1

Introduction

1.1 Definition and motivation

Objects are everywhere, natural and man-made. Geometrical data from objects are routinely collected all around us, from sophisticated medical scans in hospitals to ubiquitous smart-phone camera images. Decisions about objects are often made using their sizes and shapes in geometrical data, for example disease diagnosis, face recognition and protein identification. Hence, developing methods for the analysis of size and shape is of wide, growing importance. Locating points on objects is often straightforward and we initially consider analysing such data, before extending to curved outlines, smooth surfaces and full volumes.

Size and shape analysis is of great interest in a wide variety of disciplines. Some specific applications follow in Section 1.4 from biology, chemistry, medicine, image analysis, archaeology, bioinformatics, geology, particle science, genetics, geography, law, pharmacy and physiotherapy. As many of the earliest applications of shape analysis were in biology we concentrate initially on biological examples and terminology, but the domain of applications is in fact very broad indeed.

The word 'shape' is very commonly used in everyday language, usually referring to the appearance of an object. Following Kendall (1977) the definition of shape that we consider is intuitive.

Definition 1.1 **Shape** *is all the geometrical information that remains when location, scale and rotational effects are removed from an object.*

An object's shape is invariant under the Euclidean similarity transformations of translation, scaling and rotation. For example, the shape of a human skull consists of all the geometrical properties of the skull that are unchanged when it is translated, rescaled or rotated in an arbitrary coordinate system. Two objects have the same shape

Statistical Shape Analysis, with Applications in R, Second Edition. Ian L. Dryden and Kanti V. Mardia.

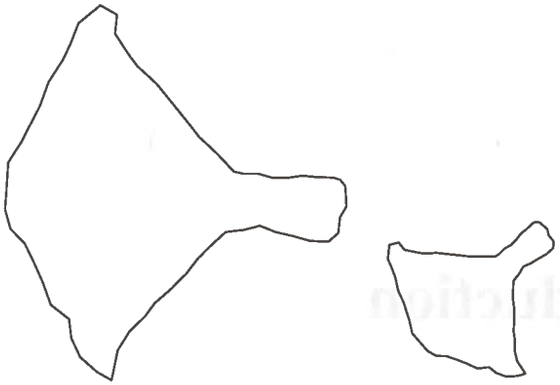


Figure 1.1 Two outlines of the same second thoracic (T2) vertebra of a mouse, which have different locations, rotations and scales but the same shape.

if they can be translated, rescaled and rotated to each other so that they match exactly, that is if the objects are similar. In Figure 1.1 the two mouse vertebrae outlines have the same shape. In practice we are interested in comparing objects with different shapes and so we require a way of measuring shape, some notion of distance between two shapes and methods for the statistical analysis of shape.

Sometimes we are also interested in retaining scale information (size) as well as the shape of the object, and so the joint analysis of size and shape (or form) is also very important.

Definition 1.2 **Size-and-shape** *is all the geometrical information that remains when location and rotational effects are removed from an object.*

Two objects have the same size-and-shape if they can be translated and rotated to each other so that they match exactly, that is if the objects are rigid-body transformations of each other. 'Size-and-shape' is also frequently denoted as **form** and we use the terms equivalently throughout the text.

A common theme throughout the text is the geometrical transformation of objects. The terms superimposition, superposition, registration, transformation, pose and matching are often used equivalently for operations which involve transforming objects, either with respect to each other or into a specified reference frame.

An early writing on shape was by Galileo (1638), who observed that bones in larger animals are not purely scaled up versions of those in smaller animals; there is a shape difference too. A bone has to become proportionally thicker so that it does not break under the increased weight of the heavier animal, see Figure 1.2. The field of geometrical shape analysis was initially developed from a biological point of view by Thompson (1917), who also discussed this application.

How should a scientist wishing to investigate a shape change proceed? Even describing an object's shape is difficult. In everyday conversation an object's shape

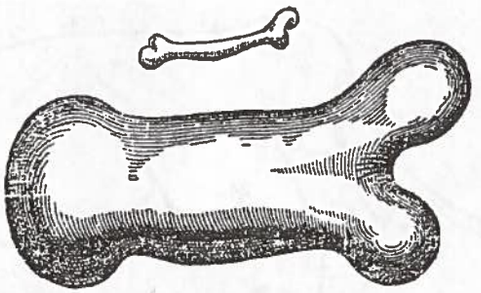


Figure 1.2 From Galileo (1638) illustrating the differences in shapes of the bones of small and large animals.

is usually described by naming a second more familiar shape which it looks like, for example a map of Italy is 'boot shaped'. This leads to very subjective descriptions that are unsuitable for most applications. A practical way forward is to locate a finite set of points on each object, which summarize the key geometrical information.

1.2 Landmarks

Initially we will describe a shape by locating a finite number of points on each specimen which are called landmarks.

Definition 1.3 *A* **landmark** *is a point of correspondence on each object that matches between and within populations.*

There are three basic types of landmarks in our applications: scientific, mathematical and pseudo-landmarks. In the literature there have been various synonyms for landmarks, including vertices, anchor points, control points, sites, profile points, 'sampling' points, design-points, key points, facets, nodes, model points, markers, fiducial markers, markers, and so on.

A **scientific landmark** is a point assigned by an expert that corresponds between objects in some scientifically meaningful way, for example the corner of an eye or the meeting of two sutures on a skull. In biological applications such landmarks are also known as **anatomical landmarks** and they designate parts of an organism that correspond in terms of biological derivation, and these parts are called homologous (e.g. see Jardine 1969). In Figure 1.3 we see some anatomical landmarks located on the skull of a macaque monkey, viewed from the side. This application is described further in Section 1.4.3. Another example of a scientific landmark is a carbon C_α atom of an amino acid on a protein backbone, as seen in Section 1.4.9.

Mathematical landmarks are points located on an object according to some mathematical or geometrical property of the figure, for example at a point of high curvature or at an extreme point. The use of mathematical landmarks is particularly useful in automated recognition and analysis.

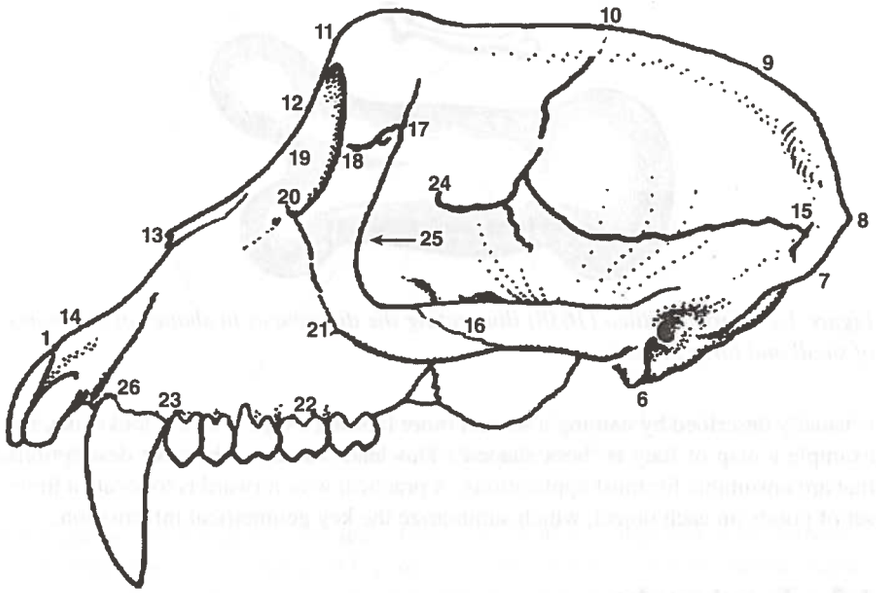


Figure 1.3 Anatomical landmarks located on the side view of a macaque monkey skull.

Pseudo-landmarks are constructed points on an object, located either around the outline or in between scientific or mathematical landmarks. For example, Lohmann (1983) took equally spaced points on the outlines of micro-fossils. In Figure 1.4 we see six mathematical landmarks at points of high curvature and seven pseudo-landmarks marked on the outline inbetween each pair of landmarks on a second thoracic (T2) mouse vertebra. Continuous curves can be approximated by a large number of pseudo-landmarks along the curve. Hence, continuous data can also be studied by landmark methods, although one needs to work with discrete approximations and the choice of spacing of the pseudo-landmarks is crucial. Examples of such approaches include the analysis of hand shapes (Grenander *et al.* 1991; Mardia *et al.* 1991; Cootes *et al.* 1992), resistors (Cootes *et al.* 1992, 1994), mitochondrial outlines (Grenander and Miller 1994), carotid arteries (Cheng *et al.* 2014; Sangalli *et al.* 2014) and mouse vertebrae (Cheng *et al.* 2016). Also, pseudo-landmarks are useful in matching surfaces, when points can be located on a grid over each surface, for example the cortical surface of the brain (Brignell *et al.* 2010) or the surface of the hippocampus (Kurtek *et al.* 2011).

Bookstein (1991) also demarks landmarks into three further types, which are of particular use in biology. **Type I landmarks** occur at the joins of tissues/bones; **type II landmarks** are defined by local properties such as maximal curvatures, and **type III landmarks** occur at extremal points or constructed landmarks, such as maximal diameters and centroids.

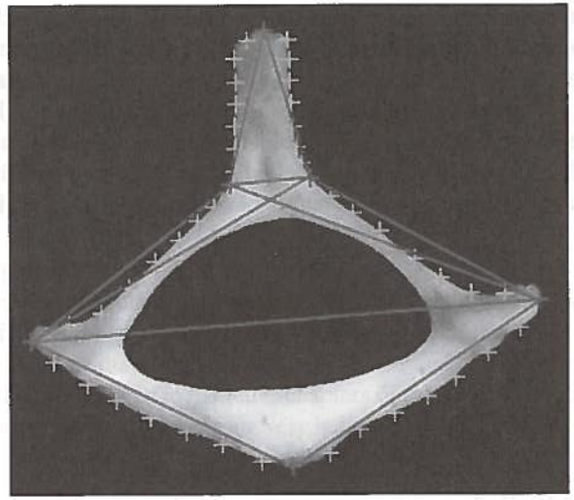


Figure 1.4 Image of a T2 mouse vertebra with six mathematical landmarks on the outline joined by lines (dark +) and 42 pseudo-landmarks (light +). Source: Dryden & Mardia 1998. Reproduced with permission from John Wiley & Sons. For a colour version of this figure, see the colour plate section.

