

SHAPE MANIFOLDS, PROCRUSTEAN METRICS, AND COMPLEX PROJECTIVE SPACES

DAVID G. KENDALL

1. Introduction.	81
2. Shape-spaces and shape-manifolds	82
3. Procrustes analysis, and the invariant (quotient) metric on Σ_2^k	87
4. Shape-measures and shape-densities	93
5. The manifold carrying the shapes of triangles	96
6. The use of the 'spherical blackboard'	100
7. The shape-density for a non-isotropic gaussian generator	106
8. Applications.	113

ABSTRACT

The shape-space Σ_m^k whose points σ represent the shapes of not totally degenerate k -ads in \mathbb{R}^m is introduced as a quotient space carrying the quotient metric. When $m = 1$, we find that $\Sigma_1^k = S^{k-2}$; when $m \geq 3$, the shape-space contains singularities. This paper deals mainly with the case $m = 2$, when the shape-space Σ_2^k can be identified with a version of $\mathbb{C}P^{k-2}$. Of special importance are the shape-measures induced on $\mathbb{C}P^{k-2}$ by any assigned diffuse law of distribution for the k vertices. We determine several such shape-measures, we resolve some of the technical problems associated with the graphic presentation and statistical analysis of empirical shape distributions, and among applications we discuss the relevance of these ideas to testing for the presence of non-accidental multiple alignments in collections of (i) neolithic stone monuments and (ii) quasars. Finally the recently introduced Ambartzumian density is examined from the present point of view, its norming constant is found, and its connexion with random Crofton polygons is established.

1. Introduction

As a background to the statistical study of random alignments in archaeology and related fields, we here select what is relevant from a general theory of "the statistics of shape"; the present account will also serve as a continuation of the author's earlier progress report [8] which dealt more particularly with "the diffusion of shape". The exposition is designed to be self-contained, and the omitted details should be readily checkable, but for the applications to random alignments which in

Received 18 April, 1983; revised 5 July, 1983.

This is an expansion of part of the text of the Whitehead Lecture delivered to the Society on 21 November, 1980.

part stimulated the work the reader is referred to Broadbent [3] and to Kendall & Kendall [9]. Some slight familiarity with these papers will be found helpful for a full understanding of the practical example chosen here as a typical application. Elaborations of the topological and differential-geometric aspects of the shape-spaces Σ_m^k ($k \geq 2, m \geq 1$) will be presented in [13] and [14], while [11] contains particulars of some other applications.

As topologists already have a theory of 'shape' (Borsuk [2]), I must apologise for using the word again with an entirely different meaning. In this paper 'shape' is used in the vulgar sense, and means what one would normally expect it to mean. Some papers of Robertson et al [18, 19, 20] are much closer to our present approach, but the thrust of their work is in another direction, even in the geometric context; nearly all our geometric constructions are different from theirs, and are motivated by statistical problems and tailored to suit statistical needs.

2. Shape-spaces and shape-manifolds

Let P_1, P_2, \dots, P_k (where k is not less than 2) be the vertices of a *labelled* k -ad of points in m dimensions, with m -vector positions summarised by $z^* = (z_i^* : i = 1, 2, \dots, k)$, and with centroid z_c^* . Degeneracies are allowed (and indeed are of considerable interest) with the sole exception that *the k -ad must not collapse to a single point*. This exception gives no trouble in the statistical applications, where it will normally correspond to an event of probability zero. (The same is true in the kinematic case, when the points are diffusing, because then we shall normally assume that $k \geq 3$ or $m \geq 2$.) When considering the configuration of the k -ad we shall ignore the effects of translations, of (proper) rotations, and of dilatations in \mathbb{R}^m , but we do attach significance to reflections. We here define 'shape' informally to be 'what is left when the differences which can be attributed to translations, rotations, and dilatations have been quotiented out'.

Because only the positions of the points relative to their centroid are of interest, it will often be convenient to work with orthogonally transformed coordinates such as

$$\begin{aligned} z_0 &= z_c^* \sqrt{k}, \\ z_j &= \{jz_{j+1}^* - (z_1^* + z_2^* + \dots + z_j^*)\} / \sqrt{(j^2 + j)} \quad (1 \leq j \leq k-1). \end{aligned} \tag{1}$$

Note that the use of the unstarred coordinates in \mathbb{R}^m is equivalent to the multiplication of the $(m \times k)$ data-matrix by a suitable *fixed* $Q \in O(k)$, acting on the right, *this being done once and for all*. Here by 'suitable' we mean that all the elements in the first column of Q are to be equal to $1/\sqrt{k}$. From any one such Q we obtain all others if we multiply it on the right by $\text{diag}(1, \tilde{Q})$, where $\tilde{Q} \in O(k-1)$. It has constantly to be borne in mind that there is nothing specially privileged about this particular choice (1) of such coordinates, and so we shall often pause (as immediately below) to check that our constructs are independent of it. However the segregation of z_0 is canonical, and it reflects our lack of interest in translations. We shall therefore frequently ignore z_0 and write $z = (z_1, z_2, \dots, z_{k-1})$, so that z will have only $k-1$ m -vector components, whereas z^* had k such components. A construction will then be canonical if and only if it does not depend on \tilde{Q} .

We can now give a precise coordinate-free definition of *shape* by saying that it is the representing point (i.e. equivalence-class) in the quotient space Σ_m^k defined by

$$\Sigma_m^k = \{(\mathbb{R}^m)^{k-1} \setminus 0\} / \text{Sim}, \quad (2)$$

where Sim denotes the group of similarities generated by the rotations and dilatations; the 'zero' is excluded here because we have dismissed totally degenerate k -ads from consideration. The totally degenerate k -ads are omitted because from one point of view they have no shape, and from another they 'almost' have every shape; in the kinematic problem (which is not our primary concern here) a stage does however arrive at which it is convenient to introduce an ideal shape-point representing the totally degenerate k -ads to serve as an entrance boundary.

We also need a definition of *size*, and for this we choose $\|z\|$, where

$$\|z\|^2 = \sum_1^{k-1} |z_j|^2 = \sum_1^k |z_i^*|^2 - k|z_c^*|^2 = \sum_1^k |z_i^* - z_c^*|^2, \quad (3)$$

and where $|\cdot| : z_i \mapsto |z_i|$ is the usual quadratic norm in \mathbb{R}^m . The last expression makes it clear that $\|z\|$ is canonical, i.e. is unaffected by our choice of \bar{Q} . We also use $\|\cdot\|$ to define a metric topology in $\mathbb{R}^{m(k-1)}$.

The quotient space Σ_m^k with the quotient topology will be called *the shape-space of order (k, m)* . For many purposes it is convenient to quotient out the dilatation group first (so replacing z by $z/\|z\|$ and then applying the rotation group $\text{SO}(m)$ afterwards). If we do this, we find that

$$\Sigma_m^k = S_m^k / \text{SO}(m), \quad (4)$$

where S_m^k denotes the unit sphere $S^{m(k-1)-1}$ with its customary metric topology, and where the rotations $R \in \text{SO}(m)$ act from the left on the columns of the $m \times (k-1)$ -matrices W representing the points W of S_m^k . Notice that $\text{Tr}(WW') = 1$. It will be convenient to call W a *pre-shape*; the *shape* itself, $\sigma = [W]$, is the corresponding equivalence class modulo $\text{SO}(m)$. The open sets in Σ_m^k are thus the images of the $\text{SO}(m)$ -saturated open sets in S_m^k , and the projection $p : W \mapsto [W]$ is continuous, so that Σ_m^k is always compact, and is connected unless $k = 2$ and $m = 1$.

An elementary group-theoretical argument now shows that these open sets are exactly the open sets determined by the quotient metric ρ defined by

$$\rho(pW, pW_0) = \inf_R d(RW, W_0), \quad (5)$$

where the infimum is taken for all $R \in \text{SO}(m)$, and where we have written

$$\begin{aligned} d(W, W_0) &= 2 \arcsin \frac{1}{2} \|W - W_0\|, \\ &= \arccos \text{Tr}(WW_0') \quad (0 \leq d \leq \pi), \end{aligned} \quad (6)$$

so that d is the great-circle metric on S_m^k .

The metric d for S_m^k is topologically equivalent to the norm-metric inherited from $\mathbb{R}^{m(k-1)}$, and as the last of the formulae (6) makes clear, it is canonical in the sense

that it does not depend on the choice of coordinates (i.e. it does not depend on \tilde{Q}). In fact the orthogonal group $O(k-1)$ (acting on the pre-shapes W from the right) is a group of isometries for S_m^k with respect to both the d -metric and the norm-metric. It commutes with $SO(m)$ (which acts on the left), and so it induces a group G of isometries on Σ_m^k . The intervention of \tilde{Q} in our preferred choice of coordinates is thus equivalent to the use, *once and for all*, of a fixed isometry to relabel the points σ of Σ_m^k .

Notice that we shall have

$$\begin{aligned} \rho(pW, pW_0) &= \inf_R \arccos \operatorname{Tr}(RWW'_0), \\ &= \arccos \sup_R \operatorname{Tr}(RWW'_0), \end{aligned} \tag{7}$$

and that the ρ -metric is bounded by π . We shall see that this upper bound to ρ can be improved if $m \geq 2$.

When m is even, we can augment the group G of symmetries to

$$G^+ = G \otimes \{1, \tau\}, \tag{8}$$

where τ is represented by multiplication *on the left* by $\operatorname{diag}(1, 1, \dots, 1, -1)$. When m is odd, however, τ is equivalent modulo $SO(m)$ to -1 , which can be identified with an element of G . We shall call the action of τ on Σ_m^k *conjugation*.

The shape-spaces Σ_m^{m+1} , which we shall call *simplicial* shape-spaces, are of special importance, because when $k = m+1$ then a k -ad determines a labelled simplex in \mathbb{R}^m with the maximal number of vertices. We now show that the shape-spaces Σ_m^k for which $k < m+1$ (called *over-dimensional* shape-spaces) are in a sense trivial. For let W be a pre-shape for such k and m . By use of $SO(m)$ we can arrange (modulo $SO(m)$) that the bottom $m-(k-1)$ rows of W consist entirely of zeros, so that the pre-shapes W for Σ_m^k can be identified with pre-shapes U for Σ_{k-1}^k supplemented by one or more rows of zeros. The presence of at least one row of zeros further ensures that each member of the pair $(U, \tau U)$ of pre-shapes for Σ_{k-1}^k corresponds to the same shape for Σ_m^k . Accordingly we have

$$\Sigma_m^k = \Sigma_{k-1}^k / \{1, \tau\} \quad (k < m+1), \tag{9}$$

or, in words, Σ_m^k for $k < m+1$ is obtainable from Σ_{k-1}^k by a process of *deconjugation*. Now A. J. Casson has proved (private communication; see [13] for an account of this) that each simplicial shape-space is *homeomorphic* to a sphere, and that the action of τ is then such that it interchanges two complementary hemispheres. Thus *each over-dimensional shape-space is homeomorphic to a hemisphere*, and therefore is contractible.

It must be emphasised that these results associated with Casson's theorem are concerned strictly with homeomorphy; isometry is not claimed, and the fact that (for $m \geq 3$) the simplicial shape-spaces are not isometric with spheres is a principal theme of [14].

The shape-spaces Σ_m^k with $k \geq m+2 \geq 5$ will be called *anomalous*; they are studied in [13], where their pathologies are exhibited, and where they are classified up to homology. Pathological or not, it is clear that the spaces Σ_3^k , at least, are of substantial practical importance.

The remaining shape-spaces Σ_m^k are those for which $m = 1$ or 2 , and these will be called *shape-manifolds*. (Of course Σ_2^2 is over-dimensional, and Σ_1^2 and Σ_2^3 are simplicial, but there is no further overlap in our classification.) These manifolds will be our main concern in this present paper.

The case $m = 1$ can be dealt with very simply; we see at once that

$$\Sigma_1^k = S^{k-2}. \quad (10)$$

In particular, $\Sigma_1^2 = S^0 = \{-1, 1\}$; this is the only shape-space which is not arc-wise connected. Also $\Sigma_1^3 = S^1$; in [13] it is shown that this is the only arc-wise connected shape-space which is not simply connected. These identifications are of course isometries, and so we see that $\max \rho = \pi$ when $m = 1$.

Now consider the case $m = 2$. We suppose that $k \geq 3$, because from (9) and (10) we already know that $\Sigma_2^2 = \text{pt}$. If we wish to make a study of the local geometry around a basic shape σ_0 , we can obviously take the full group G^+ of symmetries into account in choosing it, and thus an adequately general basic pre-shape will be

$$W_0 = \begin{pmatrix} \lambda, & 0, & 0, \dots, 0 \\ 0, & \sqrt{(1-\lambda^2)}, & 0, \dots, 0 \end{pmatrix}, \quad (11)$$

where $1/\sqrt{2} \leq \lambda \leq 1$, and therefore the nature of the local geometry depends at most on λ . We shall see below that we can improve on this statement.

More generally, if $k \geq m+1$, the singular-values decomposition enables us to write an arbitrary pre-shape as

$$W = R\Lambda S \quad (12)$$

where $R \in \text{SO}(m)$, $S \in \text{SO}(k-1)$, and

$$\Lambda = (D, 0), \quad D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{m-1}, \pm \lambda_m); \quad (13)$$

here the λ 's are the square-roots of the eigenvalues of WW' , and so satisfy $\sum \lambda_i^2 = 1$. It will be convenient always to choose R and S in such a way that $\lambda_{i_1} \geq \lambda_{i_2}$ for $i_1 \geq i_2$. When discussing shapes the factor R is always irrelevant, and it will be clear that relative to G^+ we can take the basic pre-shape to be

$$W_0 = (\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m), 0); \quad (14)$$

the \pm is not required here because of the symmetry τ , or indeed because in choosing the basic pre-shape we can apply an orthogonal factor to the right-hand side of $R\Lambda S$. Evidently (11) is just a special case of this. In this way we obtain a basic pre-shape depending on $m-1$ parameters, although when $k \geq m+1$ we have

$$\dim \Sigma_m^k = \frac{1}{2}(m-1)(m+2) + (k-1-m)m, \quad (15)$$

this being the number of parameters required to specify ΛS . (See also [13] for a homology proof of (15).)

However, when $m = 2$, we can make a further reduction in the study of the local geometry. This is because we can then write the pre-shape as $W = (z_1, z_2, \dots, z_{k-1})$

where here the z 's are complex numbers, and the equivalence relation determining the shape-space is

$$W \equiv W_0 \quad \text{when} \quad W = \alpha W_0, |\alpha| = 1. \tag{16}$$

From this it will be seen that the group G^+ of symmetries for Σ_2^k can be further extended to G^{++} which is generated by the complex conjugation operator τ and the elements of the quotient $U(k-1)/U(1)$ of the unitary group by its centre, acting on the right; these are all obviously isometries. But this implies that Σ_2^k is *homogeneous*, because the group G^{++} is transitive, and so the local geometry is in fact the same everywhere. Such metrical homogeneity does not hold when $m \geq 3$.

In this sketch of the theory of shape-spaces we are taking pains to emphasize as many as possible of the atypical aspects of the cases $m = 1$ and 2 . We now come to the most important such feature, and the one that makes the name *shape-manifold* appropriate: *for $m = 1$ and 2 , but not otherwise, Σ_m^k is a C^∞ riemannian manifold.* For the moment we shall just prove the C^∞ smoothness. We need only consider $m = 2$; when $m = 1$ there is nothing to prove (see (10)). It will be obvious that, consistently with the topology, we can cover Σ_2^k by $k-1$ coordinate patches of which the second (for example) has the coordinates featuring in

$$\sigma = (z_1, a_2, z_3, \dots, z_{k-1}),$$

where a_2 is real and strictly positive, where the z 's are complex, and where $a_2^2 + \sum |z_j|^2 = 1$. It is then a trivial matter to verify that the patch-to-patch coordinate changes are C^∞ . In [14] it will be shown that for $m \geq 3$ there is a C^∞ structure on the dense connected open subset of Σ_m^k determined by the condition,

$$WW' \text{ has not more than one vanishing eigenvalue,} \tag{17}$$

the excluded points being singular. Notice that (17) is necessarily satisfied if $m = 1$ or 2 . Note also that WW' is independent of the choice of the reference orthogonal matrix \tilde{Q} , and that its eigenvalue spectrum is a property of $\sigma = [W]$ only. When $m \geq 3$ the riemannian geometry (curvatures, geodesics, etc.) on this dense connected open subset with manifold structure is investigated in [14], where we show that the sectional curvatures are positive, are bounded away from zero, and that near the singularities they are unbounded above.

