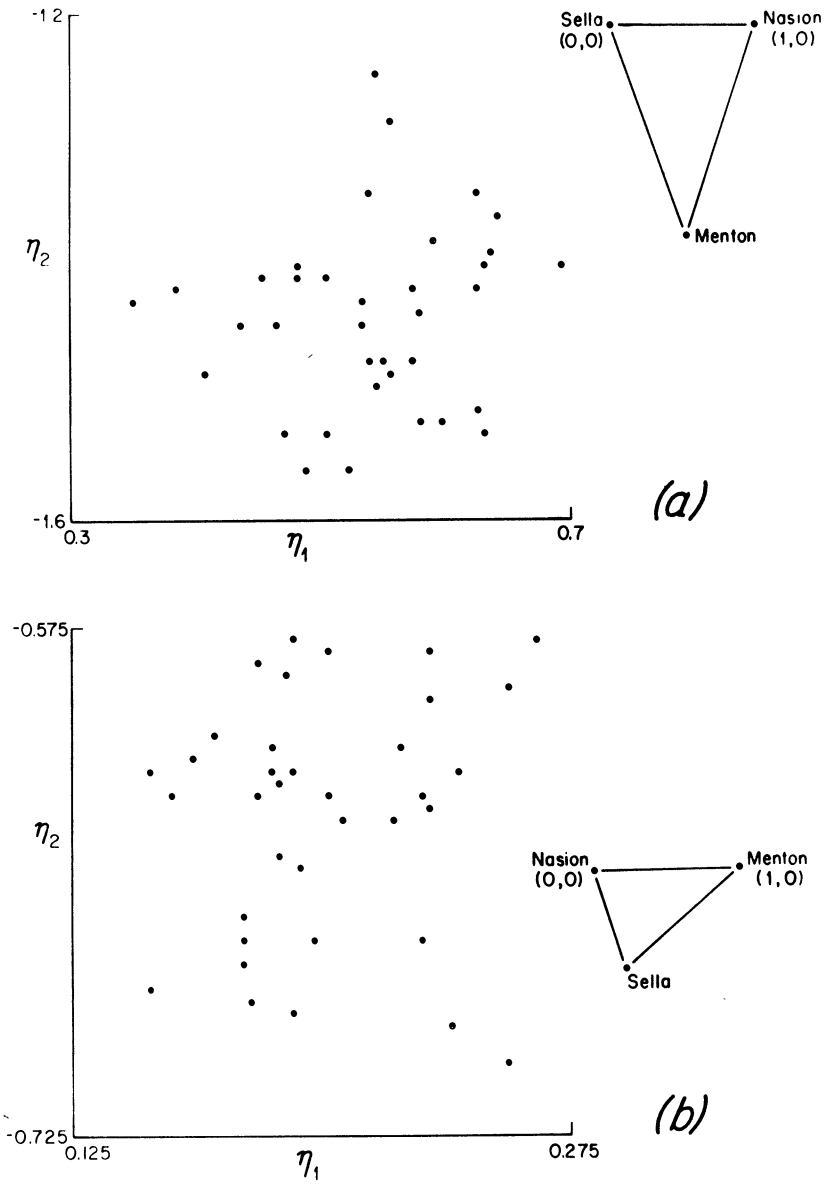


FIG. 6. *Under permutation of subscripts, a population of affine ratios* $(Z_k - Z_i)/(Z_j - Z_i)$ *undergoes very nearly a similarity transformation.* (a) *A population of* 36 *ratios* $(Z_3 - Z_1)/(Z_2 - Z_1)$ *for some data related to the example in Section* 6 ($Z_1$ = *Sella*, $Z_2$ = *Nasion*, $Z_3$ = *Menton*). (b) *The ratios* $(Z_1 - Z_2)/(Z_3 - Z_2)$ *for the same data set. Note the changes in origin and scales of the axes as well as in orientation of the scatter.*

above, merely counter-rotates and scales $dQ$. Hence, for a single triangle of three landmarks, the space of shape variables spanned by $d[(Z_k - Z_i)/(Z_j - Z_i)]$ includes the differentials of ratios of all pairs of edge lengths from the size basis for the same triangle.

In fact, because the span of $d[(Z_k - Z_i)/(Z_j - Z_i)]$ includes all the shape information about this triple of landmarks, it must include the differentials of *all* ratios

$$\frac{|a_iZ_i + a_jZ_j + a_kZ_k|}{|b_iZ_i + b_jZ_j + b_kZ_k|}$$

with $a_i + a_j + a_k = b_i + b_j + b_k = 0$, that is, all ratios of pairs of size variables whose differentials are spanned by the $d|Z_i - Z_j|$.

## The Complex Affine Ratios Span the Ratios of Size Variables

The preceding is true for more general sets of landmarks than triangles. The linearized space of all ratios of size variables

$$d\left[\frac{|\sum a_iZ_i|}{|\sum b_iZ_i|}\right], \sum a_i = \sum b_i = 0,$$

is spanned by the sets of shape coordinates $dQ(Z_i, Z_j, Z_k)$. Because for scatters bounded away from the origin

the absolute value operator is linearly equivalent to projection along a line from the origin, it is sufficient to show that the $dQ$ span the space of complex shape ratios $d[\sum a_i Z_i / \sum b_i Z_i]$ with $\sum a_i = \sum b_i = 0$.

Rewrite this latter expression as

$$d\left[\frac{\sum a_i(Z_i - Z_1)}{\sum b_i(Z_i - Z_1)}\right].$$

Ignoring a factor $(\sum b_i(W_i - W_1))^2$, the expected value of the denominator in the usual formula, the differential then becomes

$$\begin{aligned}
&[\textstyle\sum b_i(Z_i - Z_1)]\, d[\sum a_i(Z_i - Z_1)] \\
&\qquad - [\textstyle\sum a_i(Z_i - Z_1)]\, d[\sum b_i(Z_i - Z_1)] \\
&= \sum_{i=1}^{K}\sum_{j=1}^{K} (a_i b_j - a_j b_i)[(Z_i - Z_1)\, d(Z_j - Z_1) \\
&\qquad - (Z_j - Z_1)\, d(Z_i - Z_1)] \\
&= \sum_{i=1}^{K}\sum_{j=1}^{K} (a_i b_j - a_j b_i)(Z_i - Z_1)^2\, d\left[\frac{Z_j - Z_1}{Z_i - Z_1}\right] \\
&\simeq \sum_{i=1}^{K}\sum_{j=1}^{K} (a_i b_j - a_j b_i)(W_i - W_1)^2\, d\left[\frac{Z_j - Z_1}{Z_i - Z_1}\right].
\end{aligned}$$

Thus, just as the distance differentials $d|Z_i - Z_j|$ span the space of all size variables, so do the affine differentials $d[(Z_k - Z_i)/(Z_j - Z_i)]$ span the space of differentials of all ratios of size variables. For instance, the four-point ratio $(Z_k - Z_m)/(Z_j - Z_i)$, combination of the angle between two segments and the ratio of their lengths, is equal to $Q(Z_i, Z_j, Z_k) - Q(Z_i, Z_j, Z_m)$, so that its statistics may be inferred from the variances and covariances of the $Q$'s. This space of size ratios will henceforth be taken as the space in which we look for shape differences and trends; it is our space of shape variation. Like the basis for size space, for configurations of more than three landmarks this basis for shape space is redundant. The total number of geometric degrees of freedom for shape, $2K - 4$, is spanned by many sets of $K - 2$ affine ratios, such as $dQ(Z_1, Z_2, Z_k)$, $k = 3, \cdots, K$.

I noted in the previous section that Cartesian coordinates of landmarks, however registered, are not suited to be size variables; but the components of these $dQ$ are Cartesian coordinates used as shape variables. The coordinates of $Q(Z_i, Z_j, Z_k)$ describe the shape of the triangle $\Delta Z_i Z_j Z_k$, a shape determined by the distances $|Z_i - Z_j|$, etc., and thus a function of the ratios of homologously defined size measures. If we had allowed such freedom in the membership of the space of size variables, it would have included coordinates that pass through the value zero, greatly complicating the relation of size space to shape space. But the components of $Q$ lack precisely the scale information necessary to draw out a mean configuration of $W$'s. We need some single size variable to supply a single criterion of scale; otherwise, the morphometrics of landmarks can proceed wholly in shape space.

Other coordinate invariant descriptions and moments of these landmark configurations are spanned by the edge lengths and affine ratios together. For instance, the quantities $(Z_i - Z_j)$ or $(Z_k - Z_m)$ have no meaning in this biometric system, as they vary with the orientation of the configuration; likewise the product $(Z_i - Z_j)(Z_k - Z_m)$; but the product $(Z_i - Z_j)(\bar{Z}_k - \bar{Z}_m)$, inner product of the two segments $(Z_i - Z_j)$ and $(Z_k - Z_m)$ treated as Cartesian vectors, does *not* change when data are shifted or rotated, and so is expressible in terms of the two morphometric spaces:

$$\begin{aligned}
&(Z_i - Z_j)(\bar{Z}_k - \bar{Z}_m) \\
&\quad = \frac{Z_i - Z_j}{Z_k - Z_m}(Z_k - Z_m)(\bar{Z}_k - \bar{Z}_m) \\
&\quad = [Q(Z_m, Z_k, Z_i) - Q(Z_m, Z_k, Z_j)]\,|Z_k - Z_m|^2.
\end{aligned}$$

Then its differential is a linear combination of the differentials of the $Q$'s and of $|Z_k - Z_m|$ having coefficients that are functions of the $W$'s.

A few commonly used shape variables do not suit the linear methods presented here. For example, considered as a function of one movable landmark with respect to a baseline from (0,0) to (1,0), the ratio of the area of a triangle to its squared perimeter has an extremum, value $\sqrt{3}/36$, at the points $(1, \pm\sqrt{3})/2$ corresponding to the shape of an equilateral triangle. The gradient of the ratio there is zero, which is not a direction in our linear shape space. All the instances of such functions that I have seen applied treat the three landmarks symmetrically, a feature undesirable in shape analysis. But size variables may treat landmarks symmetrically, as do perimeter, area, and the special variable $S$.

## 4. THE RELATION BETWEEN SIZE AND SHAPE

### Summary

On the null model of identical circular normal perturbations at each landmark independently, the special size variable $S$ introduced in Section 2 has covariance zero with every shape variable. The existence of size allometry, associations between shape and size (or shape change and size change), may thus be tested by a single ordinary $F$ ratio, that for the multiple regression of $S$ on any basis for shape space.

### The Size Variable Independent of Shape Space

Section 1 of this essay cited Mosimann's (1970) demonstration that, under characterizations of "size" and "shape" somewhat more general than those used

here, there can exist at most one size variable stochastically independent from the space of shape variables. The usual context in which this theorem is applied (see, for instance, Mosimann and James, 1979) is the linear statistical analysis of real-valued random variables $V_1, \cdots, V_J$, perhaps the logarithms of a motley of measured distances having no particular geometrical structure. The size variables are the combinations $\sum_{i=1}^{J} a_i V_i$ with $\sum a_i = 1$, and the shape variables are the combinations $\sum a_i V_i$ with $\sum a_i = 0$. In this context the unique shape-free size variable to which we are restricted by Mosimann's theorem always exists; it is proportional to the combination $\sum c_i V_i$ for which each $c_i$ is the sum of the entries in the $i^{\text{th}}$ row of the inverse of the covariance matrix among the $V$'s. (A geometric exegesis of this result may be found in Bookstein et al., 1985, Section 2.2.2.1.)

The coefficients $c_i$ depend explicitly upon the covariances of the $V$'s. If they are to be observed from data, of course, a selection of particular size variables $V_i$ must be specified. In the present context, evidence from data may be superseded by the algebra of the null model, which provides a covariance for every pair of size measures. However, it is never necessary to compute these covariances, owing to the following very convenient result.

THEOREM 1. *On the null model, the sum of squares of the distances between landmarks in pairs, which is the special size variable $S$ of equation* (1), *has zero covariance with all shape variables that may be defined on the set of landmarks $Z_i$.*

PROOF. Because of space of shape variables is spanned by the linear projections of the $dQ(Z_i, Z_j, Z_k)$ in various directions, it is sufficient to show that on the null model the covariance of $S$ with any one of the $dQ$ is zero.

Split all the complex variables into their real and imaginary parts: write $W_i = X_i + Y_i\sqrt{-1}$, $dz_i = dx_i + dy_i\sqrt{-1}$, $(W_j - W_i)^2\, dQ = dU + dV\sqrt{-1}$. In terms of these components, we have, from equation (2),

$$(4)\quad dS \simeq \sum dx_i \left[ KX_i - \sum_m X_m \right] + \sum dy_i \left[ KY_i - \sum_m Y_m \right].$$

Also, from equation (3), the components of the numerator of $dQ(Z_i, Z_j, Z_k)$ are

$$\begin{aligned} dU \simeq\ & (X_i - X_j)\, dx_k + (X_j - X_k)\, dx_i \\ & + (X_k - X_i)\, dx_j \\ & - (Y_i - Y_j)\, dy_k - (Y_j - Y_k)\, dy_i \\ & - (Y_k - Y_i)\, dy_j, \end{aligned} \tag{5a}$$

and

$$\begin{aligned} dV \simeq\ & (X_i - X_j)\, dy_k + (X_j - X_k)\, dy_i \\ & + (X_k - X_i)\, dy_j \\ & + (Y_i - Y_j)\, dx_k + (Y_j - Y_k)\, dx_i \\ & + (Y_k - Y_i)\, dx_j. \end{aligned} \tag{5b}$$

On the null model, the $dx$'s and $dy$'s are all independent and identically distributed with variance $\sigma^2$. Then the covariances of $dU$ and $dV$ with $dS$ are proportional to the sum of products of corresponding coefficients between forms (4) and (5):

$$\begin{aligned} \sigma^{-2}\mathbf{E}(dU\, dS) \simeq\ & (X_i - X_j)[(KX_k - \textstyle\sum X_m] \\ & + (\text{cycled subscripts}) \\ & - (Y_i - Y_j)[KY_k - \textstyle\sum Y_m] \\ & - (\text{cycled subscripts}), \\ \sigma^{-2}\mathbf{E}(dV\, dS) \simeq\ & (Y_i - Y_j)[KX_k - \textstyle\sum X_m] \\ & + (\text{cycled subscripts}) \\ & + (X_i - X_j)[KY_k - \textstyle\sum Y_m] \\ & + (\text{cycled subscripts}), \end{aligned}$$

in which the cycling of subscripts replaces $(ijk)$ by $(jki)$ and then by $(kij)$. In all these expressions, the terms in $\sum_m X_m$ or $\sum_m Y_m$ can be deleted, as the coefficients of either—for instance $X_i - X_j + X_j - X_k + X_k - X_i$—sum to zero. There remain, up to a factor of $\sigma^2 K$, the expressions

$$\begin{aligned} & (X_i - X_j)X_k + (X_j - X_k)X_i + (X_k - X_i)X_j \\ & \quad - (Y_i - Y_j)Y_k - (Y_j - Y_k)Y_i - (Y_k - Y_i)Y_j \end{aligned}$$

and

$$\begin{aligned} & (X_i - X_j)Y_k + (X_j - X_k)Y_i + (X_k - X_i)Y_j \\ & \quad + (Y_i - Y_j)X_k + (Y_j - Y_k)X_i + (Y_k - Y_i)X_j. \end{aligned}$$

But each of these expressions reduces to zero by simple cancellation. Hence for all shape variables $dQ$ the covariance with size $dS$ is zero for the null model, as was to be proved.

The demonstration of this proposition relies upon the linearization of analytic relationships between $dQ$ and $dS$. Such a manipulation is less risky over substantial size ranges if $S$ is replaced by its square root, a quantity having units of linear distance like all the other members of size variable space. Because $d\sqrt{S} = dS/2\sqrt{S}$, the demonstration of Theorem 1 is unaltered.

### Testing Size Allometry

By virtue of this theorem we arrive at a rigorous statistical test for the existence of size allometry, the (linear) dependence of shape upon size, in samples

which are growing or otherwise vary in size. The usual discussions of size allometry, as criticized, for instance, in Mosimann and James (1979), are fundamentally ambiguous inasmuch as the covariance between "size" and "shape" depends on the particular pair of size and shape variables chosen. This ambiguity is especially problematic when the measures are confounded by construction—for instance, when the shape variable is a ratio with size in its denominator, or is a regression residual from which size has been "removed." Efforts to clarify this situation by refinement of shape variables have generally been unavailing (see Bookstein et al., 1985, Section 2.2.2.1).

But for data which are landmark coordinates, Theorem 1 permits us to escape all these complications. On a reasonable null hypothesis, the particular size variable $S$ is uncorrelated with the entire space of shape variables howsoever (homologously) constructed. Such a correlation is tested by the ordinary $F$ ratio for the multiple regression of $S$ (or, especially for large size ranges, $\sqrt{S}$) on any basis for shape space. If that $F$ ratio is significant, we may proceed with the estimation of loadings upon size for each dimension of shape space. These are computed as the simple regression coefficients for the shape coordinates separately upon $S$. The report of allometry thereby reverts to its traditional form, a regression of shape upon size.

In real data, even in the absence of meaningful allometry no change of form is ever exactly a change of scale. An observed change or difference in $S$ is only an estimate of scale change, with noise attributable to inconsistencies, haphazard or systematic, among the ratios by which distances between diverse pairs of landmarks actually change. The use of the term "loading" in the context of allometry reminds us that we are regressing lengths or length ratios upon a net size measure to which they have all contributed, and also that the necessary regression coefficients are simple, not multiple. We have no way of interpreting the covariances among shape coordinates except by virtue of their joint dependence upon factors of deformation.