A further type of landmark is the **semi-landmark** which is a point located on a curve and allowed to slip a small distance in a direction tangent to another corresponding curve (Bookstein 1996a,c; Green 1996; Gunz *et al.* 2005). The term 'semi-' is used because the landmark lies in a lower number of dimensions than other types of landmarks, for example along a one-dimensional (1D) curve in a two-dimensional (2D) image (see Section 16.3).

A further situation that may arise is the combination of landmarks and geometrical curves. For example, the pupil of the eye may be represented by a landmark at the centre surrounded by a circle, with the radius as an additional parameter. Yuille (1991) and Phillips and Smith (1993, 1994) considered such representations for analysing images of the human face.

Definition 1.4 *A* **label** *is a name or number associated with a landmark, and identifies which pairs of landmarks correspond when comparing two objects. Such landmarks are called* **labelled landmarks**.

The landmark with, say, label 1 on one specimen corresponds in some meaningful way with landmark 1 on another specimen. A labelling is usually known and given as part of the dataset. For example, in labelling the anatomical landmarks on a skull the labelling follows from the definition of the points. When we refer to just 'shape' of landmarks we implicitly mean the shape of labelled landmarks, that is **labelled shape**.

Unlabelled landmarks are those where no labelling correspondence is given between points on different specimens. It may make sense to try to estimate a correspondence between landmarks, although there is usually some uncertainty involved. This approach is common in bioinformatics for example, as seen in Chapter 14.

In some applications there is no natural labelling, and one must treat all permutations of labels as equivalent. The **unlabelled shape** of an object is the geometrical information that is invariant under permutations of the labels, and translation, rotation and scale.

Example 1.1 Consider the simple example in Figure 1.5. The six triangles (A, B, C, D, E and F) are constructed from triples of labelled points (1,2,3). Triangles A and

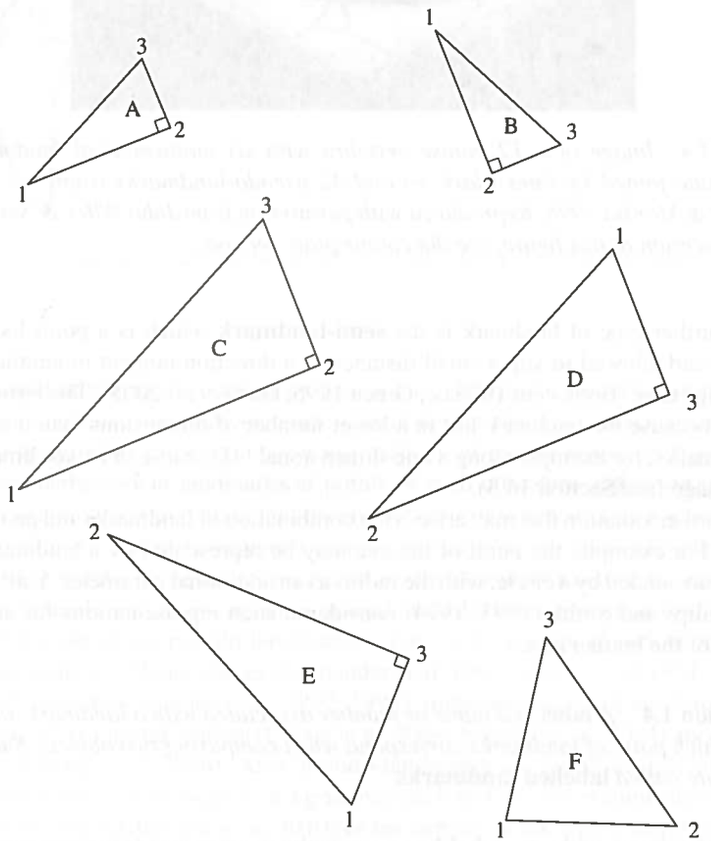


Figure 1.5 Six labelled triangles: A and B have the same size and labelled shape; C has the same labelled shape as A and B (but larger size); D has a different labelled shape but its labels can be permuted to give the same unlabelled shape as A, B and C; triangle E can be reflected to have the same labelled shape as D; triangle F has a different shape from A, B, C, D and E.

B have the same size and the same labelled shape because they can be translated and rotated to be coincident. Triangle C has the same labelled shape as A and B (but has a larger size) because it can be translated, rotated and rescaled to be coincident with A and B. Triangle D has a different labelled shape but, if ignoring the labelling, it has the same unlabelled shape as A, B and C. Triangle E has a different shape to D but it can be reflected and translated to be coincident, and so D and E have the same reflection shape. Triangle F has a different shape from all the rest. □

In the majority of this book the methodology is appropriate for landmark data or other point set data. Following Kendall (1984) our notation will be that there are k landmarks in m dimensions, where we usually have $k \geq 3$ and $m = 2$ or $m = 3$. Extensions to size and shape analysis methods for outline data, surface data and volume data are then considered in the latter chapters of the book, and many of the basic ideas from landmark shape analysis are very helpful for studying these more complex applications.

1.3 The `shapes` package in R

The statistical package and programming language R is an extremely powerful and wide ranging environment for carrying out statistical analysis (Ihaka and Gentleman 1996; R Development Core Team 2015). The progam is available for free download from `http://cran.r-project.org` and is a very widely used and popular platform for carrying out modern statistical analysis. R is continually updated and enhanced by a dedicated and enthusiastic team of developers. R has thousands of contributed packages available, including the `shapes` package (Dryden 2015), which includes many of the methods and datasets from this book. There are numerous introductory texts on using R, including Crawley (2007). For an excellent, comprehensive summary of a wide range of statistical analysis in R, see Venables and Ripley (2002).

We shall make use of the `shapes` package in R (Dryden 2015) throughout the text. Although it is not necessary to follow the R commands, we believe it may be helpful for many readers. To join in interactively the reader should ensure that they have downloaded and installed the base version of R onto their machine. Installation instructions for specific operating systems are given at `http://cran.r-project.org`. After successful installation of the base system the reader should install the `shapes` package. The reader will then be able to repeat many of the examples in the book by typing in the displayed commands.

The first command to issue is:

```
library(shapes)
```

which makes the package available for use. In order to obtain a quick listing of the commands type:

```
library(help=shapes)
```

Also, at any stage it is extremely useful to use the 'help' system, by typing:

```
help(commandname)
```

where `commandname` is the name of a command from the `shapes` package. For example,

```
help(plotshapes)
```

gives information about a basic plotting function for landmark data.

1.4 Practical applications

We now describe several specific applications that will be used throughout the text to illustrate the methodology. Some typical tasks are to study how shape changes during growth; how shape changes during evolution; how shape is related to size; how shape is affected by disease; how shape is related to other covariates such as sex, age or environmental conditions; how to discriminate and classify using shape; and how to describe shape variability. Various methodologies of multivariate analysis have been used to answer such questions over the last 75 years or so. Many of the questions in traditional areas such as biology are the same as they have always been and many of the techniques of shape analysis are closely related to those in multivariate analysis. One of the practical problems is that small sample sizes are often available with a large number of variables, and so high dimension, low sample size issues (large p, small n) are prevalent (e.g. Hall *et al.* 2005). We shall describe many new techniques that are not part of the general multivariate toolkit. As well as traditional biological applications many new problems can be tackled with statistical size and shape analysis.

1.4.1 Biology: Mouse vertebrae

In an experiment to assess the effects of selection for body weight on the shape of mouse vertebrae, three groups of mice were obtained: Large, Small and Control. The Large group contains mice selected at each generation according to large body weight, the Small group was selected for small body weight and the Control group contains unselected mice. The bones form part of a much larger study and these bones are from replicate E of the study (Falconer 1973; Truslove 1976; Johnson *et al.* 1985, 1988; Mardia and Dryden 1989b).

We consider the second thoracic vertebra T2. There are 30 Control, 23 Large and 23 Small bones. The aims are to assess whether there is a difference in size and shape between the three groups and to provide descriptions of any differences. Each vertebra was placed under a microscope and digitized using a video camera to give a grey level image, see Figure 1.4. The outline of the bone is then extracted using

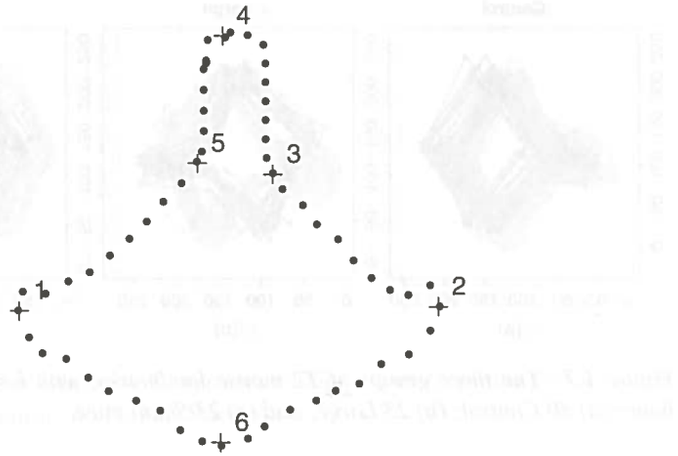


Figure 1.6 Six mathematical landmarks (+) on a second thoracic mouse vertebra, together with 54 pseudo-landmarks around the outline, approximately equally spaced between pairs of landmarks. The landmarks are 1 and 2 at maximum points of approximate curvature function (usually at the widest part of the vertebra rather than on the tips), 3 and 5 at the extreme points of negative curvature at the base of the spinous process, 4 at the tip of the spinous process, and 6 at the maximal curvature point on the opposite side of the bone from 4.

standard image processing techniques (for further details see Johnson *et al.* 1985) to give a stream of about 300 coordinates around the outline. Six landmarks were taken from the outline using a semi-automatic procedure described by Mardia (1989a) and Dryden (1989, Chapter 5), where an approximate curvature function of the smoothed outline is derived and the mathematical landmarks are placed at points of extreme curvature as measured by this function. In Figure 1.6 we see the six landmarks and also in between each pair of landmarks, nine equally spaced pseudo-landmarks are placed.

