Shape and Shape Theory

D. G. KENDALL
Churchill College, University of Cambridge, UK

D. BARDEN
Girton College, University of Cambridge, UK

T. K. CARNE
King's College, University of Cambridge, UK

H. LE
University of Nottingham, UK

JOHN WILEY & SONS, LTD
Chichester • New York • Weinheim • Brisbane • Singapore • Toronto

This Page intentionally left blank

Shape and Shape Theory

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Editors: *Vic Barnett, Noel A.C. Cressie, Nicholas I. Fisher, Iain M. Johnstone, J.B. Kadane, David G. Kendall, David W. Scott, Bernard W. Silverman, Adrian F.M. Smith, Jozef L. Teugels*
Editors Emeritus: *Ralph A. Bradley, J. Stuart Hunter*

A complete list of the titles in this series appears at the end of this volume

Shape and Shape Theory

D. G. KENDALL
Churchill College, University of Cambridge, UK

D. BARDEN
Girton College, University of Cambridge, UK

T. K. CARNE
King's College, University of Cambridge, UK

H. LE
University of Nottingham, UK

JOHN WILEY & SONS, LTD
Chichester • New York • Weinheim • Brisbane • Singapore • Toronto

Baffins Lane, Chichester,
West Sussex, PO19 1UD, England

National 01243 779777
International (+44) 1243 779777

e-mail (for orders and customer service enquiries): cs-books@wiley.co.uk

Visit our Home Page on http://www.wiley.co.uk or http://www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons, Inc., 605 Third Avenue,
New York, NY 10158-0012, USA

Wiley-VCH Verlag GmbH, Pappelallee 3,
D-69469 Weinheim, Germany

Jacaranda Wiley Ltd, 33 Park Road, Milton,
Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01,
Jin Xing Distripark, Singapore 129809

John Wiley & Sons (Canada) Ltd, 22 Worcester Road,
Rexdale, Ontario, M9W 1L1, Canada

Library of Congress Cataloging-in-Publication Data

Shape and shape theory / D.G. Kendall ... [et al.].
p. cm.—(Wiley series in probability and statistics)
Includes bibliographical references and index.
ISBN 0-471-96823-4 (alk. paper)
1. Shape theory (Topology)—Statistical methods I. Kendall, D.G. (David George), 1918- . II. Series.
QA612. 7. S48 1999 99-12433
514′ .24—dc21 CIP

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN 0-471-96823-4

Typeset in 10/12pt Times by Laser Words, Madras, India

Contents

Preface

Everyone knows what is meant by ‘shape’. However, it is not a trivial matter to define shape in a manner that is susceptible to mathematical and statistical analysis and it is only over the last two or three decades that appropriate definitions have been developed and studied. In this book we assume that the shape of an object is essentially captured by the shape of a finite subset of its points and, for the latter, we carry out much of the fundamental analysis that is likely to lie at the heart of further progress. Although this may seem a severe restriction, there is no theoretical limit to the number of points we consider and it has the significant advantage that the dimensions of the resulting shape spaces are always finite and only increase linearly with the number of points.

One of the central problems in shape theory is that it is not possible to represent the full range of possible shapes of an object in standard Euclidean coordinates without destroying our intuitive feel for the quantitative differences between them. Consequently, classical statistical methods are not always adequate or, at least, not clearly appropriate for the statistical analysis of shape and it is necessary to adapt them to work on unfamiliar spaces. We therefore need to describe the topological and geometric properties of these new spaces in some detail, as result of which this book is multidisciplinary. However, we have tried to make it accessible to as wide a range of readers as possible by giving, for each topic, more detail than the specialist in that subject might require. Where possible, we do this within the body of the text itself, with just a few of the more technical topological concepts and results reserved for the appendix.

We start with an introductory survey of the spaces in which we shall represent shapes and describe some of their more important properties and then, in Chapter 2, we investigate their global topological structure. The next three chapters lead up to a full calculation of the homology and cohomology groups of shape spaces. In the first of these we define homology theory and show how it is calculated in the special context that is adequate for our purposes because, although they are unfamiliar, shape spaces are still elementary. In Chapter 4 we examine the necessary chain complex for these computations that arises naturally from the topological structure of the spaces, and make some initial general deductions about their homology groups. Then, in Chapter 5, after giving a range of low-dimensional illustrative examples, we calculate all the groups explicitly and also derive some intriguing relationships between them.

In Chapters 6 and 7 we study the more subtle and more localised geometric properties of the spaces. Although there is only one topology, the natural quotient topology, that one can put on shape space, there is more than one metric. Here we discuss the Riemannian metric that arises from the theory of submersions, to which we shall relate any other metrics that we use. Once again, the elementary way in which shape spaces are produced enables us to prove most of the results that we require from that theory directly in our context, with little reference to the general case. In Chapter 6 we examine the geodesics, the analogues of straight lines in Euclidean space, between two shapes and find simple expressions for the distance between those shapes, as well as the distance from a shape to certain subsets of practical significance. In Chapter 7, after introducing a little more differential geometry, we are able to obtain explicit expressions for the main geometric invariants of the spaces. In particular, we are able to measure the precise extent to which they are curved. This is vital, for example, when assessing the extent to which a local linear approximation to shape space is valid. Since the curvature can be arbitrarily large, that is certainly not always the case.

In the next two chapters we turn to the probabilistic and statistical topics that were the prime motivation for the introduction of shape spaces. In Chapter 8 we investigate the distributions that arise on shape space from various standard distributions on the points that determine those shapes. We describe them generally by referring to the volume measure on shape space obtained in Chapter 7. As the initial distributions become more general, the range of shape spaces on which we give explicit formulae for the induced distributions tends to become more restricted. However, in principle, our results are quite general. Moreover, although, as was the case for the homology groups, the formulae can become quite intricate, they are still elementary and susceptible to computation. We illustrate this claim by obtaining, in the final sections of this chapter, an explicit description of the density function for the shape of a random triangle whose vertices are uniformly independently distributed in a given convex planar polygon.

In classical statistics the mean is well-defined and simple to compute. However, problems can arise both in defining a 'mean' shape in theory and also in calculating it in practice. It turns out that various 'obvious' approaches do not necessarily lead to the same results or even, in each case, to a unique result. In Chapter 9 we discuss some of the relations between different possible definitions and also identify circumstances, fortunately fairly general, in which the results we would like to take for granted are actually true.

In Chapter 10 we address the problem of visualising the first, that is, lowest dimensional, shape space that is not already familiar. That is the five-dimensional space of shapes of tetrahedra in 3-space. Although this is topologically a sphere, it is by no means a standard one, as it has a singular subset in the neighbourhood of which the curvature becomes arbitrarily large. The visualisation uses a carefully selected family of 24 two-dimensional sections that, rather surprisingly, do allow us to follow what is going on in the space. We illustrate this by describing some typical geodesics, some sample paths for a diffusion and a comparison of two distributions on the space.