It will be clear from this discussion that the properties of Σ_2^k can be put in a more familiar context by observing that if the pre-shape W is

$$(z_1, z_2, \dots, z_{k-1}),$$

then the shape $[W]$ can be described by the set of ratios

$$\sigma = (z_1 : z_2 : \dots : z_{k-1}), \tag{18}$$

and in fact that (up to a change-of-scale)

$$\Sigma_2^k = \mathbb{C}P^{k-2}, \tag{19}$$

a complex projective space. In particular, $\Sigma_2^2 = \mathbb{C}P^0 = \text{pt.}$

The missing scale factor will be identified in Section 3.

Thus if we put together the shape-manifolds, the simplicial shape-spaces, and the over-dimensional shape-spaces, we see that while only some of these are smooth manifolds (perhaps with boundary), yet all are *topological* manifolds, and all have relatively simple torsion-free homology. In [13], where Σ_m^k is examined from the topological point of view, it is shown that none of the anomalous shape-spaces are topological manifolds, and that all of them have torsion. It is in this sense that they are not only metrically but also topologically 'anomalous'.

We have now set the stage for a more detailed study of the spaces Σ_2^k (there is little more to be said about Σ_1^k , since all of those spaces are isometric with spheres). We shall develop their riemannian geometry from a statistical point of view based on a type of 'procrustes' analysis, and we shall examine them as carriers of model-determined shape-functions and shape-measures as well as of empirical data-determined shape-distributions. The groups G , G^+ , and G^{++} will play a decisive role (note, for example, that G includes the permutations associated with relabelling the k points). In order to focus attention on practical aspects, we invite the reader to consider from time to time two ways of looking at the data-set studied in [3, 9]. Here we had 52 points in 2 dimensions, determining 22,100 triangles. We can think of the shapes of these as specifying (i) 22,100 points in Σ_2^3 , or (ii) a single point in Σ_2^{52} .

3. Procrustes analysis, and the invariant (quotient) metric on Σ_2^k

If z (or z^* when we are using the starred coordinates) denotes a k -ad in \mathbb{R}^2 , then we shall write $[z]$ (or $[z^*]$) for the corresponding point on the shape-manifold Σ_2^k . For present purposes the starred coordinates will be convenient, and we shall suppose that $k \geq 3$, because we know that $\Sigma_2^2 = \text{pt}$. Suppose that we want to compare the shapes of the k -ads z^* and w^* ; then we must find the minimum value of a suitable measure of mismatch between the k -ads $\alpha z^* + \lambda$ and $\beta w^* + \mu$, where the greek letters denote complex numbers and α and β are to be non-zero, so that we allow a free choice of location, scale, and rotation for each k -ad. An appropriate measure of mismatch is obtained by associating similarly labelled vertices, summing the squares of the distances between associated vertices, and then dividing by the sum of the squares of the distances of all $2k$ vertices from their common centroid. We clearly lose no generality in assuming that each of z^* and w^* has its centroid at 0, and the quotient will then be (with \sum summing over $1 \leq i \leq k$)

$$\frac{\sum |\alpha z_i^* - \beta w_i^* + (\lambda - \mu)|^2}{\sum |\alpha z_i^* + \frac{1}{2}(\lambda - \mu)|^2 + \sum |\beta w_i^* - \frac{1}{2}(\lambda - \mu)|^2},$$

which is equal to

$$\frac{\sum |\alpha z_i^* - \beta w_i^*|^2 + k|\lambda - \mu|^2}{\sum |\alpha z_i^*|^2 + \sum |\beta w_i^*|^2 + \frac{1}{2}k|\lambda - \mu|^2};$$

this ratio is least when $\lambda = \mu$ (for given α and β) in virtue of the complex-number inequality $|\xi - \eta|^2 \leq 2|\xi|^2 + 2|\eta|^2$. We are left with the problem of performing the remaining minimisation over all admissible values of α and β , but this is now easy and we find the net *coefficient of mismatch* to be

$$\Delta([z^*], [w^*]) = 1 - \frac{|\sum (z_i^* - z_c^*)(\bar{w}_i^* - \bar{w}_c^*)|}{\sqrt{\sum |z_i^* - z_c^*|^2} \sqrt{\sum |w_i^* - w_c^*|^2}}, \quad (20)$$

where we have now adapted our formula for use with an arbitrary origin by restoring the terms in z_c^* and w_c^* .

The reader will recognise in this calculation an example of what has been called *procrustes analysis*; there is a large literature centred around this topic, and a recent review by Sibson [22] makes an excellent reference. Normally in procrustes analysis one wishes to compare two configurations (not necessarily in 2 dimensions) which may have been disguised by redundant translations, rotations, dilatations, and reflections, and one wishes to find (i) a coefficient of mismatch, as above, and (ii) the transformed version of the second configuration which most closely resembles the first configuration. In our treatment we have to be careful to exclude the use of reflections in improving the match. This introduces some important differences between our formulae and those given by Sibson and appropriate to the comparison of the output of MD-SCAL plots, where reflections are frequent, irrelevant, and must be eliminated.

We next draw attention to the fact that the expression for Δ at (20) is unaltered if we simultaneously transform z^* and w^* by complex conjugation, or if we simultaneously transform them by unitary operators acting in the subspace of \mathbb{R}^k orthogonal to $(1, 1, \dots, 1)$. That Δ depends only on the shapes of z^* and w^* is clear from the calculation itself. Thus we have a (canonical) function of the two shapes which vanishes if and only if these are identical, which depends on them in a symmetrical way, and which is unaltered if the shapes of the two k -ads are transformed by the same member of the extended group of symmetries G^{++} . Finally we observe that the maximum possible value of Δ is unity, and that this maximum degree of mismatch is attained when and only when $\sum (z_i^* - z_c^*)(\bar{w}_i^* - \bar{w}_c^*) = 0$.

Now that the canonical status of Δ has been established we can where convenient use the unstarred coordinates for its evaluation, and we find that $\Delta = 1 - \left[\sum z_j \bar{w}_j / \left[\sum |z_j|^2 \sum |w_j|^2 \right]^{1/2} \right]^2$, independently of the choice of \bar{Q} , and with \sum summing over $1 \leq j \leq k-1$. This last formula simplifies further if we introduce for $1 \leq j \leq k-1$ the variables $Z_j = z_j / \sqrt{\sum |z_h|^2}$ and $W_j = w_j / \sqrt{\sum |w_h|^2}$, for then we find that

$$\Delta([Z], [W]) = 1 - \left| \sum_1^{k-1} Z_j \bar{W}_j \right|^2, \quad (21)$$

and now the maximum mismatch of unity is attained when and only when $\sum Z_j \bar{W}_j = 0$ (e.g. when $Z_1 = 1 = W_2$). Notice that with these coordinates $\sum |Z_j|^2 = \sum |W_j|^2 = 1$, and that Δ is a C^∞ function of Z and W except when $\sum Z_j \bar{W}_j = 0$.

Consider three k -ads expressed in our latest notation as U, V , and W . Let $f = \sum V_j \bar{W}_j$, $g = \sum W_j \bar{U}_j$, and $h = \sum U_j \bar{V}_j$. The matrix with components

$$\begin{pmatrix} 1 & h & \bar{g} \\ \bar{h} & 1 & f \\ g & \bar{f} & 1 \end{pmatrix}$$

is non-negative semi-definite hermitian, and so it has a non-negative real determinant, whence

$$\begin{aligned} 1 - |f|^2 - |g|^2 - |h|^2 &\geq -2 \operatorname{Re}(fgh) \geq -2|f||g||h|, \\ (1 - |g|^2)(1 - |h|^2) &\geq (|g||h| - |f|)^2, \end{aligned}$$

and

$$|f| \geq |g||h| - \sqrt{(1-|g|^2)}\sqrt{(1-|h|^2)}.$$

We now define angles $\Theta, \Phi,$ and Ψ in the range $[0, \frac{1}{2}\pi]$, by setting

$$|f| = \cos \Theta, \quad |g| = \cos \Phi, \quad |h| = \cos \Psi;$$

it follows that $\cos \Theta \geq \cos(\Phi + \Psi)$, so that $\Theta \leq \Phi + \Psi$, and so ρ , where

$$\rho([Z], [W]) = \arccos(1 - \Delta) \quad (0 \leq \rho \leq \frac{1}{2}\pi), \tag{22}$$

is a distance-function with $\frac{1}{2}\pi$ as its (attained) maximum. From what we know already about Δ we see that this distance-function for the shape-manifold is canonical, and that it is invariant under the group of symmetries G^{++} . It is readily checked that ρ generates a metric topology which is the same as the quotient-topology already given to Σ_2^k ; indeed an elementary calculation shows that, for $m = 2,$ ρ at (22) is identical with ρ in (7), the latter formula being valid for all values of m . Evidently, then, ρ is the appropriate invariant metric for Σ_2^k .

We have seen that $\max \rho = \pi$ for each $k \geq 3$ when $m = 1$, and that $\max \rho = \frac{1}{2}\pi$ for each $k \geq 3$ when $m = 2$. We now show that $\max \rho = \frac{1}{2}\pi$ for each $k \geq 3$ and $m \geq 2$. First we observe that for $k \geq 3$ and $m \geq 2$ it is possible for WW'_0 to be the zero $m \times m$ matrix, and then $\text{Tr}(RWW'_0) = 0$ for all R , so that $\rho = \frac{1}{2}\pi$ by (7). For general W and W_0 we can write $WW'_0 = \tilde{R}D\tilde{S}$ as at (12) and (13) (where however $k-1$ is to be replaced by m , so that $\tilde{S} \in \text{SO}(m)$) and then $\rho = \arccos \sup_R \text{Tr}(RD)$.
Now

$$\text{Tr}(RD) = \sum_1^m r_{ii}d_i,$$

where the d 's can be taken to satisfy $d_1 \geq \dots \geq d_{m-1} \geq |d_m|$. From an elegant theorem of Horn [7] we know that the set of diagonals $(r_{11}, r_{22}, \dots, r_{mm})$ of the matrices R in $\text{SO}(m)$ is a compact convex set with the extreme points

$$\{(\pm 1, \pm 1, \dots, \pm 1) \quad (\text{even number of minus signs})\}.$$

The ordering and signs of the d_i 's then make it plain that

$$\cos \rho = \sup_R \text{Tr}(RD) = \text{Tr} D \geq 0,$$

so that always $\rho \leq \frac{1}{2}\pi$, this bound being attained. The exception at $m = 1$ is due to the fact that then $R = 1$ and $D = \pm 1$, so that $\text{Tr}(RD)$ can be -1 .

A by-product of this calculation is the useful fact that, for $\sigma = [W]$ and $\sigma_0 = [W_0]$,

$$\rho(\sigma, \sigma_0) = \arccos(\text{Tr} \sqrt{WW'_0 W_0 W'}) - 2\alpha s) \quad (0 \leq \rho \leq \frac{1}{2}\pi) \tag{23}$$

where α is the smallest eigenvalue of $\sqrt{WW'_0 W_0 W'}$, and

$$s = \begin{cases} 0 & \text{if } \det(WW'_0) \geq 0 \\ 1 & \text{if } \det(WW'_0) < 0. \end{cases} \tag{24}$$

The term αs in (23) distinguishes this result from one of Sibson [22]; this is because he can take the supremum over all orthogonal R , and then s can always be taken to be zero.

We now return to the case $m = 2$. Many other smooth invariant metrics exist which are equivalent to ρ ; for example $\rho/(1 + \rho)$, and also $c \sin(\rho/c)$ when $c \geq 2$. A particularly interesting such metric is that obtained when $c = 2$; this leads to the equivalent metric $\sqrt{(2\Delta)}$. However, the metric ρ is the ‘right’ one for us because of its alternative definition as the quotient metric.

To obtain the associated riemannian metric for Σ_2^k we must examine what happens when $[W]$ approaches $[W_0]$. Now

$$\cos^2 \rho = (1 - \Delta)^2 = \frac{\sum z_j \bar{w}_j \sum \bar{z}_j w_j}{(\sum z_j \bar{z}_j)(\sum w_j \bar{w}_j)},$$

and that is asymptotically $1 - \rho^2 + o(\rho^2)$ when ρ is small, while the right-hand side is homogeneous in (z_1, \dots, z_{k-1}) and in (w_1, \dots, w_{k-1}) . The mapping p from S_2^k to Σ_2^k is a submersion (this is what makes Σ_2^k a manifold) and so there is a local section of S_2^k lying above any small neighbourhood of Σ_2^k , and in this section we can do our asymptotics in terms of the z ’s and w ’s. Accordingly we get, after some simple calculations,

$$ds^2 = \frac{(\sum z_j \bar{z}_j)(\sum dz_j d\bar{z}_j) - (\sum z_j d\bar{z}_j)(\sum \bar{z}_j dz_j)}{(\sum z_j \bar{z}_j)^2}, \tag{25}$$

where \sum sums over $1 \leq j \leq k-1$, and this (in a patch on which $z_1 \neq 0$) can be written in the non-homogeneous form

$$ds^2 = \frac{(1 + \sum \zeta_j \bar{\zeta}_j)(\sum d\zeta_j d\bar{\zeta}_j) - (\sum \zeta_j d\bar{\zeta}_j)(\sum \bar{\zeta}_j d\zeta_j)}{(1 + \sum \zeta_j \bar{\zeta}_j)^2}, \tag{26}$$

where ζ_j denotes z_j/z_1 and \sum here sums over $2 \leq j \leq k-1$. This last formula (26) identifies the riemannian metric on a typical patch of Σ_2^k .

The reader will recognise in (25) and (26) the classical Fubini–Study metric [5, 26] for complex projective space $\mathbb{C}P^{k-2}$, *apart from the absence of an overall factor 4 on the right-hand side*. The absence of this 4 in (25) and (26) is associated with our norming condition, $\text{Tr}(WW') = 1$, and we shall comment further on this at a later stage.

From (25) or (26) we can compute geodesic distances δ on Σ_m^k . It is not superficially obvious that $\delta(\sigma, \sigma_0) = \rho(\sigma, \sigma_0)$ for all σ and σ_0 , and so we shall prove that this is so by a direct calculation which will also provide an exercise in the use of (26).

We first note that the value of $\delta([z], [w])$ depends only on the points Z and W (corresponding to z and w) on the unit complex sphere CS in \mathbb{C}^{k-1} . Now any two points of CS can be transformed by a unitary symmetry into the pair of points with $k-1$ complex coordinates in \mathbb{C}^{k-1} equal to

$$(1, 0, 0, \dots, 0) \quad \text{and} \quad (\alpha \cos t, \alpha \sin t, 0, \dots, 0),$$

respectively, where $0 \leq t \leq \frac{1}{2}\pi$ and where α is a complex scalar of unit modulus.

Thus any pair of points of $\mathbb{C}P^{k-2}$ can be transformed by a unitary symmetry into

$$A = \{\zeta : \zeta_j = 0 \text{ for } 2 \leq j \leq k-1\}$$

and

$$B(t) = \{\zeta : \zeta_2 = \tan t, \text{ and } \zeta_j = 0 \text{ for } 3 \leq j \leq k-1\}$$

without loss of geometric generality, although the point $B(\frac{1}{2}\pi)$ will require special consideration. Our procrustean distance (22) is

$$\rho(A, B(t)) = \arccos \left| \frac{\bar{\alpha} \cos t}{\cos^2 t + \sin^2 t} \right| = t \quad (0 \leq t < \frac{1}{2}\pi),$$

and so the parameter u for the smooth curve $\{B(u) : 0 \leq u < \frac{1}{2}\pi\}$ is just the procrustean distance from the end A at $u = 0$. The end $B(\frac{1}{2}\pi)$ is “ $z_1 = 0, z_2 = 1, z_j = 0 (j \geq 3)$ ”, and $\rho(A, B(\frac{1}{2}\pi)) = \frac{1}{2}\pi$. A similar argument shows that the procrustean distance between the points with parameters u_1 and u_2 is $|u_2 - u_1|$, and therefore the procrustean distances are additive along the whole curve from $t = 0$ to $t = \frac{1}{2}\pi$.