The dataset is available in the R package `shapes` and the three groups can be accessed by typing:

```
library(shapes)
data(mice)
```

The dataset is stored as a list with three components: `mice$x` is an array of the coordinates in two dimensions of the six landmarks for each bone; `mice$group` is a vector of group labels; and `mice$outlines` is an array of 60 points on each outline in two dimensions, containing the landmarks and pseudo-landmarks. To print the $k \times m \times n$ array of landmarks in R, type `mice$x` and to print the group labels

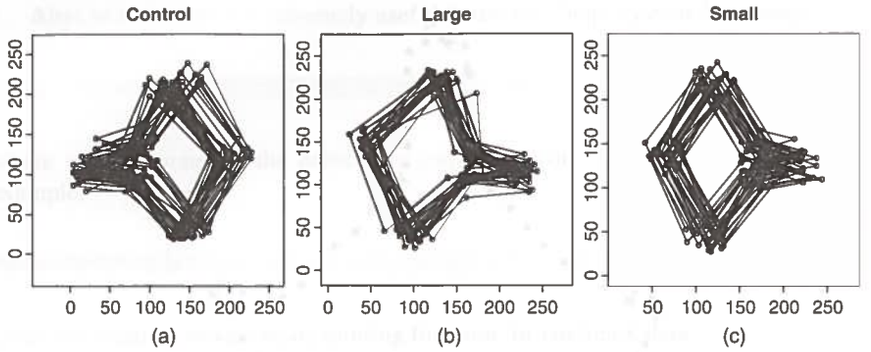


Figure 1.7 The three groups of T2 mouse landmarks, with $k = 6$ landmarks per bone: (a) 30 Control; (b) 23 Large; and (c) 23 Small mice.

type `mice$group` ('c'for Control, 'l' for Large and 's' for Small). In order to plot the landmark data we can use:

```
par(mfrow=c(1,3))
joins<-c(1,6,2:5,1)
plotshapes(mice$x[,,mice$group=="c"],joinline=joins)
title("Control")
plotshapes(mice$x[,,mice$group=="l"],joinline=joins)
title("Large")
plotshapes(mice$x[,,mice$group=="s"],joinline=joins)
title("Small")
```

Here the `plotshapes` function plots 2D (x, y) coordinates of each object, and lines are drawn between the landmarks given in the `joinline` option. Here lines are drawn from landmark 1 to 6 to 2 to 3 to 4 to 5 and finally back to 1 on each object. The plot is given in Figure 1.7.

In order to plot the outline data we can use:

```
par(mfrow=c(1,3))
joins<-c(1:60,1)
plotshapes(mice$outlines[,,mice$group=="c"],joinline=joins,col=2)
title("Control")
plotshapes(mice$outlines[,,mice$group=="l"],joinline=joins,col=2)
title("Large")
plotshapes(mice$outlines[,,mice$group=="s"],joinline=joins,col=2)
title("Small")
```

and the result is given in Figure 1.8. Here the points are drawn in red (col=2) and the lines are drawn to connect points 1 through 60 and back to 1.

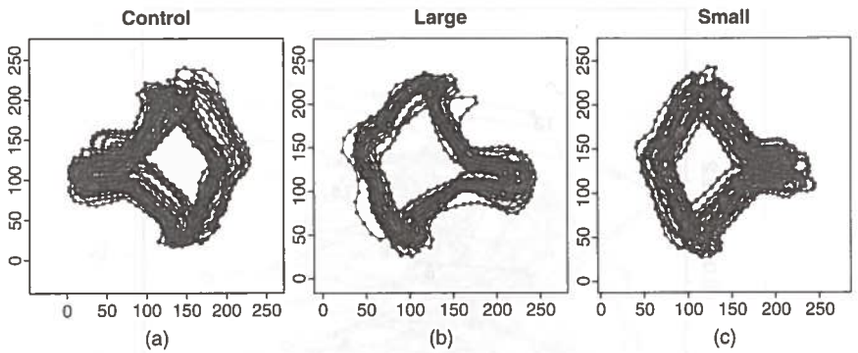


Figure 1.8 The three groups of T2 vertebra outlines, with 60 points per bone: (a) 30 Control; (b) 23 Large; and (c) 23 Small mice.

Note that the coordinates in `mice$x` are also available in the `shapes` package individually by group: `qcet2.dat`, `qlet2.dat`, and `qset2.dat`, which can be useful for short-cuts in coding.

It is of interest to examine size and shape differences in the three groups, and how shape is related to size.

1.4.2 Image analysis: Postcode recognition

A random sample of handwritten British postcodes was collected and digitized (Anderson 1997), and an example digit '3' is shown in Figure 1.9. It is of interest to classify each of the handwritten characters so that mail can be automatically sorted. The problem is a classic one in image analysis and many methods have been suggested, with varying degrees of success (e.g. see Hull 1990). The location and size of the characters are not so important for recognition but orientation information may be crucial, for example an 'M' must not be confused with a 'W'. Some successful attempts at reading handwritten numbers include Simard *et al.* (1993); Hastie and Tibshirani (1994); and Hastie and Simard (1998). A survey of relevant work is given by Plamondon and Srihari (2000), and a related topic is hand-drawn gesture recognition (e.g. see Mardia *et al.* 1993).

Anderson (1997) obtained mathematical landmarks and pseudo-landmarks on the digital images by hand, and in particular for the digit 3 there were 13 landmarks, as shown in Figure 1.9. It is of interest to examine the average shape and variability in shape in the data, which can then be used as a prior model for digit recognition from images of handwritten postcodes.

The landmark data are given the dataset `digit3.dat` in the `shapes` package. The data are displayed in Figure 1.10 using the R command:

```
plotshapes(digit3.dat,joinline=1:13)
```

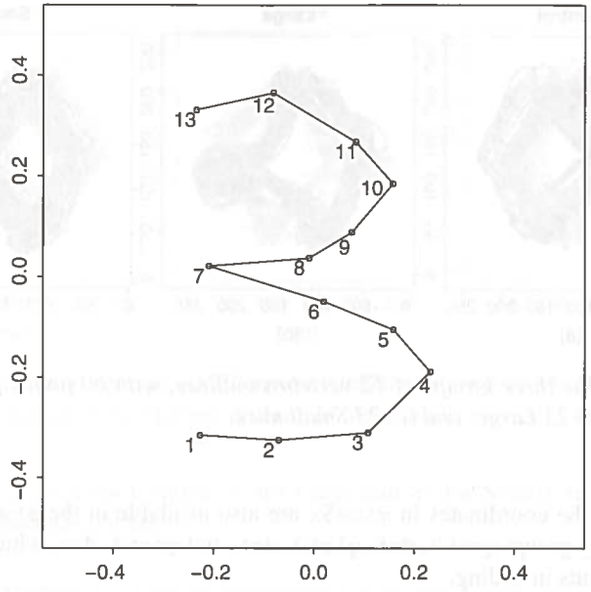


Figure 1.9 A handwritten digit '3' from the postcode dataset, with 13 labelled mathematical landmarks. Landmark 1 is at the extreme bottom left, 4 is at the maximum curvature of the bottom arc, 7 is at the extreme end of the central protrusion, 10 is at the maximum curvature of the top arc and 13 is the extreme top left point. Landmarks 2, 3, 5, 6, 8, 9, 11 and 12 are pseudo-landmarks at approximately equal intervals between the mathematical landmarks.

1.4.3 Biology: Macaque skulls

In an investigation into sex differences in the crania of a species of macaque *Macaca fascicularis* (a type of monkey), random samples of 9 male and 9 female skulls were obtained by Paul O'Higgins (Hull-York Medical School) (Dryden and Mardia 1993). A subset of seven anatomical landmarks was located on each cranium and the three-dimensional (3D) coordinates of each point were recorded.

It is of interest to assess whether there are any size and shape differences between the sexes. If there are any differences, then a description of the differences is required. An artist's impression of the 3D skull with the anatomical landmarks is given in Figure 1.11.

The data are obtained by typing:

```
library(shapes)
data(macaques)
```

The 3D landmarks are available in the $7 \times 3 \times 18$ dimensional array `macaques$x`, and the genders are in `macaques$group`.

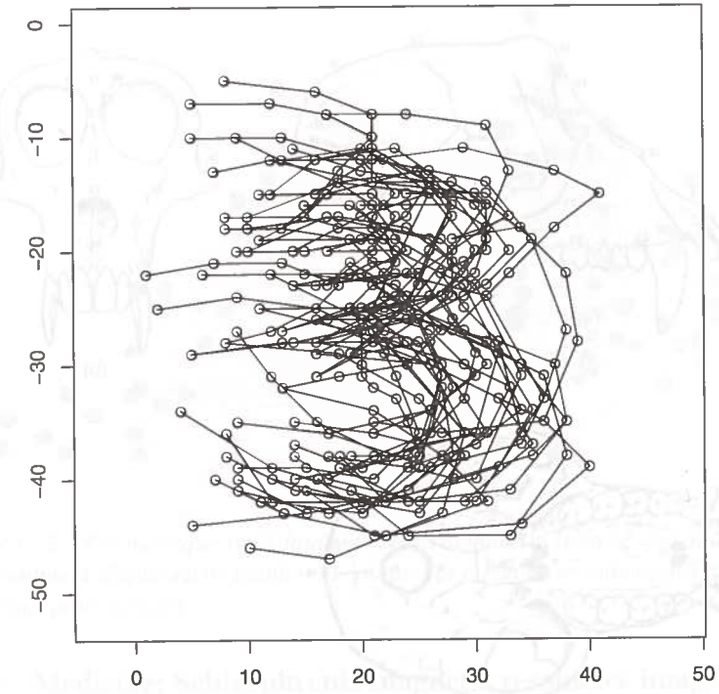


Figure 1.10 The thirty digit 3 configurations, each with 13 landmarks.