We conclude by putting our work into a broader setting where similar studies may be carried out. In particular, this enables us to look at some shape spaces related to those that have been the subject of the rest of this book. The first applications still concern finite sets of points in Euclidean space but here we study their size-and-shape, for which size is no longer quotiented out, and also an alternative metric, one having negative curvature, on the non-degenerate part of the shape space. In the final sections we consider the shapes of finite sets of points in the other standard spaces of constant curvature, the sphere and hyperbolic space, as well as some connections with the classical theory of elliptic functions.

We are, of course, indebted to all who have worked on shape theory, whether or not it lies in the area that we specifically address. Much of the material presented is previously unpublished work given in local seminars or work produced explicitly for this book, and we are grateful to all our colleagues, but especially to Marge Batchelor, for many helpful discussions over the years of gestation of this project. Thanks are also due to our publishers, particularly to Helen Ramsey for her constant encouragement and patience over our ever-receding deadlines.

This Page intentionally left blank

CHAPTER 1

Shapes and Shape Spaces

1.1 ORIGINS

There have been at least three distinct origins of what we call *shape theory*. The first approach seems to have been that of Kendall (1977) who was, at that time, concerned with 'shape' in archaeology and astronomy, but it soon became clear that the subject could profitably be studied from a more general standpoint. At about the same time Bookstein (1978a,b) began to study shape-theoretic problems in the particular context of zoology. A third early contributor was Ziezold (1977). In this present book the theory will be developed largely along the lines initiated by Kendall in his 1977 paper, but much new material will be presented.

In a typical case the calculations will be concerned with sets of, say, k *labelled* points in a Euclidean space $\mathbb{R}^m$, where $k \geqslant 2$. Normally, the centroid of the k points will serve as an origin, and the scale will be such that the sum of the squared distances of the points from that origin will be equal to unity. The basic object just described will be called the *pre-shape*, and any two configurations of k labelled points will be regarded as having the same shape *if either of their pre-shapes can be transformed into the other by a rotation about the shared centroid.* The resulting assemblage of all possible shapes will be called the *shape space* and will be denoted by Σ_m^k. Accordingly, the *shape* is defined as the *pre-shape modulo rotations*. These definitions and the related constructions provide the basis for the present book. It should be observed that the k constituent labelled points determine the shape. At a later stage we shall define 'size-and-shape' in a similar way by omitting the 'unit-sum-of-squares' standardisation.

While we do not wish to go deeply into the details of Bookstein's parallel work, it is appropriate here to stress the fact that for us *the labelled points are basic and determine the object being studied.* In Bookstein's work, however, the 'marker points' are selected from a usually two-dimensional or three-dimensional continuum. Thus, if the object in question is a planar representation of a human hand with fingers out-stretched, then the markers could be the tips of the fingers, the common roots of each pair of adjacent fingers and a few more points on the planar outline of the hand reaching down, say, as far as the wrist. Already in this simple case it is clear that the choice of markers is far from being a simple matter even in the two-dimensional case, and it would be still more difficult if the complete three-dimensional surface were to be the object being studied. To take a still more difficult example, consider the problem of coding the shape of a potato!

A further difference between the two approaches arises at the next stage. Bookstein is concerned to represent real objects, often biological ones, and the markers are chosen sufficiently well spaced to identify those objects. Thus, he is not interested in configurations in which the markers all lie in a lower-dimensional subspace or two or more of them coincide. This contrasts with Kendall's spaces, which contain the shapes of all possible configurations except those for which all the points coincide. This provides a context in which it is possible to measure the statistical significance of apparent collinearities or other degeneracies in archaeological or astronomical data.

Bookstein's work includes many delicate and important studies concerning the continuous deformation of *biological* shapes, this being a topic first studied by Thompson [1917] (1942). It is appropriate here to associate Thompson's work with that of Bower (1930) who studied 'size and form' in plants. A copy of Bower's book was given by him to Kendall, and it was this event that led many years later to the formulation of shape-theoretic studies in a general mathematical context.

1.2 SOME PRELIMINARY OBSERVATIONS

Consider the shape of a configuration of $k \geqslant 2$ labelled and not totally coincident points $\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \ldots, \boldsymbol{x}_k^*$ in a Euclidean space having $m \geqslant 1$ Euclidean dimensions. How is the *shape* of such a configuration to be represented?

Since we are not interested in the location of the k-ad, we may start by uniformly translating its component points $\boldsymbol{x}_j^*$ in $\mathbb{R}^m$ in such a way that their centroid, $\boldsymbol{x}_c^*$, is moved to the origin of the coordinates. The 'size' of this k-ad is, of course, important as an aspect of 'size-and-shape', but as far as shape alone is concerned it is of no interest, so we normally shrink or expand the size of the centred k-ad about the new origin so as to make the natural quadratic measure of 'size'

$$\sqrt{||\boldsymbol{x}_1^* - \boldsymbol{x}_c^*||^2 + \cdots + ||\boldsymbol{x}_k^* - \boldsymbol{x}_c^*||^2}$$

equal to unity. This convention makes sense because we have deliberately excluded the maximally degenerate case in which all the points $\boldsymbol{x}_j^*$ coincide.

To take the most trivial example, the only such sized-and-centred configurations when $k = 2$ and $m = 1$ are the labelled point-pairs:

$$\left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) \text{ and } \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)$$

in $\mathbb{R}^2$, so we see that Σ_1^2 consists of just two shapes, and that it can be identified with the two-point unit sphere $\{-1, 1\}$ of dimension zero.

A straightforward generalisation of this argument, which we shall give in more detail later, tells us that shape space Σ_1^k consists of all standardised k-ads of the form:

$$\left\{(x_1, x_2, \ldots, x_k) : \sum_{i=1}^{k} x_i = 0, \sum_{i=1}^{k} x_i^2 = 1\right\}$$

for $k \geqslant 2$. Thus, shape space Σ_1^k is a unit-radius $(k-2)$-sphere $\mathbf{S}^{k-2}(1)$ for all $k \geqslant 2$.

But now suppose that $m \geqslant 2$ and $k \geqslant 2$, and let us move the centroid of the k-ad to the origin and standardise the size as before. If we write $\widetilde{X}^*$ for the sized-and-centred $m \times k$ coordinate matrix with components

$$(\tilde{x}_{ij}^* : 1 \leqslant i \leqslant m, 1 \leqslant j \leqslant k),$$

then the k individual columns of the matrix can be thought of as column m-vectors specifying the positions of the k points $\boldsymbol{x}_j^* - \boldsymbol{x}_c^*$ in $\mathbb{R}^m$ where, as above, $\boldsymbol{x}_c^*$ is the centroid of the k-ad and $1 \leqslant j \leqslant k$. Then, from the shape-theoretic point of view, we will never wish to distinguish between $\widetilde{X}^*$ and $T\widetilde{X}^*$ where T is in $\mathbf{SO}(m)$. This is because it is a basic feature of our work that rotations acting on the left of $\widetilde{X}^*$ are to be regarded as irrelevant. We therefore call the sized-and-centred configuration described by $\widetilde{X}^*$ the *pre-shape*, and we define the *shape itself* to be $\widetilde{X}^*$ viewed modulo the rotations in $\mathbf{SO}(m)$ acting from the left.