As a basis for purposes of the regression of size on shape, I generally use the set of shape coordinates $Q(Z_1, Z_2, Z_k)$, $k = 3 \cdots K$, specifying the relation of all other landmarks to one particular chord $Z_1Z_2$. A suitable candidate for this "chord of reference" is a diameter of the form passing near the centroid of the set of all the landmarks. Whenever allometry is found, to draw out its explicit picture we choose a small change of size $\Delta S$ and apply it to the mean form as follows. Holding $W_1$ and $W_2$ fixed at $(0,0)$ and $(1,0)$, we displace each centroid $W_k$, $k \geq 3$, according to the predicted change $dQ(Z_1, Z_2, Z_k)$ corresponding to the change $\Delta S$ in size. Then the figure of $W_1$, $W_2$, and the displaced centroids $W_3, \cdots, W_K$ is rescaled as a whole by setting the distance between $W_1$ and $W_2$ to the predicted value of $|Z_1 - Z_2|$ corresponding to the same $\Delta S$.

Even when no statistically significant allometry is to be found, $S$ is still the best size variable for testing the statistical significance of differences and trends in size. Under the null model, $S$ is the size variable least affected by haphazard shape variation, and hence most sensitive to any real differences of scale that may exist.

## Example: Size Allometry for a Microscopic Fossil

Lohmann (1983) published outlines of *Globorotalia truncatulinoides*, a microscopic planktonic created retrieved from ooze at the ocean floor. Each of the 21 outlines itself represents a sample average based on up to 50 separate organisms; the 21 averages correspond to 21 distinct patches of ocean bottom at various latitudes in the South Indian Ocean. (The averages were computed by an interesting automated method making no explicit reference to landmark information: see Bookstein et al., 1986). The form of this organism (Figure 7a) is similar to a snail's shell. From the averaged outlines supplied, we extracted three landmarks representing the endpoint of the axis of coiling and the limits of the aperture. Figure 7b shows the scatter of shape coordinates $(\eta_1, \eta_2)$ for the top of the aperture with reference to a baseline taken from the bottom of the aperture to the axis point. At each $(\eta_1, \eta_2)$ plotted is the value of $\sqrt{S}$, the mean size of that core sample, in units of microns.

The regression of $\sqrt{S}$ on the pair of shape coordinates is highly significant. We have

$$\underset{(59)}{46}\ \eta_1 + \underset{(19)}{103}\ \eta_2, \quad R^2 = .68,$$

$$\mathbf{E}(\sqrt{S}) = 32 -$$

where the quantities in parentheses are standard errors of the coefficients above them. Hence there exists strong size allometry in this data set. The correlation between $\eta_1$ and $S$, like the coefficient for $\eta_1$ in the multiple regression, is not significant by virtue of the absence of much variance upon $\eta_1$ in this scatter. Allometry is expressed mainly in the value of the shape coordinate $\eta_2$. This coordinate is the ratio to baseline length of the height of the aperture opening perpendicular to the baseline. It is an *aspect ratio* of the form as shown in Figure 7c. The size of these creatures is strongly determined by their environment, specifically, by water temperature, and so the shape gradient we just displayed is a good index of latitude—in fact, it is a better such index than size itself. For more about this example, see Bookstein et al. (1986).

Certain departures from the assumptions of the null model will be labeled as allometry by this test procedure. For instance, if in a configuration of three landmarks two were rigidly constrained to a constant

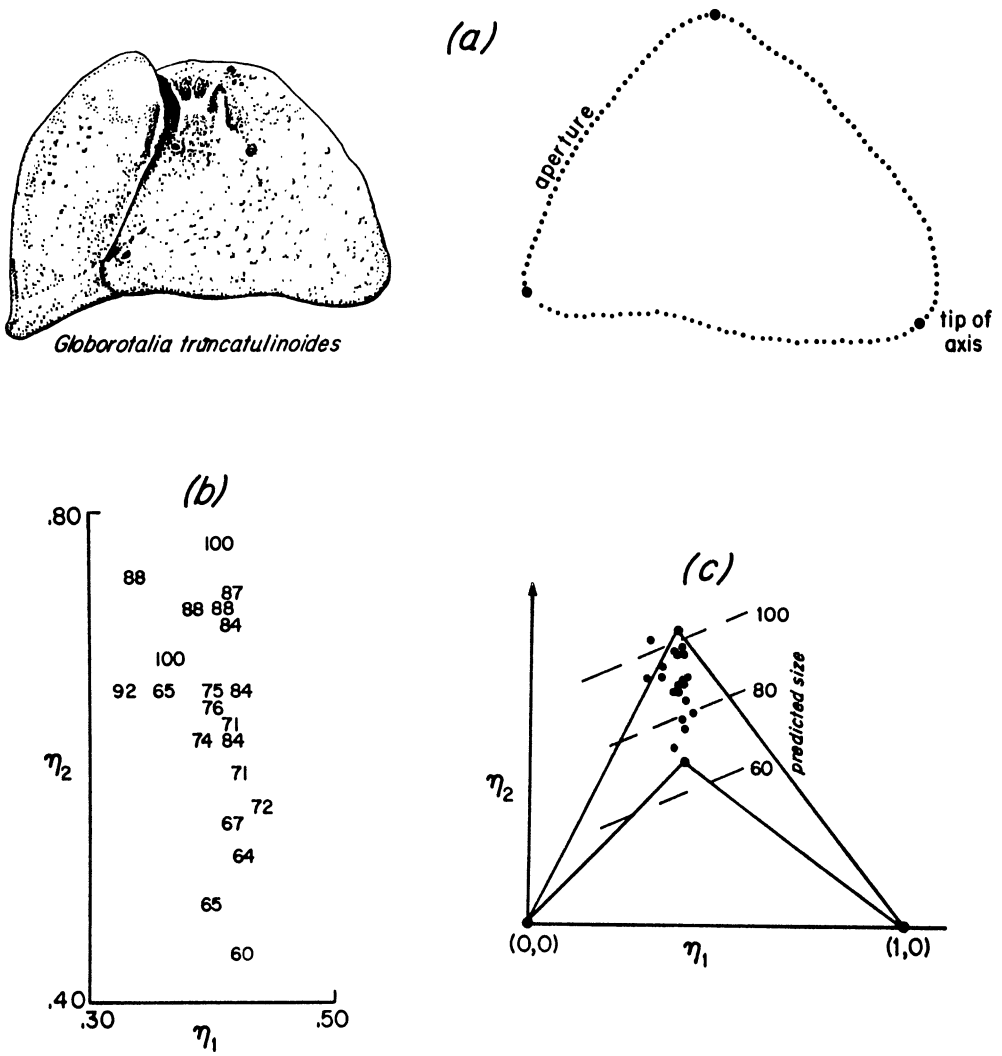


FIG. 7. *Analysis of allometry for a microfossil.* (a) *Drawing of a specimen of Globorotalia truncatulinoides, with one of the* 21 *sample averages and its landmarks.* (b) *Scatter of coordinates in shape space for* 21 *shapes representing averages of separate ocean floor samples. Each coordinate pair is coded by the mean size* $\sqrt{S}$ *of its sample, in microns.* (c) *The regression of shape on size indicates that allometry is expressed in the coordinate* $\eta_2$, *the ratio of height above this baseline to baseline length.*

separation, allometry would be found to inhere in the coordinate $\eta_2$ of the third landmark, just as was found here. But the presence of a clear axis in the $(\eta_1, \eta_2)$ scatter aligned with the computed gradient of allometry would not be expected under such an alternative.

## 5. SIZE AND SHAPE VARIABLES FOR DESCRIBING A GROUP DIFFERENCE

### Summary

For a single triangle of landmarks, the existence of any mean difference in shape between groups, mean change of shape in a single group, or mean difference of shape changes between groups may be tested by Hotelling's $T^2$. Any such difference that is statistically significant may be interpreted using a variable whose gradient in shape space is precisely along the group difference observed. That shape variable may be construed as the ratio of a pair of size variables, each a transect $| Z_i - (aZ_j + (1 - a)Z_k) |$ of the triangle, at an angle averaging 90° in the samples of forms. One size variable will bear the greatest dimensionless loading on the change in question—the greatest mean rate or ratio of change between the forms; the other will bear the least. In this way we may construe any mean shape difference as a deformation between two mean shapes and vice versa. The practical significance of the shape difference is embodied in the numerical difference between these two extreme ratios of change and in their directions upon the typical form. Analysis of differences between configurations of more than three landmarks reduces to consideration of size variables involving at most three landmarks.

## Group Differences in the Shape of a Triangle

Consider the problem of identifying a mean shape difference between two groups of triangles. Because the shapes of the triangles are embodied in their affine ratios $Q$ to any convenient baseline, we may view our task as that of choosing between the null model and an alternate hypothesis for which the groups have different centroids in $Q$ space. On the null model, the squared distance between the two centroids, divided by the square of their pooled standard error in that direction, is distributed as Hotelling's $T^2$ on two degrees of freedom for the numerator. Hence the existence of a mean shape difference, *regardless of the shape variables upon which it may be expressed*, can be tested by the single $F$ ratio which corresponds to Hotelling's $T^2$ for the net shift of means in $Q$ space. Any basis for the $dQ$'s may be used for this purpose. Should the appropriate $F$ ratio fail to achieve significance, the populations may still differ in size, but as we have just accepted the proposition that there has been no shape change, change of size must be taken as the same for all distance measures in all directions. The size variable $S$ then provides the best estimate of this change. Should the null hypothesis instead be rejected, we conclude that the groups differ in mean shape, so that the observed size difference is a function of the choice of a size variable. We now face the problem of describing that shape difference effectively in the face of this relativity of "size." This section returns to the geometric interpretation of morphometrics, the study of deformations, in order to explain how the geometry of a particular shape change leads to a selection of size and shape variables most suitable for reporting it.

## The Relation of Two Triangles As a Single Symmetric Tensor

Figure 8a shows a single displacement $dQ$ pertaining to a single affine ratio $Q$ in the shape space of a single triangle $\Delta Z_1 Z_2 Z_3$. This displacement, the replacement of one triangular form by another (Figure 8b), may be represented as a particularly simple deformation, the uniform affine transformation, or *shear*, of Figure 8c.

The mapping invoked here is that having a derivative which is constant throughout the triangle. When the fixed landmarks have their usual locations (0, 0) and (1, 0), then if $Q = (r_1, s_1)$, $Q + dQ = (r_2, s_2)$, the mapping we want is $(x, y) \rightarrow (x', y')$ with

$$x' = x + y\,\frac{r_2 - r_1}{s_1}, \qquad y' = y\,\frac{s_2}{s_1}.$$

The following exposition is the geometric equivalent of the usual *singular value decomposition* of the

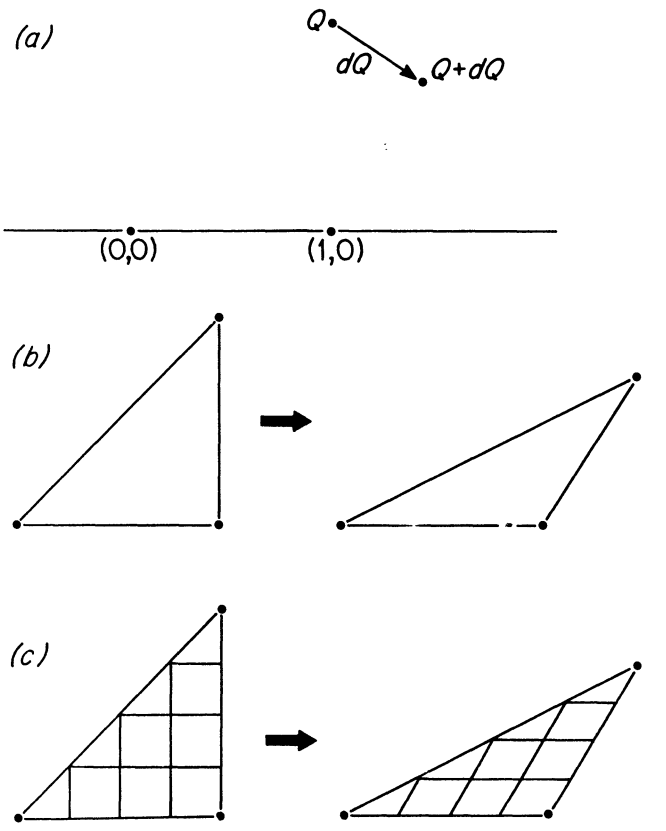


FIG. 8. *Shape displacement as deformation.* (a) *A displacement $dQ$ in shape space.* (b) *The transformation from triangle $\Delta Z_1 Z_2 Z_3$ to $\Delta Z_1' Z_2' Z_3'$ represented by $dQ$.* (c) *The uniform transformation of interior points suggested by the pair of triangles: the map taking $\sum_{i=1}^{3} c_i Z_i$ to $\sum_{i=1}^{3} c_i Z_i'$ for all combinations of coefficients $c_1, c_2, c_3$ with $\sum c_i = 1$.*

coefficient matrix

$$\begin{bmatrix} 1 & \dfrac{r_2 - r_1}{s_1} \\ 0 & \dfrac{s_2}{s_1} \end{bmatrix}.$$

Compare the algebraic treatment in Goodall (1983) or Bookstein (1978a, pp. 103–104, 154–155).

As Section 2 explained, the size measures available for this triangle of landmarks are the formulas $|\sum_{i=1}^{3} c_i Z_i|$ with $\sum c_i = 0$. We are interested in the ratios by which these distances have changed from left to right. For the uniform transformation, all distances in the same direction change in the same ratio (an assertion equivalent to the assumption that our coordinate plane is Euclidean: Bookstein (1982b), pp. 180–181). We may thus restrict our attention to an assortment of distances measured along lines through a single point (Figure 9a). To each corresponds a *strain ratio*, length on the right divided by corresponding length on the left. Note the spirit of factor analysis,

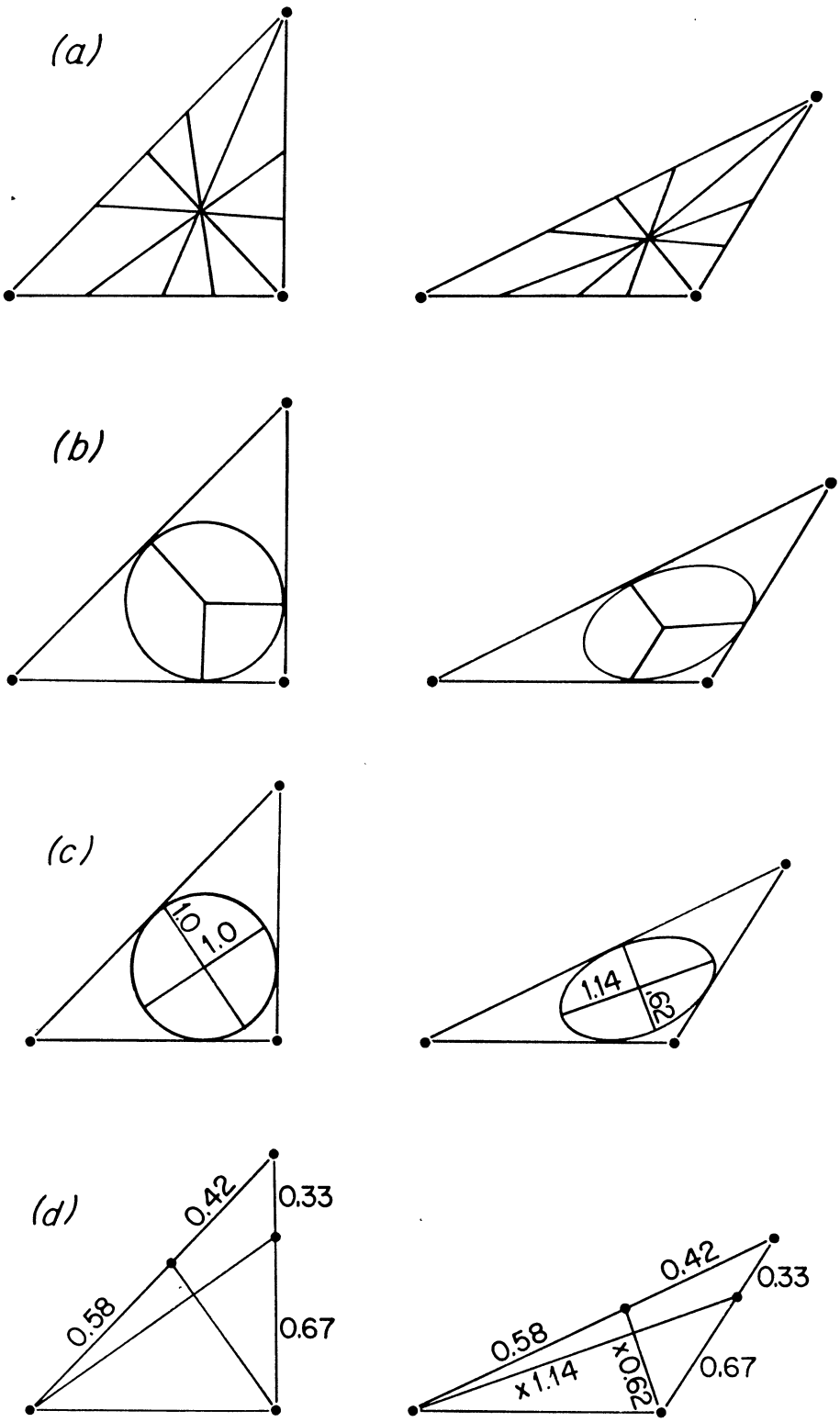


FIG. 9. *Principal strains of a single shape change.* (a) *Because the transformation shown in Figure* 8c *is uniform, we may restrict our attention to length changes for a set of lines in all directions through a single point.* (b) *The strain ratios (loadings of the logarithms of distance upon this shape change) may be read in the lengths into which a set of originally equal lengths are deformed.* (c) *The transformation indicated in Figure* 8c *takes circles into ellipses; the directions of greatest and least strain ratio correspond to the axes of the ellipse, its longest and shortest diameters. They are at* 90° *upon the ellipse, while the diameters mapped into them are at* 90° *upon the circle. Hence distances measured parallel to these principal strains are perpendicular both before and after the deformation.* (d) *Each of these distance measures may be realized as the length of a transect joining one vertex of the triangle to a point of the opposite edge.*

as promised: without yet specifying how the change will be measured, we wish to inspect the loadings (strain ratios) induced by a single underlying change (the uniform deformation as factor) upon the full space of distance measures.