We plot the data in Figure 1.12 using the command `shapes3d` as follows:

```
joins<-c(1,2,5,2,3,4,1,6,5,3,7,6,4,7)
colpts<-rep(1:7,times=18)
shapes3d(macaques$x,col=colpts,joinline=joins)
```

The command `shapes3d` uses the `rgl` library in R, which in turn uses OpenGL graphics. The 3D plots can be easily rotated and moved in the graphics window by clicking and moving the mouse, thus giving a good idea of the 3D geometry of the configuration.

Note that the coordinates in `macaques$x` are also available in the `shapes` package individually by group: `macf.dat`, and `macm.dat`.

1.4.4 Chemistry: Steroid molecules

Dryden *et al.* (2007) and Czogiel *et al.* (2011) analyse a dataset of steroids, which are small molecules with a wide variety of uses. The dataset consists of between 42 and 61 atoms for each of 31 steroid molecules. The three-dimensional coordinates, atom type, van der Waals radius and partial charges of each atom are given. The collection of steroids has been considered by a number of authors, including Wagener *et al.*

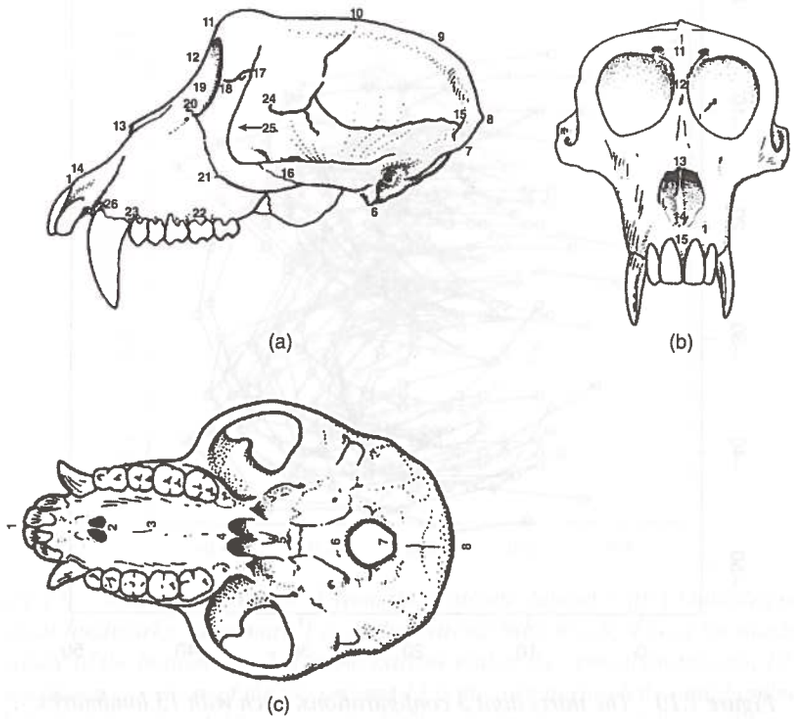


Figure 1.11 A 3D macaque skull: (a) side view; (b) frontal view; and (c) bottom view. A total of 26 landmarks are displayed on the skull and a subset of 7 was taken for the analysis. The seven chosen landmarks are: 1, prosthion*; 7,* opisthion*; 10,* bregma*; 12,* nasion*; 15,* asterion*; 16,* midpoint of zyg/temp suture*; and 17,* interfrontomalare.

(1995). This particular version of the data was constructed by Jonathan Hirst and James Melville (School of Chemistry, University of Nottingham). The steroids have different binding affinities to the corticosteroid binding globulin (CBG) receptor, and so each molecule has an activity class of either '1' high, '2' intermediate or '3' low binding affinity. It is of interest to examine how the shape ('steric') properties of the molecules are related to activity class. This dataset is quite challenging in that the molecules have different numbers of atoms, and the correspondence between atoms (labelling) is not given. However, the 17 carbon atoms in the three cyclohexane rings and one cyclopentane ring are common to all the steroids, and these do correspond in a sensible way. The carbon rings are plotted in Figure 1.13 using the following commands.

```
data(steroids)
joins<-c(1:6,1,6,5,4,7:10,5,4,7,11:14,8,14:17,13)
shapes3d(steroids$x[,,],col=rep(1:17,times=31),joinline=joins)
```

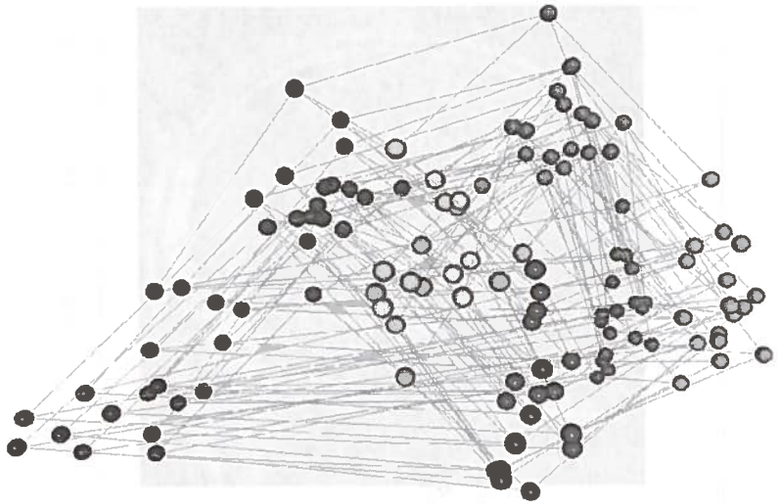



Figure 1.12 The macaque skull data with seven landmarks from 18 individuals, with each landmark displayed by a different colour. For a colour version of this figure, see the colour plate section.

1.4.5 Medicine: Schizophrenia magnetic resonance images

Bookstein (1996b) considers 13 landmarks taken on near midsagittal 2D slices from magnetic resonance (MR) brain scans of 14 control volunteers and 14 schizophrenia patients. It is of interest to study any shape differences in the brain between the two groups, either in average shape or in shape variability. If shape differences between the two groups can be established, then this should enable researchers to gain an increased understanding about the condition. In Figure 1.14 we see the 13 landmarks on a 2D slice from a scan of a schizophrenia patient. The landmarks are: 1, splenium, posteriormost point on corpus callosum; 2, genu, anteriormost point on corpus

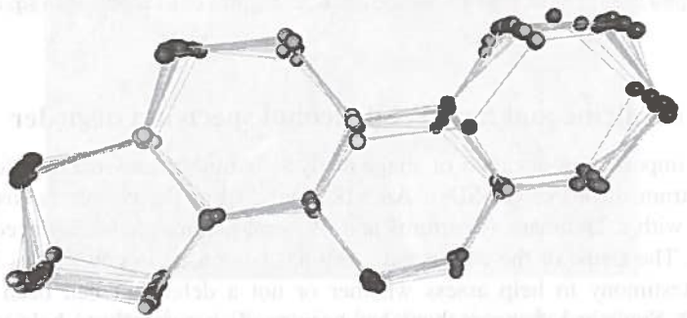


Figure 1.13 The first 17 carbon atoms in the 31 steroid molecules. For a colour version of this figure, see the colour plate section.

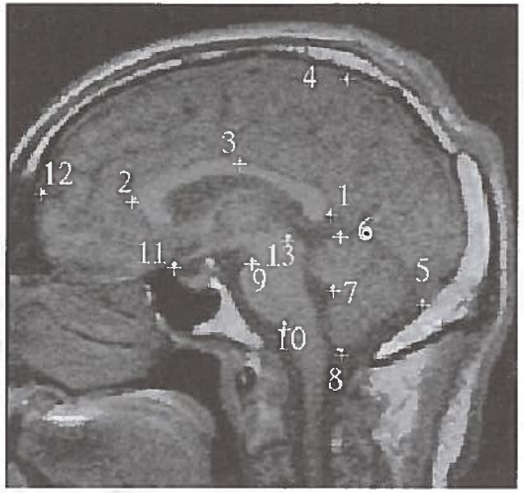


Figure 1.14 The 13 landmarks on a near midsagittal section from a brain scan of a schizophrenia patient. The landmark positions are approximately located at each cross (+). Source: Adapted from Bookstein 1996b. Reproduced with permission from Springer Science+Business Media.

callosum; 3, top of corpus callosum, uppermost point on arch of callosum (all three landmarks registered to the diameter of the callosum); 4, top of head, a point relaxed from a standard landmark along the apparent margin of the dura; 5, tentorium of cerebellum at dura; 6, top of cerebellum; 7, tip of fourth ventricle; 8, bottom of cerebellum; 9, top of pons, anterior margin; 10, bottom of pons, anterior margin; 11, optic chiasm; 12, frontal pole, extension of a line from 1 through 2 until it intersects the dura; and 13, superior colliculus.

The data are plotted in Figure 1.15 using the following commands.

```
data(schizophrenia)
plotshapes(schizophrenia$x,symbol=as.integer(schizophrenia$group))
```

1.4.6 Medicine and law: Fetal alcohol spectrum disorder

Another important application of shape analysis is in the assessment of fetal alcohol spectrum disorders (FASDs). An MR image from the corpus callosum of a prisoner with a landmark (Rostrum) and 39 semi-landmarks is displayed in Figure 1.16. The shape of the corpus callosum has been used in court cases in expert witness testimony to help assess whether or not a defendant had been affected by FASD. Statistical shape analysis has been used successfully to help waive the death penalty for many defendants. For further details see Mardia *et al.* (2013a) and Section 7.10.