It is easily checked that the complete set of all such pre-shapes is a unit sphere of dimension $m(k-1)-1$, and we call this $\mathcal{S}_m^k$. That is the *pre-shape space*, and the corresponding shape space Σ_m^k is $\mathcal{S}_m^k$ modulo $\mathbf{SO}(m)$ with the rotations acting from the left. Provided that $k \geqslant m+1$, the dimension of Σ_m^k is

$$d_m^k = m(k-1) - \tfrac{1}{2}m(m-1) - 1.$$

In this formula the first term on the right follows from the fact that, while there are m rows and k columns in the matrix $\widetilde{X}^*$, k is here reduced to $k-1$ because we want to have the centroid of the k points at the origin of the coordinates. The second term on the right arises because we must quotient out the effect of rotations of $\mathbf{SO}(m)$ acting from the left, while the final term, -1, takes account of the fact that we wish to ignore scale effects. In particular, we note that $d_1^k = k-2$ agrees with our earlier calculations. When $k \leqslant m$, a configuration of k labelled points in $\mathbb{R}^m$ lies in a $(k-1)$-dimensional subspace and so its shape lies in Σ_{k-1}^k. However, now the extra dimensions give us room to rotate the configuration onto its mirror image. This means that the pre-shape is quotiented out by $\mathbf{O}(k-1)$ rather than $\mathbf{SO}(k-1)$ to obtain the shape. Thus, for $k \leqslant m$, Σ_m^k is a 'halved' version of Σ_{k-1}^k, this being the result of identifying the shapes of configurations of k labelled points in $(k-1)$-space that are mirror images of each other. In particular, when $k \leqslant m$, Σ_m^k is 'over-dimensioned' in the sense that $d_m^k > \dim(\Sigma_m^k) = d_{k-1}^k$.

Note that, while the construction of the pre-shape is entirely elementary, the quotient operation that yields the shape itself is very far from being so, save in a few trivial cases. When $m = 1$ the shape is identical with the pre-shape $\widetilde{X}^*$ because there are no non-trivial rotations T in $\mathbf{SO}(1)$, but the corresponding situation when m is equal to two or more can be quite complicated. To illustrate the non-triviality of shape spaces in general it suffices to remark that when $k \geqslant 2$ the shape of k labelled points in the plane will turn out to be a point in the classical complex projective space $\mathbf{CP}^{k-2}(4)$, where the '4' is the appropriate value for the curvature parameter. This is its name as a classical object, but as a shape space we call it Σ_2^k. In particular, Σ_2^3 is $\mathbf{CP}^1(4)$. This, in more familiar terms, is the 2-sphere $\mathbf{S}^2(\frac{1}{2})$ of radius one-half. More details about this will be given later in Section 1.3.

Before discussing general shape spaces Σ_m^k we introduce two important diagrams shown in Tables 1.1 and 1.2. These will be useful in reminding the reader of 'what goes where' and we here mention some of their most important features.

Obviously, it is desirable that in the diagrams we should be able to recognise those shape spaces that are over-dimensioned, that is, those for which $k \leqslant m$. In Table 1.1 the over-dimensioned shape spaces are emphasised by the use of lower case σ instead of upper case Σ, which will be used elsewhere. Here, the entry at (k, m) in the table is the name of the shape space associated with k labelled points in m dimensions.

The accompanying Table 1.2 follows the same pattern, but now the entry in position (k, m) is *the dimension of the corresponding shape space*, and a bold font is used to indicate *the region* $k \geqslant m + 1$ *in which shape spaces are not over-dimensioned*. For example, Σ_3^4 in Table 1.1 is the shape space for four labelled points in three dimensions and, from the corresponding entry in Table 1.2, we see that this is a five-dimensional shape space—actually we shall find that it is a *topological* 5-*sphere* that possesses singularities.

As already mentioned, the spaces listed in the first column, where $m = 1$, are all unit spheres, while the second column also has familiar entries: Σ_2^k for each choice of k is the classical complex projective space with a complex dimension $k - 2$ and a real dimension $2k - 4$.

A striking feature of Table 1.1 is that the main diagonal consists entirely of, mainly only topological, spheres. We already know that Σ_1^2 is the two-point, zero-dimensional, metric sphere of radius unity. We shall see later that Σ_2^3 is a metric 2-sphere of radius one-half, while the further entries on the main diagonal hold the *topological* spheres of the dimensions $5, 9, 14, \ldots$, etc. indicated by the corresponding entries on the main diagonal in Table 1.2, where the mth entry is

$$d_m^{m+1} = \tfrac{1}{2}m^2 + \tfrac{1}{2}m - 1.$$

Table 1.1 The array of shape spaces (the σ-entries are 'over-dimensioned')

k\m	1	2	3	4	5	6	7	8	9	10
2	Σ_1^2	σ_2^2	σ_3^2	σ_4^2	σ_5^2	σ_6^2	σ_7^2	σ_8^2	σ_9^2	σ_{10}^2
3	Σ_1^3	Σ_2^3	σ_3^3	σ_4^3	σ_5^3	σ_6^3	σ_7^3	σ_8^3	σ_9^3	σ_{10}^3
4	Σ_1^4	Σ_2^4	Σ_3^4	σ_4^4	σ_5^4	σ_6^4	σ_7^4	σ_8^4	σ_9^4	σ_{10}^4
5	Σ_1^5	Σ_2^5	Σ_3^5	Σ_4^5	σ_5^5	σ_6^5	σ_7^5	σ_8^5	σ_9^5	σ_{10}^5
6	Σ_1^6	Σ_2^6	Σ_3^6	Σ_4^6	Σ_5^6	σ_6^6	σ_7^6	σ_8^6	σ_9^6	σ_{10}^6
7	Σ_1^7	Σ_2^7	Σ_3^7	Σ_4^7	Σ_5^7	Σ_6^7	σ_7^7	σ_8^7	σ_9^7	σ_{10}^7
8	Σ_1^8	Σ_2^8	Σ_3^8	Σ_4^8	Σ_5^8	Σ_6^8	Σ_7^8	σ_8^8	σ_9^8	σ_{10}^8
9	Σ_1^9	Σ_2^9	Σ_3^9	Σ_4^9	Σ_5^9	Σ_6^9	Σ_7^9	Σ_8^9	σ_9^9	σ_{10}^9
10	Σ_1^{10}	Σ_2^{10}	Σ_3^{10}	Σ_4^{10}	Σ_5^{10}	Σ_6^{10}	Σ_7^{10}	Σ_8^{10}	Σ_9^{10}	σ_{10}^{10}
11	Σ_1^{11}	Σ_2^{11}	Σ_3^{11}	Σ_4^{11}	Σ_5^{11}	Σ_6^{11}	Σ_7^{11}	Σ_8^{11}	Σ_9^{11}	Σ_{10}^{11}
12	Σ_1^{12}	Σ_2^{12}	Σ_3^{12}	Σ_4^{12}	Σ_5^{12}	Σ_6^{12}	Σ_7^{12}	Σ_8^{12}	Σ_9^{12}	Σ_{10}^{12}
13	Σ_1^{13}	Σ_2^{13}	Σ_3^{13}	Σ_4^{13}	Σ_5^{13}	Σ_6^{13}	Σ_7^{13}	Σ_8^{13}	Σ_9^{13}	Σ_{10}^{13}
14	Σ_1^{14}	Σ_2^{14}	Σ_3^{14}	Σ_4^{14}	Σ_5^{14}	Σ_6^{14}	Σ_7^{14}	Σ_8^{14}	Σ_9^{14}	Σ_{10}^{14}
15	Σ_1^{15}	Σ_2^{15}	Σ_3^{15}	Σ_4^{15}	Σ_5^{15}	Σ_6^{15}	Σ_7^{15}	Σ_8^{15}	Σ_9^{15}	Σ_{10}^{15}
16	Σ_1^{16}	Σ_2^{16}	Σ_3^{16}	Σ_4^{16}	Σ_5^{16}	Σ_6^{16}	Σ_7^{16}	Σ_8^{16}	Σ_9^{16}	Σ_{10}^{16}
17	Σ_1^{17}	Σ_2^{17}	Σ_3^{17}	Σ_4^{17}	Σ_5^{17}	Σ_6^{17}	Σ_7^{17}	Σ_8^{17}	Σ_9^{17}	Σ_{10}^{17}
18	Σ_1^{18}	Σ_2^{18}	Σ_3^{18}	Σ_4^{18}	Σ_5^{18}	Σ_6^{18}	Σ_7^{18}	Σ_8^{18}	Σ_9^{18}	Σ_{10}^{18}
19	Σ_1^{19}	Σ_2^{19}	Σ_3^{19}	Σ_4^{19}	Σ_5^{19}	Σ_6^{19}	Σ_7^{19}	Σ_8^{19}	Σ_9^{19}	Σ_{10}^{19}
20	Σ_1^{20}	Σ_2^{20}	Σ_3^{20}	Σ_4^{20}	Σ_5^{20}	Σ_6^{20}	Σ_7^{20}	Σ_8^{20}	Σ_9^{20}	Σ_{10}^{20}