For visualizing this quotient, it is convenient to arrange the diagram so that the lines undergoing deformation begin at constant length. To this end we may invoke the radii of any conveniently placed circle (Figure 9b); we read the strain ratios directly in the lengths of the segments into which the originally equal lengths are deformed.

Under uniform transformations, circles are mapped into ellipses. (This is a very old theorem, demonstrated in most texts of analytic geometry.) Then the *axes* of the ellipse, longest and shortest diameters, represent the directions of greatest or least ratio of change in homologous distances, that is, greatest or least loading of log distance on the factor (deformation) which is the shape change.

These axes (Figure 9c) are at 90° both before and after deformation. Because the strain ratio is a function solely of direction, the direction of either axis may be displayed, without reference to any particular circle, as a transect from some landmark through a point on the opposite edge of the triangle (Figure 9d). In the algebraic notation of Section 2, the distance referred to is the size variable $| Z_i - (aZ_j + (1 - a)Z_k) |$ for $a$ between 0 and 1 and $(ijk)$ some permutation of (123).

In shape space, these directions may be computed from observation of $dQ$ alone without explicitly drawing any ellipses. The construction (Bookstein, 1984a) is as in Figure 10a. The circle $H$ is drawn to pass through $Q$ and $Q + dQ$ with center on the real axis. Let the points at which $H$ intersects the $x$ axis be denoted $(X_1, 0)$ and $(X_2, 0)$. The angles $\angle X_1QX_2$ and $\angle(X_1(Q + dQ)X_2$ are each inscribed in a semicircle, hence both measure 90°. Because the $X$'s are on the real axis, the linear transformation represented here, which leaves (0, 0) and (1, 0) fixed, also leaves the $X$'s fixed. Then under this transformation the lines $X_iQ$, $i = 1, 2$, correspond to the lines $X_i(Q + dQ)$. Because these directions (i) correspond under the transformation, and (ii) are at 90° in both forms, they must be the principal directions of the transformation to which the construction refers—axes of the ellipse, directions of greatest and least ratio of change of size.

For small changes of shape, the concern of this essay, the circle through $Q$ and $Q + dQ$ may be approximated by the circle through $Q$ with tangent along the direction $dQ$ there. Then (Bookstein, 1984a, pp. 490–491) the directions we seek through $Q$ are at ±45° to the bisectors of the angle between $dQ$ and the real axis; they are the bisectors of the angle between the real axis and the direction perpendicular to $dQ$, as

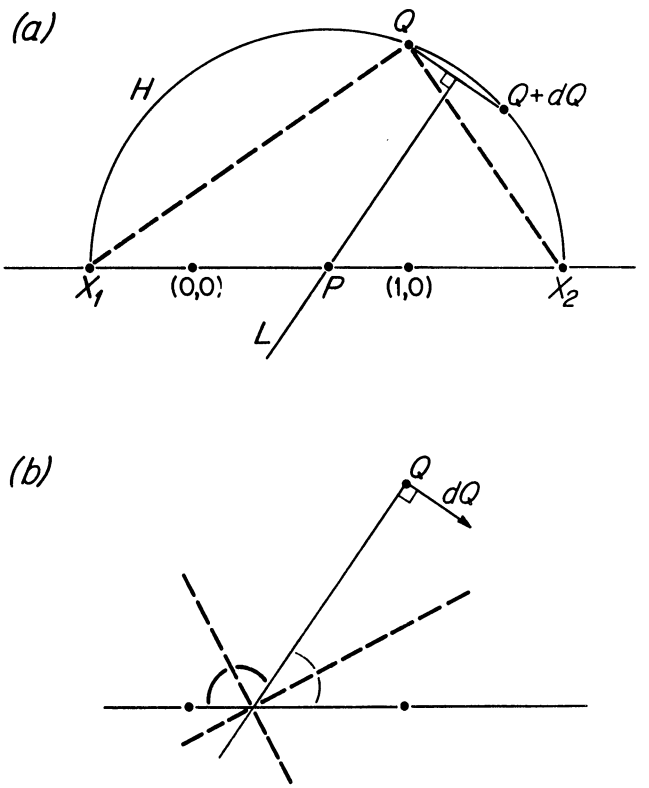


FIG. 10. *Ruler and compass construction of the principal directions corresponding to a displacement dQ in shape space.* (a) *Exact construction. Let the line L, perpendicular bisector of the segment from Q to Q + dQ, intersect the real axis at P. About P draw the circle H through Q, intersecting the real axis at $X_1$ and $X_2$. The principal axes of the deformation, referred to the triangle on Q, lie along $X_1Q$ and $X_2Q$.* (b) *Approximation for small $|dQ|$. The previous construction results in directions bisecting the angle between the perpendicular to dQ and the real axis.*

in Figure 10b. Furthermore, if $\delta_1$ and $\delta_2$ are the strain ratios in the two principal directions, we have, to first order terms,

$$\delta_1 - \delta_2 = \frac{|dQ|}{\text{Im } Q}.$$

This is the *anisotropy* of the transformation, the greatest divergence of specific rates of change (difference of loadings) in any pair of distances.

The geometrical object just introduced is a representation of homogeneous deformation that is familiar to the mathematician or engineer: a *symmetric tensor* formally independent of any choice of coordinate system a priori. In this form it appears frequently in mathematical discussions of growth and in cartography, geology, and other sciences of position.

## Mean Shape Difference As the Deformation between Mean Shapes

This construction applies directly to sample scatters of points $Q$ or displacements $dQ$ in shape space. The maximum and minimum mean ratios of distance lie

in directions determined by this construction as applied to the centroid of the observed group difference $dQ$ acting at the centroid of the observed points $Q$ (Figure 11). If the mean $dQ$ is significantly different from the zero vector, by $T^2$ test, these directions are empirically determinate; if the centroid of the $dQ$ is not significantly different from (0, 0), then mean ratios of distance must be considered to be the same in all directions.

This notion of a mean ratio of distances is not the ratio of the mean distances contributing to it. Size change along baselines is scaled out first, so that diverse changes of net size do not contribute so disparately to determining the principal directions. In exchange for this complication, we arrive at a sense of "mean shape" analogous to "mean size" for the computation of differences and changes. The size variable of greatest "mean" ratio of difference, in this sense, becomes the size variable showing the greatest loading upon the single transformation associated with the mean differential $dQ$ at the mean point $Q$, as in Figure 11. That is, just as the mean difference between two samples of scalars is the difference of their means separately, so the mean difference between two samples of triangular landmark shapes is here construed, and measured, as the transformation between the means of their shapes separately.

In this way, for any set of three landmarks we can extract by purely geometric operations the size variables relatively most increasing or most decreasing over any comparison of biometric interest. The shape variable showing the greatest proportional change—the "optimal" shape variable—is the ratio of these two size variables. Thus for a single triangle, a single geometric construction which began with *arbitrary* size and shape variables (the baseline, together with one pair of shape coordinates with respect to it) results in a selection of size and shape descriptors customized for the mean difference actually observed. The sense in which these descriptors are optimal is here one of *net* proportional difference, raw loading of the logarithm of distance upon the deformation in question.

For a study design comparing two groups of triangles observed once only, the appropriate statistic is the ordinary two-group Hotelling's $T^2$; the report which results identifies the distances showing the greatest and least mean ratio between the groups; and the shape variable which is the ratio of these distances shows the greatest mean difference (on a log scale) between the groups. If the data follow a single group of triangles observed through time, or other matched design, then the statistic is a *matched* Hotelling's $T^2$ and the distances identified are those showing the greatest and least mean ratios of change over the period of observation. (Once again, change in length of the baseline is scaled out before changes in other lengths are averaged.) Variations in elapsed time in this design may be corrected (Bookstein, 1984a) by normalizing the vector $dQ$ of shape difference to a standard interval of time. Finally, if the purpose of one's study is the comparison of observed growth in two groups of triangles followed over time, then Hotelling's $T^2$ is applied between the scatters of the change scores in shape space, leading to a report of

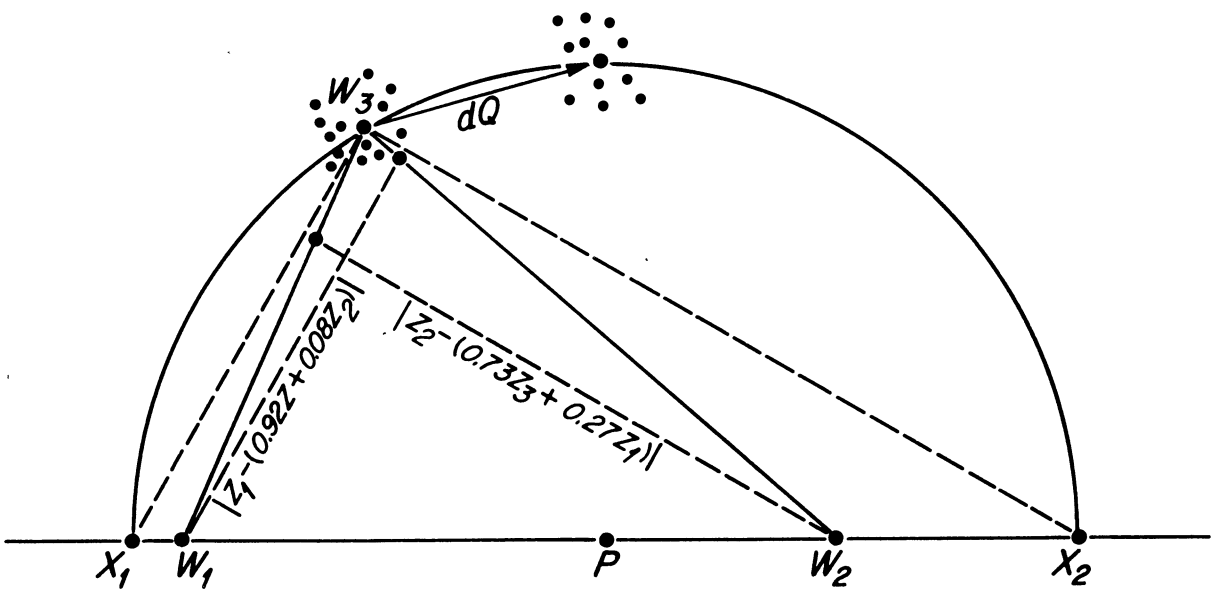


FIG. 11. *The mean shape difference describes the relation between two mean shapes as a deformation. It is reported by reference to its directions of greatest and least mean ratios of length. Here "mean" refers to a mildly nonlinear mode of averaging in which differences in length of the baseline are divided out first. The centroids* $W_1$ *and* $W_2$ *have been moved to* (0, 0) *and* (1, 0) *so that we can identify landmark space with shape space for the purpose of this diagram.*

the distances showing the algebraically greatest and least differences of mean growth rate between the groups. In all cases, the two distance measurements having extremal loadings on the contrast under study will tend to be at approximately 90° in all the forms of the series.

### Analyses of More Than Three Landmarks

It might be imagined that for configurations of four or more landmarks the extraction of optimal mean strain ratios would lead to geometrical constructions distinctly more complicated than these. Surprisingly, no further complexity is necessary, owing to the following finding. (Recall that admissible size variables are distances between weighted combinations of landmarks $\sum_{c_i>0} c_i Z_i$, $\sum_{c_i<0} -c_i Z_i$ for which the convex hulls of the landmark sets $\{Z_i \mid c_i > 0\}$, $\{Z_i \mid c_i < 0\}$ are disjoint.)

THEOREM 2. *For any set of landmarks, the admissible size variables showing the maximum or minimum mean ratio of difference between two configurations are lengths of segments from one vertex through the proper linear combination of two others, transects* $|Z_i - (aZ_j + (1-a)Z_k)|$, $0 \le a \le 1$, *as in Figure 9d.*

However many the landmarks of a configuration observed in two populations, the theorem asserts, any distance showing the largest or smallest mean ratio between the populations may be measured using at most three of the landmarks. In other words, all the distances we use in our reports of shape change are either direct length measures between landmarks or else transects of some triangle as exemplified in Figure 9d. The complexity of our search for distances with optimal loadings upon a single comparison is thus limited to a combinatorial search through $\binom{K}{3}$ triangles. Because the principal strains relating a pair of mean triangular shapes can be computed in closed form, the search for these distances does not require any empirical numerical optimization with respect to even a single continuous parameter.

This theorem is principally geometrical, not statistical. Its proof proceeds by reducing the problem to a matter of the ratio between two positive definite quadratic forms. Write $W_i$, $W_i'$ for the (unobservable) mean landmark configurations in the two groups under comparison. Let $|\sum c_i Z_i|$ be the size variable we are seeking, with $\sum c_i = 0$, and construct a formal squared distance $R = |\sum c_i W_i|^2$ pertaining to means of measures made upon the configurations of $Z_i$. Write $R' = |\sum c_i W_i'|^2$, the homologous quantity defined upon the configuration of the $W_i'$. Even though the $W$'s are unobservable, $R$ and $R'$ are estimable, since, because the $c_i$ are real,

$$|\sum c_i W_i|^2 = |\sum c_i (W_i - W_1)|^2$$
$$= (\sum c_i (W_i - W_1))(\sum c_i (\bar{W}_i - \bar{W}_1))$$
$$= \sum_{i=2}^{K} \sum_{j=2}^{K} c_i c_j (W_i - W_1)(\bar{W}_j - \bar{W}_1).$$

But each product $(W_i - W_1)(\bar{W}_j - \bar{W}_1)$ is observable: it is estimated as the mean of the products $(Z_i - Z_1)(\bar{Z}_j - \bar{Z}_1)$ of two Euclidean vectors themselves observed in a single coordinate system. Thus the squared length $R$ we seek is a non-negative definite quadratic form in the coefficients $c_1, \cdots, c_K$.

The theorem deals with extrema of the ratio of $R'$ to $R$ over all sets of coefficients $c_i$ summing to zero. This ratio may be construed, in light of the corresponding remark for triangles, as a (squared) loading of the log distance measure represented by the $c_i$ upon the shape difference considered as a factor; we seek to specify, prior to measuring them, the distances that have the highest and lowest loadings upon this factor. The ratio is a function only of the proportions among the $c$'s—multiplying them all by the same constant multiplies both numerator and denominator of the ratio $R'/R$ by the square of this constant, leaving their ratio unchanged. Then by reordering terms, as in Section 2, we may interpret $R$ as the squared distance between a pair of constructed landmarks in the mean configuration of the $W_i$, and $R'$ as the comparable squared distance for the $W_i'$.

From this point on, the proof of the Theorem follows by an induction argument using two geometrical lemmas. It is presented in the Appendix.

## 6. AN EXTENDED EXAMPLE FROM CRANIOFACIAL BIOMETRICS

### Summary

The techniques of the previous sections are exemplified in a study of growth of the head in 62 normal Ann Arbor youth. Using the landmarks Sella, Nasion, Gonion, and Menton of the lateral cephalogram, we rigorously test for sexual dimorphism in form at age 8, mean shape change from age 8 to 14, and presence and dimorphism of size allometry. By canonical correlations analysis of the shape coordinates, we extract the shape variable of highest autocorrelation over this age interval. All analyses are interpreted as smooth tensor fields over the interior of the quadrilateral of landmarks. Each is visualized by the method of biorthogonal grids, an orthogonal coordinate system customized for the depiction of particular deformations.

## The Lateral Cephalogram

The algebraic and statistical machinery is now in place for averaging populations of deformations and subsequently expressing their central tendencies by way of particularly helpful size and shape variables. This section illustrates the tensor biometrics in a typical study design combining growth and group differences.

Most craniofacial biometrics begins with x-ray images of bony crania and jaws positioned for exposure in a standardized fashion. The subject's head is placed some 6 feet from the x-ray tube and a few inches from a film cassette; the central beam of rays is made to pass along the line joining his ear holes, and then intersects the film plane at 90°. In the images which result, edges of anatomical structures can be reliably traced in a conventional abstraction of normal anatomy which includes the midsagittal structures traced in Figure 1 as well as many more. Two kinds of curves are used—projections of true space curves, and edges of regression of bony surfaces; and landmark "points" may be true anatomical loci or intersections of shadows of structures which do not actually abut in the skull. Figure 12 shows a stereotyped tracing of this so-called *lateral cephalogram*, with indications of the landmarks used in the course of the example.

The data for the example are landmark locations from cephalograms taken annually in the course of the University of Michigan University School Study. The summary of that study (Riolo et al., 1974) includes operational definitions of these and many other conventional landmark points. The full sample is of about 100 Ann Arbor schoolchildren followed over various age ranges in the 1950's and 1960's; for a subset of 36 males and 26 females, there are serial records of the four landmark locations at ages 8 years ± 6 months and 14 years ± 6 months. The statistical demonstrations to follow may be read in the context of earlier, simpler analyses of these same data (Bookstein, 1983, 1984a, b).