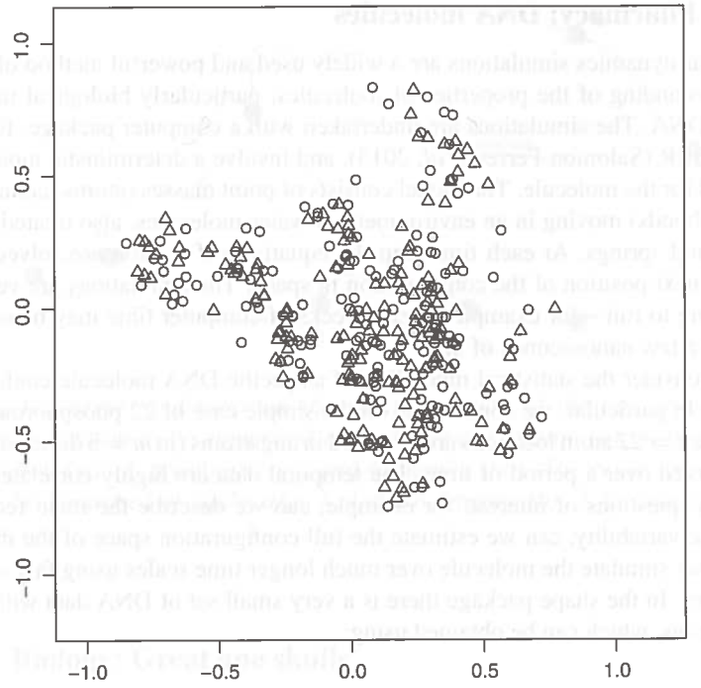


Figure 1.15 The dataset of 13 landmarks per individual from the schizophrenia study, with circles for controls and triangles for patients.

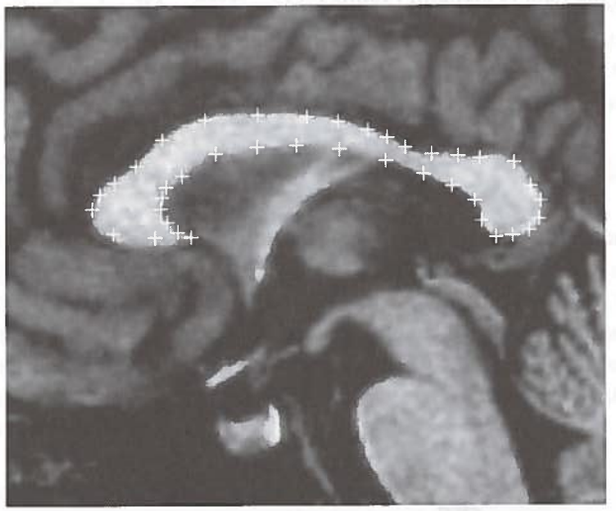


Figure 1.16 A landmark and 39 semi-landmarks on the outline of the corpus callosum from an MR image of a prisoner. Source: Mardia et al. *2013a. Reproduced with permission from John Wiley & Sons.*

1.4.7 Pharmacy: DNA molecules

Molecular dynamics simulations are a widely used and powerful method of gaining an understanding of the properties of molecules, particularly biological molecules such as DNA. The simulations are undertaken with a computer package, for example AMBER (Salomon-Ferrer *et al.* 2013), and involve a deterministic model being specified for the molecule. The model consists of point masses (atoms) connected by springs (bonds) moving in an environment of water molecules, also treated as point masses and springs. At each time step the equations of motion are solved to provide the next position of the configuration in space. The simulations are very time-consuming to run – for example several weeks of computer time may be needed to generate a few nanoseconds of data.

We consider the statistical modelling of a specific DNA molecule configuration in water. In particular, we concentrate on the simple case of 22 phosphorous atoms, where the $k = 22$ atom locations are recorded in angstroms (in $m = 3$ dimensions) and are observed over a period of time. The temporal data are highly correlated. There are many questions of interest, for example, can we describe the main features of geometric variability, can we estimate the full configuration space of the molecule, and can we simulate the molecule over much longer time scales using fast statistical techniques. In the shape package there is a very small set of DNA data with $n = 30$ observations, which can be obtained using:

```
data(dna.dat)
```

A plot of the data is given in Figure 1.17.

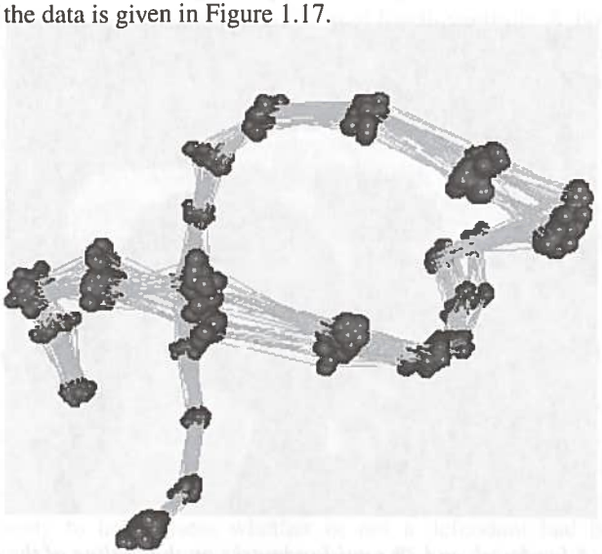


Figure 1.17 A small dataset of 22 phosphorous atoms from a DNA molecule at n = 30 time points. For a colour version of this figure, see the colour plate section.

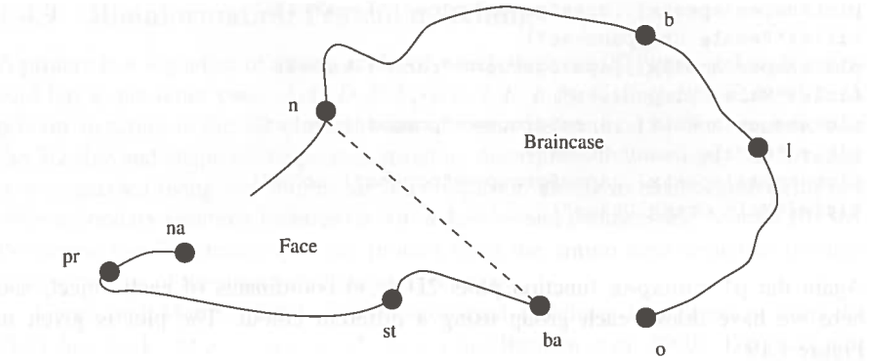


Figure 1.18 Eight landmarks on the midline section of the ape cranium. The face region is taken to be comprised of landmarks: 7, nasion *(n); 4,* basion *(ba); 5,* staphylion *(st); 1,* prosthion *(pr); and 6,* nariale *(na). The braincase region is taken to be comprised of landmarks: 7, 4 and 8,* bregma *(b); 2,* lambda *(l); and 3,* opisthion *(o).*

1.4.8 Biology: Great ape skulls

In an investigation to assess the cranial differences between the sexes of ape, 29 male and 30 female adult gorillas (*Gorilla gorilla*), 28 male and 26 female adult chimpanzees (*Pan*), and 30 male and 24 female adult orangutans (*Pongo*) were studied. The data are described in detail by O'Higgins (1989) and O'Higgins and Dryden (1993). Eight landmarks are chosen in the midline plane of each skull as shown in Figure 1.18. The landmarks are anatomical landmarks and are located by an expert biologist.

The dataset is available in the R package `shapes` and can be accessed by typing:

```
data(apes)
```

The dataset is stored as a list with two components: `apes$x` is an array of coordinates in eight landmarks in two dimensions for each skull; and `apes$group` is a vector of group labels. To print the $k \times m \times n$ array of landmarks in R, type `apes$x` and to see the group labels type `apes$group` ('gorf'for female gorillas, 'gorm' for male gorillas, 'panf' for female chimpanzees, 'panm' for male chimpanzees, 'pongof' for female orangutans, and 'pongom' for male orangutans). In order to plot the landmark data we can use:

```
par(mfcol=c(2,3))
plotshapes(apes$x[,,apes$group=="gorf"],col=1)
title("Female Gorillas")
plotshapes(apes$x[,,apes$group=="gorm"],col=2)
title("Male Gorillas")
```

```
plotshapes(apes$x[,,apes$group=="panf"],col=3)
title("Female Chimpanzees")
plotshapes(apes$x[,,apes$group=="panm"],col=4)
title("Male Chimpanzees")
plotshapes(apes$x[,,apes$group=="pongof"],col=5)
title("Female Orang Utans")
plotshapes(apes$x[,,apes$group=="pongom"],col=6)
title("Male Orang Utans")
```

Again the `plotshapes` function plots 2D (x, y) coordinates of each object, and here we have drawn each group using a different colour. The plot is given in Figure 1.19.

It is of interest to assess whether there is a size difference between the sexes and whether there are any shape differences between the sexes in the face and braincase regions. A biologist would also be interested in geometrical descriptions of the shape difference, and how shape relates to size and other covariates.

Note that the coordinates in `apes$x` also available in the `shapes` package individually by group: `gorf.dat`, `gorm.dat`, `panf.dat`, `panm.dat`, `pongof.dat` and `pongom.dat`.

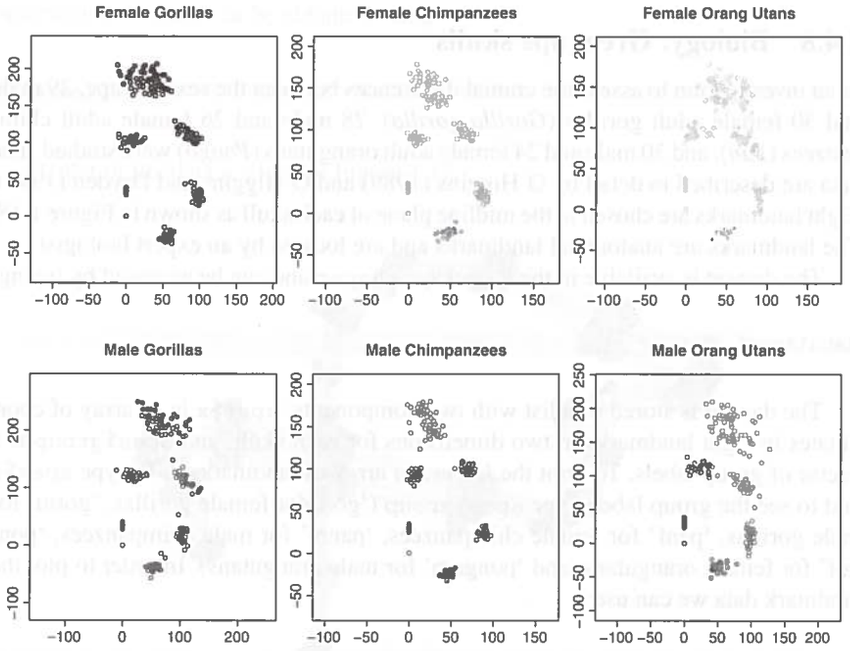


Figure 1.19 The six groups of great ape skull landmarks: (left column) female and male gorillas; (middle column) female and male chimpanzees; and (right column) female and male orangutans. For a colour version of this figure, see the colour plate section.