This fact was first observed by Casson (1977, private communication), and it plays an important role in many later contexts. Casson's proof of this theorem will be given at the end of this chapter and alternative proofs will be given elsewhere in this book. We shall also show, in Chapter 6, that none of these spheres is a *metric* sphere except for the first two.

Another important feature arises from the fact that in the kth row of Table 1.1 the spaces beyond the main diagonal are those Σ_m^k for which $k \leqslant m$ so, as we have seen, they are all *identical* to halved versions of Σ_{k-1}^k, which is the corresponding diagonal entry. In fact, a simple modification of Casson's proof would show that they are all topological balls, and we shall give an alternative proof in Chapter 2. To take the simplest example, the row labelled $k = 2$ starts with a two-point, zero-dimensional space, its two points being separated by two units, and this two-point space is followed on the right by an infinite part-row of one-point spaces. Similarly, the row $k = 3$ begins with a one-dimensional circle of radius one, followed by a 2-sphere of radius one-half, this in its turn being followed by an infinite part-row of metric hemispheres having the same dimension and radius.

Table 1.2 Shape space dimensions

k\m	1	2	3	4	5	6	7	8	9	10
2	0	0	0	0	0	0	0	0	0	0
3	**1**	**2**	2	2	2	2	2	2	2	2
4	**2**	**4**	**5**	5	5	5	5	5	5	5
5	**3**	**6**	**8**	**9**	9	9	9	9	9	9
6	**4**	**8**	**11**	**13**	**14**	14	14	14	14	14
7	**5**	**10**	**14**	**17**	**19**	**20**	20	20	20	20
8	**6**	**12**	**17**	**21**	**24**	**26**	**27**	27	27	27
9	**7**	**14**	**20**	**25**	**29**	**32**	**34**	**35**	35	35
10	**8**	**16**	**23**	**29**	**34**	**38**	**41**	**43**	**44**	44
11	**9**	**18**	**26**	**33**	**39**	**44**	**48**	**51**	**53**	**54**
12	**10**	**20**	**29**	**37**	**44**	**50**	**55**	**59**	**62**	**64**
13	**11**	**22**	**32**	**41**	**49**	**56**	**62**	**67**	**71**	**74**
14	**12**	**24**	**35**	**45**	**54**	**62**	**69**	**75**	**80**	**84**
15	**13**	**26**	**38**	**49**	**59**	**68**	**76**	**83**	**89**	**94**
16	**14**	**28**	**41**	**53**	**64**	**74**	**83**	**91**	**98**	**104**
17	**15**	**30**	**44**	**57**	**69**	**80**	**90**	**99**	**107**	**114**
18	**16**	**32**	**47**	**61**	**74**	**86**	**97**	**107**	**116**	**124**
19	**17**	**34**	**50**	**65**	**79**	**92**	**104**	**115**	**125**	**134**
20	**18**	**36**	**53**	**69**	**84**	**98**	**111**	**123**	**134**	**144**

However, when $k = 4$ the row starts off with a unit metric 2-sphere, followed by a four-dimensional complex projective space of complex curvature 4, and then by a five-dimensional *topological* sphere, this last space having singularities. That space lies on the diagonal and is followed on its right by an infinite sequence of *identical topological hemispheres*, or, equivalently, topological balls, which are precisely 'halves' of the 5-sphere on the diagonal. The situation is highlighted in Table 1.1 by the use of upper-case and lower-case sigmas.

After inspecting Table 1.1 the reader will notice that we have still to describe the shape spaces Σ_m^k in the infinite triangular region determined by the inequalities $m \geqslant 3$ and $k \geqslant m + 2$, that is, those that lie to the right of the column $m = 2$ and to the left of the spheres on the diagonal of the array. These particular shape spaces are truly peculiar in that they appear not to have occurred in any earlier contexts. They have not yet been determined up to homeomorphism, but in due course we will present the integral homology for each one and describe its global geodesic geometry as well as the Riemannian metric and associated curvature tensors.

Since they have different dimensions, *no two* of the 'diagonal' shape spaces are the same, but to the left of the diagonal the dimension alone is not sufficient to distinguish between them. For example, Σ_7^9 and Σ_5^{10} each have dimension 34. In later chapters we will make use of a 'topological recurrence' that provides useful structural information about all of them and, in principle, leads to a complete characterisation of the whole family of shape spaces. It will be shown, in particular, that the shape spaces in the infinite triangular region mentioned above possess the following interesting properties:

(i) no one of these is a sphere, and indeed no one is even a homotopy sphere or a manifold,
(ii) all of these spaces have torsion in homology,
(iii) no two of them share the same homology, even at the $\mathbb{Z}_2$-level.