Now the curve we have been considering is part of a closed geodesic, the whole loop being parametrised by $-\frac{1}{2}\pi \leq t \leq \frac{1}{2}\pi$. To see that this is so it suffices to take $2k-4$ real coordinates by writing $\zeta_j = \xi^j + i\xi^{k-2+j}$ ($j = 2, \dots, k-1$), and then to calculate the Riemann-Christoffel coefficients Γ_{jk}^i in the differential equations for geodesics,

$$\ddot{\xi}^i + \Gamma_{jk}^i \dot{\xi}^j \dot{\xi}^k = 0 \quad (j = 2, \dots, 2k-3).$$

We only have to show that $\xi^2 = \tan t$ satisfies these equations when all the other real coordinates are set identically to zero, and so it is enough to obtain the evaluations $\Gamma_{22}^2 = -\sin 2t$ and $\Gamma_{22}^j = 0$ (for $j \neq 2$) at all points of this particular curve; the calculations are straightforward using (26), and the geodesic character of the curve then follows. Indeed one can see in this way that the parameter must be an affine function of the arc-length, and the fact that it actually is the arc-length can be checked by integration: the arc-length from A to $B(t)$ according to the riemannian metric (26) is

$$\int_0^t \sqrt{\left\{ \frac{(1 + \tan^2 u)(\sec^2 u du)^2 - (\tan u \sec^2 u du)^2}{(1 + \tan^2 u)^2} \right\}},$$

which is equal to t , as required. In particular it follows that the length of a closed geodesic loop is π .

To get the geodesic metric we must examine the “focussing” properties of the geodesics. Now it is clear that we can get a complete family of geodesics through A by applying symmetries from the group $1 \times U(k-3)$ to the geodesics

$$G_\beta^0 = \{\zeta(u) : \zeta_2 = \beta \cos \theta \tan u, \zeta_3 = \beta \sin \theta \tan u, \\ \zeta_j = 0 \ (4 \leq j \leq k-1); 0 \leq u \leq \frac{1}{2}\pi\}$$

where $0 \leq \theta \leq \frac{1}{2}\pi$ and where β is a complex scalar of unit modulus (i.e. $\beta \in S^1$). Now consider G_β^0 or any one of its symmetry-transforms. This will pass through A and it

will hit $B(t)$ for some t ($0 < t < \frac{1}{2}\pi$) only if both $\sin \theta = 0$ and $\beta = 1$, in which case it coincides with the original curve. The only other possibility is that the intersection occurs at $t = \frac{1}{2}\pi$, and then the necessary and sufficient condition is that $\sin \theta = 0$. In this last case we have a one-parameter family of geodesics through A labelled by the parameter $\beta \in S^1$ which are again concurrent at $B(\frac{1}{2}\pi)$, the length $\frac{1}{2}\pi$ being the same for all (these geodesic arcs can in fact be matched up in pairs so that each pair makes up one of the closed geodesics through A and $B(\frac{1}{2}\pi)$ of length π). From this analysis it follows that $\delta(A, B(t)) = t$ is always the *minimum* geodesic distance from A to $B(t)$ (though in the case $t = \frac{1}{2}\pi$ it can be realised in an infinity of ways). Because A and $B(t)$ were adequately representative we have proved

THEOREM 1. *The procrustean (quotient) metric distance ρ between two shapes represented by two points on $\Sigma_2^k = \mathbb{C}P^{k-2}$ is equal to the shortest distance along a geodesic path of the particular Fubini–Study riemannian metric (25). There is only one such path unless the shapes are maximally remote ($\rho = \frac{1}{2}\pi$), in which case there is an infinity of paths identifiable with points on S^1 and obtainable by halving the closed geodesics of total length π passing through the two shapes.*

The shapes $[w]$ maximally remote from $[z]$ are those for which $\sum z_j \bar{w}_j = 0$; they form a subspace isometric with $\mathbb{C}P^{k-3}$, and in particular there is only one such shape if $k = 3$.

It is important not to confuse the geodesics on the shape-manifold with the trajectories generated by the algorithms of procrustean theory and practice; these are in fact complementary to one another. The procrustean algorithms are concerned with two shapes, each having a pre-image which is an equivalence class. Their purpose is to enable one to travel efficiently inside the first equivalence class so as to reach the point in that class nearest to the second equivalence class. In contrast to this, the geodesics linking the two shapes enable one to travel efficiently from the class of one to the class of the other through a set of intervening classes.

Some attention in the procrustean literature has been devoted to the problem of converting measures of mismatch to distance-functions. So far as Δ is concerned we have now found the definitive answer to this question, and with its aid it is easy to generate other solutions. For example, we have noted above that $c \sin(\rho/c)$ will be a distance-function whenever $c \geq 2$. If we take $c = 2$, we find at once that $\sqrt{(2\Delta)}$ is a distance-function. While ρ is clearly superior to all other distance-functions for theoretical purposes, there may be occasions in practice when $\sqrt{\Delta}$ has advantages.

We close this section by examining the important special case $k = 3, m = 2$. From (26) we know that there is a coordinate patch covering Σ_2^3 with the exception of one point for which

$$ds^2 = \frac{(1 + |\zeta|^2)|d\zeta|^2 - |\zeta d\zeta|^2}{(1 + |\zeta|^2)^2}$$

is the riemannian metric. Let A denote $\zeta = 0$; the excluded point B corresponds to the compactification point $\zeta = \infty$ for the complex ζ -plane. Now put $\zeta = re^{i\phi}$ and we get

$$ds^2 = \frac{dr^2 + r^2 d\phi^2}{(1 + r^2)^2} \quad (r \geq 0, \phi \in S^1).$$

Finally put $r = \tan \frac{1}{2}\theta$ and we get

$$ds^2 = \frac{1}{4}(d\theta^2 + \sin^2 \theta d\phi^2) \quad (0 \leq \theta < \pi, \phi \in S^1).$$

Now this is the riemannian metric for the punctured two-dimensional sphere $S^2(\frac{1}{2})$ of radius $\frac{1}{2}$, and our general arguments about geodesics and homogeneity show at once that each of $\Sigma_2^3 \setminus A$ and $\Sigma_2^3 \setminus B$ is isometric to $S^2(\frac{1}{2})$ punctured at one point, whence we have the basic

THEOREM 2. Σ_2^3 is isometric with $S^2(\frac{1}{2})$.

As a statement that $\mathbb{C}P^1 = S^2(\frac{1}{2})$ this is not new, but its practical significance, that $S^2(\frac{1}{2})$ provides the natural home for the shapes of triangles, is far-reaching. The value $\frac{1}{2}$ for the radius of the sphere Σ_2^3 is of course related to the absence of the customary factor 4 in (25). We have here obtained Theorem 2 as a corollary to Theorem 1, but a direct proof is possible, and instructive; for this, see [12].

4. Shape-measures and shape-densities

The projection p mapping from S_m^k to Σ_m^k is measurable (from Borel sets to Borel sets), and so if P_1, P_2, \dots, P_k in \mathbb{R}^m form a random k -ad such that

$$\text{pr}(P_1 = P_2 = \dots = P_k) = 0, \tag{27}$$

then the shape σ of the k -ad will be a random variable on Σ_m^k , and as such will be distributed there according to a *shape-measure* determined by the joint law of (P_1, P_2, \dots, P_k) . The most interesting case at this stage of the development of the theory is that in which the points (P_j) are IID in \mathbb{R}^m with individual law μ , say, and we now confine our attention to this. We shall write μ^k for the joint law of (P_1, P_2, \dots, P_k) . The *induced shape-measure* μ_*^k will be unchanged if we vary the parameters in μ controlling location, scale, and rotation of axes, so we write $[\mu]$ for the equivalence-class of measures μ modulo such transformations, and we then call $[\mu]$ the *generator* of the induced shape-measure μ_*^k . We shall suppose throughout that μ is absolutely continuous with respect to Lebesgue measure in \mathbb{R}^m ; this will guarantee that (27) holds. Because $\mu_*^k = \mu^k p^{-1}$, it is clear that μ_*^k will be absolutely continuous with respect to v_*^k whenever μ is absolutely continuous with respect to v . Now let γ denote an m -spherically symmetric gaussian distribution in \mathbb{R}^m . Then any μ_*^k will be absolutely continuous with respect to γ_*^k , and so can be represented by its Radon–Nikodym derivative

$$m^k = \frac{d\mu_*^k}{d\gamma_*^k} \tag{28}$$

which we shall call the k -th *shape-density* associated with μ or $[\mu]$.

We can now formulate a number of questions concerning the invertibility of the mapping $[\mu] \rightarrow \mu_*^k$. These are of interest for any k , but are especially so for $k = m + 1$.

- (1) Is $[\mu]$ uniquely determined if μ_*^k is given?

- (2) Let $\lambda \mid K$ denote the uniform distribution over the compact convex set K in \mathbb{R}^m . If $\mu = \lambda \mid K$ and $\mu_*^k = \nu_*^k$, must we have $[\nu] = [\lambda \mid L]$ for some compact convex L ; also if $[\nu] = [\lambda \mid L]$, must K and L be equivalent modulo the similarity group?
- (3) (The corresponding questions for compact convex polyhedra K and L .)
- (4) If a probability measure θ on the Borel sets of Σ_m^k is given, what are the necessary and sufficient conditions for the existence of a μ such that $\theta = \mu_*^k$?
- (5) (The corresponding questions when μ is required to be of the form $\lambda \mid K$, with K a compact convex set, or a compact convex polyhedron.)

When $m = 1$ these reduce to questions about characteristic functions. The case $m = 2$ has been investigated by Small; a large amount of information is contained in his doctoral thesis [25], parts of which are summarised informally below. When $m \geq 3$, virtually nothing is known.

For $m = 2$, Small first tackles question (1), and he shows that the answer is ‘no’; more precisely, if $k \geq 3$ then there exist two different circularly symmetric generators $[\mu]$ and $[\nu]$ such that $\mu_*^k = \nu_*^k$. Here k is fixed; in a complementary theorem he shows that we must have $[\mu] = [\nu]$ (now without any assumptions of circular symmetry) if $\mu_*^k = \nu_*^k$ for all k . This is proved by first establishing a continuity theorem of independent interest. In between these two results Small proves intermediate theorems which tell us that with pairs of integers (m, n) , both not less than 3, there are associated classes of generators $S(m, n)$ such that

$$\text{if } [\mu] \in S(m, n), \text{ then } \mu_*^{mn} \text{ determines } [\mu]. \tag{29}$$

In particular it turns out that the gaussian generator $[\gamma]$ belongs to $S(3, 3)$, so that no generator $[\mu] \neq [\gamma]$ satisfies $\mu_*^9 = \gamma_*^9$ on $\mathbb{C}P^7$. Finally Small obtains conditions on μ_*^4 which (in the presence of some technical side conditions) are necessary and sufficient for μ to belong to $(\lambda \mid K : K \text{ compact convex})$. It should be added that Small presents these results both in the context of shape-spaces (where the problems he solves were originally formulated) and also in the equally appropriate context of maximal invariant statistics.

We now turn to the explicit determination of the ‘gaussian’ shape-measure γ_*^k when $m = 2$; the corresponding result for general m will be given in [14]. Let f be any bounded measurable function

$$f : \Sigma_2^k \rightarrow \mathbb{R},$$

let $z = (z_1, z_2, \dots, z_{k-1})$ be as in (1), so that the z ’s are complex numbers and $\|z\|^2 = \sum |z_j|^2$ has not yet been normalised to unity, let $\sigma = \tilde{p}z$ be the associated shape, and let g be any element of the symmetry group G^{++} . We note that

$$g\sigma = g\tilde{p}z = \tilde{p}gz. \tag{30}$$

Suppose that z arises (via (1)) from a labelled sample of size k from the planar gaussian distribution with circular symmetry. Then

$$\mathbb{E}\{f(g\sigma)\} = \int f(\tilde{p}z)\gamma^k(dz)$$

because of the symmetries of the gaussian distribution γ^k , and if we write $z_j = R_j \exp(i\Phi_j)$ for $j = 1, 2, \dots, k-1$, we shall have

$$\mathbb{E}\{f(g\sigma)\} = \int f(\tilde{p}z) \prod_1^{k-1} \exp(-\frac{1}{2}R_j^2) d(\frac{1}{2}R_j^2) d\left(\frac{\Phi_j}{2\pi}\right).$$

Now make the change of variables

$$r = R_1, \quad r_h = R_{h+1}/r, \quad \phi = \Phi_1, \quad \phi_h = \Phi_{h+1} - \phi \pmod{2\pi},$$

where $1 \leq h \leq k-2$; we find that

$$\mathbb{E}\{f(g\sigma)\} = (k-2)! \int f(\tilde{p}z) \left(1 + \sum_1^{k-2} r_h^2\right)^{-(k-1)k-2} \prod_1^{k-2} d(r_h^2) d\left(\frac{\phi_h}{2\pi}\right).$$

But all of r_h and ϕ_h ($h = 1, 2, \dots, k-2$) are shape coordinates, and in fact $\zeta_{h+1} = r_h \exp(i\phi_h)$ ($h = 1, 2, \dots, k-2$) are the usual coordinates for the shape-patch defined by $z_1 \neq 0$. We know however that $\text{pr}(z_1 = 0)$ vanishes for this generator, so that the induced shape-measure can be defined unambiguously over the whole of Σ_2^k by

$$d\gamma_*^k = (k-2)! \left(1 + \sum_1^{k-2} r_h^2\right)^{-(k-1)k-2} \prod_1^{k-2} d(r_h^2) d\left(\frac{\phi_h}{2\pi}\right), \quad (31)$$

and it is evidently invariant under G^{++} .

From the invariant riemannian metric (26) we get another invariant probability measure on Σ_2^k ; this is

$$d\tilde{\omega}^k = C_k \sqrt{(\det(g_{ij} : 1 \leq i, j \leq 2k-4))} \prod_{q=2}^{2k-3} d\xi^q, \quad (32)$$

where g_{ij} is the coefficient of $d\xi^i d\xi^j$ in (26), and C_k is a constant chosen to make $\int d\tilde{\omega}^k = 1$. That C_k can be so chosen follows from the compactness and smoothness of Σ_2^k via a calculation involving a partition of unity. Now γ_*^k and $\tilde{\omega}^k$ are both absolutely continuous with respect to $\gamma_*^k + \tilde{\omega}^k$ (which again is invariant), so have Radon-Nikodym derivatives which are invariant and thus (because G^{++} is transitive for $m = 2$) constant. From this remark we deduce

THEOREM 3. *There is a unique G^{++} -invariant probability measure on the Borel sets of Σ_2^k ; it coincides with the 'gaussian' shape-measure γ_*^k and also with the differential-geometric 'volume' measure $\tilde{\omega}^k$, and it is given explicitly by (31).*

It is an interesting exercise to compute $\tilde{\omega}^k$ directly from (26), which we can rewrite in the form

$$(1 + \sum r^2)^2 ds^2 = (1 + \sum r^2) \sum (dr^2 + r^2 d\phi^2) - (\sum r dr)^2 - (\sum r^2 d\phi)^2. \quad (33)$$

When $m \geq 3$, G^{++} is no longer transitive. It should therefore come as no surprise that $\gamma_*^k \neq \tilde{\omega}^k$ when $m \geq 3$ (this will be proved in [14]). We thus have a genuine choice between γ_*^k and $\tilde{\omega}^k$ when agreeing on the reference measure in the

definition of m^k at (28), and we shall see later that there are considerable advantages in agreeing to use γ_*^k .

When $k = 3$ and $m = 2$ we shall see that G^{++} translates into the group of proper and improper rotations of $S^2(\frac{1}{2})$. Thus it follows at once from Theorem 3 that $\tilde{\omega}^3 = \gamma_*^3$ can be none other than the classical surface measure on $S^2(\frac{1}{2})$, normed so as to have total mass unity. (So far as $\tilde{\omega}^3$ is concerned this is obvious from the calculations preceding the statement of Theorem 2.) We shall often wish to display shape-functions, shape-measures, and empirical shape-distributions on Σ_2^3 , and a useful device for accomplishing this will be developed in the next section.

5. The manifold carrying the shapes of triangles

We now know that there is a natural isometry between Σ_2^3 , the space whose points are the shapes of labelled triangles, and the sphere $S^2(\frac{1}{2})$. The patch on which $z_1 \neq 0$ can be given a coordinate map with a single complex coordinate $\zeta = z_2/z_1$. For computational purposes we write $\zeta = re^{i\phi}$. We now explore the geometrical details of the mapping.

When measuring up a triangle ABC with $A \neq B$ our standard procedure will be to set $\frac{1}{2}|AB|$ equal to the unit of length, and to let the direction from A to B be the reference direction. With these conventions let the vector from the mid-point of AB to the vertex C be $Me^{i\Psi}$. To calculate r and ϕ we use the fact that the matrix-product

$$\begin{pmatrix} -1, & 1, & M \cos \Psi \\ 0, & 0, & M \sin \Psi \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}}, & -\frac{1}{\sqrt{2}}, & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}}, & \frac{1}{\sqrt{2}}, & -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{3}}, & 0, & \frac{\sqrt{2}}{\sqrt{3}} \end{pmatrix}$$

must be equivalent to

$$\begin{pmatrix} *, & 1, & r \cos \phi \\ *, & 0, & r \sin \phi \end{pmatrix},$$

modulo similarities (cf. equation (1)).

