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.

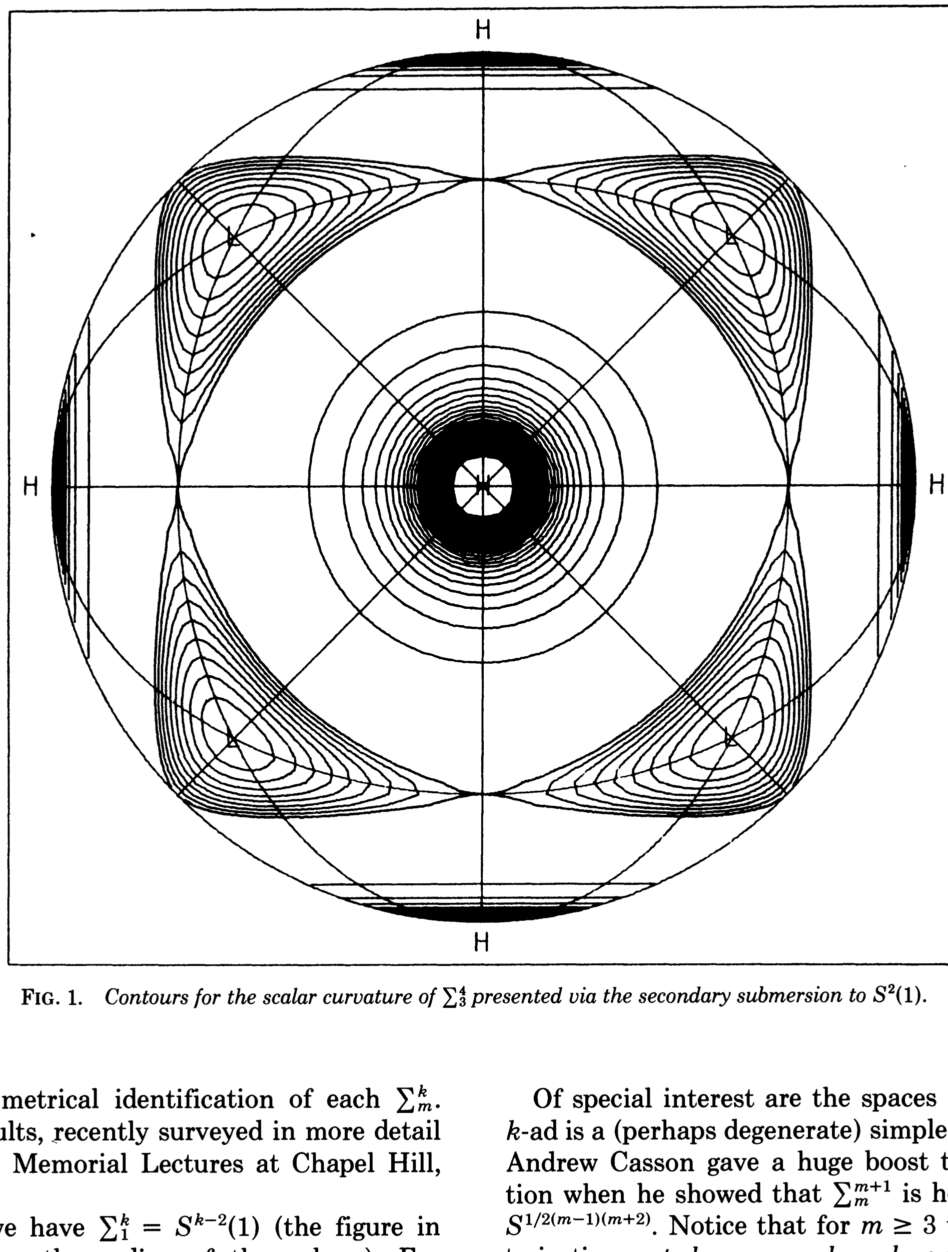

FIG. 1. *Contours for the scalar curvature of $\sum_3^4$ presented via the secondary submersion to $S^2(1)$.*

devoted to the metrical identification of each $\sum_m^k$. Some of the results, recently surveyed in more detail in my Hotelling Memorial Lectures at Chapel Hill, are as follows.