These facts were first established by making use of the exact sequences for shape space homology that we introduce in Chapter 5 and that results from the above-mentioned topological recurrence. Note, in particular, an important consequence implied by (iii): *the shape spaces located in the infinite triangular region of Table 1.1 are all topologically distinct from one another.* They are also distinct from those in the first two columns so that apart from Σ_1^4 and Σ_2^3, which are different sized copies of the 2-sphere, all shape spaces with $k \geqslant m$ are topologically distinct.

1.3 A MATRIX REPRESENTATION FOR THE SHAPE OF A k-ad

Let us consider a labelled set of k points in $\mathbb{R}^m$, where $k \geqslant 2$, whose coordinates $\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \ldots, \boldsymbol{x}_k^*$ we shall write as the columns of the matrix X^*. We recall that degeneracies are allowed except that we insist that the points are not totally coincident, and that the *shape* of the k-ad is what is left when all effects attributable to translation, rotation and dilatation have been quotiented out.

We now orthogonally transform the k-ad X^* as follows:

$$\boldsymbol{x}_0 = \sqrt{k}\boldsymbol{x}_c^* = \frac{1}{\sqrt{k}}(\boldsymbol{x}_1^* + \boldsymbol{x}_2^* + \cdots + \boldsymbol{x}_k^*)$$

and

$$\tilde{\boldsymbol{x}}_j = \frac{1}{\sqrt{j+j^2}}\{j\boldsymbol{x}_{j+1}^* - (\boldsymbol{x}_1^* + \boldsymbol{x}_2^* + \cdots + \boldsymbol{x}_j^*)\}$$

for $1 \leqslant j \leqslant k-1$. We can see that the matrix $(\sqrt{k}\boldsymbol{x}_c^* \; \tilde{\boldsymbol{x}}_1 \; \cdots \; \tilde{\boldsymbol{x}}_{k-1})$ representing the new k-ad is obtained from X^* by multiplying on the right by a special $k \times k$ matrix Q_k. The second equation also shows that, for each $j > 0$, $\tilde{\boldsymbol{x}}_j$ is a scalar multiple of

$$\boldsymbol{x}_{j+1}^* - \frac{\boldsymbol{x}_1^* + \cdots + \boldsymbol{x}_j^*}{j},$$

and a striking feature of this construction is the progressive re-centring of $\boldsymbol{x}_{j+1}^*$ relative to its predecessors $\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_j^*$. This follows from the form of Q_k and it provides the main justification for its use. As an example we present the matrix

Q_k in the particular case $k = 6$ as follows:

$$\begin{pmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{20}} & -\frac{1}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{20}} & -\frac{1}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{12}} & -\frac{1}{\sqrt{20}} & -\frac{1}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & 0 & \frac{3}{\sqrt{12}} & -\frac{1}{\sqrt{20}} & -\frac{1}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & 0 & 0 & \frac{4}{\sqrt{20}} & -\frac{1}{\sqrt{30}} \\ \frac{1}{\sqrt{6}} & 0 & 0 & 0 & 0 & \frac{5}{\sqrt{30}} \end{pmatrix}.$$

It should be noted that

(i) in each column of the array the squares of the entries sum to unity,
(ii) each column is orthogonal to all of the other columns,
(iii) the integers, the square roots of which appear in the denominators in the second and later positions in row one, are as follows:

$$2 = 1 \times 2, \quad 6 = 2 \times 3, \quad 12 = 3 \times 4,$$
$$20 = 4 \times 5, \quad 30 = 5 \times 6, \quad \ldots$$

these entries being repeated in the rows below in the same horizontal locations until just before the main diagonal is reached.

In fact, these properties suffice to specify the matrix Q_k in the general case up to the sign of each column. For our particular choice it turns out that Q_k is a rotation. Indeed, the fact that it is orthogonal is immediate from properties (i) and (ii) above. However, it remains for us to show that Q_k has determinant $+1$ rather than -1. In order to do this we start by adding to the top row of the matrix the sum of all the subsequent rows. This yields a new top row consisting of $\sqrt{k}$ followed by zeros. It follows that the value of the determinant is

$$\sqrt{k} \times 1 \times 2 \times 3 \times \cdots \times (k-1)$$

divided by

$$\{(1 \times 2) \times (2 \times 3) \times (3 \times 4) \times \cdots \times ((k-1) \times k)\}^{1/2},$$

which reduces to $+1$ as required.

For most purposes it is convenient to shift the configuration so that its centroid is moved to the origin of coordinates after which the matrix X^* will have all its row sums equal to zero. If we now examine the product

$$(\boldsymbol{x}_1^* \ \boldsymbol{x}_2^* \ \cdots \ \boldsymbol{x}_k^*) Q_k$$

we find that it has the form

$$(0\ \tilde{\boldsymbol{x}}_1\ \tilde{\boldsymbol{x}}_2\ \cdots\ \tilde{\boldsymbol{x}}_{k-1})$$

because the first column of Q_k is 'constant' and each row-sum of the matrix

$$\boldsymbol{x}_1^* + \boldsymbol{x}_2^* + \cdots + \boldsymbol{x}_k^*$$

is equal to zero. We also note that our normalisation for size, dividing by

$$\{||\boldsymbol{x}_1^* - \boldsymbol{x}_c^*||^2 + \cdots + ||\boldsymbol{x}_k^* - \boldsymbol{x}_c^*||^2\}^{1/2},$$

now corresponds to dividing by $\left\{\sum_{i=1}^{k} ||\boldsymbol{x}_i^*||^2\right\}^{1/2}$ and, since Q_k is orthogonal, to dividing by $\left\{\sum_{i=1}^{k-1} ||\tilde{\boldsymbol{x}}_i||^2\right\}^{1/2}$.

The result of this normalisation will be a matrix

$$(0\ \boldsymbol{x}_1\ \boldsymbol{x}_2\ \cdots\ \boldsymbol{x}_{k-1})$$

and, if we throw away the zero first column, we can represent the pre-shape by

$$X = (\boldsymbol{x}_1\ \boldsymbol{x}_2\ \cdots\ \boldsymbol{x}_{k-1}),$$

and then the shape itself is represented by this $m \times (k-1)$ array modulo $\mathbf{SO}(m)$ acting on the left. We also note that, if we identify the space of $m \times (k-1)$ real matrices with Euclidean $(m \times (k-1))$-space, our normalisation implies that this pre-shape will lie on the unit $(m(k-1)-1)$-sphere in that space.