Here the entries at * are of no interest, and the relation between (M, Ψ) and (r, ϕ) is to be determined by comparing the last matrix with

$$\begin{pmatrix} \frac{M \cos \Psi}{\sqrt{3}}, & \sqrt{2}, & \frac{M\sqrt{2} \cos \Psi}{\sqrt{3}} \\ \frac{M \sin \Psi}{\sqrt{3}}, & 0, & \frac{M\sqrt{2} \sin \Psi}{\sqrt{3}} \end{pmatrix}.$$

If we discard the first column and then compare the remaining columns we obtain the *basic formulae*,

$$r = M/\sqrt{3}, \quad \phi = \Psi. \tag{34}$$

We next invoke the transformation $r = \tan \frac{1}{2}\theta$ used at the end of Section 3. If we take one of the usual angular coordinate systems for $S^2(\frac{1}{2})$, the corresponding point on the sphere will be

$$\left(\frac{1}{2} \cos \theta, \frac{1}{2} \sin \theta \cos \phi, \frac{1}{2} \sin \theta \sin \phi\right),$$

but for computational purposes it is preferable to drop the factor $\frac{1}{2}$ and to write

$$\left. \begin{aligned} X &= \cos \theta = \frac{1-r^2}{1+r^2}, \\ Y &= \sin \theta \cos \phi = \frac{2r \cos \phi}{1+r^2}, \\ Z &= \sin \theta \sin \phi = \frac{2r \sin \phi}{1+r^2}, \end{aligned} \right\} \quad (35)$$

so that $X^2 + Y^2 + Z^2 = 1$ and the pole of the (θ, ϕ) -coordinates is at $X = 1$. This will be permissible so long as we remember that *great-circle distances on the (X, Y, Z) -sphere must be halved*, so that (for example) a complete great circle has length $\frac{1}{2}(2\pi) = \pi$, as required by Theorem 1.

When it is necessary to work in the other direction, starting with (X, Y, Z) , we easily compute (r, ϕ) , and then the vertices of a typical triangle ABC of the specified shape will be given by

$$\begin{pmatrix} 0, & 1, & r \cos \phi \\ 0, & 0, & r \sin \phi \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}}, & \frac{1}{\sqrt{3}}, & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}}, & \frac{1}{\sqrt{2}}, & 0 \\ -\frac{1}{\sqrt{6}}, & -\frac{1}{\sqrt{6}}, & \frac{2}{\sqrt{3}} \end{pmatrix}.$$

As we are not interested in size we can multiply the product by $\sqrt{6}$, and so we get as vertices of a triangle of the specified shape

$$\left. \begin{aligned} A &= -\sqrt{3} - re^{i\phi}, \\ B &= \sqrt{3} - re^{i\phi}, \\ C &= 2re^{i\phi}. \end{aligned} \right\} \quad (36)$$

Formulae (34), (35), and (36) will be used continuously in what follows. The reader should now test his understanding of the geometry by constructing triangles ABC for which

$$(i) \ r = 0, \quad (ii) \ r = \frac{1}{\sqrt{3}}, \ \phi = \pm \pi, \quad (iii) \ r = 1, \ \phi = \frac{1}{2}\pi,$$

and locating the corresponding shape-points on the sphere. He should also verify (iv)

that ABC is right-angled when $r = 1/\sqrt{3}$, whatever ϕ may be, and (v) determine at which vertex the right angle then lies. Finally (vi) he should determine the locus on $S^2(\frac{1}{2})$ of the points representing the shapes of *all* right-angled triangles. The solutions to these exercises will be found below.

Most readers will appreciate that these transformation formulae are connected with classical results about geometrical inversion and the Hopf submersion, but because those features do not suitably generalise when $m \neq 2$, we prefer the present approach based on the procrustean differential geometry of Section 3.

The above discussion does not cover the unique point of Σ_2^3 omitted from the patch-map. We can think of this as $\zeta = \infty$, or indeed as $r = \infty$, which maps to $(-1, 0, 0)$ on the sphere. From symmetry arguments (as below), or by continuity, we see that this is the shape of the 'triangle' for which A and B coincide.

It is of minor interest that in the representation on $S^2(\frac{1}{2})$, the coefficient of mismatch Δ is equal to twice the 'sagitta', this being the greatest distance in \mathbb{R}^3 between the shorter geodesic connecting the two points on $S^2(\frac{1}{2})$ representing the two shapes, and the chord joining them.

We need to identify the action of the group of symmetries G^{++} on the sphere. As the argument is essentially classical [21, Ch. III] we shall not set it out in full detail. The main points are (a) the elements of $U(2)/U(1)$ correspond to the (proper) rotations of the sphere (note that $O(2)$ gives only a subset of the proper rotations), and (b) when we adjoin also the conjugation operator τ , this has the effect of adding the improper rotations. Indeed, τ itself is just the operation of changing the sign of ϕ , that is, reflecting the sphere in the equatorial plane $Z = 0$. Alternatively we can say that τ replaces a shape by a mirror-image of itself (this is easily checked from (36)).

Within G^{++} we have as a subgroup the symmetric group of order six which permutes the labels of the vertices; its action is determined by the following rules.

If we wish to look at shape BCA instead of shape ABC , we must rotate our point of observation about the axis OZ through an angle $+2\pi/3$; if we wish to look at shape BAC instead of shape ABC , we must switch our point of observation from (X, Y, Z) to $(X, -Y, -Z)$.

For most practical and theoretical purposes the labelling of the vertices is unimportant. Thus we can usefully identify points on $S^2(\frac{1}{2})$ which correspond to one another under the relabelling group G_6 . The result of this is that $S^2(\frac{1}{2})$ is split into six equivalent lunes (see Figure 1). As each of the two semicircular boundaries of a lune are subject to the identification $(X, Y, Z) = (X, Y, -Z)$, by repeated application of the above rules, it will be seen that the quotient of $S^2(\frac{1}{2})$ by the relabelling group G_6 is again *topologically* a 2-sphere. Figure 1 shows the upper halves ($Z \geq 0$) of the six lunes. If we fix attention on one of these as it is seen in the Figure, it has two 'straight' boundaries and one 'curved' boundary. In fact each such 'straight' boundary coincides under the identifications with the corresponding 'straight' boundary on the lower half ($Z \leq 0$) of $S^2(\frac{1}{2})$, and the two 'straight' boundaries and the 'curved' boundary together are to be thought of as an 'equator' for the *topological* sphere $S^2(\frac{1}{2})/G_6$.

In some contexts the distinction between a shape and its reflexion is very important, and when this is so the whole lune, with its identifications, or some space equivalent to this, must be used, and topologically the reduced representation space will then be a sphere as explained above. *If we want isometry, however, we must work with the lune itself.*

But in many problems the effects of reflexion can also be omitted without significant loss, and then an upper half-lune will suffice. From what we have already said it will be clear that this is *topologically* a hemisphere, and that there will be no residual identifications at the boundary. If we want isometry we must always work with the *geometrical* half-lune itself, which we can write as $S^2(\frac{1}{2})/G_{12}$, where G_{12} is the group of order 12 formed by adjoining the conjugation operator τ to G_6 .

In what follows we shall make considerable use of $S^2(\frac{1}{2})/G_{12}$, which shall be called *the spherical blackboard*, in the display of theoretical shape-functions and shape-measures as well as of empirical shape-distributions. Figure 2 shows an 'opened out' cylindrical projection from $S^2(\frac{1}{2})/G_{12}$ onto a right circular cylinder which touches $S^2(\frac{1}{2})$ along a great circle. If α denotes the acute angle between OZ and the axis of the cylinder, it will be found convenient to take

$$\tan \alpha = 2/\sqrt{3}, \quad (37)$$

and this choice will be used throughout, while it will further be arranged that the plane containing OZ and the axis of the cylinder cuts the sphere $S^2(\frac{1}{2})$ in the median great circle of the (half) lune being employed.

It is of course necessary to state which lune is being used, and we can best do this by specifying the three vertices of the curvilinear triangle LMN which is the result of the projection process. These are

- (1) $L = (r = 1/\sqrt{3}, \phi = \pi)$, A and C coincident;
- (2) $M = (r = 0, \phi \text{ unspecified})$, $C = \text{midpoint of } AB$;
- (3) $N = (r = 1, \phi = \frac{1}{2}\pi)$, ABC equilateral.

In Figure 1, this is the upper half of the lune marked ABC .

Inspection of Figure 2 will show where the various shapes now lie. Notice that the 'collinear' shapes are those on the bottom bounding arc LM , while the 'isosceles' shapes are those which lie on the two lateral bounding arcs LN and MN . Accordingly the isosceles triangles are here separated into two sets; those with base angles greater than $\pi/3$ and less than $\pi/3$ respectively.

In theoretical studies the value of the spherical blackboard plot is that it is 'area-true'; i.e. it preserves (up to a factor 12) the invariant measure $\tilde{\omega}^3 (= \gamma_*^3)$, and while the projection is not an isometry the metrical distortions are reasonably controlled. In the context of data-analysis or of simulations there is the further dramatic advantage of an automatic 12-fold increase in the effective size of the sample. A final advantage is the prominence given in the projection to

- (i) near-equilateral shapes (these will lie near the vertex N),

and

- (ii) near-collinear shapes (these will lie near the arc LM).

We now give a few further illustrations in order to enable the reader to become thoroughly familiar with the geometry. Given a shape ABC , one of the most interesting shape-functions is the max-angle function, $\max(A, B, C)$. Contours for this, viewed in three different ways, will be seen in Figures 3, 4, and 5. The reader should notice the location in each of these displays of the shapes of those triads ABC for which the max-angle exceeds 175° . These pictures show how the alignment studies in Kendall & Kendall [9] can be reduced to spherical trigonometry. All the

contours on $S^2(\frac{1}{2})$ for the max-angle function are composed of arcs of 'small' circles, and it is precisely because they are not 'great' circles that the alignment problem is so complicated. In Figure 4 two singularities of the max-angle function will be seen; one of these is the point L (at which $C = A$), and there are two others, one in view (at which $B = C$) and the third (at which $A = B$) lying out of sight at the 'back' of $S^2(\frac{1}{2})$. Careful inspection will reveal a 'small' circle passing through the two visible max-angle singularities and concentric with the circular outline of $S^2(\frac{1}{2})$. This is part of the locus $\max(A, B, C) = \frac{1}{2}\pi$; the complete locus consists of three such 'small' circles touching one another in pairs.

It may have occurred to the reader that the familiar antipodal involution of the sphere is not here enjoying the prominent role that might have been expected for it. Reference back to Theorem 1 will show that the antipode σ^* of a shape σ is that unique shape which is 'most unlike' σ on a procrustean basis, the distance between the two shapes being $\pi = \frac{1}{2}(2\pi)$. We may call σ^* the *procrustean transform* of σ ; such transforms enjoy the following elegant property.

PROPOSITION A. *Given three distinct points $A, B,$ and $C,$ let $A^*, B^*,$ and C^* be defined so that $[BCA^*], [CAB^*],$ and $[ABC^*]$ are the procrustean transforms of $[BCA], [CAB],$ and $[ABC],$ respectively. Then the triangles $ABC^*, AB^*C,$ and A^*BC are similar.*

This is a simple consequence of the geometry we have developed for $S^2(\frac{1}{2})$, for if R denotes a rotation about OZ by $+2\pi/3,$ and if -1 is the antipodal map, we have $[A^*BC] = R^2[BCA^*] = -R^2[BCA] = -[ABC],$ and $[AB^*C] = R[CAB^*] = -R[CAB] = -[ABC],$ and $-[ABC] = [ABC^*],$ which completes the proof.

If we 'translate' the proposition by inserting a concrete construction for the procrustean transforms, we obtain a non-obvious (but equivalent) proposition in euclidean plane geometry. Figure 6 illustrates this, using the construction:

if C^+ lies on CQ produced, Q being the midpoint of $AB,$ and if $CQ \cdot QC^+ = \frac{3}{4}AB^2,$ then $C^+ = C^.$*

Proof. The procrustean transform maps z to $-1/\bar{z}.$

6. The use of the 'spherical blackboard'

We next discuss some examples of the use of the 'spherical blackboard' when presenting and assessing information about shapes derived from calculation, from empirical data, or from Monte Carlo simulation.

First of all it is desirable to investigate the extent to which errors in the coordinates of the vertices of a k -ad give rise to induced errors in the location of the shape-point on the shape-manifold. We shall do this for general $k.$ Suppose then that the planar coordinates of the k vertices are subject to small independent errors with mean zero and with a given variance-covariance matrix (say with principal variances d_1^2 and d_2^2). Because of the orthogonal relationship at (1) it follows that the coordinates z_1, z_2, \dots, z_{k-1} will be subject to uncorrelated errors with these same first and second-order moment properties. From (3) and (25) it is easy to work out the expectation $\mathbb{E}(D^2),$ where here D is the geodesic arc-displacement of the shape-point

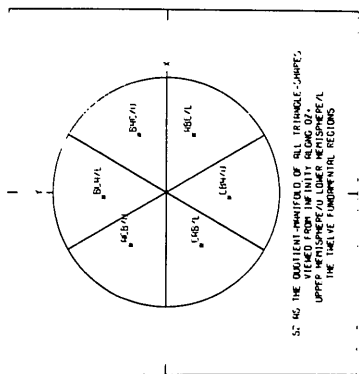


FIG. 1 The action of the re-labelling group of order 6, which divides $S^2(\frac{1}{2})$ into six lunes (here viewed from the point at plus infinity on OZ). Identifications at the boundaries give to each lune the topology of the 2-sphere.

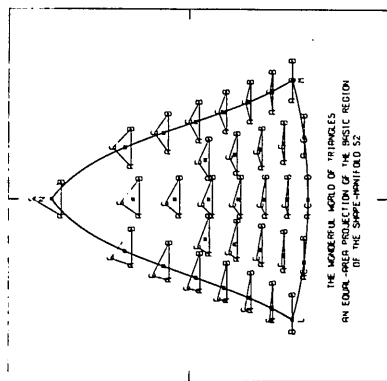


FIG. 2 The 'spherical blackboard', with 32 typical triangle-shapes 'at home'. (Reproduced from Kendall [10].)

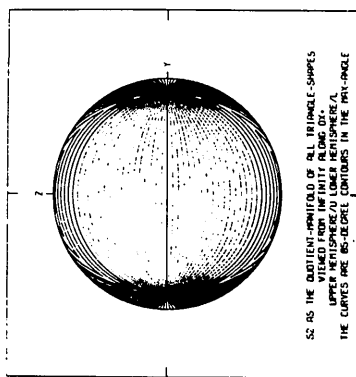


FIG. 4 $S^2(\frac{1}{2})$ viewed from the point at plus infinity on OX, showing 5° contours for the max-angle function.

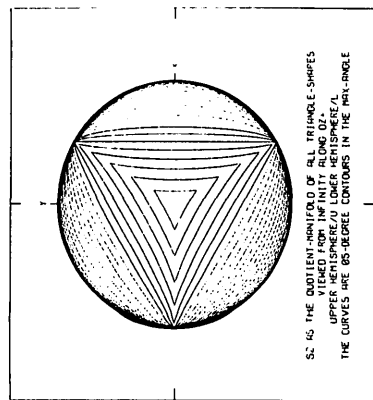


FIG. 3 $S^2(\frac{1}{2})$ viewed from the point at plus infinity on OZ, showing 5° contours for the max-angle function.

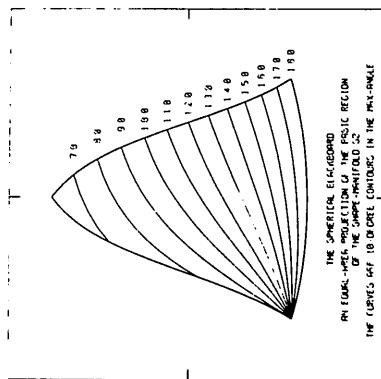


FIG. 5 The 'spherical blackboard', showing 10° contours for the max-angle function.