## A Splanchnocranial Quadrilateral

When we view a person in profile, one of the features to which we attend is the position and angle of the

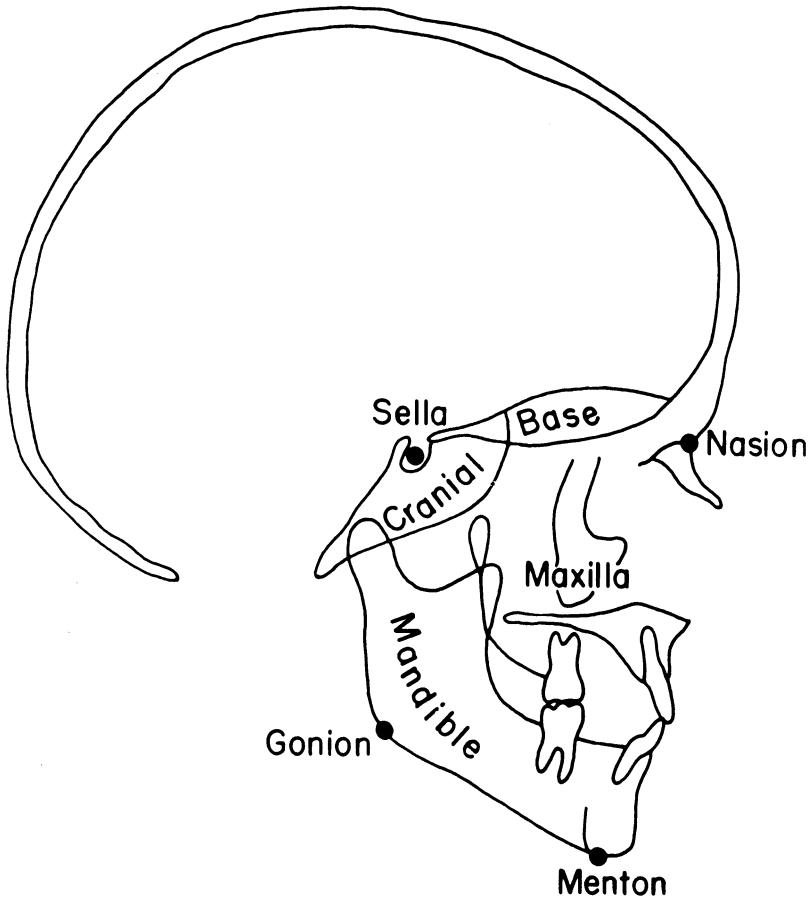


FIG. 12. *Schematic of a lateral cephalogram, indicating the quadrilateral of landmarks used in the example. Sella, center of the pituitary fossa (sella turcica) in the sphenoid bone; Nasion, junction of the frontonasal and metopic sutures; Gonion, "midpoint" of the angle of the mandible (for details, see Riolo et al., 1974, p. 17); Menton, junction of symphyseal outline with the lower border of the chin.*

lower jaw, or *mandible*, relative to the rest of the head. We may record the lie of the mandible by two of its landmarks: *Menton*, one of several alternate definitions of the point of the chin, and *Gonion*, the palpable, often visible corner of the mandible underneath the rear molars. The join of these two points is the so-called *mandibular plane.*

The trained observer perceives this pair of landmarks with respect to a superior baseline, *Frankfurt Horizontal*, a straight line running over the top of the earhole and under the lower rim of the eye socket. A configuration of landmarks under better biological regulation is the quadrilateral relating the mandibular plane to the *anterior cranial base*, the front part of the floor of the braincase. Landmarks for this segment, also present in Figure 1, are conventionally taken as *Sella* (the seat of the pituitary gland, a little circular fossa visible on the x-ray) and *Nasion* (the concavity at the bridge of the nose). The quadrilateral formed by these four landmarks (Figure 12) straddles most of the *splanchnocranium*, that part of the head which deals with breathing, smelling, and chewing rather than with protection of the brain.

In a study of only four landmarks, it is feasible, even though redundant, to tabulate the sample means and mean changes of all distances between landmarks and of one affine ratio for each triangle of landmarks. These are presented in Tables 1 and 2. The roster of interlandmark distances is augmented by the additional size variable $S$, sum of the squares of all the interlandmark distances.

TABLE 1
*Statistics for a size basis, quadrilateral Sella-Nasion-Menton-Gonion*

| Size measure[a] | Age 8 | | Age 14 | | Change | | |
|---|---|---|---|---|---|---|---|
| | Mean | S.D. | Mean | S.D. | Mean | % | S.D. |
| Males ($N = 36$) | | | | | | | |
| \| Sel-Nas \| | 2.951 | .124 | 3.156 | .149 | .205 | 6.9 | .052 |
| \| Sel-Men \| | 4.541 | .176 | 5.276 | .250 | .735 | 16.2 | .165 |
| \| Sel-Gon \| | 2.739 | .169 | 3.267 | .217 | .528 | 19.3 | .120 |
| \| Nas-Men \| | 4.462 | .216 | 5.116 | .297 | .654 | 14.7 | .155 |
| \| Nas-Gon \| | 4.362 | .174 | 4.976 | .221 | .613 | 14.1 | .128 |
| \| Men-Gon \| | 2.597 | .122 | 3.052 | .151 | .455 | 17.5 | .124 |
| $S$ | 82.7 | 5.8 | 109.0 | 9.0 | 26.4 | | 5.7 |
| Females ($N = 26$) | | | | | | | |
| \| Sel-Nas \| | 2.825 | .123 | 3.003 | .139 | .178 | 6.3 | .039 |
| \| Sel-Men \| | 4.370 | .195 | 4.994 | .220 | .624 | 14.3 | .099 |
| \| Sel-Gon \| | 2.614 | .184 | 3.036 | .226 | .422 | 16.1 | .092 |
| \| Nas-Men \| | 4.284 | .182 | 4.821 | .234 | .537 | 12.5 | .120 |
| \| Nas-Gon \| | 4.176 | .168 | 4.689 | .188 | .512 | 12.3 | .075 |
| \| Men-Gon \| | 2.524 | .160 | 2.952 | .174 | .429 | 17.0 | .071 |
| $S$ | 76.2 | 5.6 | 97.3 | 7.2 | 21.2 | | 3.1 |

[a] The units of $S$ are square inches; of the other size measures, inches.

## Findings

The tests appropriate for a study of growth and growth dimorphism are as follows.

*Sexual dimorphism at age 8: Size.* The mean difference in $S$ corresponds to a $t$ ratio of 4.35, $p < 0.0001$. There is about a 4% difference in mean linear dimension $\sqrt{S}$, the boys tending to be larger. *Shape.* The overall $T^2$ on four degrees of freedom for net difference of means in shape space is not significant. Triangle by triangle, no $T^2$ for differences in mean $Q$ attains even an $F$ ratio of 1.0. Thus there are no reportable differences in starting shape between the sexes.

*Allometry at age 8.* Separately by sex, *static allometry* for the 8-year-old samples may be tested by a single $F$ ratio for the multiple regression of $S$ on shape space. Using the shape basis derived from the first two triangles—real and imaginary parts of $dQ$(Sel, Men, Nas) and $dQ$(Sel, Men, Gon)—we find $F$'s of 1.87 for the males and 0.75 for the females, neither significant.

*Growth from age 8 to age 14.* The loadings of interlandmark distances upon this factor appear under the heading "%" in Table 1. *Males.* The change in $S$ averages 26.4 with a standard deviation of 5.7; this

TABLE 2
*Statistics for a shape basis, quadrilateral Sella-Nasion-Menton-Gonion*

| Baseline, movable point | Age 8 | | | Age 14 | | | Change | | | \|Chg\| / Im $Q$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | Re $Q$ | Im $Q$ | S.D.[a] | Re $Q$ | Im $Q$ | S.D. | Re $Q$ | Im $Q$ | S.D. | |
| Males ($N = 36$) | | | | | | | | | | |
| Sel-Men, Nas | .23 | .61 | .03 | .21 | .56 | .04 | −.019 | −.047 | .019 | .08 |
| Sel-Men, Gon | .52 | −.31 | .02 | .52 | −.33 | .03 | .006 | −.021 | .015 | .07 |
| Sel-Gon, Nas | −.19 | 1.06 | .07 | −.20 | .95 | .07 | −.005 | −.115 | .038 | .11 |
| Gon-Nas, Men | .15 | −.57 | .03 | .16 | −.59 | .04 | .005 | −.017 | .021 | .03 |
| Females ($N = 26$) | | | | | | | | | | |
| Sel-Men, Nas | .23 | .60 | .03 | .21 | .56 | .04 | −.013 | −.043 | .016 | .08 |
| Sel-Men, Gon | .51 | −.31 | .03 | .51 | −.33 | .03 | −.002 | −.022 | .014 | .07 |
| Sel-Gon, Nas | −.19 | 1.06 | .08 | −.21 | .97 | .08 | −.011 | −.095 | .034 | .10 |
| Gon-Nas, Men | .16 | −.58 | .03 | .17 | −.61 | .04 | .013 | −.022 | .020 | .04 |

[a] The average of the population standard deviations for Re $Q$ and Im $Q$ separately.

corresponds to a $t$ ratio above 20. For all triangles, all $T^2$ ratios for testing shape change are at least 30. Both change of general size and changes of shape in all parts of the form are therefore inferred to be real. The $F$ ratio for *growth allometry*, the regression of change in size on change in shape, is 8.15, itself highly significant. *Females.* The change in $S$ averages 21.2 with a standard deviation of 3.1. The change in shape of each triangle bears a $T^2$ of at least 50. The $F$ ratio for growth allometry in this change is 4.06, $p < 0.015$. For both sexes, the greatest mean change of proportion (about 10%) is in the triangle Sella-Gonion-Nasion; the least is for triangle Menton-Nasion-Gonion.

*Sexual dimorphism in growth.* Triangle by triangle, $F$ ratios corresponding to Hotelling's $T^2$ range from 1.33–2.86, none quite significant at the 5% level. The dimorphism most nearly approaching significance is for the shape change of the triangle Sella-Menton-Gonion. In an analysis of covariance of $dS$ upon two $dQ$'s and sex, the $F$ ratio for equality of slopes is 0.31, so that the patterns of allometry may be taken as geometrically the same in the two sexes, differing only in the extent to which $dS$ actually varies.

### Visualizing Mean Changes by Biorthogonal Grids

Triangle by triangle, the mean growth changes just tabulated may be diagrammed as mean shifts of single landmarks with respect to a baseline, together with information about the change in size of that baseline. For either sex there are 12 such vector interpretations, some of which are shown in Figure 13 along with the directions of greatest and least mean size change that may be derived from the vectors by the construction of Figure 11. In these latter figures the change of scale of the baseline, ignored in the construction of shape space, has been restored, so that the principal directions now bear true mean fractions of growth as observed over the 6 years of the study. They may be compared to analogous quantities differently averaged, the mean change as a fraction of mean starting form, in Table 1.

The net change in the entire configuration, all four landmarks, may be diagrammed as a combination of the shape changes corresponding to any two of these triangles, together with one scale change along their common edge. For instance, we may use $dQ$(Sel, Men, Nas) = (−.019, −.047), $dQ$(Sel, Men, Gon) = (.005, −.021), $d\,|\,\text{Sel-Men}\,|$ = 16.2% for males, and the corresponding values of (−.013, −.043), (−.002, −.022), 14.3% for females. There result the two displacement plots of Figure 14, the changes of mean polygonal shape for the two sexes.

For triangles, the mean shape change of a configuration could be graphed in two ways: as a displacement $dQ$ of one vertex with respect to a fixed baseline or as a deformation distributed continuously (and uniformly) over the interior of the form. (Recall Figure 8.) The same option is available in the study of polygons of higher order. We may treat the observed correspondence of vertices as a sample from a *computed* homology mapping that is linear along edges of the form between landmarks and is smoothly interpolated within. This map is drawn by its effect on an arbitrary set of interior points: the square grid on the left, deforming into the more general structure on the right. This calculation is illustrated in Figure 15a for the pair of forms representing the mean shape of our 36 growing males at ages 8 and 14. The deformation as thus distributed cannot be uniform, because it must blend the different linear transformations that are appropriate near each corner separately. An algorithm for such interpolations has been published previously (Bookstein, 1978b; 1978a, pp. 108–118); for quadrilaterals, a simplification has been described more recently (Bookstein, 1985). The resulting figures are reminiscent of D'Arcy Thompson's, but have been computed to an objective criterion. Other examples of this computation were presented in Figure 1.

In Figure 15a it may be seen that the segments in the righthand mesh corresponding to lines Sella-Menton, Nasion-Gonion—the diagonals of the quadrilateral—in the left form are in fact slightly curved. In the statistical treatment, points $a$(Sella) + $(1 - a)$(Menton) correspond, so that the line Sella-Menton, like the external edges of the form, is formally perfectly straight; but this constraint precludes the sort of interpolation we are now seeking—the intersection of the diagonals, for instance, is not homologous from form to form—and must be relaxed.

At every point of a smoothly interpolated mapping one may extract a pair of local principal strains from the derivative of the mapping. These correspond to the geometry of Figure 9 as applied to very small triangles. Figure 15b draws a sample of these tensors at the points of the original square mesh (Figure 15a), but the crosses are in fact computable as a smooth *symmetric tensor field* defined everywhere throughout the interior of the landmark polygon.

The visual effect of representing the deformation by this field is greatly simplified and enhanced if we replace the original square mesh, having no relation to the facts of the comparison under study, with a new mesh computed so as to expressly accord with that change. This is the *biorthogonal grid pair* of Figure 15c, the coordinate mesh at 90° in both forms. It consists of a selection of integral curves of the symmetric tensor field in Figure 15b; an algorithm for this integration has been published elsewhere (Bookstein, 1978a, pp. 118–120). The intersections of the curves of the grid correspond, left to right, under the interpolation with which we began (Figure 15a). In other

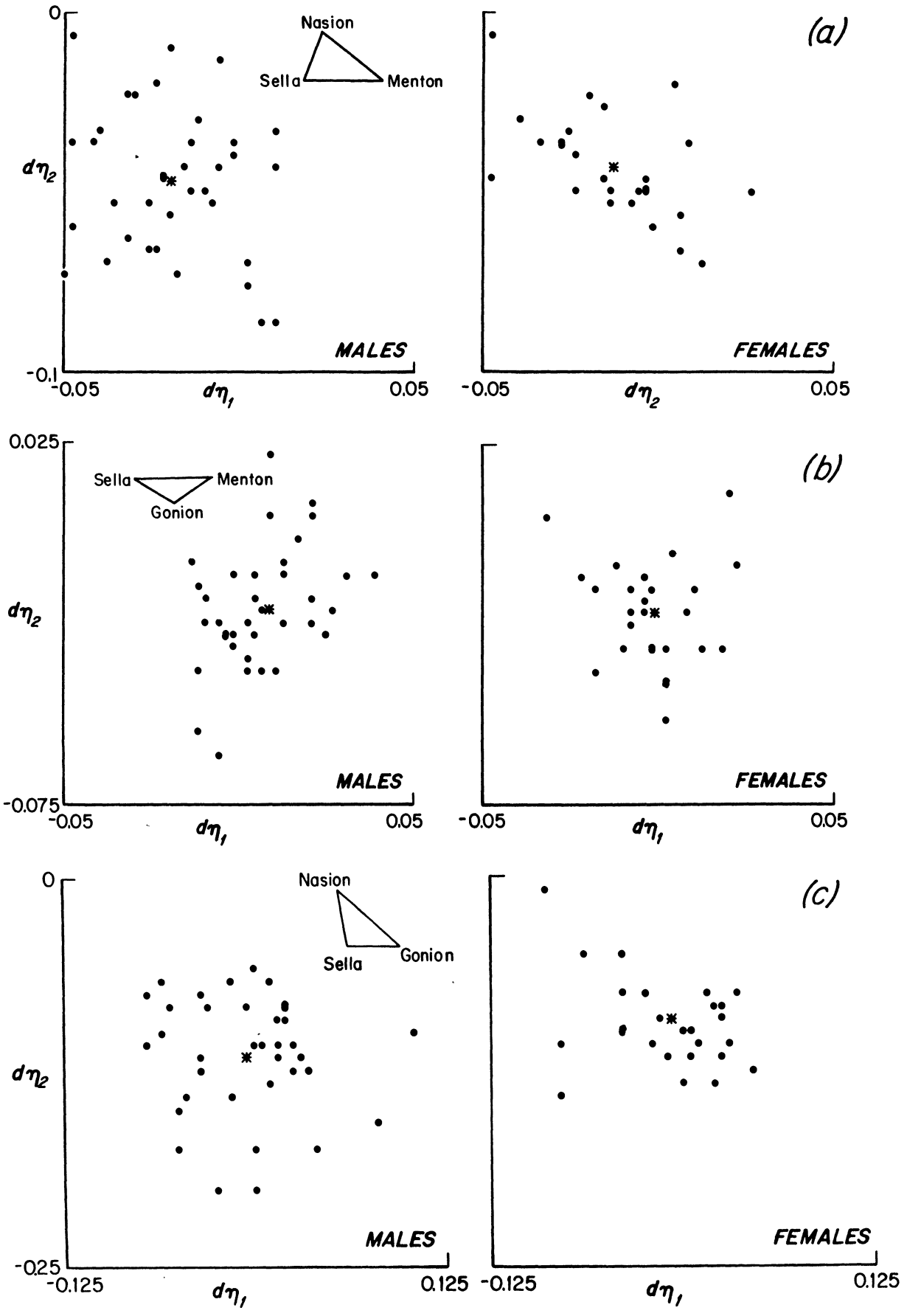
0
Nasion
Sella
Menton
$d\eta_2$
MALES
-0.1
-0.05
$d\eta_1$
0.05
(a)
FEMALES
-0.05
$d\eta_2$
0.05
0.025
Sella
Menton
Gonion
$d\eta_2$
MALES
-0.075
-0.05
$d\eta_1$
0.05
(b)
FEMALES
-0.05
$d\eta_1$
0.05
0
Nasion
Sella
Gonion
$d\eta_2$
MALES
-0.25
-0.125
$d\eta_1$
0.125
(c)
FEMALES
-0.125
$d\eta_1$
0.125

FIG. 13a–c.