1.4.9 Bioinformatics: Protein matching

A protein is a sequence of amino acids, of which there are 20 types, and each amino acid has a one-letter code: $A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y$. A protein structure is the 3D configuration of atoms determined by the sequence and the 3D size and shape of the protein structure determines its function. The structure is summarized using key atoms, such as α-carbon atoms in each amino acid, and other secondary structure features such as α-helices and β-sheets (e.g. Mardia 2013b). Predicting the 3D structure of the protein from the amino acid sequence (protein folding) is one of the grand challenges in science.

Green and Mardia (2004, 2006) considered matching two proteins from the PDB data bank `http://www.rcsb.org/pdb` (Berman *et al.* 2000). The particular pair of proteins that were compared had PDB codes `1a27` (Human Type I 17Beta-Hydroxysteroid Dehydrogenase) (Mazza 1997) with 63 points; and `1cyd` (Mouse lung carbonyl reductase) (Tanaka *et al.* 1996) with 40 points. The points are the active sites of the proteins and consist of the coordinates of the centres of gravity of the amino acids that make up the nicotinamide adenine dinucleotide phosphate (NADP) binding sites of two proteins. The 20 amino acid types of each active site are labelled in one of four groups as follows:

- Group 1: Hydrophobic. A, F, I, L, M, P, V.
- Group 2: Charged. D, E, K, R.
- Group 3: Polar. C, H, N, Q, S, T, W, Y.
- Group 4: Glycine. G.

Representations of the full proteins are given in Figure 1.20.

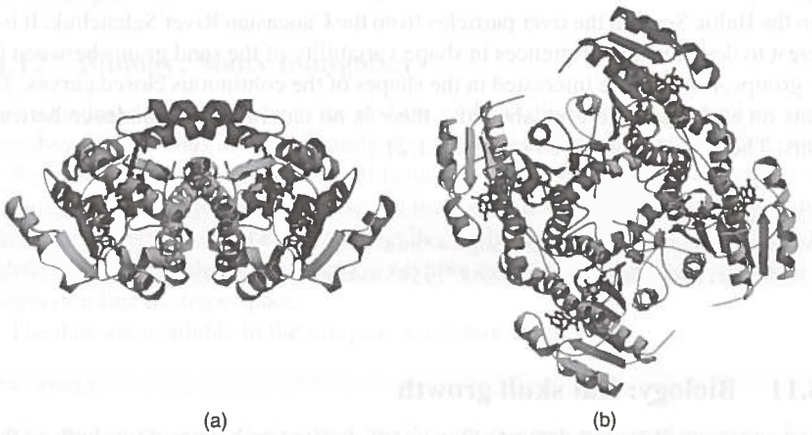


Figure 1.20 The proteins 1a27 (a) and 1cyd (b) from the PDB databank (Tanaka et al. *1996; Mazza 1997; Berman* et al. *2000). For a colour version of this figure, see the colour plate section.*

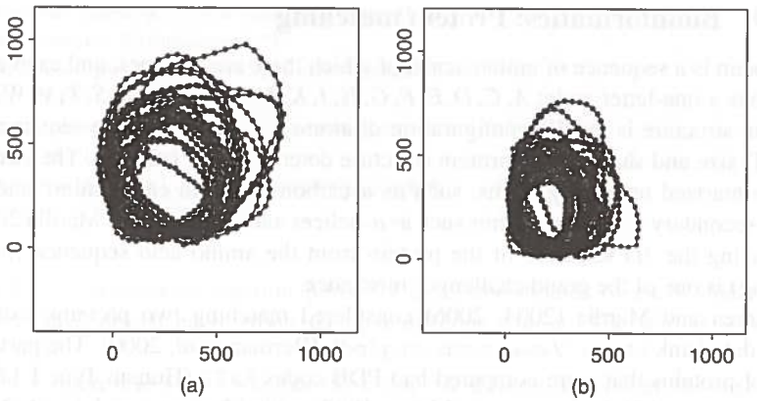


Figure 1.21 The sand particle outlines: (a) sea sand; and (b) river sand.

It is of interest to align the two molecules in order to find common similar geometrical parts. The size and shape of the proteins are of interest, and scale invariance is not required here. In this application the landmarks are unlabelled, and estimation of the correspondence between subsets of the active sites of each protein is required.

1.4.10 Particle science: Sand grains

A dataset of the curved outlines of sand grain profiles is available. There are 24 sea sand and 25 river sand grain profiles in two dimensions. The original data were provided by Dietrich Stoyan (Stoyan and Stoyan 1994; Stoyan 1997). On each particle outline there are 50 points, which were were extracted at approximately equal arc lengths by the method described in Kent *et al.* (2000, section 8.1). The sea particles are from the Baltic Sea and the river particles from the Caucasian River Selenchuk. It is of interest to describe the differences in shape variability of the sand grains between the two groups. Here we are interested in the shapes of the continuous closed curves. The points on each outline are unlabelled – there is no natural correspondence between points. The data are displayed in Figure 1.21.

```
data(sand)
plotshapes(sand$x[,,sand$group=="sea"],
   sand$x[,,sand$group=="river"],joinline=c(1:50,1))
```

1.4.11 Biology: Rat skull growth

We consider a well-known dataset of landmarks located on X-rays of rat skulls as they grow. The data are described in Bookstein (1991) and studied by several other authors including Goodall and Lange (1989); Monteiro (1999); Le and Kume (2000a); Kent *et al.* (2001); and Kenobi *et al.* (2010). The rats were carefully X-rayed at ages 7,

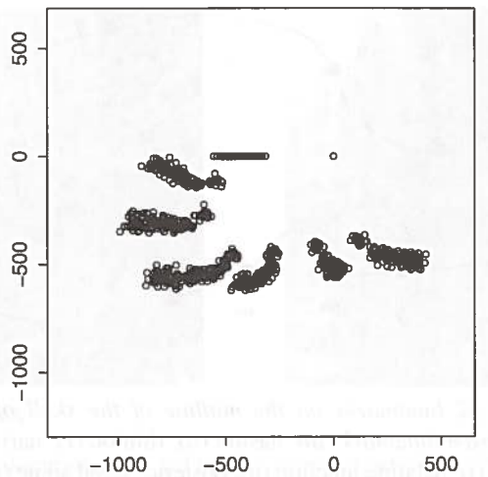


Figure 1.22 The eight landmarks on the 18 rat skulls, observed at eight time points.

14, 21, 30, 40, 60, 90 and 150 days, and there are 18 rats with complete sets of eight landmarks in two dimensions at each age. The X-rays were taken of the skulls of the same rats recorded throughout their lifetimes, and it is of interest to describe the size and shape changes as the rats undergo growth.

The data for the 18 rats with complete data are displayed in Figure 1.22 and are available in the `shapes` package using:

```
data(rats)
```

with the landmark coordinates in `rats$x`, identification number in `rats$no` and the time in `rats$time`.

1.4.12 Biology: Sooty mangabeys

Twelve landmarks are taken from the midline of the skulls of a type of monkey, sooty mangabey (*Cercocebus atys*), in a study described by O'Higgins and Dryden (1992), see Figure 1.23. The specimens ranged from young juveniles to an adult female and an adult male. The objective is to describe the size and shape differences in the individuals in the series from the young juveniles to the older juveniles, and then to the adults. A further problem is to examine whether the individuals can be modelled by a regression line in shape space.

The data are available in the `shapes package` using:

```
data(sooty)
```

The data is an array of size $10 \times 2 \times 7$, with the first two configurations being centred, rotated coordinates for the smallest juvenile and the male adult, and then the last five observations contain the original data for the three juveniles, then the female adult and the male adult.