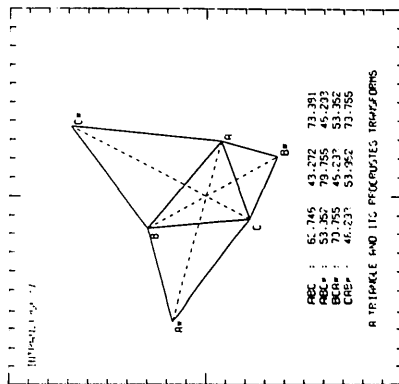


FIG. 6 Procrustean transforms.

on the manifold due to the above random errors. We find that, asymptotically for small d ,

$$\mathbb{E}(D^2) = \frac{\|z\|^2(k-1)2d^2 - \|z\|^2 2d^2}{\|z\|^4} = \frac{k-2}{k-1} \frac{d^2}{\tau^2},$$

where $2d^2 = d_1^2 + d_2^2$, and $\tau^2 = \|z\|^2/(2k-2)$, so that τ is the (linear) standard deviation of the pattern formed by the k points when calculated with $2k-2$ degrees of freedom (appropriate here because of the removal of each component of the centroid). Thus we have

THEOREM 4. *If we assess accuracy of position on the shape-manifold in terms of the geodesic root-mean-square displacement, using the riemannian metric, then asymptotically this will be $\sqrt{\{(k-2)/(k-1)\}}$ multiplied by the ratio d/τ measuring the size of the errors in the positions of the vertices relative to the size of the k -ad.*

The theorem is closely connected with the stochastic-kinematic results described in an earlier paper (Kendall [8]), but we shall not dwell on these here beyond remarking that for $d_1 = d_2$ the local error distribution for shapes is likewise isotropic. It could be used as a rational basis for choosing a suitable size for the symbol to be printed at each shape-point in the representation. In the illustrations to this paper, however, symbol-size will be chosen solely with regard to legibility.

We now return to the case $k = 3$. While it is quite interesting to see representations of data on orthogonal projections of the whole sphere, such as those used in the construction of Figures 3 and 4, it is often far more instructive to concentrate attention on the basic region LMN . In this way we can quotient out what will often be irrelevant symmetries and so concentrate the information on an adequately representative smaller region. As the obvious reference case will be that in which the distribution of the vertices of the k -ad is of the gaussian form (though not always necessarily isotropic), it is most convenient that the 'spherical blackboard' gives us a plane representation of the region LMN on which the transformed shape-distribution is uniform for an isotropic gaussian generator. The effect of non-isotropy of the generator will be of great importance to us, and we shall want to be able to recognise this, when present, from an inspection of the 'spherical blackboard' representation.

We therefore show in Figures 7 and 8 scatter plots for 1000 independent shapes of triangles having IID gaussian vertices with generator $N(0; \sigma_1^2, \sigma_2^2)$, first in the case of circular symmetry $\sigma_1 = \sigma_2$, and then with $\sigma_1 = 5\sigma_2$. In Figure 7, as our theory indicates should be so, the scatter plot is approximately uniform, while in Figure 8 there is a general retreat of points from the vertex N towards the arc LM ; this is another way of saying that ellipticity of the generator produces a higher proportion of nearly collinear triads. From Section 7 it will appear that the induced 'gaussian' shape-density when $\sigma_1 \neq \sigma_2$ is a function of the cartesian coordinate Z only, and Figure 8 bears that out. (Remember that $Z = 0$ on LM , and that Z increases as we move up towards N). We shall also see in Section 7 that the 'Broadbent factor',

$$f = \frac{1}{2} \left(\frac{\sigma_1}{\sigma_2} + \frac{\sigma_2}{\sigma_1} \right), \quad (38)$$

gives an appropriate indication of the degree of ellipticity, and indeed (Kendall & Kendall [9]) there is a usefully wide range of cases in which this factor correctly

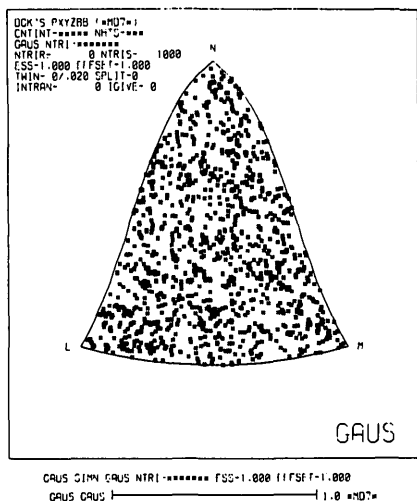


FIG. 7 One thousand triangle-shapes; IID vertices with generator = $N(0; \sigma_1^2, \sigma_2^2)$ ($\sigma_1 = \sigma_2$).

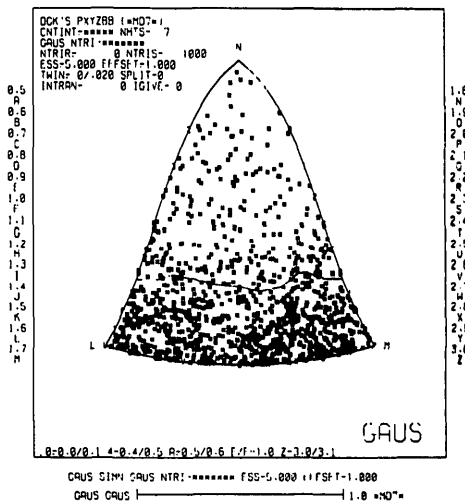


FIG. 8 One thousand triangle-shapes; IID vertices with generator = $N(0; \sigma_1^2, \sigma_2^2)$ ($\sigma_1 = 5\sigma_2$). The contour line is drawn at level 1.0 for the empirical shape-density.

adjusts the expected number of triads with $\max(A, B, C) \geq \pi - \epsilon$ (ϵ small). In Figure 8 we have $f = 2.6$, and we shall see later that for $k = 3$ the theoretical shape-density m^3 at (28) is constant along the arc LM for gaussian generators, and is uniformly enhanced there by this factor 2.6 when σ_1/σ_2 is increased from 1 to 5.

Another class of models which has been much studied is that in which the IID vertices are uniformly distributed over a compact convex set K (see Kendall & Kendall [9], and Small [24]). It is now appropriate to rewrite f at (38) in the form

$$f = \frac{1}{2} \left(\frac{\kappa_1}{\kappa_2} + \frac{\kappa_2}{\kappa_1} \right), \tag{39}$$

thus obtaining an equivalent 'dynamical' measure of ellipticity. Here κ_1 and κ_2 are the radii of gyration for a thin plate of uniform density and of form K , the gyration being about the principal axes of inertia through the centroid and in the plane of the plate. Even when $K =$ circular disk, the shape-density at (28) is no longer constant along the arc LM , but we do have the interesting fact that

$$\frac{d(\lambda | \text{ellipse})_*^3}{d(\lambda | \text{circle})_*^3} = f \text{ at all points of } LM. \tag{40}$$

This is a special case of one of Small's results, and his generalisation of it is very important for the alignment problem. Small also computed and drew contours for the associated shape-densities using numerical integration. Recently (Kendall [12]) it has proved possible to evaluate the shape-density $\frac{d(\lambda | \text{circle})_*^3}{d\gamma_*^3}$ in closed form.

Figure 9 shows contours for this density, which agree well with those obtained from Small's numerical integrations. The same analytical approach could also in principle be applied when K is a polygon, but the details would be rather unpleasant, and this has not yet been done.

One suspects that the form of the non-uniformity of the induced shape-density $d(\lambda | K)_*^3/d\gamma_*^3$ may betray in a perhaps far from obvious way the shape of ∂K (here a circle), and the uniqueness questions of Section 4 become relevant. For some theoretical studies of this problem see Small ([24, 25]).

It is easy to make a scatter-plot of a large number of independent shape-points obtained by simulation when K is an $(s:1)$ -rectangle. The non-uniformity is not particularly marked in a scatter-plot for $s = 1$ (K a square), but it can be detected by vastly increasing the number of shape-points and drawing rough density contours automatically. This is done in Figure 10 with $s = 1$ (and in Figure 11 with $s = 1.661$) for samples of size 100,000. The long contours are thought to be meaningful, while the small ones and the minor crinkles exhibit the statistical noise level.

If we take as a 52-ad the data-set analysed in Broadbent [3] and Kendall & Kendall [9], it supplies us with a total of 22,100 triangle shapes, and contours for the empirical shape-density are shown in Figure 12. It should be borne in mind that these 22,100 shapes are not independent, although they are 'dissociated' (Silverman & Brown [23]). There is a slight ellipticity in the distribution here ($\hat{\sigma}_2/\hat{\sigma}_1 = 1.661$, $f = 1.13$); thus we can fairly compare Figure 12 with Figure 11, at least as a first approximation. It appears that the increased density near the vertex L in Figure 12 may be attributed to the occurrence of a few pairs of near neighbours in the generating pattern, while the change in the prevailing course of the long contours may reflect the tendency of the 52 generating points to segregate into two groups. Figure 13 shows contours for 100,000 shape-points derived from a comparable bimodal mixed-gaussian generator with an artificial supply of about the right number of closed pairs, and this supports the interpretation just offered.

A striking example of a theoretical model in which the generating mechanism is not one based on IID sampling is provided by the Poisson-Delaunay process. Here we start with an infinite 2-dimensional Poisson process, and construct the Delaunay tessellation (Delone [4]), and

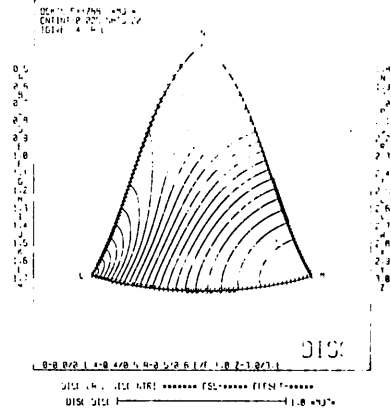


FIG. 9 Contours for $d(\lambda | \text{circle})_*^3/d\gamma_*^3$ computed from the explicit formula in (Kendall [12]). The minimum density $3/4$ occurs at M , and the maximum $4/3$ occurs at L . The density at N is $3 - (27\sqrt{3})/(8\pi) = 1.1393$.

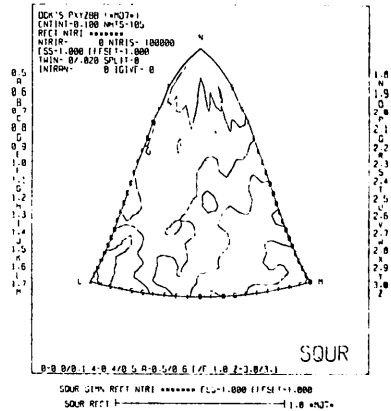


FIG. 10 Simulation contours for $d(\lambda | \text{rectangle})_*^3/d\gamma_*^3$ (sample size = 100,000; $s = 1$).

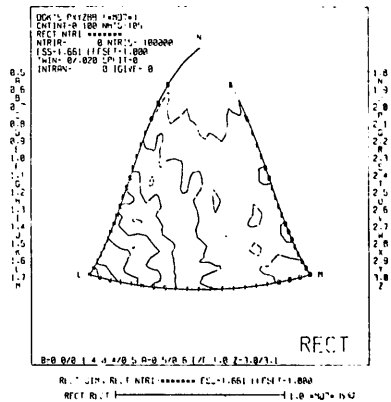


FIG. 11 Simulation contours for $d(\lambda | \text{rectangle})_*^3/d\gamma_*^3$ (sample size = 100,000; $s = 1.661$).

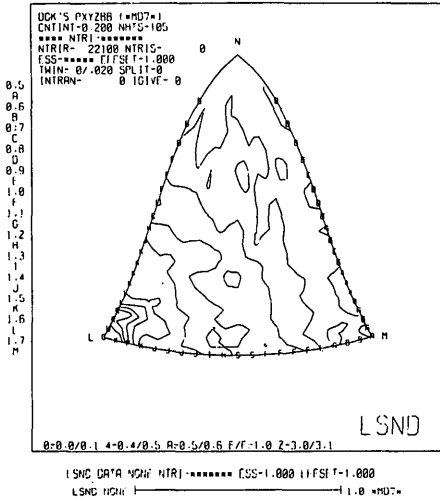


FIG. 12 Contours for an empirical shape-density (sample size 22,100; $\hat{s} = 1.661$): data from (Broadbent [3]).

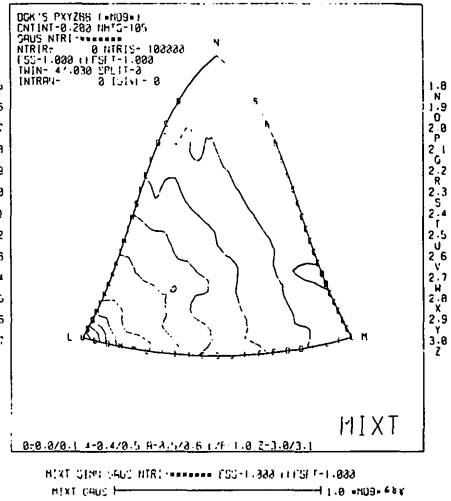


FIG. 13 Simulation contours for a bimodal mixed-gaussian generator (sample size = 100,000; $s = 1.661$, 2 groups, 4 close pairs).

then observe the shape of the triangular 'tile' containing the origin. Repetition of this procedure yields a random sample from an extraordinary theoretical shape-distribution discovered and first studied by R. E. Miles, though not from the present point of view. I have recently shown [11] that its shape-density can be written in the form

$$\frac{4}{9}(\sin^2 A + \sin^2 B + \sin^2 C)^2, \tag{41}$$

where $A, B,$ and C are the angles at the vertices of the triangle. This is the Radon–Nikodym density relative to $\gamma_*^3 (= \tilde{\omega}^3)$; it is more appropriate for our purposes than the earlier form for the shape-density obtained by Miles relative to an ad hoc measure $dBdC$. Figure 14 shows 5000 such Poisson–Delaunay random shapes, and Figure 15 shows contours for the function at (41). (The contour-drawing routine fails

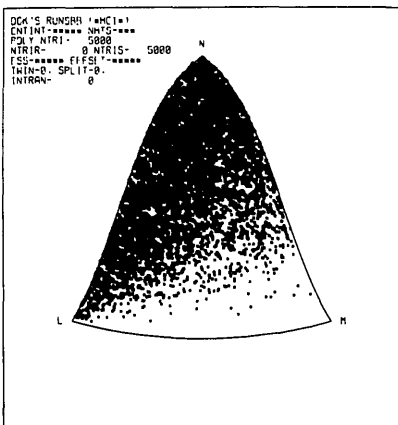


FIG. 14 Five thousand Poisson–Delaunay shapes (from Kendall [11]).

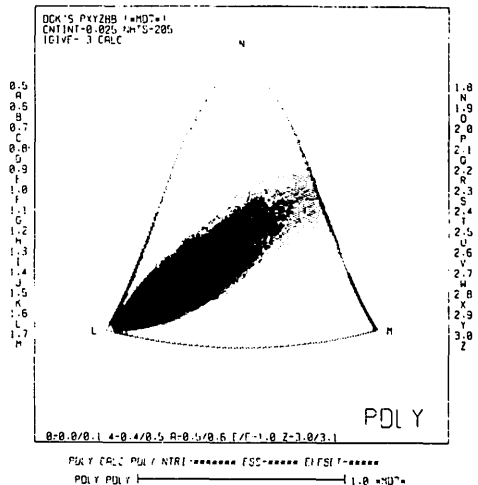


FIG. 15 Contours for the theoretical Poisson–Delaunay shape-density (from Kendall [11]). It is discontinuous at the point marked L .

in the vicinity of L because the function at (41) is not continuous at this point.) It is interesting to compare these pictures with shape-plots for all the Delaunay triangles associated with a *finite* Poisson sample, e.g. with a set of N points IID and uniform in a square. In that case one obtains the theoretical Poisson–Delaunay distribution of about $2N$ triangle-shapes, supplemented by $O(\sqrt{N})$ others associated with ‘false’ Delaunay triangles at the boundary, and unless N is very large these tend to fill in the void near the arc LM . For further details see Kendall [11].

7. *The shape-density for a non-isotropic gaussian generator*

We now return to Σ_2^k carrying the invariant measure $\gamma_*^k (= \tilde{\omega}^k)$; we shall determine the shape-measure $\gamma_*^{k,s}$ induced by the bivariate gaussian distribution with principal component variances s and $1/s$; here the Broadbent coefficient is defined to be

$$f = \frac{1}{2} \left(s + \frac{1}{s} \right) \quad (1 \leq s < \infty). \tag{42}$$

We must not expect $\gamma_*^{k,s}$ to be invariant under G^{++} for $s > 1$, for then it would have to coincide with $\gamma_*^k = \gamma_*^{k,1}$. For this reason we use a more direct method than that employed in Section 4.