Thus, the shape is to be identified with the equivalence class or 'orbit' associated with the left action of $\mathbf{SO}(m)$ on the pre-shape, and we shall be free to represent each such class by any one of its members. In particular, we can if we wish transform the pre-shape matrix

$$X = (\boldsymbol{x}_1\ \boldsymbol{x}_2\ \cdots\ \boldsymbol{x}_{k-1})$$

by using one of the rotations in $\mathbf{SO}(m)$ to perform various 'left-hand' tidying-up operations. Thus, we can exploit these procedures to yield

(i) an upper semi-diagonal matrix that has a strictly positive sign for the first non-zero entry in each of the first $m-1$ rows, all the entries in the last row being 0 or $\pm$, as in

$$\begin{pmatrix} 0 & + & \pm & \pm & \pm & \pm & \pm & \pm & \pm & \pm & \pm \\ 0 & 0 & 0 & + & \pm & \pm & \pm & \pm & \pm & \pm & \pm \\ 0 & 0 & 0 & 0 & \pm & \pm & \pm & \pm & \pm & \pm & \pm \end{pmatrix},$$

or

(ii) an upper semi-diagonal matrix with a strictly positive sign for the first non-zero entry in each of the first $j \leqslant m-2$ rows, with 'zero' rows below this, as in the further example with $j = 2$:

$$\begin{pmatrix} 0 & 0 & + & \pm & \pm & \pm & \pm & \pm & \pm & \pm & \pm & \pm \\ 0 & 0 & 0 & + & \pm & \pm & \pm & \pm & \pm & \pm & \pm & \pm \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Such 'tidying-up' operations involving an $(m \times m)$-rotation on the left will often be useful. Of course, on replacing each $\pm$ by the actual numerical entry we get a tidied-up version of the original data. In fact, when we come to perform our mathematical computations on these matrices it will sometimes be more convenient to have all the potentially non-zero elements at the beginning of each row.

Another presentation of the pre-shape $X = (\boldsymbol{x}_1\ \boldsymbol{x}_2\ \cdots\ \boldsymbol{x}_{k-1})$ is based on a 'pseudo-singular values decomposition' of X. This allows us to present the pre-shape in the three-factor form

$$U(\Lambda\ 0)V.$$

Here, U is an element of $\mathbf{SO}(m)$, V is an element of $\mathbf{SO}(k-1)$ and Λ is the $m \times m$ diagonal matrix

$$\operatorname{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_m\}$$

with

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_{m-1} \geqslant |\lambda_m|.$$

In this formula the sum of the squares of the λ's is equal to unity, and $\lambda_m \geqslant 0$ unless $k = m + 1$.

We can re-write this decomposition in either of the forms

$$U(\Lambda D\ 0)\ \operatorname{diag}\{D^{-1}, 0\}V,$$

or, equivalently,

$$U\,D(\Lambda\ 0)\ \operatorname{diag}\{D^{-1}, 0\}V,$$

where D is any diagonal $(m \times m)$-matrix of the form

$$\operatorname{diag}\{\pm 1, \pm 1, \ldots, \pm 1\}$$

with an even number of minus signs.

Accordingly, UD is a left-rotation and can be dismissed when we are only interested in the shape, so that in that case we are left with

$$(\Lambda D^{-1}\ 0)V.$$

These transformations will be useful later.

1.4 'ELEMENTARY' SHAPE SPACES Σ_1^k AND Σ_2^k

We have already claimed that

$$\Sigma_1^k = \mathbf{S}^{k-2}(1),$$

and that

$$\Sigma_2^k = \mathbf{CP}^{k-2}(4),$$

and in this section we provide the evidence for these assertions. We begin with Σ_1^k.

When $k = 2$ we start with a non-degenerate point-pair (x_1^*, x_2^*), and we carry out our standard reduction using Q_2 to yield the singleton

$$\tilde{x}_1 = \tfrac{1}{2}(x_2^* - x_1^*).$$

Because, with our conventions, x_1^* and x_2^* must be distinct when $k = 2$, it is clear that we can divide out the size $|x_2^* - x_1^*|/\sqrt{2}$ to get $+1$ when $x_2^* > x_1^*$, and -1 when $x_2^* < x_1^*$. Accordingly, we find that $\Sigma_1^2 = \{-1, 1\}$, and this is $\mathbf{S}^0(1)$ as already noted. Of course, quotienting on the left by $\mathbf{SO}(1)$ is here irrelevant because $\mathbf{SO}(1)$ is the trivial group.

Next, suppose that $k = 3$. We then find, using Q_3, that

$$\tilde{x}_1 = \tfrac{1}{\sqrt{2}}(x_2^* - x_1^*) \quad \text{and} \quad \tilde{x}_2 = \tfrac{1}{\sqrt{6}}\{2x_3^* - (x_1^* + x_2^*)\},$$

and after dividing by the size s, where

$$s = \sqrt{\tilde{x}_1^2 + \tilde{x}_2^2},$$

the components $x_1 = \tilde{x}_1/s$ and $x_2 = \tilde{x}_2/s$ of the shape will satisfy the equation

$$x_1^2 + x_2^2 = 1,$$

so that

$$\Sigma_1^3 = \mathbf{S}^1(1).$$

This argument extends to general k, and it tells us that $\Sigma_1^k = \mathbf{S}^{k-2}(1)$, confirming our claim for these spaces.

We turn next to the identification of the shape spaces Σ_2^k. Here, we consider a not totally coincident k-ad of points

$$\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \ldots, \boldsymbol{x}_k^*$$

in two dimensions. Assuming that these are the coordinates after we have moved the centroid to the origin and normalised the size, we construct the pre-shape in the form

$$\boldsymbol{x}_j = \frac{1}{\sqrt{j + j^2}}\{j\boldsymbol{x}_{j+1}^* - (\boldsymbol{x}_1^* + \boldsymbol{x}_2^* + \cdots + \boldsymbol{x}_j^*)\}, \qquad j = 1, 2, \ldots, k-1,$$

following the specification of Q_k. For $m = 2$ it is, of course, natural to think of each $\boldsymbol{x}_j$ as a complex number, z_j, so we can think of the $2 \times (k-1)$ matrix X as identified with the ordered set of complex numbers $z = (z_1, z_2, \ldots, z_{k-1})$. To get the shape from this pre-shape we still have to quotient out the action of $\mathbf{SO}(2)$ acting on the left, which is just scalar multiplication of $z = (z_1, z_2, \ldots, z_{k-1})$ by the group $\{e^{i\alpha} : \alpha \in [0, 2\pi)\}$ of complex numbers of unit modulus. The resulting quotient space is known as the complex projective space $\mathbf{CP}^{k-2}(4)$, where the '4' is the value of the complex curvature constant that is determined by our unit-size convention.

Since we have excluded the totally coincident k-ad, not all the z_j will be zero. Then, if $z_j \neq 0$ for some particular j, the ordered set of complex numbers

$$(z_1/z_j, z_2/z_j, \ldots, z_{j-1}/z_j, 1, z_{j+1}/z_j, \ldots, z_{k-1}/z_j)$$

is invariant under the above action of $\mathbf{SO}(2)$. Ignoring the redundant entry '1', this provides us with a local coordinate system, which we shall employ from time to time, on all the shape space except for the points where $z_j = 0$. These excluded points, in fact, form a subspace isometric with $\mathbf{CP}^{k-3}(4)$. We have thus confirmed the identification of Σ_2^k in the column for $m = 2$ in Table 1.1. Note, in particular, that $\Sigma_2^2 = \mathbf{CP}^0(4)$, this being a one-point space.