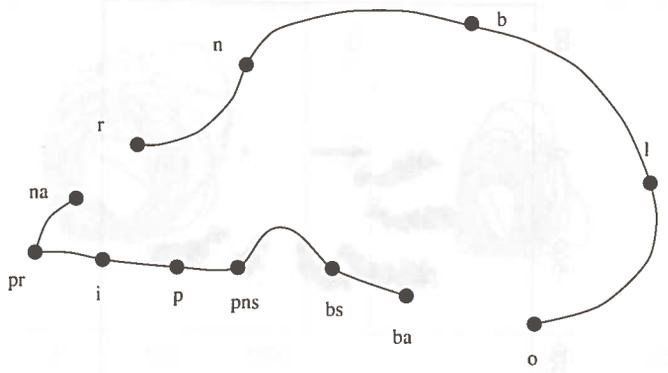


Figure 1.23 The 12 landmarks on the midline of the skull of a juvenile sooty mangabey. The chosen landmarks are nasion *(n),* rhinion *(r),* nariale *(na),* prosthion *(pr),* incisive canal *(i),* palatine junction *(p),* posterior nasal spine *(pns),* basisphenoid *(bs),* basion *(ba),* opisthion *(o),* lambda *(l) and* bregma *(b).*

1.4.13 Physiotherapy: Human movement data

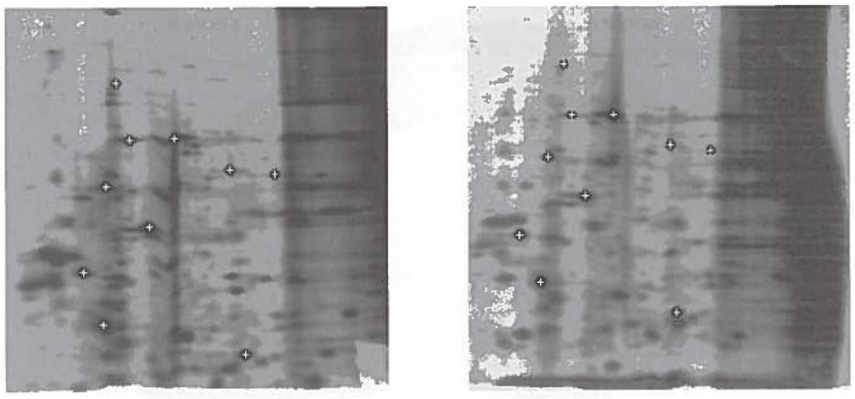
Kume *et al.* (2007) consider an application in the study of human movement data, consisting of $k = 4$ landmarks (lower back, shoulder, wrist and index finger) moving in time. The landmark coordinates are obtained by recording the 3D locations of small reflective markers using a system of seven video cameras. Each individual in the study sat at a table and was asked to move his or her index finger towards a target point which was positioned straight, or to the right or or to the left at different angles. The data were collected by Dr James Richardson, Université Paris Sud, France. We are interested in the shapes of the landmark configurations, and we concentrate on the shapes of the configurations in the plane of the table and this subset was considered by Kume *et al.* (2007). We shall concentrate on a subset of the data where five curves are available, which are labelled a, b, c, d, e. The dataset consists of the projected view of the movements in the plane of the table at which the subject is sitting. For each of the individuals we have 10 equally spaced time points (after initially linearly transforming the times to $[0, 1]$ for each curve). The data are available as a four-dimensional array in the `shapes` package using:

```
data(humanmove)
```

The array is of dimension $4 \times 2 \times 10 \times 5$ and represents the coordinates of the $k = 4$ landmarks in $m = 2$ dimensions with 10 time points for each of the 5 movements.

1.4.14 Genetics: Electrophoretic gels

A technique for the identification of proteins involves the comparison of electrophoretic gel images (Horgan *et al.* 1992). Two examples of such images are gel



(a) (b)

Figure 1.24 The electrophoretic gel images from (a) gel A and (b) gel B. The invariant spots are marked with a '+' in both images. Source: Adapted from Horgan et al. *1992.*

A and gel B shown in Figure 1.24. The images were obtained from particular strains of parasites which carry malaria. The objective is to use the gel image to identify the strain of parasite.

In each gel there are a number of black spots, where each spot can be one of two types – invariant or variant. The invariant spots are present for all parasites and the arrangement of variant spots enables identification of the parasite. A problem with the technique when used in the field is that the gels are prone to deformations and so the gel images first need to be 'registered' (transformed) so that direct comparisons can be made.

In this application, 10 invariant spots have been picked out by an expert, as shown in Figure 1.25. The spots are available in the `shapes` package using:

```
data(gels)
```

The invariant spots are used to match gel A to gel B, either by a similarity transformation or by a more complicated transformation. A question of interest is: can a matching procedure be made resistant to some outlier points, for example mislabelled points?

1.4.15 Medicine: Cortical surface shape

In a study of the shape of cortical surfaces of schizophrenia patients and normal volunteers, some structural MR images were taken of the brain (Brignell *et al.* 2010). The dataset consists of $n = 68$ 3D MR images of the brain from 29 male healthy controls, 25 male schizophrenia patients, 9 female healthy controls, and 5 female

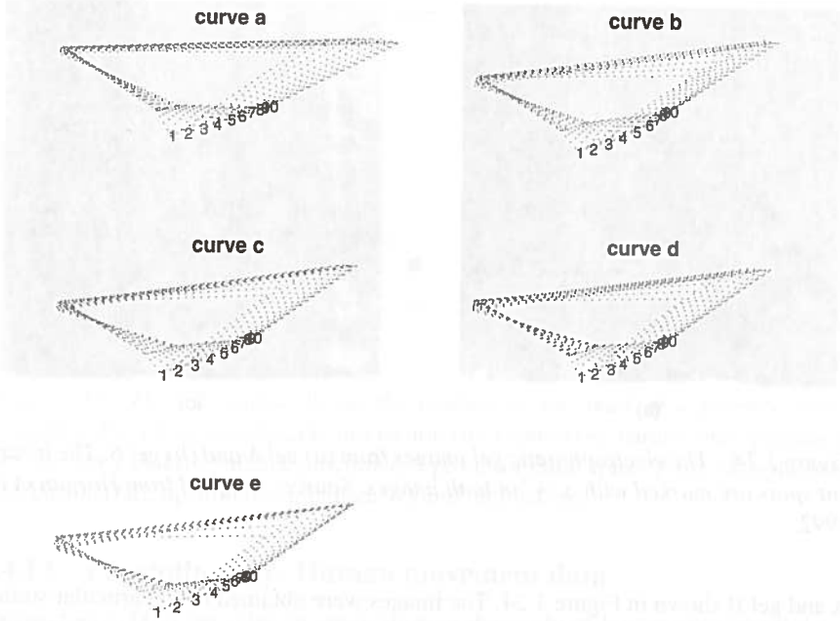


Figure 1.25 The five series of configurations projected into the plane of the table. Each series consists of 10 quadrilaterals observed at equal fractions of the time taken to carry out the pointing movement. Source: Kume et al. *2007. Reproduced with permission of Oxford University Press. For a colour version of this figure, see the colour plate section.*

schizophrenia patients. The MR images are proton density weighted images and were collected by Sean Flynn at the University of British Columbia, Canada. Each volunteer's image consists of $256 \times 256 \times 256$ voxels (3D pixels of size 1 mm^3). The cortical surface was extracted from each dataset using image processing techniques [specifically the brain extraction tool of Smith (2002)]. Finally 62 501 points were located on the surface of the cortex above a transverse plane passing through the anterior and posterior commisures, with the brain midline plane being located in a sagittal plane. We see an example of the cortical surface points of one of the subjects in Figure 1.26. It is of interest to describe the cortical surface shape, and whether there are any differences in shape between the schizophrenia group and the control group.

A subset of the data are available in the `shapes` package:

```
data(cortical)
```

which contains the 250 points on the outline intersecting the axial slice containing the commisures (the bottom-most axial cross-section in Figure 1.26).

Figure 1.26 A set of 62 501 cortical surface points in three dimensions. The colouring indicates the ordering of the points. For a colour version of this figure, see the colour plate section.

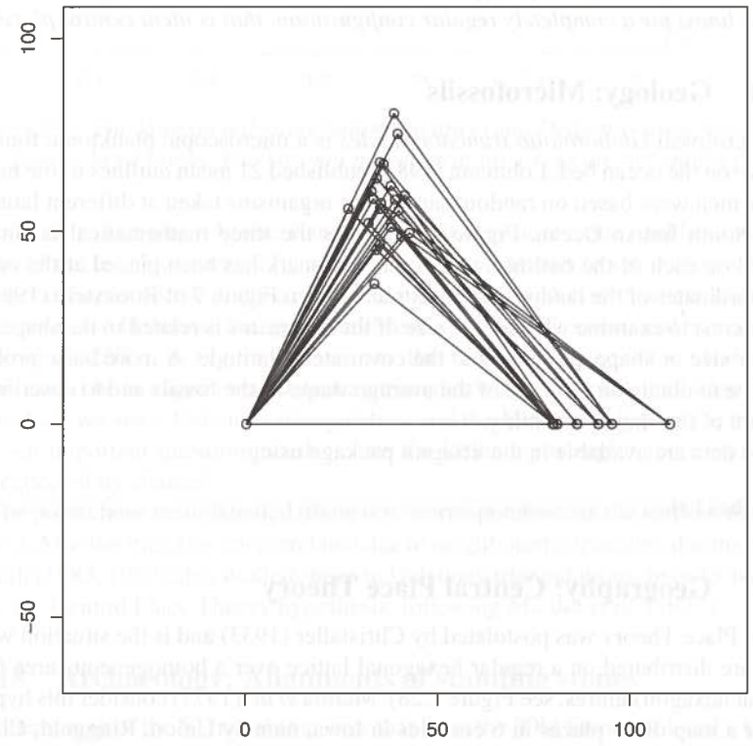


Figure 1.27 Landmarks from 21 mean outlines of microfossils.

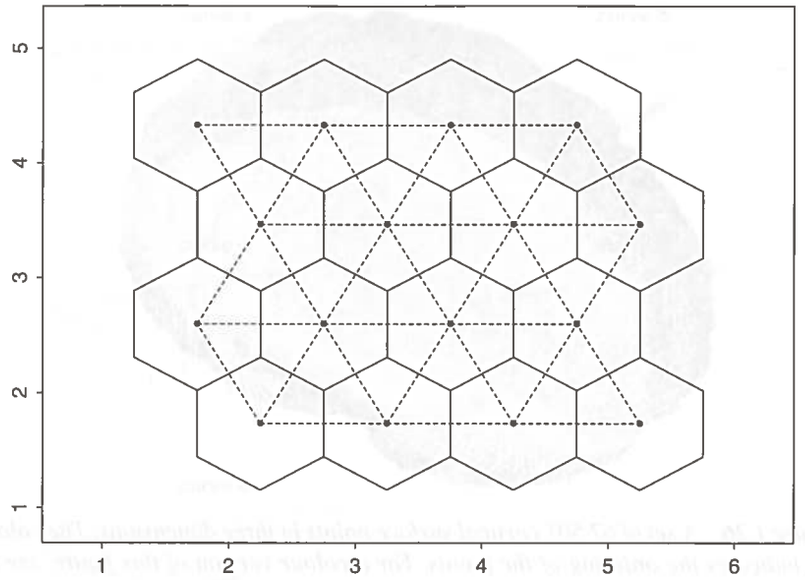


Figure 1.28 The Voronoi polygons (unbroken lines) and Delaunay triangulation (broken lines) for a completely regular configuration, that is ideal central places.