Because of the real-orthogonal nature of the transformation (1) we can assume that all of z_1, z_2, \dots, z_{k-1} have the generating gaussian distribution, and that they are independent. As in Section 4 we write

$$z_1 = r e^{i\phi}, z_{h+1} = r r_h e^{i(\phi + \phi_h)} \quad (h = 1, 2, \dots, k-2),$$

and we shall perform our initial calculations in the coordinate-patch of $\mathbb{C}P^{k-2}$ for which $z_1 \neq 0$. We can write the joint distribution of z_j ($j = 1, 2, \dots, k-1$) as

$$\exp \left\{ -\frac{1}{2} r^2 Q_0 - \frac{1}{2} r^2 \sum_1^{k-2} r_h^2 Q_h \right\} \left(\frac{1}{2} r^2 \right)^{k-2} d\left(\frac{1}{2} r^2 \right) d\left(\frac{\phi}{2\pi} \right) \prod_1^{k-2} d(r_h^2) d\left(\frac{\phi_h}{2\pi} \right),$$

where $Q_h = s^{-1} \cos^2(\phi + \phi_h) + s \sin^2(\phi + \phi_h)$, and where Q_0 is defined similarly with the argument ϕ instead of $\phi + \phi_h$. If we re-arrange this and then integrate out the redundant coordinates r and ϕ , we get

$$d\gamma_*^{k,s} = (k-2)! Q(\zeta_1, \zeta_2, \dots, \zeta_{k-2}) \prod_1^{k-2} d(r_h^2) d\left(\frac{\phi_h}{2\pi} \right),$$

when

$$Q(\zeta_1, \zeta_2, \dots, \zeta_{k-2}) = \frac{1}{2\pi} \oint (A \cos^2 \phi + 2H \sin \phi \cos \phi + B \sin^2 \phi)^{-(k-1)} d\phi,$$

and

$$A = s^{-1} (1 + \sum x_h^2) + s \sum y_h^2,$$

$$H = (-s^{-1} + s) \sum x_h y_h,$$

$$B = s^{-1} \sum y_h^2 + s (1 + \sum x_h^2);$$

here we have written $\zeta_h = r_h e^{i\phi_h} = x_h + iy_h$. Now

$$A \cos^2 \phi + 2H \sin \phi \cos \phi + B \sin^2 \phi = \frac{1}{2}(A+B) + \frac{1}{2}(A-B) \cos 2\phi + H \sin 2\phi,$$

and here $(A-B)^2 + 4H^2 < (A+B)^2$, so that we can write

$$A \cos^2 \phi + 2H \sin \phi \cos \phi + B \sin^2 \phi = \lambda \{v + (v^2 - 1)^{\frac{1}{2}} \cos(2\phi + \varepsilon)\}$$

with $\lambda > 0$ and $v \geq 1$; in fact we shall have $\lambda = \sqrt{(AB - H^2)}$ and $v = (A+B)/(2\lambda)$. Notice that $v = 1$ and $\lambda = 1 + \sum (x_h^2 + y_h^2)$ when $s = 1$.

We now call in Laplace's 'second' integral representation formula for the Legendre polynomial $P_n(v)$ and obtain

THEOREM 5. *The Radon-Nikodym derivative of $\gamma_*^{k,s}$ with respect to $\gamma_*^{k,1}$, which is the factor correcting the invariant measure $\gamma_*^{k,1} = \gamma_*^k = \tilde{\omega}^k$ on Σ_2^k for the effect of the ellipticity s , is given by*

$$\frac{d\gamma_*^{k,s}}{d\gamma_*^{k,1}} = \left(\frac{2v}{s+s^{-1}} \right)^{k-1} P_{k-2}(v), \quad (43)$$

where

$$1 - \frac{1}{v^2} = \left(\frac{s^2 - 1}{s^2 + 1} \right)^2 \left(\frac{\left| \sum_{j=1}^{k-1} z_j^2 \right|}{\sum_{j=1}^{k-1} |z_j|^2} \right)^2. \quad (44)$$

Notice that v is a symmetric zero-degree complex-homogeneous function of $(z_1, z_2, \dots, z_{k-1})$; this confirms that it is a shape-function, and that our proof is generally valid although worked out in the special case $z_1 \neq 0$. When the generator is isotropic we shall have $s = 1$ and $v = 1$, and the right-hand side of (43) then reduces to unity as it ought to do.

From Laplace's other ('first') integral for $P_n(v)$ we have

$$P_n(v) = \frac{1}{2\pi} \int_0^\pi \left\{ (v + (v^2 - 1)^{\frac{1}{2}} \cos \phi)^n + (v - (v^2 - 1)^{\frac{1}{2}} \cos \phi)^n \right\} d\phi,$$

and if we now expand the binomials, the only terms that remain are positive multiples of integrals of

$$v^{n-r} (v^2 - 1)^{\frac{1}{2}r} \cos^r \phi = v^n (1 - v^{-2})^{\frac{1}{2}r} \cos^r \phi,$$

for even r , so it is clear that $P_n(v)$ increases monotonically for $1 \leq v \uparrow \infty$. It follows that the density (43) is least when $v = 1$, that is (for $s > 1$) when $\sum z_j^2 = 0$, that is, when the k -ad has dynamical symmetry. On this set the density (43) has the value $1/f^{k-1}$ where f is the Broadbent factor (42).

In the same way we see that the density (43) has its largest value when v is greatest, that is, when the ratios

$$z_1 : z_2 : \dots : z_{k-1}$$

are all real, so that $v = f$. The set of shapes on which this happens is the real projective space $\mathbb{R}P^{k-2}$ naturally imbedded in $\mathbb{C}P^{k-2}$, and the value of (43) is then

$$P_{k-2}(f). \tag{45}$$

This is the factor by which a degree of ellipticity s increases the chance of an approximate multiple alignment, because the Radon–Nikodym derivative (43) is C^α (indeed analytic) everywhere on Σ_2^k , and so is asymptotically equal to $P_{k-2}(f)$ on any small neighbourhood of the multiple-alignment set $\mathbb{R}P^{k-2}$. It follows (cf. Kendall & Kendall [9], and W. S. Kendall [15]) that in the gaussian case all multiple alignment studies ‘factorize’ into (i) trigonometry on $\mathbb{C}P^{k-2}$ (evaluating the $\tilde{\omega}^k$ -measure of the prescribed neighbourhood of $\mathbb{R}P^{k-2}$), and (ii) multiplication of the ‘isotropic’ probability obtained at (i) by the correcting factor (45). Small [24] noted that there is a similar phenomenon when the generator is the uniform distribution over a circular disk (elliptic disk).

Another corollary to Theorem 5 is that the Radon–Nikodym derivative (43) when considered as a function defined over Σ_2^k depends on the shape-point only via the shape-function

$$\Gamma = \left| \frac{\sum_1^{k-1} z_j^2}{\sum_1^{k-1} |z_j|^2} \right|, \tag{46}$$

and this remark enables us to determine the subgroup of G^{++} with respect to which $\gamma_*^{k,s}$ is invariant when $s > 1$. First of all it is clear that τ leaves (46) unchanged, as also does $O(k-1)$; note however that we must here replace $O(k-1)$ by $SO(k-1)$ when k is even because then $\text{diag}(-1, -1, \dots, -1)$ can be absorbed into $SO(2)$ (acting on the left). In fact we have

THEOREM 6. *The ‘skew gaussian’ measure $\gamma_*^{k,s}$ on Σ_2^k is invariant under (and only under) the subgroups*

$$O(k-1) \otimes \{1, \tau\} \text{ when } k \text{ is odd,}$$

$$SO(k-1) \otimes \{1, \tau\} \text{ when } k \text{ is even.}$$

For let $U = A + iB$ be an element of $U(k-1)$ which leaves (46) invariant; plainly it must map the set $\{\sum z_j^2 = 0\}$ into itself, and so must map all (row) z -vectors of the form $(0, \dots, 0, 1, \dots, 0, \pm i, 0, \dots, 0)$ into this set. Thus if the complex vectors λ and μ denote different rows of U , we must have $\sum (\lambda_j \pm i\mu_j)^2 = 0$ for either uniform choice of the ambiguous sign, and therefore

$$\sum \lambda_j^2 = \sum \mu_j^2, \quad \text{and} \quad \sum \lambda_j \mu_j = 0.$$

This however implies that UU' is a multiple ρI of the unit matrix with $\rho = e^{i\theta}$. Without loss of generality we can divide out by $e^{i\theta/2}$ (which can be absorbed into $SO(2)$), and so we get $UU' = I = \bar{U}U'$, whence $U = \bar{U}$, and therefore U is real and belongs to $O(k-1)$ as was claimed.

When $k = 3$, we find that

$$\Gamma^2 = \frac{|1 + r^2 e^{2i\theta}|^2}{(1 + r^2)^2} = 1 - Z^2,$$

where Z is the third coordinate in (35), and for a gaussian generator with $s = 1$ we know that $|Z|$ has the uniform distribution on $[0, 1]$. Moreover Z is unaltered by $SO(2)$ acting on the *right* (which rotates $S^2(\frac{1}{2})$ about the Z -axis), and so it depends only on λ_1 and λ_2 in (12). In fact

$$Z = \frac{2\lambda_1\lambda_2}{\lambda_1^2 + \lambda_2^2},$$

and

$$\Gamma^2 = 1 - 4 \frac{\lambda_1^2 \lambda_2^2}{(\lambda_1^2 + \lambda_2^2)^2}. \quad (47)$$

This suggests that we should try to find a transformation like this when $k > 3$ which will enable us to replace Γ by a statistic having a simple law of distribution in the null case when $s = 1$.

Now (47) as it stands is true independently of the value of k , because

$$\Gamma^2 = 1 - 4 \frac{\sum x_j^2 \sum y_j^2 - (\sum x_j y_j)^2}{(\sum x_j^2 + \sum y_j^2)^2},$$

where now we have written $z_j = x_j + iy_j$ for $1 \leq j \leq k-1$, and this leads immediately to (47) when we notice that λ_1^2 and λ_2^2 are the eigenvalues of WW' , which is essentially the sample variance-covariance matrix. Obviously we are now very close to classical arguments associated with the Wishart distribution.

We accordingly look for a connexion with work by Mauchly [16] and by Girshick [6] on 'testing for sphericity'. Mauchly proposed the use of

$$L = 2 \frac{|\lambda_1 \lambda_2|}{\lambda_1^2 + \lambda_2^2}, \quad (48)$$

as a test-statistic for use in a sphericity problem arising in connexion with terrestrial magnetism; note that

$$\Gamma^2 = 1 - L^2. \quad (49)$$

He obtained the law of distribution

$$(k-2)L^{k-3} dL \quad (0 \leq L \leq 1) \quad (50)$$

in the null case when $s = 1$. (When $k = 3$ this reduces to the uniform distribution of $|Z|$, as noted above.) Girshick then gave the L -distribution in the non-null case ($s > 1$ in our notation) in infinite series form as

$$\frac{\mathcal{L}^{k-1} L^{k-3}}{(k-3)!} \sum_{m \geq 0} \frac{(2m+k-2)!}{2^{2m}(m!)^2} \{(1-\mathcal{L}^2)(1-L^2)\}^m dL, \quad (51)$$

where \mathcal{L} is the parameter

$$\mathcal{L} = 2 \frac{\sqrt{(s \cdot s^{-1})}}{s + s^{-1}} = \frac{1}{f}. \quad (52)$$

It is clear that if we now multiply the Mauchly distribution (50) by our likelihood ratio (43), we must get the Girshick distribution (51), but in a new form involving a

closed expression, instead of his infinite series! On carrying out this calculation and comparing the two results we obtain the mathematical identity,

$$v^{n+1} P_n(v) = \frac{1}{n!} \sum_{m \geq 0} \frac{(2m+n)!}{(m!)^2} \left(\frac{v^2-1}{4v^2} \right)^m, \tag{53}$$

which is easily proved independently by expanding the integrand of

$$\frac{1}{\pi} \int_0^\pi \left\{ 1 + \sqrt{\left(\frac{v^2-1}{v^2} \right) \cos \phi} \right\}^{-(n+1)} d\phi$$

in a power-series, rejecting the odd powers of $\cos \phi$, performing the term-by-term integration, evaluating the gamma functions, and using the duplication formula.

Those readers who are intrigued by these interrelationships may like to try to derive the Mauchly distribution (50) directly from our formula (31), of which it must be a consequence.

At least three different methods have been suggested for testing a single k -ad for suspiciously close ‘alignment’, or an n -ad for the presence of suspiciously many multiple close ‘alignments’ of the k -th order ($3 \leq k < n$). The present discussion points rather strongly to yet another method, the use of the statistic

$$U = L^{k-2} \tag{54}$$

in this role, with $L = 0$ corresponding to perfect ‘alignment’. Its relevance is obvious, it is uniformly distributed in the null (gaussian) case, and we also know its law of distribution in the non-null (gaussian) case; indeed something is also known about non-gaussian cases (see the report on these matters in Muirhead [17]). The Mauchly distribution is sufficient if we are testing a single k -ad, for it will give us the exact P -value in the null gaussian case.

When we are testing a sample of n points for too many close ‘alignments’ of the k -th order, then we have, as was pointed out by Kendall & Kendall [9], an extra complication in that we must either (i) test the number $N(u)$ of k -ads for which $U < u$, where u is a *fixed* small positive number, *given in advance*, or (ii) test

$$\max_{u \in J} N(u), \tag{55}$$

where J is a *fixed* interval $[u_1, u_2]$ (with $0 < u_1 < u_2 < 1$, u_2 small), *given in advance*. The first case is straightforward, and the second can be resolved by the ‘pontogram’ technique and the use of laterally perturbed data-based simulations as developed by Kendall & Kendall.

The obvious snag in the use of U and $N(u)$ as test statistics is that there is no immediate intuitive interpretation for U , but this would not, I think, be an obstacle in practice. An example will be found in the next section.

To round off the present discussion we make the following observation which goes some way towards conferring intuitive status on U .

We have already seen that we can adequately study an arbitrary point of $\mathbb{C}P^{k-2}$ (relative to the symmetry group G^+ acting on the right) by confining attention to

those shape-points of the form $[z] = [(z_1, z_2, \dots, z_{k-1})]$ for which

$$z_1 = \cos \theta, \quad z_2 = i \sin \theta, \quad \text{all other } z_j = 0, \quad (56)$$

where $0 \leq \theta \leq \frac{1}{4}\pi$. (This is just another way of phrasing the argument at (14) in the special case when $m = 2$.) We want $[z]$ not to correspond to an aligned k -ad, so we here take $\theta > 0$. Also we can write a general shape-point in the 'aligned' subset $\mathbb{R}\mathbb{P}^{k-2}$ as $[w]$, where

$$w = (w_1, w_2, \dots, w_{k-1}),$$

every w_j is real, and $\sum w_j^2 = 1$. The distance ρ from $[z]$ to $[w]$ can be calculated from (22) and (21); it is

$$\arccos |w_1 \cos \theta + i w_2 \sin \theta| = \arccos \sqrt{(w_1^2 \cos^2 \theta + w_2^2 \sin^2 \theta)}.$$

The (uniquely attained) minimum of this distance is θ (corresponding to the equivalent choices $w_1 = \pm 1$ and all other $w_j = 0$). Thus θ is the shortest geodesic distance from such a point $[z]$ to the alignment set, and $[(1, 0, \dots, 0)]$ is the (unique) nearest point in the alignment set. If we now take into account the fact that

$$|\Gamma| = \cos 2\theta \quad \text{and} \quad L = \sin 2\theta, \quad (57)$$

we are left with

THEOREM 7. *If a shape-point σ on $\Sigma_2^k = \mathbb{C}\mathbb{P}^{k-2}$ has 'ellipticity' $L(\sigma)$, as defined at (48), where λ_1^2 and λ_2^2 are the eigenvalues of WW' , then its shortest geodesic distance from the alignment set $\mathbb{R}\mathbb{P}^{k-2}$ is $\frac{1}{2} \arcsin L(\sigma)$, with maximum possible value $\frac{1}{4}\pi$ corresponding to $L(\sigma) = 1$, $\lambda_1 = |\lambda_2|$ and a shape σ having 'dynamical symmetry'.*

If the k -ad is generated by an isotropic gaussian distribution for the k IID vertices, then $U(\sigma) = L(\sigma)^{k-2}$ is uniformly distributed on $[0, 1]$, while if this gaussian generator is anisotropic with parameter s , then that distribution of $L(\sigma)^{k-2}$ has to be modified by the likelihood-ratio factor

$$[1 + (f^2 - 1)L^2]^{-\frac{1}{2}(k-1)} P_{k-2}(f/\sqrt{[1 + (f^2 - 1)L^2]}), \quad (58)$$

where

$$f = \frac{1}{2}(s + s^{-1}).$$

The reader may wonder why L (or U) can be used both as a test for sphericity ($\lambda_1 = |\lambda_2|$), and also as a test for collinearity ($\lambda_2 = 0$). Further study of the geometry of $\mathbb{C}\mathbb{P}^{k-2}$ provides an illuminating explanation of this. Let us write Sph for the (compact) 'sphericity set' where $|\lambda_2| = \lambda_1$, and Coll for the (compact connected) collinearity set where $\lambda_2 = 0$; let R denote the remaining region $\mathbb{C}\mathbb{P}^{k-2} \setminus (\text{Sph} \cup \text{Coll})$. Then we have

THEOREM 8. *Each point $\sigma \in R$ has a unique nearest point P in Coll and a unique nearest point Q in Sph, and σ lies on the shortest geodesic arc from P to Q , with distances*

$$\rho(\sigma, P) = \frac{1}{2} \arcsin L(\sigma)$$

and

$$\rho(\sigma, Q) = \frac{1}{4}\pi - \frac{1}{2} \arcsin L(\sigma),$$

the whole arc PQ having length $\frac{1}{4}\pi$.