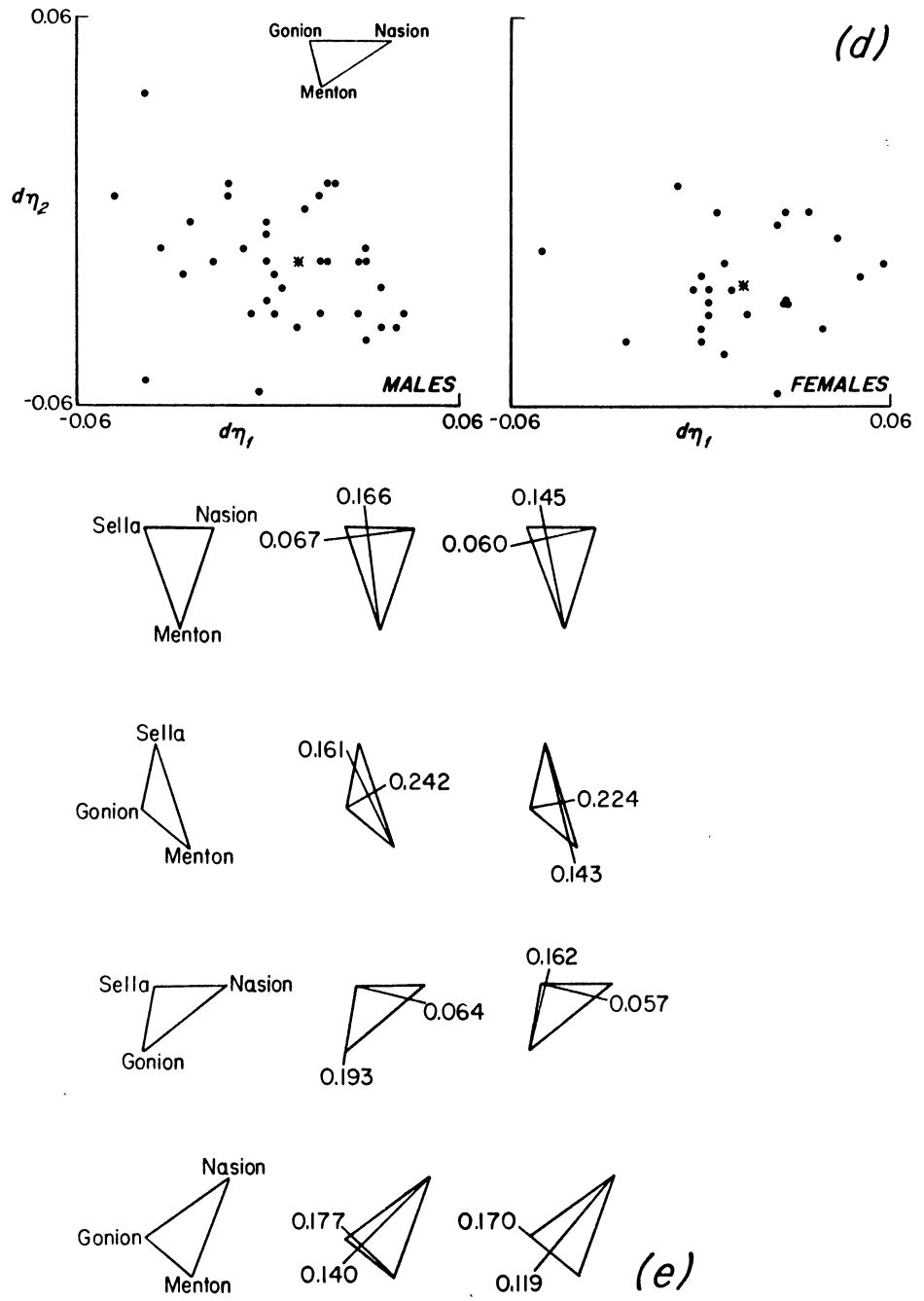


FIG. 13. *Growth and form for four triangles of landmarks.* (a–d) *Scatters, and their centroids, representing changes in shape coordinate space for all triangles using the landmarks in Figure* 12. *The sexes are plotted separately. Baselines were assigned arbitrarily.* (e) *Principal mean growth strains corresponding to these shape displacements, with scale change along the baselines restored. The display is now wholly independent of any baseline.*

words, this diagram depicts the same homology map as Figure 15a, but in a more appropriate coordinate system. The selection of curves displayed is arbitrary, but not their orientation: each curve must lie parallel to one arm or the other of the little cross of principal directions at every point through which it passes.

The grid of Figure 15c combines changes in all regions of the quadrilateral into a single geometric scheme of principal strains that are graded as they curve through the form. Some of these strain ratios are indicated on the left in the figure. Each quantity is an actual ratio of the length between neighboring mesh intersections on the right to the corresponding length on the left, an "actual" rate of increase of length (for typical male growth between ages 8 and 14) as imputed to a little length element aligned with the

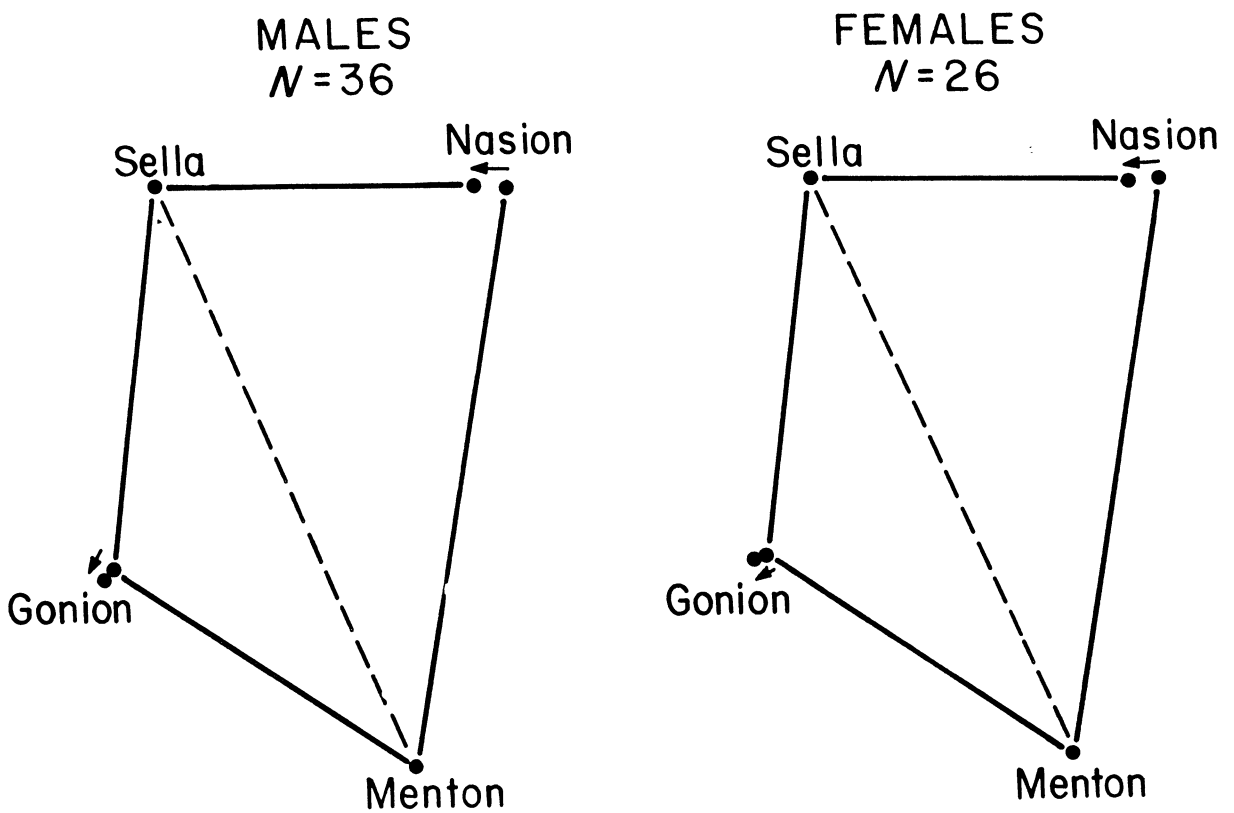


FIG. 14. *Mean shape changes, by sex, for the quadrilateral Sella-Nasion-Gonion-Menton from age 8 to 14 years. Each change is represented by the displacements of Nasion and Gonion referred to a Sella-Menton baseline.*

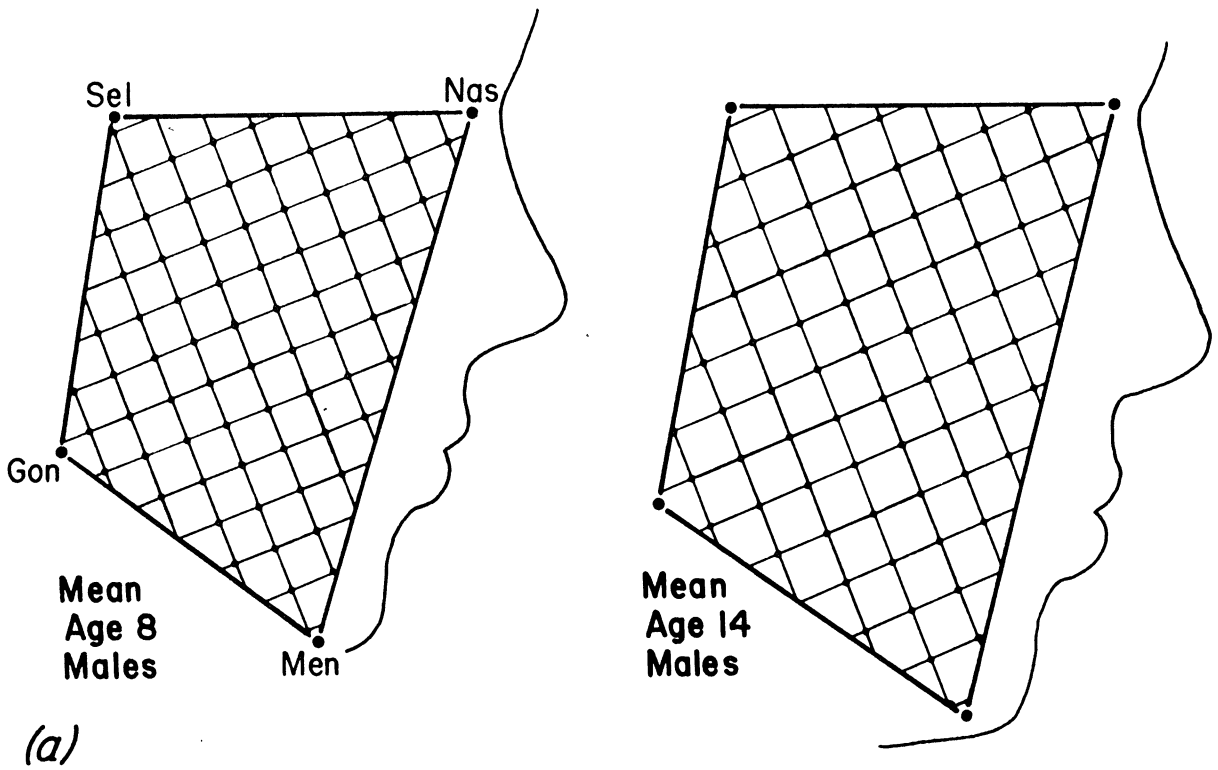


FIG. 15a.

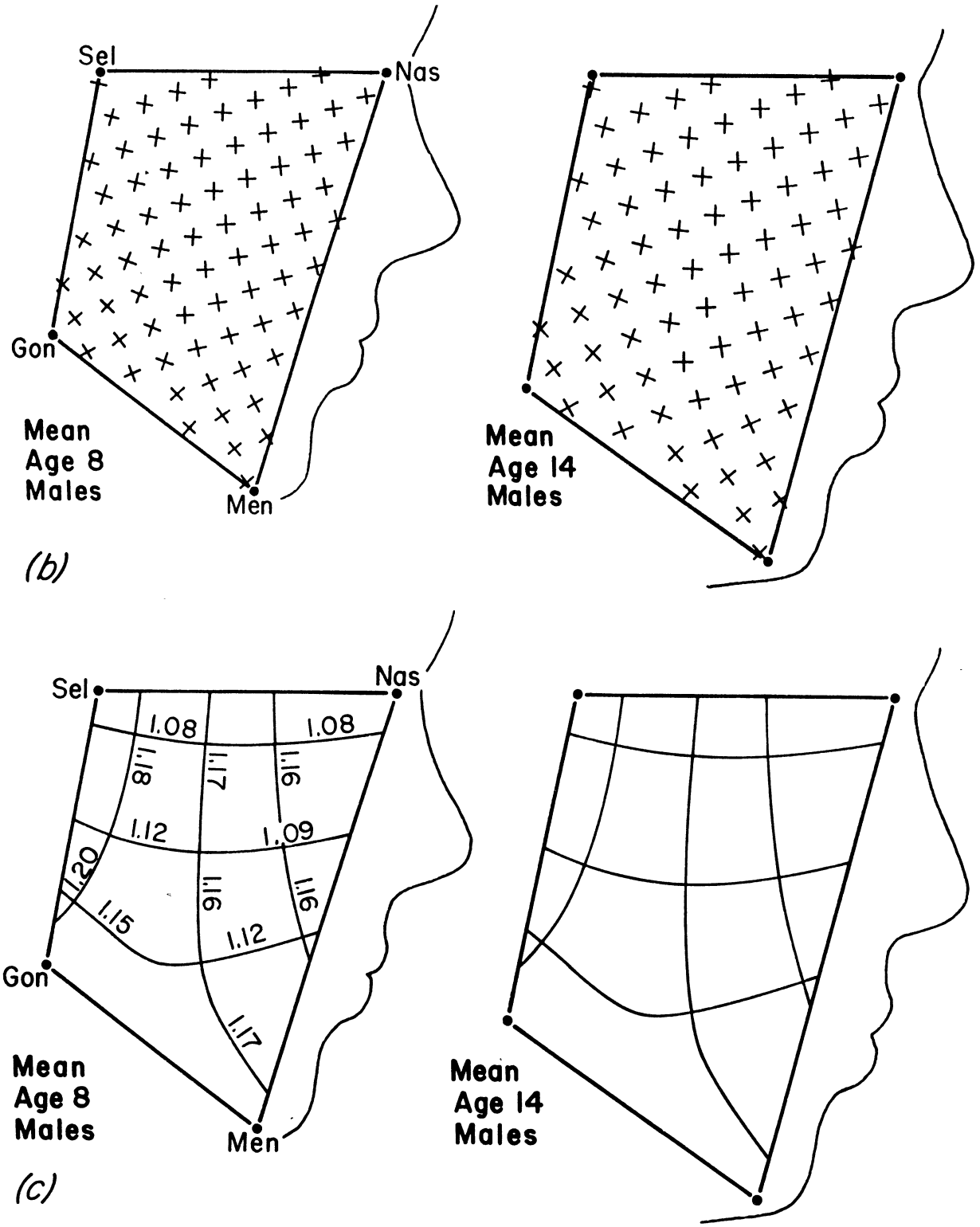


FIG. 15. *The method of biorthogonal grids.* (a) *Mean shape change for males, from the preceding figure, with baseline size change restored. The change has been modeled as a computed homology, a smooth deformation of the (mathematical) interior of the quadrilateral.* (b) *At every point of a smooth map, the construction of Figure 9 applies to yield a pair of directions that are at 90° in both forms. Along one direction, the ratio of change in length is greatest, and, along the other, least, among the set of such ratios in all directions through the point considered. The set of all these crosses, not just those at mesh points but also those everywhere in between, represents the deformation as a symmetric tensor field.* (c) *We replace the original, arbitrary coordinate mesh with a new coordinate system each of whose curves is aligned with one arm or the other of these crosses at every point through which it passes. There results an essentially unique computed coordinate system, the biorthogonal grid, which lies at 90° in both forms. Corresponding intersections of perpendicular curves match from one grid to the other according to the interpolation of frame* (a). *The ratios printed upon the 8-year-old mean form are an arbitrary sample of rates of growth from age 8 to 14 for line elements in the locations and along the directions indicated.*

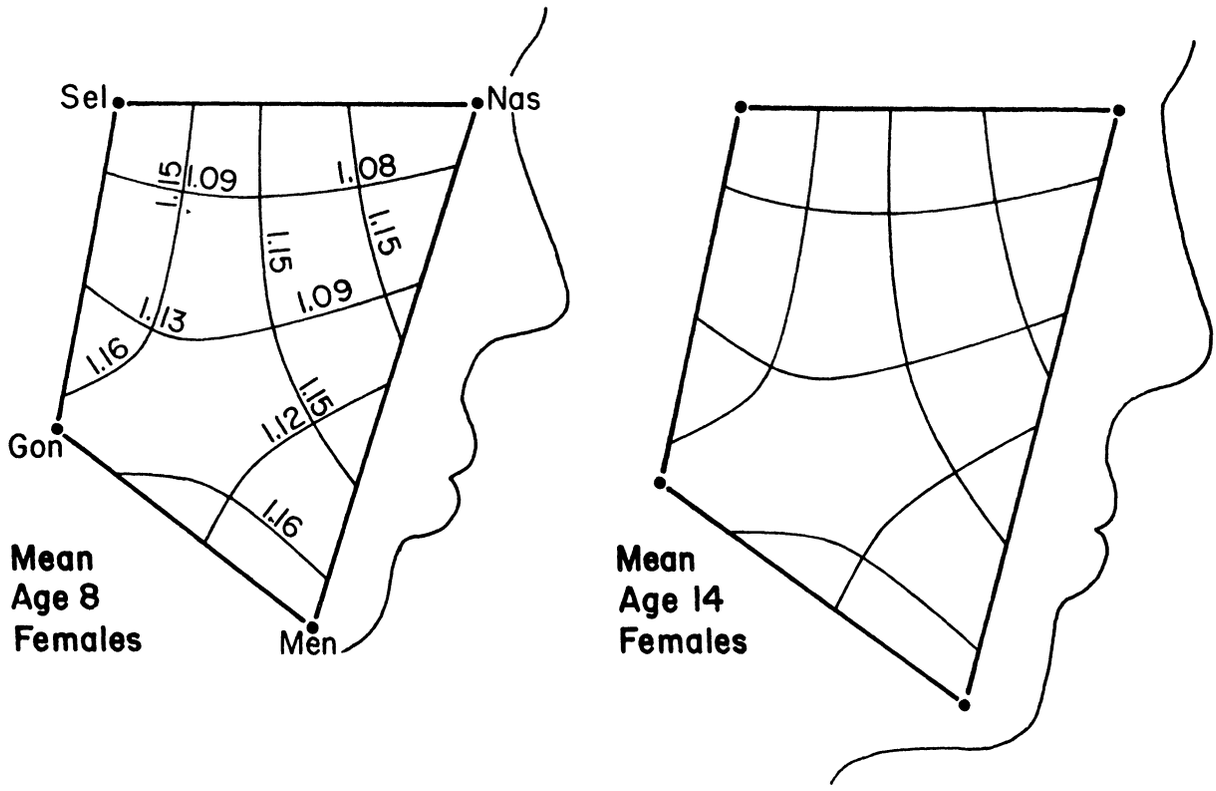


FIG. 16. *Biorthogonal grids for the same 6 years of growth in 26 females. Compare Figure* 15c.

grid. These quantities, once again, are analogous to factor loadings representing the expected response of each distance when there is a change in the factor (deformation) called chronological age. At the top of the figure, near the cranial base, change in size is predominantly vertical at a rate of 17% over the 6-year interval. Perpendicular to this is the direction of least rate of growth, here, 7% along the cranial base. Toward the mandibular plane the horizontal rate of growth increases until it matches the vertical rate. There results a *singularity*, a point at which change of size is at the same rate in all directions; the ellipse of Figure 9 is a circle, and so the principal directions are undetermined.