For $m = 1$, we have $\sum_1^k = S^{k-2}(1)$ (the figure in parentheses being the radius of the sphere). For $m = 2$ we have $\sum_2^k = \mathbb{C}P^{k-2}(4)$ (the figure in parentheses being a measure of curvature). Note that $\mathbb{C}$ means *complex* and P means *projective*, so this is the $(2k - 4)$-dimensional complex projective space much used by algebraic geometers. Having already gotten rid of location and scale, we can code the remaining configuration by $k - 1$ complex numbers, and because we do not care about rotations, only the *ratios* $z_1:z_2: \cdots : z_{k-1}$ are significant, whence the projective context. The metrical aspects of the identification (developed in Kendall, 1984) are somewhat more technical. In a few simple cases these sophistications can be bypassed. Thus $\mathbb{C}P^0 = pt$, and $\mathbb{C}P^1(4) = S^2(½)$; I have given an elementary proof that $\sum_2^3 = S^2(½)$ (Kendall, 1985).

Of special interest are the spaces $\sum_m^{m+1}$; here each k-ad is a (perhaps degenerate) simplex, and my friend Andrew Casson gave a huge boost to the investigation when he showed that $\sum_m^{m+1}$ is homeomorphic to $S^{1/2(m-1)(m+2)}$. Notice that for $m \geq 3$ this is a characterization *up to homeomorphy only*, and does not identify the metrical structure. Casson also showed that if the $m + 1$ points lie in more (say $m + r$) dimensions, then $\sum_{m+r}^{m+1}$ is homemorphic to the corresponding hemisphere.

We are left with the problem of describing the spaces $\sum_m^k$ when

$$(1) \qquad m \geq 3, \quad k \geq m + 2.$$

I have recently been able to show that (i) each of these spaces is a finite CW complex, i.e., it can be regarded as a finite collection of cells of various dimensions, the boundary of each cell being appropriately identified with a subset of the union of the cells of lower dimension; (ii) no one of them is even a topological manifold, i.e., each one contains some points with