For the preceding argument shows that

$$P = [(1, 0, \dots, 0)]$$

is the unique point in Coll whose geodesic distance from

$$\sigma = [(\cos \theta, i \sin \theta, 0, \dots, 0)]$$

has the minimum possible value $\frac{1}{2} \arcsin L(\sigma)$, and the connecting minimal geodesic arc is the stretch $0 \leq \alpha \leq \theta$ of the geodesic whose general point is

$$[(\cos \alpha, i \sin \alpha, 0, \dots, 0)],$$

where the range $0 \leq \alpha \leq \pi$ gives the whole of this geodesic (a closed loop of length π). We now write

$$Q = \left[\left(\frac{1}{\sqrt{2}}, \pm \frac{i}{\sqrt{2}}, 0, \dots, 0 \right) S \right]$$

for a generic point in Sph, where S is in $\text{SO}(k-1)$ and only the top two rows of S actually contribute to the definition of Q .

Notice that we do not need the optional sign when $k \geq 4$, because it is then adequately catered for by the freedom in the choice of S . The triviality of the case $k = 3$ (when Coll is the equator $Z = 0$, and Sph consists of its two poles $Z = \pm 1$) therefore allows us to omit the optional sign altogether in what follows.

Now the distance from σ to Q is given by (22) and (21), and if we write $S = (s_{ij})$ then this is

$$\rho = \arccos \left| \sum_1^{k-1} Z_j \bar{W}_j \right|,$$

where

$$Z_1 = \cos \theta, \quad Z_2 = i \sin \theta, \quad Z_j = 0 \quad (j \geq 3),$$

and

$$\bar{W}_j = (s_{1j} - i s_{2j}) / \sqrt{2} \quad (1 \leq j \leq k-1),$$

so that $\rho = \arccos \sqrt{U}$, where

$$\begin{aligned} 2U &= (\cos \theta s_{11} + \sin \theta s_{22})^2 + (-\cos \theta s_{21} + \sin \theta s_{12})^2 \\ &= \cos^2 \theta (s_{11}^2 + s_{21}^2) + 2 \sin \theta \cos \theta (s_{11} s_{22} - s_{21} s_{12}) + \sin^2 \theta (s_{22}^2 + s_{12}^2). \end{aligned}$$

For a minimal distance we must make U maximal, and this requires

$$s_{11}^2 + s_{21}^2 = s_{11} s_{22} - s_{21} s_{12} = s_{22}^2 + s_{12}^2 = 1.$$

Notice however that $\text{SO}(2)$ acting on the left here commutes with $(1/\sqrt{2}, i/\sqrt{2}, 0, \dots, 0)$ in the sense that

$$\begin{pmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 0 & 0, \dots, 0 \\ 0 & 1/\sqrt{2} & 0, \dots, 0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \cos \phi & \sin \phi & 0, \dots, 0 \\ -\sin \phi & \cos \phi & 0, \dots, 0 \end{pmatrix},$$

and so we can ‘rotate’ the top two rows of S as convenient, replacing them for example by an array of the form

$$\begin{pmatrix} s_{11}, & s_{12}, & s_{13}, & \dots, & s_{1,k-1} \\ 0, & s_{22}, & s_{23}, & \dots, & s_{2,k-1} \end{pmatrix}$$

with $s_{11} \geq 0$ and $s_{21} = 0$. Accordingly the minimality conditions become

$$s_{11} = 1, \quad s_{21} = 0, \quad s_{22} = 1, \quad s_{12} = 0,$$

and all other elements in the top two rows of S now necessarily vanish, leaving us with the unique point

$$Q = [(1/\sqrt{2}, i/\sqrt{2}, 0, \dots, 0)]$$

in Sph at the minimum distance $\frac{1}{4}\pi - \theta$ from σ . This point Q lies on the geodesic described above at the point $\alpha = \frac{1}{4}\pi$, and $\text{arc } PQ = \frac{1}{4}\pi$.

When $k = 3$ the set Sph is disconnected into two isolated points, but otherwise there is no change. The arc PQ now becomes the half-meridian through σ .

8. Applications

8.1. *The collinearity-constants λ, μ , and ν .* In Kendall & Kendall [9] an asymptotic expression was found for the variance of the number $N(\varepsilon)$ of triads aligned at the angular level ε when the n points contributing to the triads form an IID sample from a given generating law in \mathbb{R}^2 . This expression depends on that law through three numerical parameters there called λ, μ , and ν , and of these λ was shown in a wide range of anisotropic cases to be proportional to the Broadbent coefficient f . When the generator is gaussian, the arguments of Section 7 tell us that ν must be proportional to $P_2(f) = \frac{1}{2}(3f^2 - 1)$; this follows from our remark at (45) for $k = 4$. We can make no comparable statement about μ , because here the associated calculations take us well out into the complement of $\mathbb{R}P^{k-2}$ in $\mathbb{C}P^{k-2}$.

8.2. *The Z-distribution.* Some of the formulae of Section 7 simplify considerably when $k = 3$. We then have $L = |Z|$, and so in the anisotropic gaussian case the Z -distribution is

$$\frac{\frac{1}{2}f}{[1 + (f^2 - 1)Z^2]^{3/2}} dZ \quad (-1 \leq Z \leq 1); \tag{59}$$

note that this reduces to $\frac{1}{2}dZ$ when $s = 1$. If Ψ denotes a rotation angle about OZ , then Ψ and Z are independent and Ψ has a uniform distribution on $[0, 2\pi)$ for all $s \geq 1$.

The cumulative law for Z is sometimes useful; it is

$$\text{pr}(Z \leq t) = \frac{1}{2} + \frac{1}{2}ft/\sqrt{[1 + (f^2 - 1)t^2]}. \tag{60}$$

Other useful results are

$$\mathbb{E}|Z| = \frac{1}{f+1},$$

$$\mathbb{E}Z^2 = \frac{f}{(f^2-1)^{3/2}} \log [f + \sqrt{(f^2-1)}] - \frac{1}{f^2-1}.$$

The last formula is badly behaved near $s = 1, f = 1$, and then we do better to write $s = e^g, f = \cosh g$, and

$$\mathbb{E}Z^2 = \frac{g \cosh g - \sinh g}{\sinh^3 g} = \frac{1}{3} - \frac{2}{15}g^2 + \dots \quad (g \text{ small}).$$

These results may be illustrated by considering the situation treated in Figure 8, where $s = 5, f = 2.6, g = 1.6094, \sinh g = 2.4, \mathbb{E}|Z| = 0.2778, \mathbb{E}Z^2 = 0.1291$, and $\text{rms}(Z) = 0.3593$. The shape-density 1.0 is attained when $Z = 0.3933$; the empirical contour line for that level is shown and the reader can check its position by referring to Figure 19.

The law (59) also makes sense for $0 \leq f \leq 1$, when it gives a bipolar distribution on the sphere, instead of one with an equatorial mode. Figure 16 shows the Z -density for $f = 0.4(0.05)1.0(0.5)6.0$.

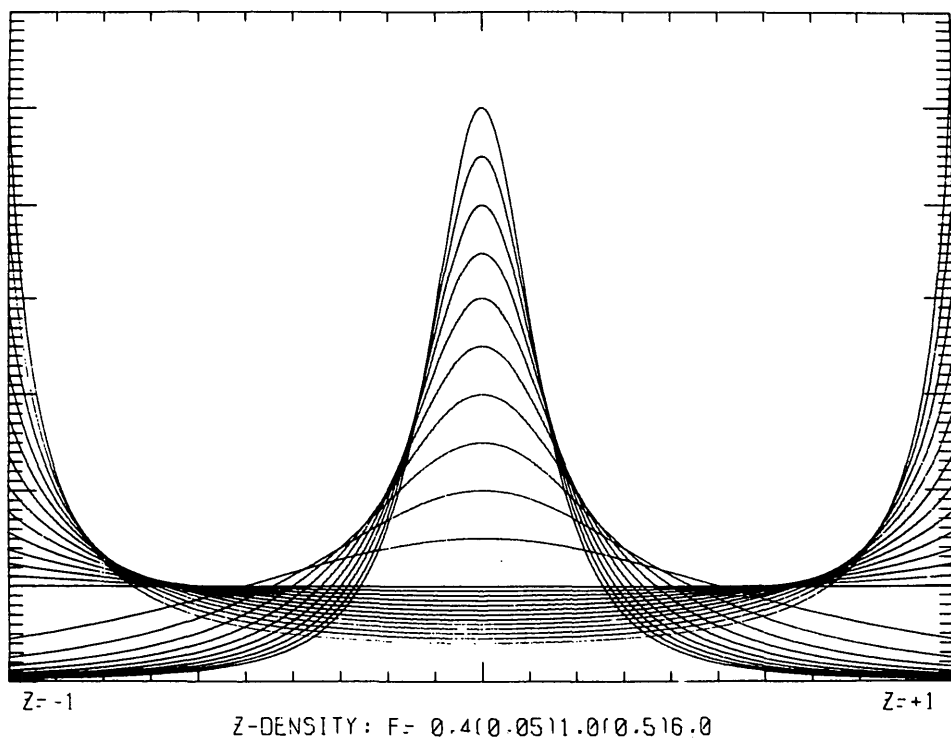


FIG. 16 The Z -distribution for $0.4 \leq f \leq 6.0$.

Let us write, for $0 < g < \infty$,

$$T_g t = \frac{gt}{\sqrt{[1+(g^2-1)t^2]}} \quad (-1 \leq t \leq 1); \tag{61}$$

notice that $T_g t$ increases when t increases, and $T_g(\pm 1) = \pm 1$. Also $T_1 = I$,

$$T_g T_h = T_{gh}, \quad \text{and} \quad (T_g)^{-1} = T_{1/g}. \tag{62}$$

Thus we have

THEOREM 9. *If Z has the distribution (59) with parameter f , then $T_g Z$ has the distribution (59) with parameter f/g , and so $T_f Z$ is uniform on $[-1, 1]$.*

We leave the (easy) proof of Theorem 9 to the reader.

Thus if we are to compare two empirical shape-distributions with anisotropies f_1 and f_2 , and if we wish to discount differences associated with $f_1 \neq f_2$, we can do this by applying $T_{\sqrt{(f_1/f_2)}}$ to the first distribution and $T_{\sqrt{(f_2/f_1)}}$ to the second.

8.3. 'Collinear' quasars? When examining empirical data on the 'spherical blackboard' it is often useful to superimpose the image of a suitable curvilinear orthogonal fret. Thus Figures 17, 18, and 19 show the positions and shapes of triangles in the famous set of 6 quasars discovered by Arp & Hazard (see Zuiderwijk [27]), superimposed in Figure 18 on a fret based on contours for the max-angle C and the ratio-of-sides AC/BC , and in Figure 19 (quasars X, Y, Z, B only) on a fret based on contours for Z and Ψ . There are 3 marked triple collinearities:

- (i) XYZ , $\epsilon = 0^\circ.46$,
- (ii) XYB , $\epsilon = 0^\circ.47$,
- (iii) ABC , $\epsilon = 0^\circ.59$.

Zuiderwijk [27] examined this situation by making use of the methods of Broadbent [3] and Kendall & Kendall [9]. He found it not exceptional when viewed as one of the 1000 or so quasar fields which might have been studied.

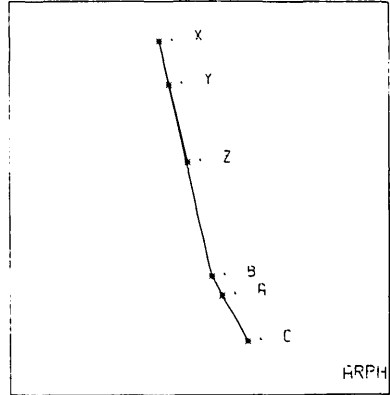


FIG. 17 The positions of the six Arp & Hazard quasars, with three close alignments marked. (Coordinates from Zuiderwijk [27].)

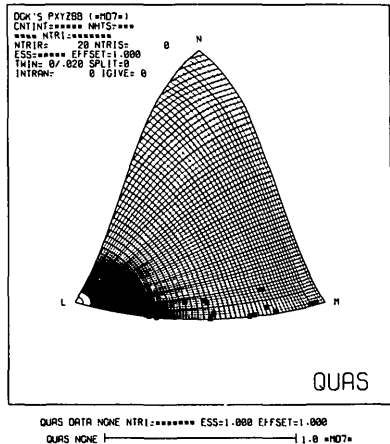


FIG. 18 The quasar triangle-shapes: (C, AC/BC)-fret. (Quasars A, B, C, X, Y, Z.)

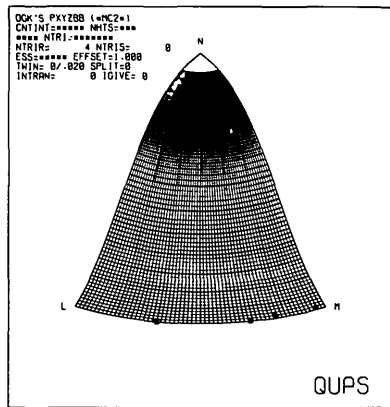


FIG. 19 The quasar triangle-shapes: (Z, Ψ)-fret. (Quasars X, Y, Z, B.)

However he missed the alignment (ii), and so underestimated the significance. Let us here avoid the uncertainty in the choice of the 'selection factor' 1000, and base our argument on the fact that Arp & Hazard first investigated this plate because they noticed the collinearity (iii); they then *subsequently* noticed (i), but not (ii). So it seems fair to test (i) and (ii), with the collinearity (iii) *omitted* (because it was the 'signal' determining the selection). We therefore ask, how remarkable is the *quartet* $XYZB$ as one of the 12 quartets which one can form from the set of six quasars *when we omit the 3 quartets involving all of A, B, and C?*

We can answer this question using Theorem 7. We find that $L = 0.01327$, and $U = 0.0001761$, so the P -value on this basis is

$$15 \times 0.0001761 = 0.00264.$$

(Here we have used $15 = \binom{6}{4}$ as a cautious overestimate of the combinatorial factor.)

The situation is therefore quite surprising, and the astronomers were right to take it seriously. Notice however that the significance would be eroded if more quasars were found on the plate. Such an erosion sets in quite quickly; for example $\binom{24}{4} = 10626$. It would be therefore wise to allow for 'optional stopping' as well as for 'selection'.

8.4. *Connexions with integral geometry.* Ambartzumian has recently [1] drawn attention to a remarkable shape-measure on Σ_2^3 which we shall call Amb , and whose differential for right-handed triangle-shapes, i.e. on the upper hemisphere of Σ_2^3 , takes the form

$$d \text{ Amb} = K \frac{\sin A \sin B \sin C}{(\sum \sin A)^4} dBdC, \quad (63)$$

where the vertical angles A , B , and C of the triangle are non-negative and sum to π . The measure Amb is to have exactly the same form for left-handed shapes, and so the constant K is to be chosen to make the integral of (63) over the 'right-handed' domain equal to one half. This formula (63) forms the basis of Ambartzumian's attack on the problem of deciding by statistical tests whether a given finite planar set of points has been obtained by observing an infinite planar Poisson process through some unspecified 'window'. We now investigate Amb on Σ_2^3 from the standpoint of the present paper, and in particular we seek a sampling process yielding Amb as the resulting shape-measure.

Now in [11] I show that the Miles distribution for the shapes of the Poisson-Delaunay triangles,

$$\frac{1}{2} \frac{8}{3\pi} \sin A \sin B \sin C dBdC$$

(with the same conventions about 'handedness' as before), can be identified with the shape-measure

$$\frac{4}{9} (\sum \sin^2 A)^2 d\gamma_*^3,$$

on the whole of Σ_2^3 , so that we can write

$$d \text{ Amb} = \frac{1}{3} K \pi \frac{(\sum \sin^2 A)^2}{(\sum \sin A)^4} d\gamma_*^3, \quad (64)$$

and it will be more convenient to work with (64), which is now to be understood as defining Amb over the whole of the shape-space.