The analogous grid pair showing the field of mean growth for females is shown in Figure 16. The geometry of deformation is very nearly the same; the vertical component of this shape change is 1 or 2% less than that for males throughout, while the horizontal component bears nearly the same strain ratios. The shift in the location of the singularity is not important.

### Visualizing Size Allometry

The nature of size allometry for the observed growth pattern may be displayed by another grid of this sort. A few paragraphs above, it was mentioned that the regression of $dS$ on our chosen basis for shape change space was highly significant. The effect of $dS$ on the expected shape change is the pattern of changes in predicted shape coordinates $dQ$ from the inverse regressions, covariances of $dS$ with the separate components of the $dQ$. As explained in Section 4, these are exactly analogous to the ordinary notion of factor loadings for the indicators of shape space upon the factor of size change.

The coefficients of the multiple regression of $dS$ on the components of $dQ$(Sel, Men, Nas) and $dQ$(Sel, Men, Gon), for the male subsample, are −136, −181, 0, −29; of these, the first two are statistically significant, the other two not. (These quantities are nearly the same in the comparable regression for females, and also in the analysis of covariance pooling the sexes.) Regressions of the components of the $dQ$'s separately upon $dS$ yield loadings as follows: for Re $dQ$(Sel, Men, Nas), −.002; for Im $dQ$(Sel, Men, Nas), −.004; for Re $dQ$(Sel, Men, Gon), −.0002; for Im $dQ$(Sel, Men, Gon), −.0005; for $d$ | Sel-Men |, .03. There is no evidence of any failure of the assumptions underlying these regressions: all scatters of $dS$ against components of the $dQ$ appear consistent with the model of bivariate normality, and no cases have particularly high leverage on the regression coefficients.

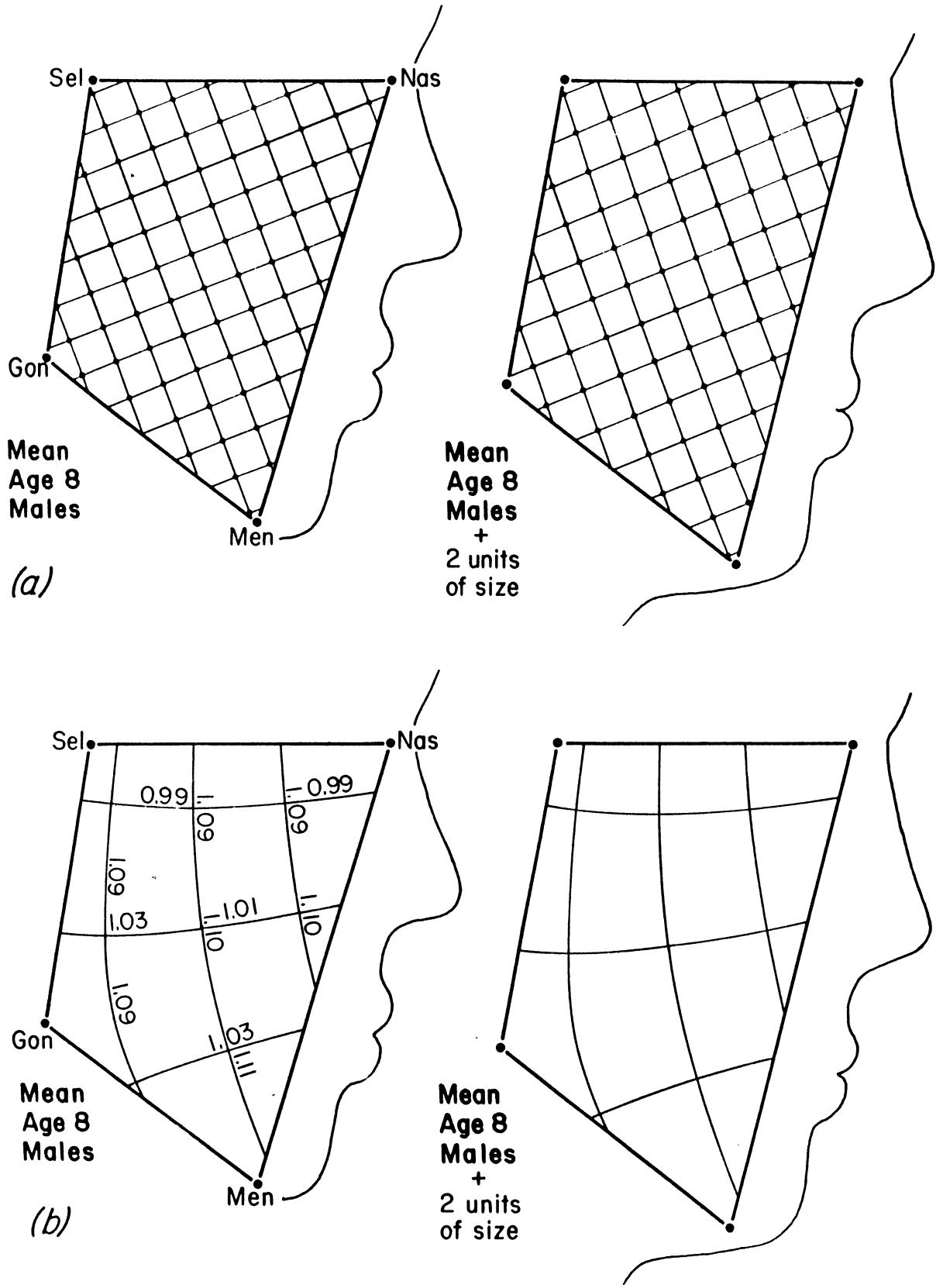


FIG. 17. *Size allometry in males.* (a) *We construe size allometry as a deformation from the mean starting form* (left) *to a form altered by the effect upon these coordinates of a size change corresponding to two standard deviations of size* (right). *Here the deformation is drawn as a simple interpolation.* (b) *The same, now expressed by its biorthogonal grid pair. Note the absence of size loading along the cranial base.*

We visualize this effect by augmenting the mean male starting form by 15 units of size change, an amount corresponding to two standard deviations of size. The resulting displacements of two $dQ$'s and rescaling by | Sel-Men |, explained in Section 4, lead to the grid of Figure 17. In effect, having altered a factor score by 2 units, we examine the expected effect on all the variables that load on that factor.

This grid resembles the two preceding, which dealt with average growth. Change of net size has no covariance with change of cranial base length Sella-Nasion: the strain differs from 1.00 by only .01. Even in the mandible, horizontal growth rate loads hardly at all upon net size change. Instead, size allometry seems to be effective in only one direction, that which showed the greatest mean ratio of change over the age interval of this study. Because change of size is not susceptible to forecasting in the clinical setting, this phenomenon of allometry appears to orthodontists as the concentration of craniofacial unpredictability along the direction of population mean change, the "growth axis."

At this point in the analysis, we cannot yet tell the difference between two alternate morphological interpretations of this allometry. In one interpretation, the growth of the cranial base segment Sella-Nasion is constrained, while the other two landmarks Gonion and Menton are free to vary on the usual circular model. In the other interpretation, the displacements of Sella and Nasion relative to the rest of the configuration are not constrained, but there is additional vertical (noncircular) displacement, variable in amount, applied jointly at Gonion and at Nasion. To discriminate between these two hypotheses requires additional evidence from the serial aspect of this design, which we shall now explore.

### Visualizing Shape Stability

Figures 15–17 dealt with contrasts; changes from age 8 to 14, sexual dimorphism, or the shape change associated with small versus large change of size. Another class of questions, of considerable import for craniofacial biology, has the opposite thrust: what aspects of shape change *least* over the age interval we are studying?—which geometric relations are most stable? This is not to be construed as a search for measures which do not change in mean; in fact, for any triangle of landmarks we can locate a variety of measures which show precisely zero mean change (Bookstein, 1983). Instead we ask which shape variables, regardless of any mean shift, are most highly correlated from age 8 to age 14.

This form of the question suggests that conventional *canonical correlations analysis* be applied to relate the pair of shape spaces, one for the age 8 forms, one for those at age 14. We may carry out this aim by an ordinary analysis of any pair of shape basis elements $dQ$ at the two ages.

We rely, as before, on a shape basis consisting of $dQ$(Sel, Men, Nas) and $dQ$(Sel, Men, Gon). The first canonical variates for these spaces bear a correlation of .938 (versus .873 for the second pair) and canonical coefficients which are nearly identical at the two ages (.58, .07, .59, −.41) and (.63, .11, .55, −.30). As usual, we convert these coefficients to loadings, simple regression coefficients of the shape basis vectors one by one on their canonical variate. This dimension of shape stability is thereby modeled by two landmark displacements, each embodying the effect of changing the canonical score upon the predicted location of a landmark $Q$ while the baseline Sella-Menton remains fixed. The collection of displacements may be reinterpreted as the deformation of Figure 18. The strain ratios presented are now the ratios of change of distance induced by change of canonical variate score. As these are, once again, analogues of ordinary factor loadings for the stable aspect of shape we are teasing out, we may take Figure 18 as the diagram of a *shape stability factor.*

By Theorem 2, distance measures which increase or decrease most over a particular change of configuration need not involve more than three landmarks. The shape variable loading most upon the contrast at hand, assembled from the numerator and denominator with the greatest divergence of loading among size measures, will involve more than three landmarks if those terms are measured upon different triangles. In the grid for shape stability (Figure 18), the distance with the highest loading on the stability factor is the distance Nasion-Menton, and the one with the lowest loading is Sella-Gonion. The simple shape ratio having the greatest covariance with the shape stability factor is thus the ratio | Sella-Gonion | / | Nasion-Menton |, bearing a correlation of .927 between the waves of observation. Orthodontists are familiar with this indicator in the guise of the *mandibular plane angle* between the segments Sella-Nasion and Gonion-Menton.

Recall the pair of alternate hypotheses accounting for the allometry uncovered in the preceding section. The stability of this mandibular plane angle is not consistent with the first of those alternatives, the hypothesis of restricted growth between Sella and Nasion; but it is consistent with the other interpretation, which augmented the null model by a vector of parallel displacement of the mandibular plane away from the cranial base. If this data set had not permitted the study of shape stability as well as shape change—if it were not a matched design—we could not have discriminated between these two explanations of allometry.

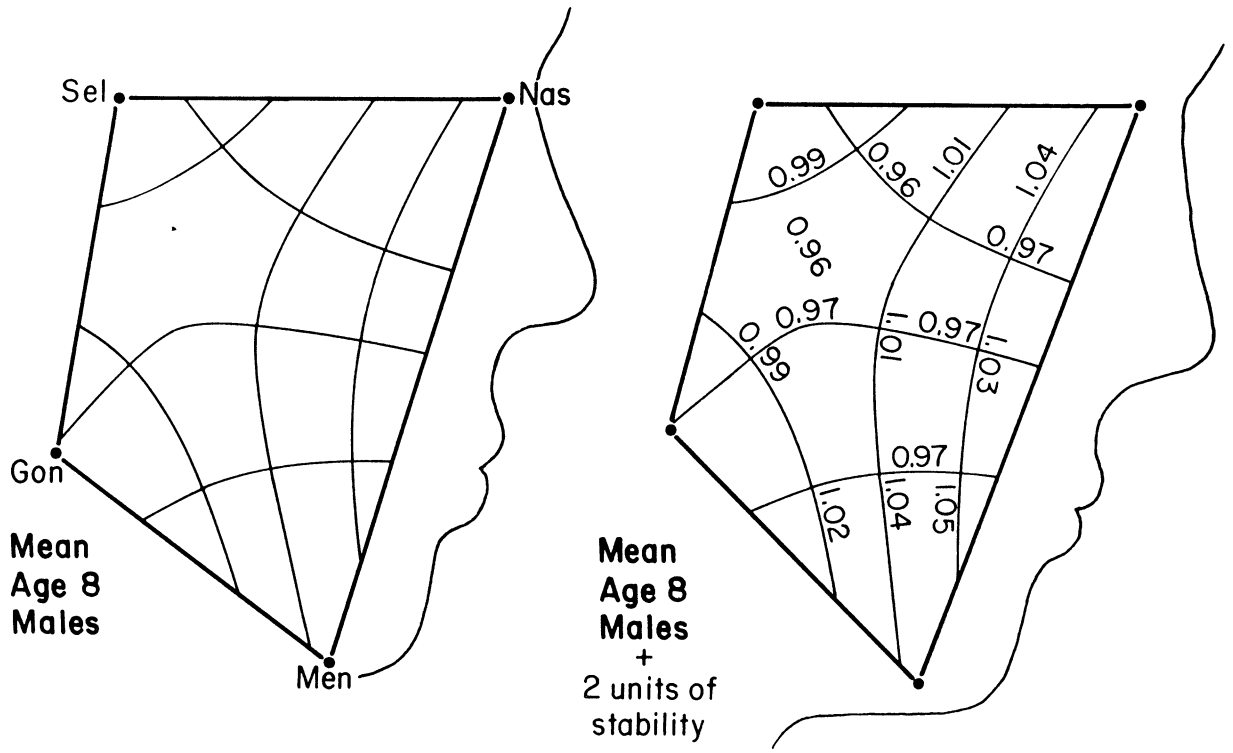


FIG. 18. *The aspect of shape variation showing the greatest autocorrelation from age 8 to 14 in males. The variation in question is realized as a deformation from the 8-year-old mean to a form differing from it by two standard deviations on the shape score. We display it by means of its biorthogonal grid. The ratio of anterior to posterior facial height, | Sella-Gonion |/| Nasion-Menton |, is a good, simple proxy for this dimension: an ordinary shape scalar of unusual stability over time.*

## 7. RECAPITULATION AND CONCLUDING REMARKS

In this essay I have attempted to unify the two principal schools of modern morphometrics, the multivariate and the geometric. The theorems, tests, and constructions presented here allow the biometrician easily to pass back and forth between the alternate modes of description. The fundamental link between the two styles is the spatial structure of size and shape measures: the analogy between directional derivatives of a deformation (a notion from the geometric school) and factor loadings of distance measures (a notion from the multivariate school), yielding an expected change in every size or shape variable whether or not actually measured. By this algebraic machinery we may interpret shape change, shape stability, and the dependence of shape upon size using a considerably broader variety of shape variables than those actually involved in the multivariate analyses.

The tensor biometric techniques introduced here are amenable to generalization in many directions.

(1) For data in three dimensions, the formalism of complex arithmetic is not available. The deformation of a tetrahedron continues to be described by a pair of distances at 90°, one bearing the highest loading, one the lowest; but there is an intermediate rate as well, the "stationary" eigenvalue. A suitable null model for this scenario might be spherical normal variation of observed landmarks about (unobservable) centroids in three dimensions. Do size and shape spaces for three-dimensional data, spanned by the interlandmark distances and their ratios as characterized in Section 2 and 3, still have so tight a relationship? Does the size variable $S$ still have covariance zero with all shape variables (Theorem 1)? Does the analysis of complicated landmark sets still reduce to analysis of tetrahedra (Theorem 2)?

(2) We need the capabilities of testing and, if necessary, relaxing the assumption of independent and identically distributed circular perturbations at each landmark. In reality, different landmarks have different measurement error ellipses as observed explicitly in fixed coordinate frames, and also different ellipses of true biological noise between individuals or over growth stages. Observed size-shape covariances, instead of representing "allometry," might instead be due to anisotropic patterns of nonsystematic variation. We need formal protocols for detecting such specification errors from geometric features of the size and shape spaces.

(3) We need ways of averaging information on the

curving of form in between landmarks. This information, easily retrieved at the same time that landmarks are located, can be handled by the method of biorthogonal grids two forms at a time (Bookstein, 1978a); but it is not clear how to extend the notion of a landmark configuration so as to permit the averaging of this additional information in a geometrically meaningful way. Sampson (1981) offers a technique for averaging arcs of conic sections used to model boundary arcs between landmarks; Bookstein et al. (1986) experiment instead with the use of a single "pseudolandmark" halfway along such an arc.

(4) Related to this is the problem of deciding whether particular landmarks, perhaps those thus dubiously extracted from boundary curvature information, contribute to the representation of the homology map or to the detection of group differences or allometric trends. When several landmarks are available in the same region of a form, or when landmarks vary so much in relative position that the assumption $|Z_i - W_i| \ll |W_j - W_i|$ of the null model is untenable, we need a way of choosing the most useful landmark from a set of alternatives. When landmarks are available inside or adjacent to regions delineated by other landmarks, we need to know if they add to the precision with which we are characterizing the transformation.