A particular identification already referred to is

$$\Sigma_2^3 = \mathbf{CP}^1(4) = \mathbf{S}^2(\tfrac{1}{2}).$$

Now the metric identification of $\mathbf{CP}^1(4)$ with $\mathbf{S}^2(\frac{1}{2})$ is a classical theorem, but for the sake of completeness we will set out the details below. We shall make frequent use of this both as it stands and also in a flat, but not isometric, version obtained by stereographic projection from a point of the sphere $\mathbf{S}^2(\frac{1}{2})$ onto the tangent plane at the antipodal or some other point.

Before entering into a more detailed study of Σ_2^k, it will be helpful to note a few common features of the general shape spaces Σ_m^k. For every k and m, Σ_m^k is $\mathcal{S}_m^k$ modulo $\mathbf{SO}(m)$, where $\mathcal{S}_m^k$ denotes the *pre-shape* unit sphere $\mathbf{S}^{m(k-1)-1}(1)$ and the rotations T in $\mathbf{SO}(m)$ act from the left on the matrices X that are the points of $\mathcal{S}_m^k$. We write

$$\pi : \mathcal{S}_m^k \longrightarrow \Sigma_m^k; \quad X \mapsto \pi(X)$$

for the quotient mapping. Although, in some papers, the shape is denoted by $[X]$ rather than $\pi(X)$, we shall use the latter notation throughout this book. It is natural to choose the customary metric topology on $\mathcal{S}_m^k$ and the corresponding quotient topology on Σ_m^k for which the open sets are the images of the $\mathbf{SO}(m)$-saturated open sets in $\mathcal{S}_m^k$. Then, the mapping π is continuous. It follows that, with this topology, shape space Σ_m^k is *compact* and that it is *connected* whenever $\mathcal{S}_m^k$ is connected, that is, whenever (k, m) is not $(2, 1)$.

Equivalently, we can say that the open sets in Σ_m^k are those determined by the quotient metric ρ defined by the fundamental formula

$$\rho(\pi(X), \pi(Y)) = \min_{T \in \mathbf{SO}(m)} d(X, TY), \tag{1.1}$$

in which each of X and Y is a pre-shape and, when $m \geqslant 2$,

$$d(X, Y) = 2 \arcsin\left(\tfrac{1}{2}||X - Y||\right),$$

which is equal to

$$\arccos \operatorname{tr}(YX^t) \tag{1.2}$$

and so is, in fact, *the great-circle metric on the unit sphere* $\mathcal{S}_m^k$. This d-metric on $\mathcal{S}_m^k$ is topologically equivalent to the norm-metric or 'chordal metric' inherited from $\mathbb{R}^{m(k-1)}$.

When (1.1) and (1.2) are combined we see that for $m \geqslant 2$ *the distance* ρ *between two shapes* $\pi(X)$ *and* $\pi(Y)$ *is given by*

$$\rho(\pi(X), \pi(Y)) = \min_{T \in \mathbf{SO}(m)} \arccos \operatorname{tr}(TYX^t), \tag{1.3}$$

where we always have $0 \leqslant \rho \leqslant \pi$. *This important result underlies the whole discussion of the metric geometry of shape spaces for all values of* $m \geqslant 2$.

We now return to the special, but particularly interesting, case Σ_2^3 and, as before, we take $\boldsymbol{x}_1^*, \boldsymbol{x}_2^*$ and $\boldsymbol{x}_3^*$ to be the labelled vertices of a not totally degenerate triangle in $\mathbb{R}^2$. We then have a number of different ways to represent the shapes of such triangles. If for the moment we leave on one side the special case when $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ coincide, then the triangle $(\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \boldsymbol{x}_3^*)$ with O as the mid-point of $\boldsymbol{x}_1^*\boldsymbol{x}_2^*$ can be arranged to have the vector $O\boldsymbol{x}_2^*$ horizontal and of unit length without altering the shape situation. In this way, all except the excluded shape are represented uniquely by the resulting position of $\boldsymbol{x}_3^*$ in the plane.

Alternatively, we may reduce the specification to $(\tilde{\boldsymbol{x}}_1, \tilde{\boldsymbol{x}}_2) \in \mathbb{R}^2 \times \mathbb{R}^2$ or $(z_1, z_2) \in \mathbb{C}^2$ in the manner already explained. Then, as explained above, because of the freedom to rotate and rescale the data we can specify the shape either by

(i) $(1, \zeta)$ where $\zeta = z_2/z_1$ and $z_1 \neq 0$,

or by

(ii) $(\zeta, 1)$ where $\zeta = z_1/z_2$ and $z_2 \neq 0$.

It may be checked that the coordinate ζ in (i) is $\boldsymbol{x}_3^*/\sqrt{3}$ when $\boldsymbol{x}_3^*$ is obtained as in the previous paragraph. Alternatively, if we normalise $O\boldsymbol{x}_2^*$ to have length $1/\sqrt{3}$ in the previous paragraph, then the resulting $\boldsymbol{x}_3^*$ will have coordinate precisely ζ.

When a more symmetrical parameterisation of the shapes in Σ_2^3 is required, then it is convenient to use

$$\frac{(z_1, z_2)}{\sqrt{|z_1|^2 + |z_2|^2}}$$

for the pre-shape, and similarly for points in Σ_2^k with general k.

When $k = 3$ and $z_1 \neq 0$ another useful coding of the shape is

$$\frac{1}{\sqrt{1 + r^2}} \begin{pmatrix} 1 & r\cos\theta \\ 0 & r\sin\theta \end{pmatrix} = \frac{(1, \zeta)}{\sqrt{1 + |\zeta|^2}},$$

where $r = |z_2/z_1|$. That version is often convenient for calculation and it leads to informative pictorial representations. When $z_1 = 0$, that is, when $\boldsymbol{x}_1^*$ coincides with $\boldsymbol{x}_2^*$, we can find out what happens by letting r tend to infinity. There the coded shape reduces to

$$\begin{pmatrix} 0 & \cos\theta \\ 0 & \sin\theta \end{pmatrix},$$

or, equivalently,

$$(0,\ e^{i\theta}),$$

and this, modulo $\mathbf{SO}(2)$ acting on the left, is the shape point (0, 1) in complex coordinates, which is the only shape excluded from the coordinate representation (i) above.

Whichever of the above representations of the shapes is chosen, however, we must not lose sight of the fact that *the shape really is the ratio* z_2/z_1, and similar remarks apply when working with shape spaces Σ_2^k for general k.

Now suppose that we are interested in *two* shapes, these being identified, say, as

$$\zeta_1 = r_1 e^{i\theta_1}$$

and

$$\zeta_2 = r_2 e^{i\theta_2},$$

where r_1 and r_2 are real, non-negative and finite. To find the shape-theoretic distance $\rho(\zeta_1, \zeta_2)$ between these two shapes we calculate the trace of the triple matrix-product

$$\begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} 1 & r_1\cos\theta_1 \\ 0 & r_1\sin\theta_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ r_2\cos\theta_2 & r_2\sin\theta_2 \end{pmatrix}$$

divided by $\sqrt{(1+r_1^2)(1+r_2^2)}$, which reduces to

$$\frac{\{1 + r_1 r_2\cos(\theta_1 - \theta_2)\}\cos\alpha + r_1 r_2 \sin(\theta_1 - \theta_2)\sin\alpha}{\sqrt{(1+r_1^2)(1+r_2^2)}}.$$

We now need the maximum value of this ratio for $0 \leqslant \alpha < 2\pi$: namely,

$$\sqrt{\frac{1 + r_1^2 r_2^2 + 2r_1 r_2\cos(\theta_2 - \theta_1)}{(1+r_1^2)(1+r_2^2)}}.$$

The value of the inter-shape distance $\rho(\zeta_1, \zeta_2)$ is then the arc-cosine of the above expression.