Ambartzumian arrived at (63) as the fourth (shape) factor in a decomposition of the differential form for the vertices,

$$dP_1 dP_2 dP_3 = \prod_{j=1}^3 dx_j dy_j,$$

into four factors, the other three accounting respectively for location, rotation, and size. He measures size by the perimeter h of the triangle $P_1 P_2 P_3$, where $P_j = (x_j, y_j)$, so that $h = \sum |P_1 P_2|$. He is only concerned with measure-theoretic questions, and does not topologise the shape-space, but we shall of course identify it with $S^2(\frac{1}{2})$ in accordance with Theorem 2. The angles B and C do not form an ideal pair of shape-coordinates, but this does not matter in the present measure-theoretic discussion, in which all the associated singularities in the description of $S^2(\frac{1}{2})$ will form a set of measure zero.

As usual we fix attention on one vertex P_1 and one side $P_1 P_2$. Let this side be of length $2u$ and have direction θ . For the six-dimensional volume element $dP_1 dP_2 dP_3$ we find the expression

$$dx_1 dy_1 \cdot d\theta \cdot 4u^3 du \cdot 3r dr d\phi,$$

where now r and ϕ refer to our standard notation as explained in Section 5.

Here, without incurring any additional Jacobian factor, we can replace $dP_1 = dx_1 dy_1$ by $d\bar{P} = d\bar{x} d\bar{y}$ whenever we wish, \bar{P} being the centroid (\bar{x}, \bar{y}) . Also from (31) for $k = 3$ we have

$$d\gamma_*^3 = \frac{r dr d\phi}{\pi(1+r^2)^2}. \quad (65)$$

The variables u and h both describe size, but as usual we prefer for metrical reasons to work with $\$$, where $\$^2 = \sum |P_1 P_2|^2 = 3||z||^2$; we have

$$\$^2 = 6u^2(1+r^2), \quad \text{and} \quad h = \$ \frac{\sum \sin A}{\sqrt{\sum \sin^2 A}}.$$

In this way we get our preferred form of Ambartzumian's decomposition, namely

$$dP_1 dP_2 dP_3 = \frac{1}{3}\pi \cdot d\bar{x} d\bar{y} \cdot d\theta \cdot \$^3 d\$ \cdot d\gamma_*^3, \quad (66)$$

and also an equivalent form for it,

$$dP_1 dP_2 dP_3 = \frac{1}{3}\pi \cdot dx_1 dy_1 \cdot d\theta \cdot h^3 dh \cdot \frac{(\sum \sin^2 A)^2}{(\sum \sin A)^4} d\gamma_*^3, \quad (67)$$

which corresponds to the one given by Ambartzumian up to a numerical constant. The reader will find it interesting to check that (66) integrates to unity when multiplied by $g(P_1)g(P_2)g(P_3)$, where $g(x, y) = (2\pi)^{-1} \exp(-\frac{1}{2}x^2 - \frac{1}{2}y^2)$, and also that limiting the integration to the subset of variables $(\bar{x}, \bar{y}, \theta, \$)$ confirms the identity of γ_*^3 with the induced 'gaussian' shape-density.

We now construct a sampling situation appropriate to the decomposition (67), which we can write as

$$\frac{1}{K} \cdot dx_1 dy_1 \cdot d\theta \cdot h^3 dh \cdot d \text{Amb} , \tag{68}$$

in order to interpret *Amb* in an operational way. We invest the plane \mathbb{R}^2 simultaneously *and independently* with two probabilistic structures; a Poisson point-process of intensity α , and a Poisson line-process of intensity β , so that for any rectifiable Jordan curve Γ with interior G we shall have a probability $\exp(-\alpha|G|)$ of there being no random point in G , and a probability $\exp(-\beta|\Gamma|)$ of there being no random line which cuts Γ . Here $|G|$ is the area of G , and $|\Gamma|$ is the length of Γ .

We then define a random scoring function $T(f; F)$ whose value is $\sum f(\sigma)$, where f is a bounded non-negative Borel map from \sum_2^3 to the reals, and where the summation is over the shapes σ of all triangles $P_1 P_2 P_3$ such that

- (i) each vertex belongs to the realisation of the point-process,
- (ii) P_1 lies in a Borel set F of finite positive Lebesgue measure $|F|$,
- (iii) none of the realised lines of the line-process cut the perimeter of the triangle $P_1 P_2 P_3$.

We shall evaluate $\mathbb{E}T(f; F)$ in two different ways, using Fubini's theorem, and then equate the results.

In the first evaluation we begin by averaging over the randomness in the line-process, keeping the realisation of the point-process fixed. This contributes a factor $e^{-\beta h}$. We then average over the randomness in the point-process, and thus get

$$\mathbb{E}T(f; F) = \frac{\alpha^3}{K} \cdot |F| \cdot 2\pi \cdot \frac{6}{\beta^4} \cdot \text{Amb}(f) ,$$

where $\text{Amb}(f)$ denotes integration of f with respect to the measure *Amb*.

In the second evaluation we reverse the order of the two averaging operations. The realised line-process will divide the plane up into countably many compact convex 'Crofton polygons' Π , each of which has associated with it a shape-measure for triangles, $(\lambda | \Pi)_\star^3$. We hold the realisation of the line process fixed for the moment, and let M be the (random) number of (point-process) points in $\Pi \cap F$, and N the number of such points in all of Π . There are thus $M(N-1)(N-2)$ ways of choosing P_1, P_2 , and P_3 in Π , with P_1 also in F , and this is the number of triangles associated with Π which contribute to the score. (Remember that rule (iii) forces all of the triangles to lie in a single Crofton polygon!) Each one of these triangles will yield a shape σ and a value for $f(\sigma)$. Of course $M = 0$ if Π does not meet F , and almost surely M will vanish for all but finitely many Π 's. Plainly we shall have

$$\mathbb{E}T(f; F) = \mathbb{E} \sum_{\Pi \text{ meets } F} M(N-1)(N-2) (\lambda | \Pi; P_1 \in F)_\star^3(f)$$

in an obvious extension of the notation of Section 4 which includes the additional requirement ' $P_1 \in F$ '. The symbol \mathbb{E} on the right denotes the operation of averaging

over M , N , and Π . So long as the line-process is held fixed we shall have

$$\begin{aligned} \mathbb{E}\{M(N-1)(N-2) \mid \Pi\} &= \frac{|F \cap \Pi|}{|\Pi|} \mathbb{E}\{N(N-1)(N-2) \mid \Pi\} \\ &= \alpha^3 |F \cap \Pi| |\Pi|^2 \end{aligned}$$

because of the formulae for Poisson factorial moments. We therefore have

$$\mathbb{E}T(f; F) = \alpha^3 \mathbb{E} \sum |F \cap \Pi| |\Pi|^2 (\lambda \mid \Pi; P_1 \in F)_*^3(f),$$

and so

$$\text{Amb}(f) \int_F dP_1 = \frac{K}{12\pi} \beta^4 \mathbb{E} \sum |F \cap \Pi| |\Pi|^2 (\lambda \mid \Pi; P_1 \in F)_*^3(f).$$

Now $(\lambda \mid \Pi; P_1 \in F)_*^3(f)$ is equal to

$$\int_F 1_{\Pi} dP_1 \iint_{\Pi} f(\sigma) dP_2 dP_3 / (|F \cap \Pi| |\Pi|^2),$$

and so, because two equal absolutely convergent indefinite integrals must have ‘almost’ the same integrand, we see that

$$\text{Amb}(f) = \frac{K}{12\pi} \beta^4 \mathbb{E} \int_{\Pi_1} \int_{\Pi_1} f(\sigma) dP_2 dP_3 \quad (69)$$

for any fixed P_1 , where now Π_1 denotes that Crofton polygon Π which contains this fixed point $P_1 \in F$. We do not need to say ‘for almost all P_1 ’, because both sides are now independent of the position of P_1 . Also we can now ignore F . Obviously it does not matter where P_1 is, so we may as well fix it at the origin, referring thereafter to 0 and Π_0 instead of P_1 and Π_1 .

In particular (69) holds if $f \equiv 1$, and so we are able to evaluate the Ambartzumian constant as

$$K = \frac{12\pi}{\beta^4 \mathbb{E}|\Pi_0|^2} = \frac{21}{\pi} = 6.684508, \quad (70)$$

where we have used the evaluation $\mathbb{E}|\Pi_0|^2 = \mathbb{E}|\Pi|^3 / \mathbb{E}|\Pi| = 4\pi^2 / (7\beta^4)$. Of course we can check this value for K by numerical integration of (63), and this has been done.

If we substitute the first expression for K in (69) we now get

$$\text{Amb} = \frac{\mathbb{E}\{|\Pi_0|^2 (\lambda \mid \Pi_0; P_1 = 0)_*^3\}}{\mathbb{E}|\Pi_0|^2}.$$

However we can here appeal to the rather delicate principle (a consequence of stationarity) that ‘relative to Π_0 , supposed given, the origin is uniformly distributed within Π_0 ’, and so $(\lambda \mid \Pi_0; P_1 = 0)_*^3$ is the same as $(\lambda \mid \Pi_0)_*^3$. Also the ‘anchored’ character of Π_0 , requiring it to be that Crofton polygon which contains the origin,

can be dropped providing that we select a Crofton polygon Π using a selection rule which weights the cells into which the plane is divided according to their areas. These considerations give

THEOREM 10. *The Ambartzumian shape-measure at (64) can be expressed in the form*

$$\text{Amb} = \mathbb{E}\{|\Pi|^3 \cdot (\lambda | \Pi)_*^3\} / \mathbb{E}(|\Pi|^3). \tag{71}$$

Here Π denotes a 'typical' cell of the Crofton dissection of the plane, and $\mathbb{E}(\cdot)$ denotes the limit of a 'spreading average' over all the cells of the dissection.

The weighting in Theorem 10 is important, and the possibility of interaction between the cell area $|\Pi|$ and shape-variables such as $(\lambda | \Pi)_*^3$, which can be expressed in terms of the shape $\sigma(\Pi)$ of Π , can be illustrated by noting that

$$\mathbb{E}(V|\Pi) = \frac{\pi}{2\beta^2} > \frac{4}{\pi\beta^2} = \mathbb{E}(V) \cdot \mathbb{E}(|\Pi|)$$

when V denotes the number of vertices of Π .

If we insert the value of K at (70) into (68), to get

$$dP_1 dP_2 dP_3 = \frac{\pi dx_1 dy_1 d\theta}{21} \cdot h^3 dh \cdot d \text{Amb}, \tag{72}$$

we can now formulate other operational interpretations of Amb . For example let $P_1, P_2,$ and P_3 be distributed uniformly and independently in the circular disk of radius R centred at the origin, and let them be selected subject to the restrictions,

- (a) $\sqrt{(x_1^2 + y_1^2)} < R - \frac{1}{2}H,$
- (b) $h < H.$

This procedure, for any $0 < H < 2R,$ will generate triangles with Amb as shape-measure; it has already been outlined in [1], and exhibits Amb in close relation to $(\lambda | \text{circular disk})_*^3,$ but differing from the latter because of the restrictions at (a) and (b).

Contour lines for the density $d \text{Amb} / dy_*^3$ are shown in Figure 20.

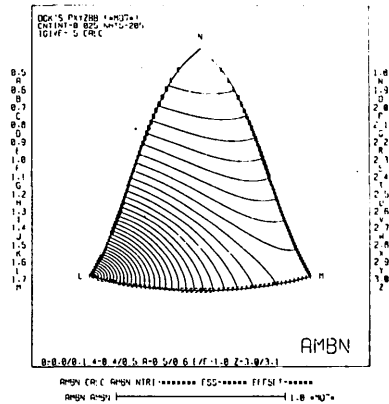


FIG. 20 Contours for the Ambartzumian density.

In closing this paper I should like to acknowledge helpful advice from many colleagues, especially Frank Adams, Adrian Baddeley, Simon Broadbent, Andrew Casson, Wilfrid Kendall, and Christopher Small.

References

1. R. V. AMBARTZUMIAN, 'Random shapes by factorization', *Statistics in theory and practice* (ed. B. Ranney, Umeå, undated).
2. K. BORSUK, *Theory of shape* (PWN, Warsaw, 1975).

3. S. R. BROADBENT, 'Simulating the ley-hunter', *J. Roy. Statist. Soc. (A)*, 143 (1980), 109–140.
4. B. N. DELONE (= B. DELAUNAY), 'Sur la sphère vide', *Proc. Int. Math. Congr. Toronto*, (1) (1928), 695–700.
5. G. FUBINI, 'Sulle metriche definite da una forma Hermitiana', *Atti Istit. Veneto*, 6 (1903), 501–513.
6. M. A. GIRSHICK, 'The distribution of the ellipticity statistic L_e when the hypothesis is false', *Terrest. Magn. & Atmosph. Elec.*, 46 (1941), 455–457. [The version of Girshick's distribution in N. L. Johnson & S. Kotz, *Distributions in statistics: continuous multivariate distributions* (Wiley, New York, 1972), contains two misprints.]
7. A. HORN, 'Doubly stochastic matrices and the diagonal of a rotation matrix', *Amer. J. Math.*, 76 (1954), 620–630.
8. D. G. KENDALL, 'The diffusion of shape', *Adv. in Appl. Probab.*, 9 (1977), 428–430.
9. D. G. KENDALL and W. S. KENDALL, 'Alignments in two-dimensional random sets of points', *Adv. in Appl. Probab.*, 12 (1980), 380–424.
10. D. G. KENDALL, 'The statistics of shape', *Interpreting multivariate data* (ed. V. Barnett, Wiley, 1981), pp. 75–80.
11. D. G. KENDALL, 'The shape of Poisson–Delaunay triangles', *Studies in probability and related topics in honour of Octav Onicescu* (eds. M. C. Demetrescu and M. Iosifescu, Nagard, Montreal, 1983), pp. 321–330. For the original form of the Miles distribution see R. E. Miles, 'On the homogeneous planar Poisson point process', *Math. Biosciences*, 6 (1970), 85–127.
12. D. G. KENDALL, 'The exact distribution for the shape of a random triangle with IID vertices uniformly distributed inside a circle'. (To be published.)
13. D. G. KENDALL, 'Foundations of a theory of random shape, I. Topology'. (In preparation.)
14. D. G. KENDALL, 'Foundations of a theory of random shape, II. Differential geometry' (In preparation.)
15. W. S. KENDALL, 'Random gaussian triangles'. (To be published.)
16. J. W. MAUCHLY, 'A significance test for ellipticity in the harmonic dial', *Terrest. Magn. & Atmosph. Elec.*, 45 (1940), 145–148. [See also C. T. Hsu, *Ann. Math. Statist.*, 11 (1940), 410–426.]
17. R. J. MUIRHEAD, *Aspects of multivariate statistical theory* (Wiley, New York, 1982).
18. S. A. ROBERTSON and S. CARTER, 'On the platonic and archimedean solids', *J. London Math. Soc.* (2), 2 (1970), 125–132.
19. S. A. ROBERTSON, S. CARTER and H. R. MORTON, 'Finite orthogonal symmetry', *Topology*, 9 (1970), 79–95.
20. S. A. ROBERTSON, 'Classifying triangles and quadrilaterals', *Math. Gaz.*, 61 (1977), 38–49.
21. H. SCHWERTFEGGER, *Geometry of complex numbers* (Oliver and Boyd, 1962.)
22. R. SIBSON, 'Studies in the robustness of multidimensional scaling: procrustes statistics', *J. Roy. Statist. Soc. (B)*, 40 (1978), 234–238.
23. B. W. SILVERMAN and T. C. BROWN, 'Short distances, flat triangles, and Poisson limits', *J. Appl. Probab.*, 15 (1978), 815–825.
24. C. G. SMALL, 'Random uniform triangles and the alignment problem', *Math. Proc. Cambridge Philos. Soc.*, 91 (1982), 315–322.
25. C. G. SMALL, 'Distributions of shape, and maximal invariant statistics', Ph.D. thesis, University of Cambridge, 1981.
26. E. STUDY, 'Kürzeste Wege im komplexen Gebiete', *Math. Ann.*, 60 (1905), 321–377.
27. E. J. ZUIDERWIJK, 'Alignment of randomly distributed objects', *Nature* 295 (1982), 577–578.

(Figure 2. and Figures 14 and 15. are reproduced from [10] and [11] respectively.)

Statistical Laboratory,
16 Mill Lane,
Cambridge CB2 1SB,
U.K.