(5) Cartesian coordinates are not all the geometric data of morphometrics. Another of its active areas is the analysis of biological point processes and their geometric generalizations: counts of cells of various sorts over a region of tissue; stereological estimations of mean perimeter and area or mean optical density of selected objects; and quantification of texture, branching patterns, or reticulation of networks in two dimensions or three. Modern techniques of image processing greatly ease the tedium of gathering such data (see, for instance, the various biological examples in Serra, 1982).

For reports of the effect of gross interventions, averages of these counts and indices may be sufficient: "the treatment doubled the number of dividing cells." But for studies of subtle or regional effects, especially abnormalities, and of developing systems generally, we need to compare these fields at corresponding points of different organisms, or at the "same" points at different ages. That is, the machinery of deformation analysis is to be run in reverse, to *unwarp* all the organisms of a sample to a standard form so that we can carry out conventional statistical analyses of quantities like cell density at corresponding points across the sample. Because this correspondence must be set in terms of landmarks at some distance, we need the statistical machinery of our size and shape spaces to compute standard errors for that unwarping. In short, tensor biometrics is one of the many tools that will be necessary for effective application of image analysis in biology.

This generalized problem notwithstanding, I believe that the majority of statistical problems encountered in applications of morphometrics will be geometric. Encouraged by enormous technique advances in medical imaging and image processing, the biomedical community is now gathering far more geometric data than it is capable of analyzing. In addition to images of normal human growth, such as those underlying the data of Section 6, there exist countless images of beating hearts, breathing lungs, and other natural cycles; records of explicit geometric interventions, such as surgery, and of the body's response to these; associations of anatomy (form) with symptoms (function); and extensive archives of body form and form changes in organisms other than man. Much of biometrics has traditionally been concerned with contrasts and trends among such geometric records, but our literature does not, at present, offer satisfactory protocols for extracting the relevant measurements. If shape data are to be analyzed efficiently, modern multivariate statistics must be made to penetrate further into the special combination of two types of information—geometric and biological—making up morphometrics.

## ACKNOWLEDGMENTS

Preparation of this essay was partially underwritten by Research Grant DE-03610 to R. E. Moyers and Research Grant DE-05410 to F. L. Bookstein from the National Insitutes of Health. I began it as the result of a friendly challenge from Melvin Moss and Richard Skalak of Columbia University. Many of my arguments were clarified in response to comments from the Editor of this journal, an anonymous reviewer, and the late Morphometrics Study Group at the University of Michigan: Barry Chernoff, Ruth Elder, Julian Humphries, Gerry Smith, and Richard Strauss.

## APPENDIX: PROOF OF THEOREM 2

Theorem 2 states that any admissible size variable bearing an extremal loading on the difference between two mean configurations can be expressed as a distance involving no more than three landmarks. The proof of this theorem proceeded in the text as far as showing that the size variable sought may be taken as a distance between two constructed landmarks. Completion of the proof involves three geometrical lemmas about distances bearing extremal loadings and defined on configurations of four landmarks. Lemma 2 asserts that if a transect of a quadrilateral of landmarks bears an extremal loading, one of the four landmarks need not be involved in the transect. Lemma 3 asserts that

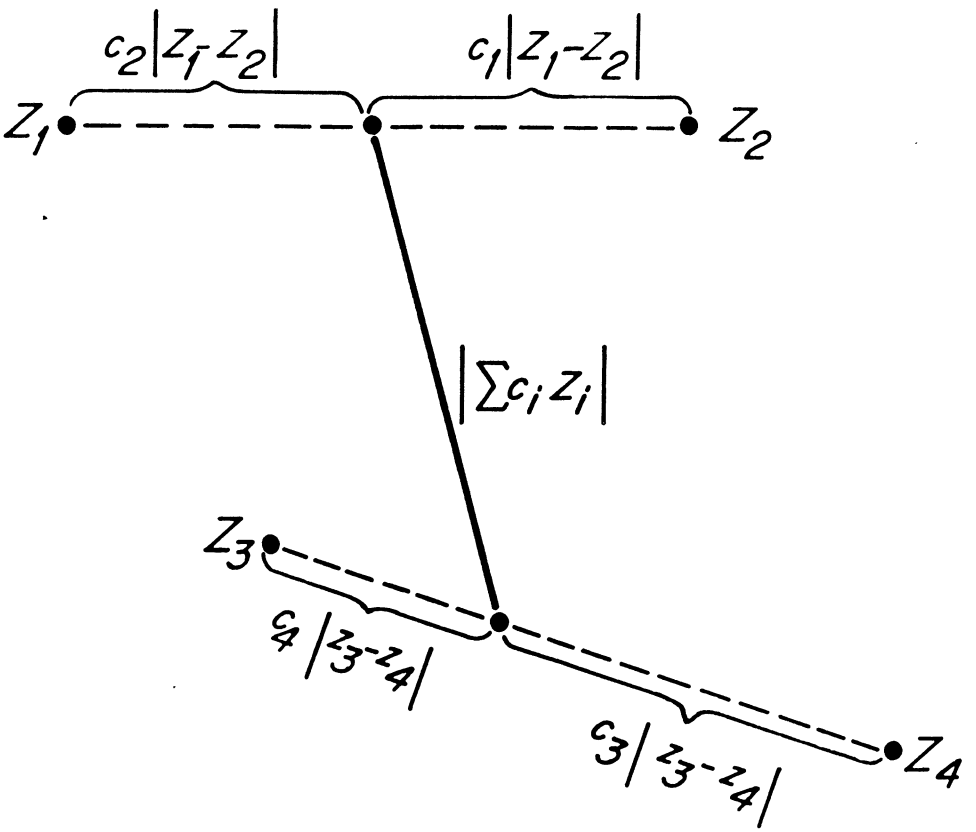


FIG. 19. *The distance* $|\sum_{i=1}^{4} c_i Z_i|$ *with* $c_1 + c_2 = -(c_3 + c_4) = 1$: *length of the segment shown. Lemma* 2 *states that the ratio* $R'/R$ *of this distance between two sample means has no proper maximum or minimum properly interior to both edges.*

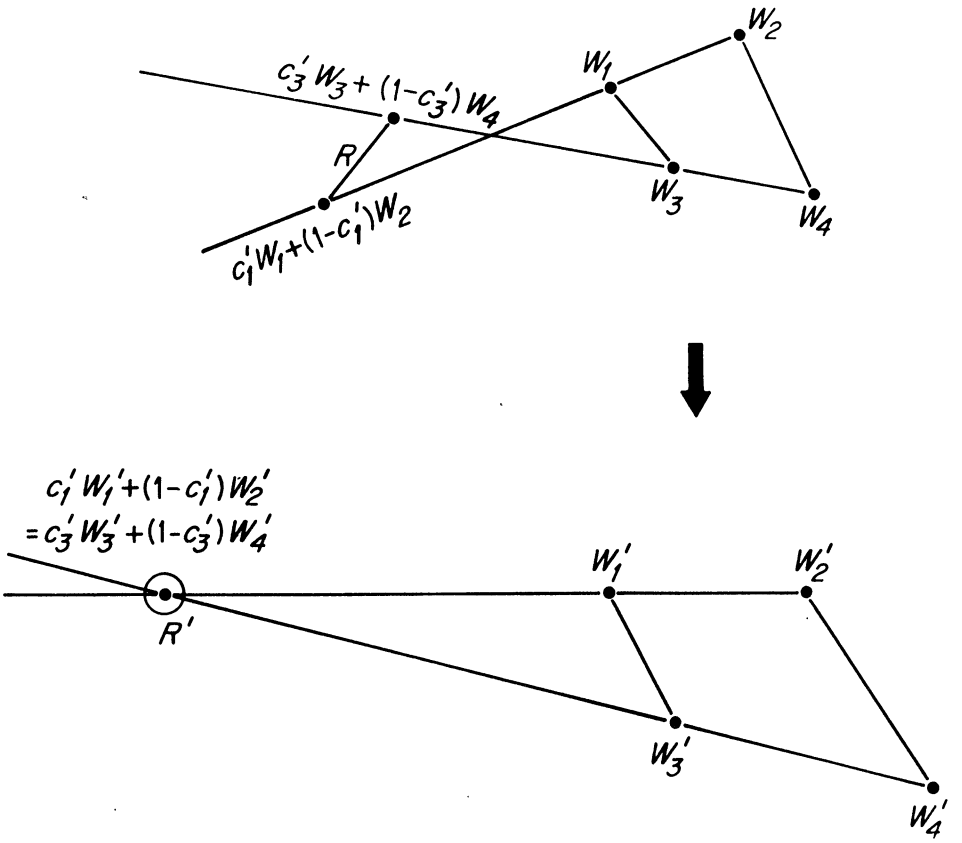


FIG. 20. *The pair* $(c_1', c_3')$ *for which* $R'$ *is smallest. The minimum of the ratio* $R'/R$ *between mean configurations is near this segment.*

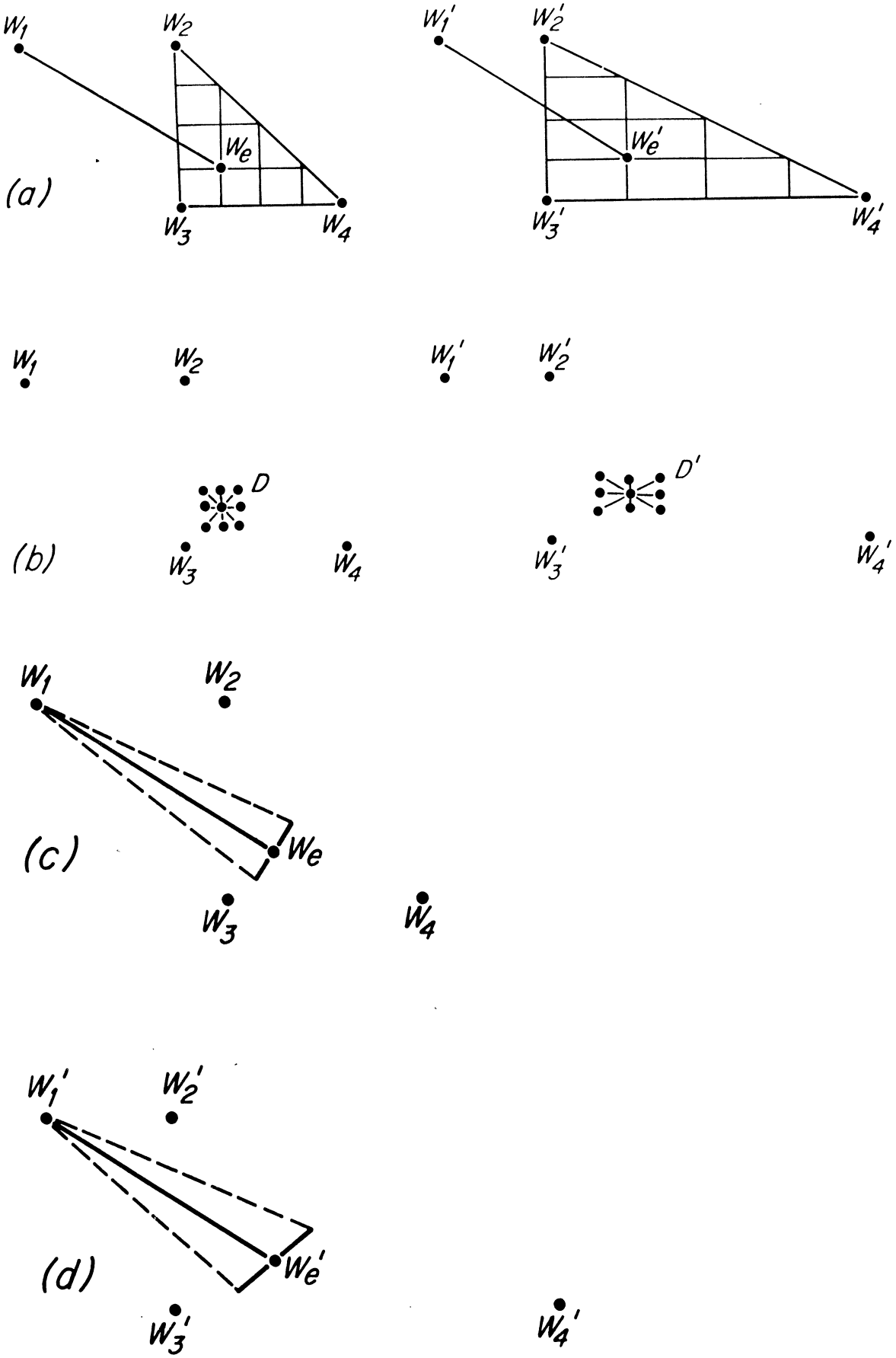
$W_1$
$W_2$
$W_e$
(a)
$W_3$
$W_4$
$W_1'$
$W_2'$
$W_e'$
$W_3'$
$W_4'$
$W_1$
$W_2$
$W_1'$
$W_2'$
D
D'
(b)
$W_3$
$W_4$
$W_3'$
$W_4'$
$W_1$
$W_2$
(c)
$W_e$
$W_3$
$W_4$
$W_1'$
$W_2'$
(d)
$W_e'$
$W_3'$
$W_4'$

FIG. 21a–d.

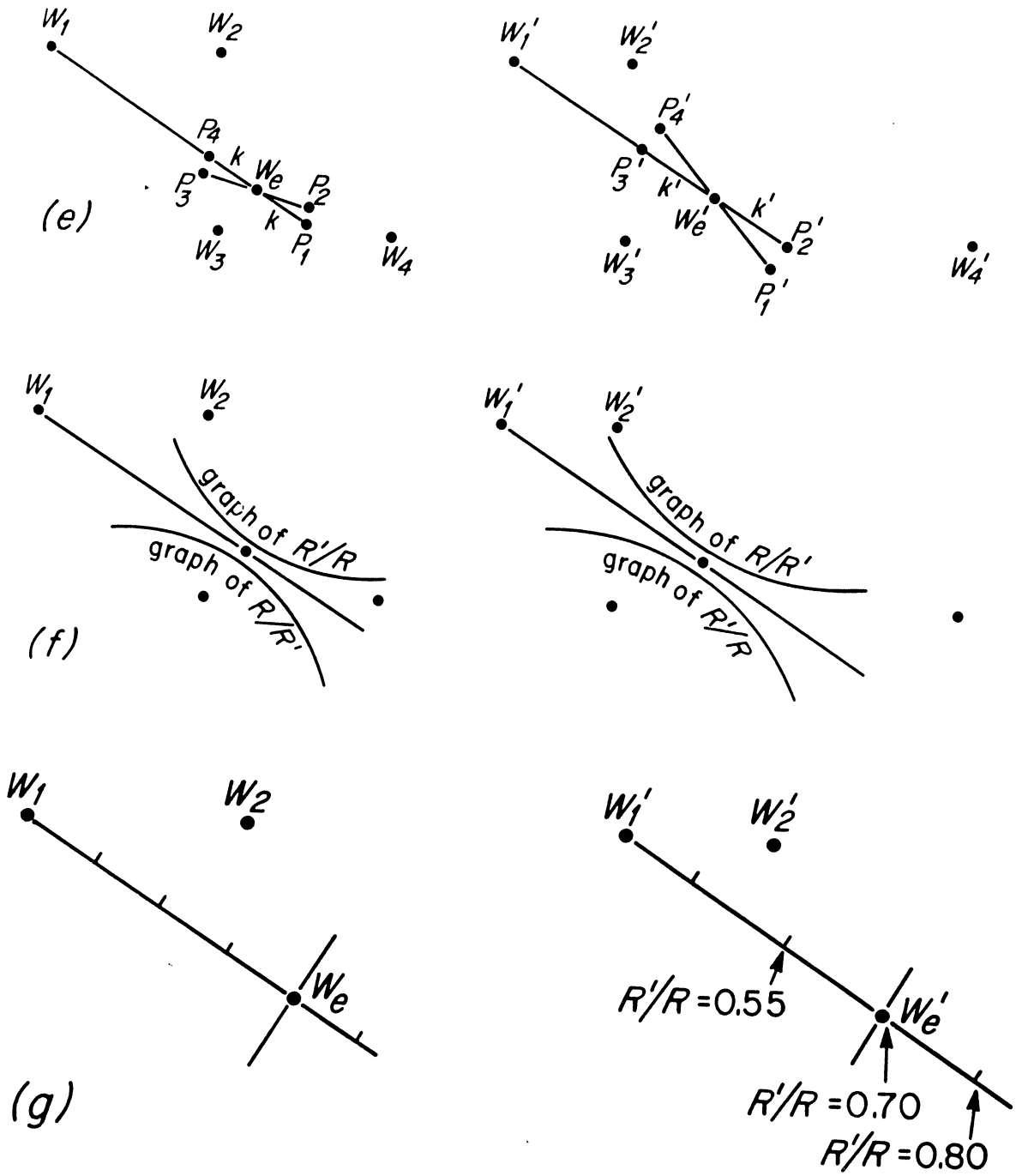


FIG. 21. *Diagrams for the proof of Lemma* 3. (a) *The distance which the Lemma states cannot be a proper extremum for the ratio (squared loading) $R'/R$ between mean configurations: join of one landmark $W_1$ to a proper linear combination $W_e = \sum_{i=2}^{4} c_i W_i$ of three others. The grids indicate the affine transformation $T$ relating the two triangles.* (b) *Consider a set of small displacements $D$ connecting the point $W_e$ to neighbors $\sum_{i=2}^{4} c_i W_i$, and the homologous segments $D'$ connecting $W_e'$ to $\sum_{i=2}^{4} c_i W_i'$.* (c) *Perturbation of the point $W_e$ by displacements $D$ perpendicular to the segment $W_1 W_e$ does not alter the length of that segment.* (d) *If the homologues $D'$ of these perpendicular increments $D$ were oblique to the segment $W_1' W_e'$, the ratio $R'/R$ of their lengths between the images would bear a gradient in the vicinity of $W_e$, contradicting the optimality posited by the Lemma.* (e) *Whenever displacements $D$ parallel to the optimal segment in the lefthand form do not map into displacements $D'$ parallel to the homologue of the optimal segment in the right-hand, primed configuration, define points $P_1 \cdots P_4$ by distances $k$, $k'$ as shown.* (f) *The ratio $R'/R$ of distances along $W_1 P_1$ is a convex upward function of $k'$; but then so must be the ratio $R/R'$ of distances along $W_1' P_2'$ as a function of $k'$. Hence, if $P_1 \neq P_2$, the ratio $R'/R$ may have only a saddle point, not an extremum, at $W_e$.* (g) *If the increment in length of $W_1 W_e$ is not the same as the increment in length of little segments $D$ in its direction around $W_e$, then the ratio $R'/R$ has a nonzero gradient at $W_e$. Hence* (d, e, f, g), *the landmark $W_1$ is displaced according to the same affine transformation governing the reconfiguration of $\Delta W_2 W_3 W_4$, so that the distance having the extremal loading may be specified using at most three landmarks, as the Lemma asserts.*

if a line joining one landmark to a weighted average of three others bears an extremal loading, one of the three in the average is redundant. Lemma 4 states that in the statement of Lemma 3, the word "three" may be replaced by "many."