This formula for the shape-distance, which is necessarily always finite, simplifies if we introduce super-abundant shape variables $(\tilde{a}, \tilde{b}, \tilde{c})$ that are related to the variables (r, θ) by the formulae

$$\tilde{a} = \frac{r\cos\theta}{1+r^2},$$
$$\tilde{b} = \frac{r\sin\theta}{1+r^2},$$
$$\tilde{c} = \frac{r^2-1}{2(1+r^2)},$$

for then it will be seen that the new shape coordinates $(\tilde{a}, \tilde{b}, \tilde{c})$ satisfy the equation

$$\tilde{a}^2 + \tilde{b}^2 + \tilde{c}^2 = \tfrac{1}{4}.$$

Thus, $(\tilde{a}, \tilde{b}, \tilde{c})$ *is a point on the sphere* $\mathbf{S}^2(\frac{1}{2})$. Moreover, this one-to-one correspondence between the points of that sphere and the shapes in Σ_2^3 is an *isometric correspondence*, since

$$2\cos^2\rho(\zeta_1, \zeta_2) - 1 = \cos 2\rho(\zeta_1, \zeta_2) = 4\langle(\tilde{a}_1, \tilde{b}_1, \tilde{c}_2), (\tilde{a}_2, \tilde{b}_2, \tilde{c}_2)\rangle.$$

Accordingly, we see that *the shape space* Σ_2^3 *is indeed the sphere* $\mathbf{S}^2(\frac{1}{2})$, with the point $(0, 0, \frac{1}{2})$ replacing the compactification point at infinity. This takes care of the otherwise excluded point $(0, 0, \frac{1}{2})$ on the sphere.

To visualise how the various triangular shapes lie upon this sphere we note first that $\mathbf{SO}(3)$ acts naturally on it as a group of isometries. In particular, there is the group of six isometries induced by the permutations of the labels for the vertices of the original triangles. The result of this is that $\mathbf{S}^2(\frac{1}{2})$ is split into six equivalent lunes with their common 'upper' and 'lower' vertices corresponding to the two possible shapes of equilateral triangles. Then, for each unlabelled triangle, the six shapes obtained by labelling the vertices will occur at corresponding points in the six lunes. Thus, if we are only interested in the shapes of unlabelled triangles, these may be specified in just one of these lunes. In addition, the mapping of each triangle to its reflection induces an isometry ι_2 of the shape space, which, in the above representation, corresponds to mapping the 'upper' half of each lune onto its 'lower' half. As a result, up to labelling and reflection, all possible shapes may be represented in such a half-lune. If we now project this lune onto a right circular cylinder that touches $\mathbf{S}^2(\frac{1}{2})$ along a great circle and then open out that cylinder onto a plane, then we obtain, for a suitable choice of cylinder, the outline in Figure 1.1 on which a selection of shapes are indicated at their representative points on the projection. It is important to note that although this projection is not isometric it *is* measure-preserving and the metric distortions are reasonably controlled. In this diagram the shapes of degenerate triangles, that is, those in which the three vertices are collinear, lie along the base curve and the shape of the regular, equilateral triangles at the upper vertex is maximally remote from them. The shapes of isosceles triangles appear along the other two

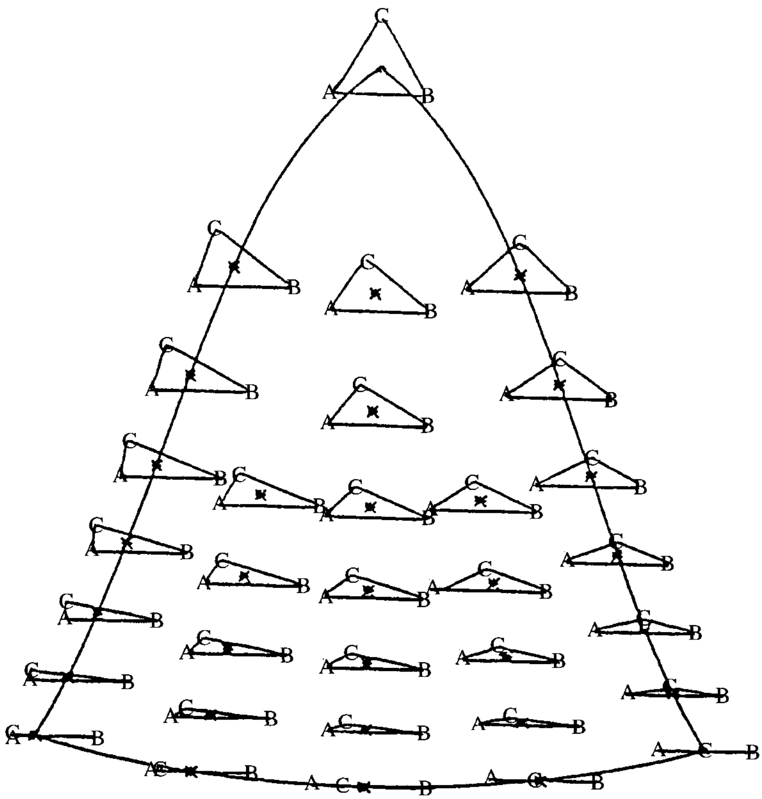

Figure 1.1 Reproduced with permission from D.G. Kendall, The statistics of shape, in V. Barnett (ed), *Interpreting Multivariate Date*. Copyright John Wiley & Sons Ltd

boundary arcs with those whose third angle is less than $\pi/3$ on the left-hand arc and the third angle progressively increasing to π as they move down the right-hand arc.

To compute the metric on the shape space sphere $S^2(\frac{1}{2})$ in terms of the coordinate ζ in (i) above we look first at the way in which stereographic projection links the points on the shape sphere to the points in the compactified plane. We start by translating the above sphere of radius one-half upwards along the $\tilde{c}$-axis so that it sits on the plane spanned by $O\tilde{a}$ and $O\tilde{b}$ with the point of contact at the origin O and $O\tilde{c}$ passing through the centre of the sphere. We label the new coordinates a, b, c, so that the equation of the translated sphere is $a^2 + b^2 + c^2 = c$. The point N at the top of the sphere with coordinate $(0, 0, 1)$ corresponds to the shape of those triangles that have their first two vertices coincident, and the point of contact $(0, 0, 0)$ between the sphere and the plane corresponds to the shape of those triangles whose third vertex lies midway between the first two. If $(x, y, 0)$ is an arbitrary point of the supporting plane and if (a, b, c), where $c \neq 1$, is the