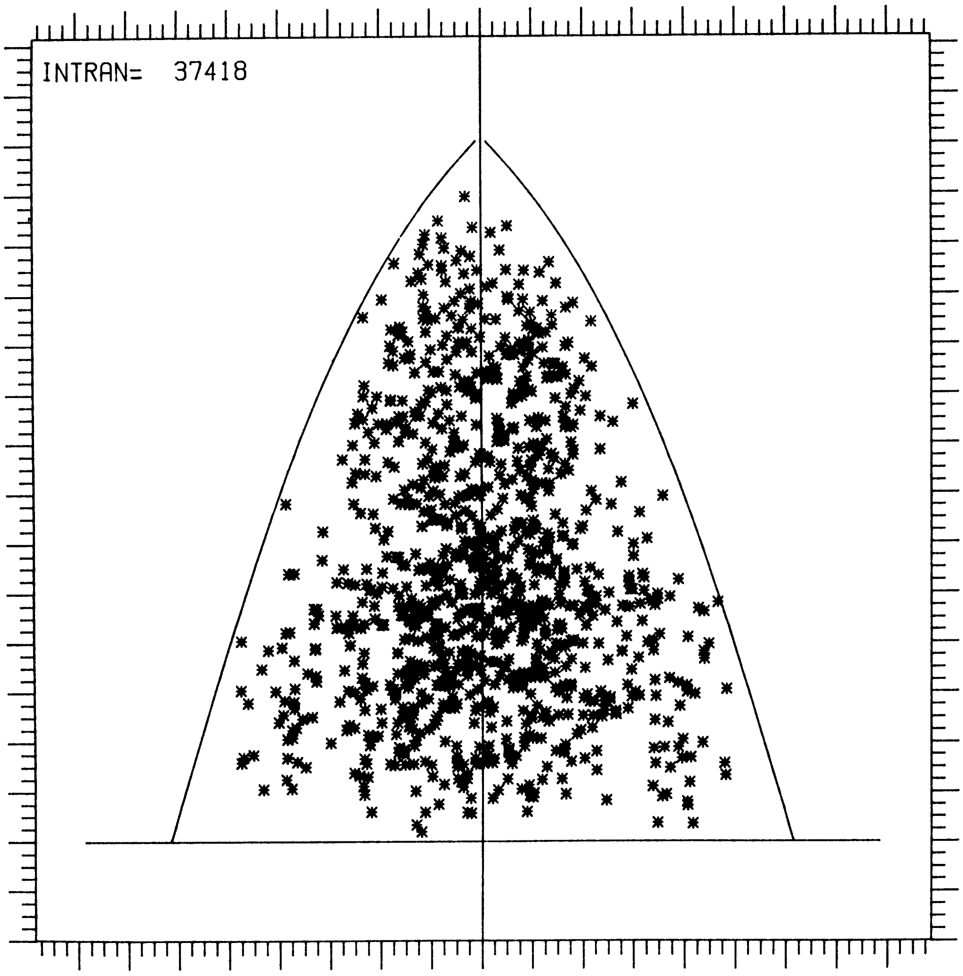

FIG. 2. *The shapes of* 1001 *random gaussian tetrahedra presented via the secondary submersion to* $S^2(1)$.

topologically non-Euclidean neighborhoods; (iii) each of them has torsion in its integral homology; and (iv) for $(k, m) \neq (k', m')$, and m, m', k, and k' satisfying the above inequalities (1), the spaces $\sum_m^k$ and $\sum_{m'}^{k'}$ are distinct even at the homology level, i.e., even in the coarsest topological classification.

Notice that the bad behavior sets in as soon as the k points lie in a space of dimension $m \geq 3$, yet $m = 3$ is obviously a situation of great practical importance.

Notice also that either one of (ii) and (iii) above is enough to show that none of the shape spaces for which (1) holds can be even remotely like a sphere, so that there can be no question of an extension of Casson's theorem.

Turning now to metrical considerations, we have quite generally a dense open subset of $\sum_m^k$ which has a natural smooth riemannian structure. The complementary (singular) set is void for $m = 1$ and 2. This situation can best be investigated by exploiting O'Neill's theory of riemannian submersions. I have investigated $\sum_3^4$ (which is homeomorphic to S^5) in especial detail; the scalar curvature explodes to $+\infty$ as we approach any point on the singular set (which can here be characterised as the seat of the collinear tetrads in $\mathbb{R}^3$). This is a startling confirmation of the metrical asphericity of $\sum_3^4$.

Collinearity studies first become interesting when $k = 3$, and for $m = 2$ we then have a nice planar representation for $\sum_2^3$ if we are content to omit one shape point (say $P_1 = P_2 \neq P_3$), which we can think of as lying at infinity. This planar representation is very useful for displaying real or simulated data. If the points (P_1, P_2, P_3) form an independent and identically distributed sample from a two-dimensional generating law μ, we get an induced shape measure μ^* on $\sum_2^3$, and the density $d\mu^*/d\gamma^*$ (where γ is the symmetrical gaussian generator) can be displayed by plotting contours on the planar representation. Many illustrations of this technique can be found (Kendall, 1984).

Multiple collinearities for plane configurations must be studied on the complex manifold $\sum_2^k = \mathbb{C}P^{k-2}(4)$ with $2k - 4$ real dimensions. The links with Bookstein's work are now very close; mostly his calculations appear to take place in a tangent space of $\sum_2^k$. The