LEMMA 2. *Let $W_1$, $W_2$, $W_3$, $W_4$ be four landmark centroids such that the segments $W_1W_2$, $W_3W_4$ are not the interior diagonals of their quadrilateral in two mean configurations. Consider the set of distances joining a point of the segment $W_1W_2$ to a point of the segment $W_3W_4$ (Figure 19): that is, a distance $\sum_{i=1}^{4} c_iW_i$ with $c_1 + c_2 = -(c_3 + c_4) = 1$ and all $c_i > 0$. The ratio $R'/R$ of this squared distance between the mean configurations has no proper extrema within this set of segments.*

PROOF. Under the assumptions stated, the ratio $R'/R$ of the squares of homologous distances between the mean configurations is a ratio of two quadratic forms in two variables $c_1$ and $c_3$. As such, it has three stationary points: a maximum, a minimum, and an intermediate saddle point that is not an extremum at all.

But there is a pair $(c_1', c_3')$ for which the numerator $R'$ of this ratio is nearly zero: the parameter pair corresponding to the intersection of the lines $W_1'W_2'$, $W_3'W_4'$ in Figure 20. (The value of $R'$ for this pair is not exactly zero. As it is the population mean square of the corresponding distances observed in the actual data, it is positive but small.) Hence the minimum of the ratio $R'/R$ is achieved nearby, outside the biometrically usable range of segments. This inference is blocked only if $R$ is also zero at the pair $(c_1', c_3')$—but in that case the transformation is affine over all four points, so that the extremum of distance ratio is not proper.

Likewise, for another pair $(c_1, c_3)$ corresponding to the intersection of the lines $W_1W_2$, $W_3W_4$, the distance $R$ in the denominator is nearly zero and the ratio $R'/R$ approaches its maximum (a very large value). But this point is likewise outside the biometrically useful range.

There are no remaining extremal points to be had, and hence no pair $(c_1, c_3)$ corresponding to a distance along a segment inside the quadrilateral can yield a proper (nondegenerate) optimum for the ratio of the $R$'s, that is, an extremal loading on the mean shape change, even if it corresponds to the third stationary point of the ratio.

LEMMA 3. *Suppose that landmark $W_1$ is outside the triangle $\Delta W_2W_3W_4$ in two mean configurations, and that for three non-negative parameters $c_2$, $c_3$, $c_4$ with $\sum_{i=2}^{4} c_i = 1$, the distance $|W_1 - \sum_{i=2}^{4} c_iW_i|$ yields an extremal ratio between the mean configurations. Then either (i) one of the c's is zero, or (ii) the same extremum can be achieved with one of the c's equal to zero.*

PROOF. Consider the transformation $T$ (Figure 21a) taking the interior of the triangle $W_2W_3W_4$ homologously onto the interior of $W_2'W_3'W_4'$. This is the map taking $\sum_{i=2}^{4} c_iW_i$ to $\sum_{i=2}^{4} c_iW_i'$ for $\sum_{i=2}^{4} c_i = 1$—the same uniform linear transformation of triangles we have been using all along. Write $W_e$, $W_e' = \sum_{i=2}^{4} c_iW_i$, $\sum_{i=2}^{4} c_iW_i'$ for the $c$'s that yield the distance bearing the extremal ratio.

Consider little displacements $D$ joining $W_e$ to its neighbors inside $\Delta W_1W_2W_3$ in the lefthand form (Figure 21b). They represent variation of the coefficients $c$ about their optimal values, those used for $W_e$, and thus correspond (by the transformation $T$) to perturbations $D'$ around $W_e'$. To segments $D$ perpendicular to the optimal distance on the left (Figure 21c) must correspond homologues perpendicular to the optimal segment on the right, else (Figure 21d) there would be a gradient of the distance ratio between the images along this direction, contradicting the assumption that the segment drawn bears an extremum for that ratio. Likewise (Figure 21e), corresponding to little line elements *along* the optimal segment $W_1W_e$ on the left, $T$ must yield little segments aligned with the extremum $W_1'W_e'$ on the right, for otherwise (Figure 21f), there would be a saddle-point of the ratio between corresponding distances. Hence the distance $W_1W_e$ is aligned with a principal direction of $T$. (It is in these two steps that the assumption about the position of $W_1$ entered. The points $W_1$ and $W_e$ must be distinct if the segment joining them is to have a direction.)

Finally, the increments in length measured all the way from $W_1$ must be at the same rate as those determined by $T$ in the immediate vicinity of $W_e$ (Figure 21g), else, once again, the ratio of corresponding lengths from $W_1$ between forms would bear a gradient at $W_e$.

Thus the point $W_1$ must be transformed according to the same homogeneous affine transformation $T$ that applies to the triangle $\Delta W_2W_3W_4$ of the remaining landmarks, and the distance we are considering is the ordinary principal strain of that homogeneous transformation. But we already know that the principal distance ratios of such transformations may be realized along transects from one vertex through points of the edges opposite, distances which involve only three landmarks.

Hence either the point described by the $c$'s is not interior to this triangle—that is, one of the coefficients $c_i$ is zero—or else the transformation of quadrilaterals is the uniform strain $T$, in which case one of the four landmarks is superfluous in the formula for the optimal distance ratio. This is what was to be proved.

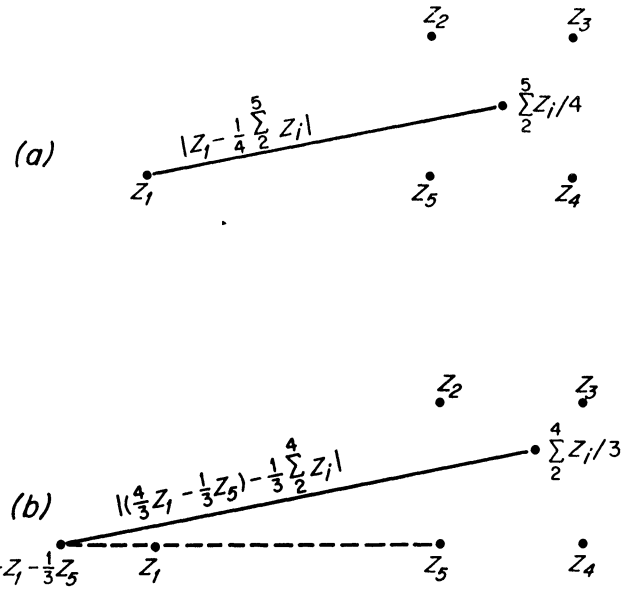


FIG. 22. *Reducing sets of more than four landmarks to four.* (a) *A distance* $|W_1 - \sum_{i=2}^{5} \frac{1}{4}W_i|$. (b) *Dividing through by* $1 - \frac{1}{4}$ *yields the statistically equivalent distance measure* $|[\frac{4}{3}W_1 - \frac{1}{3}W_5] - \sum_{i=2}^{4} \frac{1}{3}W_i|$, *meeting the requirements of Lemma* 3.

LEMMA 4. *The same is true for an optimal distance of formula* $|W_1 - \sum_{i=2}^{J} c_i W_i|$, $\sum_{i=2}^{J} c_i = 1$, *for any* $J > 4$.

PROOF. Apply Lemma 3 to the expression

$$\left(W_1 - \sum_{i=5}^{J} c_i W_i\right) - \sum_{i=2}^{4} c_i W_i$$

after its coefficients have been normalized by division by $1 - \sum_{i=5}^{J} c_i = \sum_{i=2}^{4} c_i$ (see Figure 22).

PROOF OF THEOREM 2. Let the segment $\sum c_i W_i$ whose distance ratio is optimal be rewritten in the form

$$\sum_{c_i>0} c_i W_i - \sum_{c_i<0} -c_i W_i$$

with coefficients normalized so that $\sum_{c_i>0} c_i = \sum_{c_i<0} -c_i = 1$ and with nonoverlapping convex hulls for the two sets of $W_i$.

Treating $\sum_{c_i>0} c_i W_i$ as a single constructed landmark, apply Lemma 3 or Lemma 4 as many times as necessary to eliminate all but two terms of the second sum. The assumption regarding admissibility guarantees that $\sum_{c_i>0} c_i W_i$ will always be outside the polygon formed by any terms of the second sum.

Likewise, treating the two remaining terms of the second sum as a single constructed landmark, apply Lemma 3 or Lemma 4 as many times as required to eliminate all but two terms of the first sum.

Finally, apply Lemma 2 to the two pairs of terms that remain, so as to eliminate one landmark of the four. The assumption regarding admissibility guarantees that the condition of the Lemma applies. In the expression for the optimal distance which remains, there will appear at most three landmarks, as was to be proved.

## REFERENCES


BENSON, R. H., CHAPMAN, R. E. and SIEGEL, A. F. (1982). On the measurement of morphology and its change. *Paleobiology* **8** 328–339.

BOOKSTEIN, F. L. (1978a). *The Measurement of Biological Shape and Shape Change.* Lecture Notes in Biomathematics **24**. Springer, New York.

BOOKSTEIN, F. L. (1978b). Linear machinery for morphological distortion. *Comput. Biomed. Res.* **11** 435–458.

BOOKSTEIN, F. L. (1982a). Foundations of morphometrics. *Annu. Rev. Ecology Systemat.* **13** 451–470.

BOOKSTEIN, F. L. (1982b). On the cephalometrics of skeletal change. *Amer. J. Orthodont.* **82** 177–198.

BOOKSTEIN, F. L. (1983). The geometry of craniofacial growth invariants. *Amer. J. Orthodont.* **83** 221–234.

BOOKSTEIN, F. L. (1984a). A statistical method for biological shape comparisons. *J. Theoret. Biol.* **107** 475–520.

BOOKSTEIN, F. L. (1984b). Tensor biometrics for changes in cranial shape. *Ann. Human Biol.* **11** 413–437.

BOOKSTEIN, F. L. (1985). Transformations of quadrilaterals, tensor fields, and morphogenesis. In *Mathematical Essays on Growth and the Emergence of Form* (P. L. Antonelli, ed.) 221–265. Univ. of Alberta Press, Edmonton.

BOOKSTEIN, F. L. (1986). Soft modeling and the measurement of biological shape. In *Theoretical Empiricism: a General Rationale for Scientific Model-Building* (H. Wold, ed.). Paragon Press, in press.

BOOKSTEIN, F. L, CHERNOFF, B., ELDER, R., HUMPHRIES, J., SMITH, G. and STRAUSS, R. (1985). *Morphometrics in Evolutionary Biology. The Geometry of Size and Shape Change, with Examples from Fishes.* Philadelphia, Academy of Natural Sciences of Philadelphia.

BOOKSTEIN, F. L., LOHMANN, G. P. and SCHWEITZER, P. (1986). Landmarks from eigenshapes: Ecophenotypy in *Globorotalia truncatulinoides*, revisited. Unpublished manuscript.

CHAYES, F. (1971). *Ratio Correlation.* Univ. of Chicago Press.

GOODALL, C. R. (1983). The statistical analysis of growth in two dimensions. Doctoral dissertation, Department of Statistics, Harvard Univ.

HOFER, H. (1954). Die cranio-cerebrale Topographie bei den Affen und ihrer Bedeutung für die menschliche Schädelform. *Homo* **5** 52–72.

HOFER, H. (1965). Die morphologische Analyse des Schädels des Menschen. *Menschliche Abstammungslehre* (G. Heberer, ed.) 145–226. Gustav Fischer, Stuttgart.

LOHMANN, G. P. (1983). Eigenshape analysis of microfossils: A general morphometric procedure for describing changes in shape. *Math. Geol.* **15** 659–672.

MOSIMANN, J. E. (1970). Size allometry: Size and shape variables with characterizations of the lognormal and generalized gamma distributions. *J. Amer. Statist. Assoc.* **65** 930–948.

MOSIMANN, J. E. and JAMES, F. C. (1979). New statistical methods for allometry with application to Florida red-winged blackbirds. *Evolution* **33** 444–459.

OXNARD, C. (1984). *The Order of Man.* Yale Univ. Press, New Haven, Conn.

REYMENT, R. A., BLACKITH, R. E. and CAMPBELL, N. A. (1984). *Multivariate Morphometrics.* 2nd ed. Academic, New York.

RIOLO, M. L., MOYERS, R. E., MCNAMARA, J. A. and HUNTER, W. S. (1974). *An Atlas of Craniofacial Growth.* Monograph No. 2, Craniofacial Growth Series, Center for Human Growth and Development, Univ. of Michigan.

SAMPSON, P. D. (1981). Dental arch shape: a statistical analysis using conic sections. *Amer. J. Orthodont.* **79** 535–548.

SERRA, J. P. (1982). *Image Analysis and Mathematical Morphology.* Academic, New York.

SIEGEL, A. F. and BENSON, R. H. (1982). A robust comparison of biological shapes. *Biometrics* **38** 341–350.

SNEATH, P. H. A. (1967). Trend-surface analysis of transformation grids. *J. Zool. (London)* **151** 65–88.

SPRENT, P. (1972). The mathematics of size and shape. *Biometrics* **28** 23–37.

STRAUSS, R. E. and BOOKSTEIN, F. L. (1982). The truss: body form reconstruction in morphometrics. *Systematic Zool.* **31** 113–135.

THOMPSON, D'A. W. (1961 [1917, 1942]). *On Growth and Form* (J. T. Bonner, ed. ab.). University Press, Cambridge.

TOBLER, W. R. (1978). The comparison of plane forms. *Geograph. Anal.* **10** 154–162.

# Comment

**David G. Kendall**

The reading of this paper gave me great pleasure. I wish to congratulate the Editor for throwing open the windows and allowing a welcome draught of fresh air to enter the hitherto hermetically sealed publications department of the IMS. I hope this wise policy will be a pattern for the future.

Of course Bookstein mentions D'Arcy Thompson's work in zoology, and to this name I should like to add that of F. O. Bower (1930), whose book "Size and Form in Plants" played a similar role in botany. Bower gave me a copy of this in 1943, and expressed his conviction that mathematicians would eventually find it food for thought. We can now see that he was right.

Despite their name, interdisciplinary studies have to begin within the author's own discipline, or one adjacent to it, and so it takes some time for such work to diffuse into other disciplines in which it has potential relevance. Thus it is not surprising that most of Bookstein's work was new to me, and I do not think he has even heard of mine, originating as it did in response to requests from archaeologists and astronomers. A summary of this therefore seems in order.

I have been concerned with constructing and adapting for statistical purposes the natural representation space for the *shape* of a labeled set of $k$ points in $m$ dimensions. This is only interesting if $k \geq 3$, and the $m$ values of practical importance are 1, 2, and 3, together with some higher values like 15 for cosmologists interested in space-time foams and so on. We exclude the situation in which the points totally coincide. Obviously the points could be Bookstein's landmarks, but they could also be archaeological sites or quasars.

"Shape is what remains when location, size, and rotational effects are filtered out," so it is natural to begin by taking the centroid as origin and changing the scale to make $\sum \sum x_{ij}^2 = 1$ (this is equivalent to adopting Bookstein's preferred size measure up to a constant factor). We now have what I call a *preshape*, which we can think of as a generic point on a unit sphere of dimension $m(k-1)-1$, the "points" of this sphere being labeled by $m \times (k-1)$ matrices. The sphere of preshapes comes with a natural metric topology based on geodesic arc length, and a big group of symmetries generated by a subgroup of the orthogonal group acting on the *right* which includes the relabeling symmetries, together with the reflexion symmetry. To get to what I have called the *shape space* $\sum_m^k$ we quotient out the rotation group acting on the *left*. Natural quotient constructions endow this space with a metric topology and its own rich group of symmetries. This then is the stage on which shape statistical transactions are acted out, and we need to become familiar with it.

The quotient mapping from preshapes (points on the unit sphere of dimension $m(k-1)-1$ to shapes (points in $\sum_m^k$) is what is called a submersion when $m = 1$ or 2; that is, it is a smooth mapping "onto" whose Jacobian has everywhere a rank equal to the dimension of $\sum_m^k$. When $m$ is greater than or equal to 3, we still have a submersion if we exclude a *singular set* in $\sum_m^k$ (and its preimage on the sphere). As will be seen below, the existence of this singular set when $m \geq 3$ adds considerable interest (and difficulty) to the discussion of these higher dimensional situations.

A great deal of my time in the last 8 years has been

*David G. Kendall is Professor at the Statistical Laboratory, University of Cambridge, 16 Mill Lane, Cambridge CB2 1SB, England.*