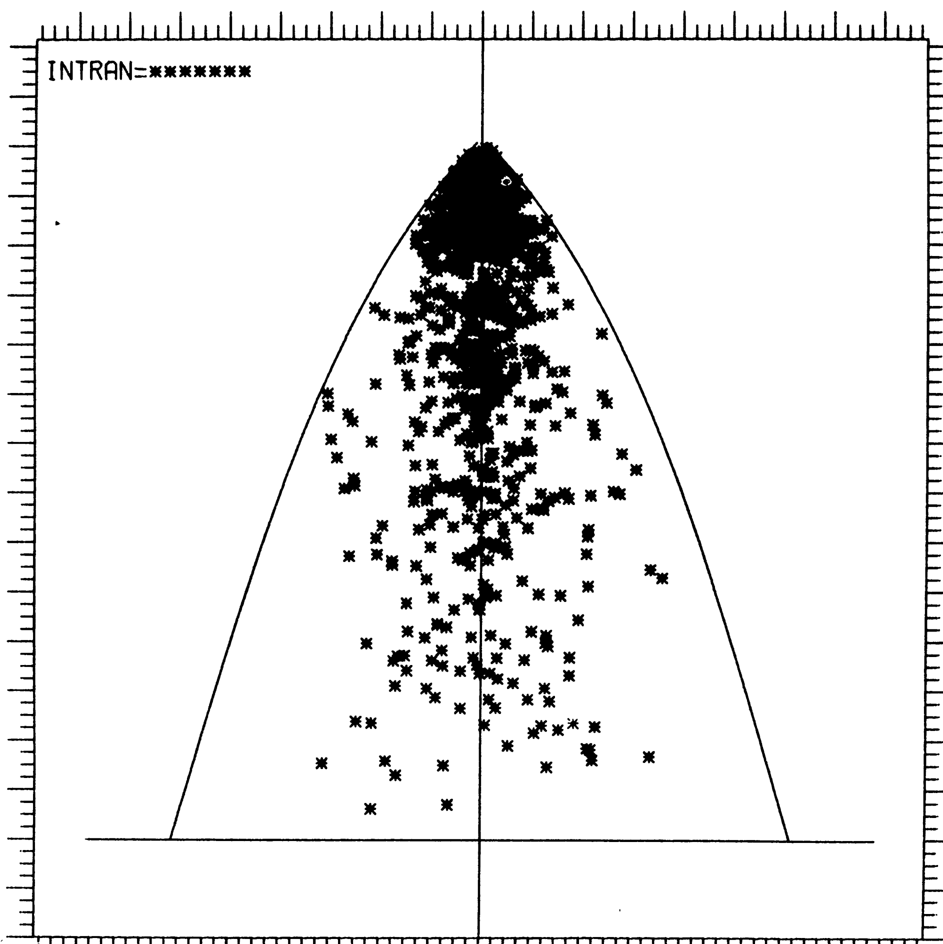

FIG. 3. *The shapes of all* 1001 *tetrads drawn from the* 14 *major objects in the solar system on* 1 *January* 1980 *presented via the secondary submersion to* $S^2(1)$.

exact multiple collinearities here constitute a $(k - 2)$-dimensional subspace which can be identified with *real* projective space $\mathbb{R}P^{k-2}$. If σ is an arbitrary point of Σ_2^k, then we write $L(\sigma)$ for the shortest geodesic distance from σ to this $\mathbb{R}P^{k-2}$; this is an attractive test statistic for assessing near collinearity, and I have found its law of distribution for *all* gaussian generators, so in the null and also in the non-null situation. This result illustrates the rewards which follow naturally from such global studies.

With increasing k and m the dimension of the shape space increases, and data presentation is less easy. We can however make use of a secondary submersion from the nonsingular part of Σ_m^k onto a spherical triangle of the unit sphere $S^{m-1}(1)$ by a singular values decomposition argument, and this is especially attractive when $m = 3$.

To illustrate this, Figure 1 shows contours for the scalar curvature of Σ_3^4, emphasising the fact that in metrical terms this topological sphere loses its sphericity. Were it metrically spherical, the scalar curvature would be constant! The secondary submersion has here been extended by symmetry to all of $S^2(1)$; in data analytic studies it is better to work with the exact range of the submersion, a spherical triangle. Thus Figure 2 shows in that manner the shapes of 1001 random tetrads in $\mathbb{R}^3$ with independent and identically distributed gaussian vertices. To read this picture note that all co-planar tetrads map onto the median vertical line, all collinear tetrads map onto the vertex at the top, and equilateral tetrads of righthanded and left-handed symmetry map to the two vertices at the bottom.

For a contrasting scene we may look at Figure 3, which shows the shapes of the 1001 tetrahedra determined by all the subsets of size 4 drawn from the 14 principal objects in the solar system as they were at the outset of 1980. We know that the solar system consists of *nearly* co-planar objects, and this fact is illustrated here by the concentration along the median line. But there is a much stronger concentration near the vertex at the top, and that illustrates the peculiar event in 1980 when most of these objects appeared to us to be located in the same corner of the sky.

(Astrologers were interested in this phenomenon!) All of these are shape theoretic features, and give examples of the way in which such presentations of the shape space can be instructive.

I hope that these remarks will help to fill out the fascinating paper by Bookstein, and will at the same time indicate the additional advantages which can be gained from global geometrical studies, in supplementation of the less exacting but equally relevant local ones.

ADDITIONAL REFERENCES

BOWER, F. O. (1930). *Size and Form in Plants.* Macmillan, London.

KENDALL, D. G. (1984). Shape manifolds, procrustean metrics, and complex projective spaces. *Bull. London Math. Soc.* **16** 81–121.

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

Comment

Noel Cressie

Professor Bookstein has ably presented a synthesis of the "size or shape variables" approach and the "deformation" approach within the field of morphometrics. He has developed a number of results which can be used to test for association between shape and size, and for shape differences between groups. These results exploit the relative spatial locations of landmarks (well defined "summary" points that are biologically apparent in each specimen under study), whence one is able to summarize the geometry of the specimen. This is clearly a desirable direction to take. Professor Bookstein has shown us the inadequacy of the first approach, which uses very little spatial information, and has given us a way of making statistical inferences within the second approach, which up to now has been essentially descriptive.

There are other ways to analyze size and shape; the approach of the so-called "mathematical morphology" school of Matheron and Serra (Serra, 1982) in France, is trying to solve a different problem. Automated cytological examinations of a smear enable rapid objective detection of deformed cells, whose frequency of occurrence is typically only a minute fraction of the total number of cells examined. The use of landmark techniques here is clearly not appropriate, although one is still interested in (perhaps more gross) deformations of cells. Here the geometry of the specimen (represented as a set) is of paramount importance; morphological transformations that take set into set do not destroy the spatial nature of the problem. After a suitable series of transformations (e.g., smoothing, convexifying, etc.) on the set, the *final* stage is to calculate real valued summaries (e.g., volume, surface area, connectivity number, etc.).

The great advantage of mathematical morphology is that it is applicable in $\mathbb{R}^n$, $n \geqq 1$, whereas Bookstein is not yet able to make the jump from $\mathbb{R}^2$ to $\mathbb{R}^3$. With this in mind, there is the notion of "the skeleton" (see Serra, 1982, Chapter XI) of an object, which may be useful in the analysis of landmark data. It can essentially be obtained (in $\mathbb{R}^2$) in the following graphic way. Imagine the object to be a bounded field of grass, and a fire is started at the same instant at every point on the boundary. The fire burns inward at constant speed, so that there is a line in the field where the fire is reached simultaneously from at least two different points on the boundary. When applying this idea to jawbones for example, it may suggest landmarks (the various "junctions" of the skeleton) and curvatures (the part of the skeleton linking the "junctions"). This may be an answer to the question of the statistical sufficiency of landmarks to describe the specimen, and at the same time of the generalizability to $\mathbb{R}^n$.

There is another possibility for generalization to higher dimensions contained in the article by Kendall (1984). He considers k-ads in m dimensions (e.g., here we have $k = 3$, $m = 2$), and constructs shape spaces and shape measures on these spaces. The mathematics is formidable, but there is the formalism for dealing with higher dimensional landmark data.

I would like to join Professor Bookstein in his concern that the null statistical model of identical uncorrelated normal perturbations at each landmark, is unrealistic. Would he comment on what happens when the bivariate Gaussian distribution is correlated, perhaps differently from landmark to landmark, and also when another distribution is controlling the perturbations? (Kendall's shape measures would generate a very general type of model.) I am worried about the robustness of the approach, and I think finding exploratory ways of checking the assumptions made about these spatial data is an open and interesting problem.

Noel Cressie is Professor of Statistics, Department of Statistics, Snedecor Hall, Iowa State University, Ames, Iowa 50011.

ADDITIONAL REFERENCE

KENDALL, D. G. (1984). Shape manifolds, procrustean metrics, and complex projective spaces. *Bull. London Math. Soc.* **16** 81–121.