1.4.16 Geology: Microfossils

The microfossil *Globorotalia truncatulinoides* is a microscopic planktonic found in the ooze on the ocean bed. Lohmann (1983) published 21 mean outlines of the microfossil which were based on random samples of organisms taken at different latitudes in the South Indian Ocean. Figure 1.27 shows the three mathematical landmarks selected on each of the outlines, where one landmark has been placed at the origin. The coordinates of the landmarks are extracted from Figure 7 of Bookstein (1986). It is of interest to examine whether the size of the organisms is related to the shape, and whether size or shape are related to the covariate of latitude. A more basic problem would be to obtain an estimate of the average shape of the fossils and to describe the structure of the shape variability.

The data are available in the `shapes` package using:

```
data(shells)
```

1.4.17 Geography: Central Place Theory

Central Place Theory was postulated by Christaller (1933) and is the situation where towns are distributed on a regular hexagonal lattice over a homogeneous area (with towns at hexagon centres, see Figure 1.28). Mardia *et al.* (1977) consider this hypothesis for a map of 44 places in 6 counties in Iowa, namely Union, Ringgold, Clarke, Decatur, Lucas and Wayne.

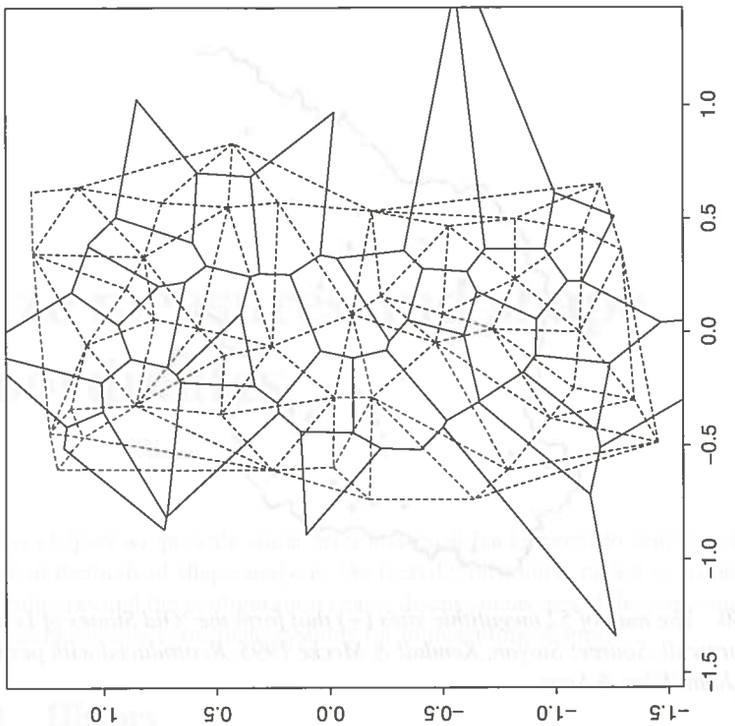


Figure 1.29 The Voronoi polygons (unbroken lines) and Delaunay triangles (broken lines) for the Iowa towns. The Voronoi polygons at the edges are not shown fully.

In order to examine whether Central Place Theory holds, one could examine the shapes of the triangles formed by a town and its neighbours to see if they are more equilateral than expected under a randomness hypothesis. A convenient triangulation of the towns is a Delaunay triangulation (Mardia *et al.* 1977; Green and Sibson 1978; Okabe *et al.* 2000). In Figure 1.28 we see that Voronoi polygons for ideal central places would be hexagons and Delaunay triangles would be equilateral triangles. In Figure 1.29 we see a Delaunay triangulation and the Voronoi polygons for the Iowa data. An important question to ask is: are the Delaunay triangles more equilateral than expected by chance?

The points here are unlabelled (there is no correspondence in the vertices of the triangles). Also the triangles are correlated due to neighbouring triangles sharing points. Kendall (1983, 1989) also studied shape in Delaunay triangulations, in order to investigate the Central Place Theory hypothesis, following Mardia *et al.* (1977).

1.4.18 Archaeology: Alignments of standing stones

Consider a map of the 52 megalithic sites that form the 'Old Stones of Land's End' in Cornwall, UK, given in Figure 1.30. It was proposed by Alfred Watkins, in the early

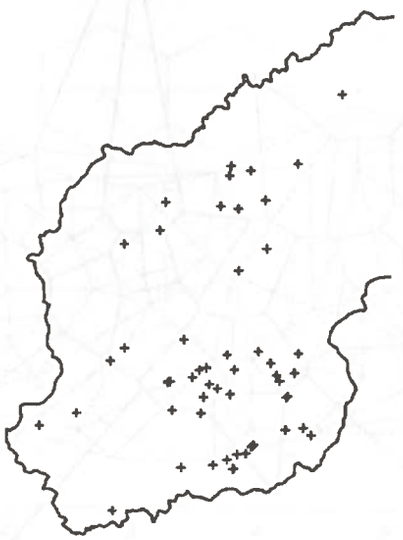


Figure 1.30 The map of 52 megalithic sites (+) that form the 'Old Stones of Land's End' in Cornwall. Source: Stoyan, Kendall & Mecke 1995. Reproduced with permission from John Wiley & Sons.

1920s, that these and other megalithic sites were placed in deliberate straight lines, called ley lines. One approach is to consider the shapes of all possible triangles and to see if there are more 'flat' triangles (triangles with the largest angle close to 180°) than expected under a randomness hypothesis. The points are unlabelled, and in this dataset there are $\binom{52}{3}$ triangles in two dimensions.

This dataset is particularly important in the history of shape analysis because it motivated D.G. Kendall's pioneering work. Analysis of these data is considered by Broadbent (1980), Kendall and Kendall (1980), Small (1988) and Stoyan *et al.* (1995) among others.

2

Size measures and shape coordinates

In this chapter we provide some brief historical background to traditional and geometrical methods of shape analysis. We then define what is meant by a configuration of landmarks and the configuration space, discuss measures of the size, and then consider some simple coordinate systems for representing shapes.

2.1 History

Shape analysis has a long history, especially in biology. In order to analyse size or shape, a biologist traditionally selects ratios of distances between landmarks or angles, and then carries out multivariate analysis (e.g. Rao 1948). This approach has been called 'multivariate morphometrics' in biology and a review is given by Reyment *et al.* (1984, p. 120). Early researchers using such methods include Pearson (1926), who studied a measure of similarity between skulls based on many lengths between landmarks. In fact, various studies have been cited in the journal *Biometrika*, starting from studies in volume 1 itself (Fawcett and Lee 1902). The data were usually distances between landmarks (such as lengths and widths) or angles between landmarks, rather than landmarks themselves. The definitions of landmarks on skulls follow the Frankfurt Concordat (*Frankfurter Verständigung*) of the 13th Congress of the German Anthropological Society which met in Frankfurt-am-Main, 14–17 August 1882 (e.g. see Trevor 1950).

Another large area of work in biology has been in the study of allometry, that is differences in shape associated with size (Hopkins 1966; Sprent 1972). The notion of allometry was introduced by Huxley (1924, 1932) and often involves fitting

Statistical Shape Analysis, with Applications in R, Second Edition. Ian L. Dryden and Kanti V. Mardia.

simple non-linear equations to length measurements. We consider a discussion of some related distance-based methods later in Chapter 15.

In the studies of multivariate morphometrics one deals exclusively with positive variables (lengths, angles and ratios of lengths). However, considering only distances and angles can be inferior to using the actual coordinates of the landmarks, because the geometry is often thrown away when using the former. Ratios of distances can easily be calculated from coordinates whereas the converse is not generally true. However, if enough distances are taken, then a configuration can be reconstructed up to a reflection and analysis can be carried out using multidimensional scaling (MDS) or Euclidean distance matrix analysis (see Kent 1994; Lele and Richtsmeier 2001; Dryden *et al.* 2008a), see Chapter 15.

Since the 1980s there have been many key developments in shape analysis that allow us to work on the landmark coordinates directly. Also, the advances in technology of measuring landmarks have been helpful, for example landmarks from digitized objects. Of course if there were no constraints from registration, then we could use standard multivariate analysis in Euclidean space, but in general the statistical methodology for shape is inherently non-Euclidean.

The main idea of the geometrical approach to shape analysis is that, rather than working with quantities derived from organisms, one works with the complete geometrical object itself (up to similarity transformations). As pointed out by Bookstein (1978) the approach is very much in the spirit of Thompson (1917) who considered the geometric transformations of one species to another (e.g. see Figure 2.1). D'Arcy Thompson's key ideas will be discussed in more detail in Chapter 12, but the important point to note is that he worked with geometrical pictures of organisms rather than derived quantities.

Throughout the text it will be observed that pictures of the objects under study can always be easily constructed and it is this fact that embodies the geometrical approach to shape analysis. In many applications the statistical goal is inference, for example testing for shape difference. However, the scientific goal is to depict or describe the size and shape changes in a study, and this is a major strength of the geometrical methods that we describe.

We shall consider a shape space obtained directly from the landmark coordinates, which retains the geometry of a point configuration. This approach to shape analysis has been called 'geometric shape analysis' or 'geometric morphometrics' by various authors and the subject progressed rapidly around the late 1970s/early 1980s. A glossary of geometrical morphometrics is given by Slice *et al.* (1996).

Following Thompson (1917) some of the earlier work in geometrical shape analysis included that by Medawar (1944) and Sneath (1967). However, the research area became firmly established with the pioneering work of D.G. Kendall and F.L. Bookstein, who independently developed many of the key ideas, in very different styles. Also, some very important mathematical work in the area was also given by Ziezold (1977). Bookstein (1994a) has summarized the history of geometrical shape analysis, mainly through applications in biology. Kendall (1989) has reviewed shape theory and its development from a different, more theoretical, viewpoint, with applications in archaeology, astronomy and geography. Kendall (1989, 1995) provides

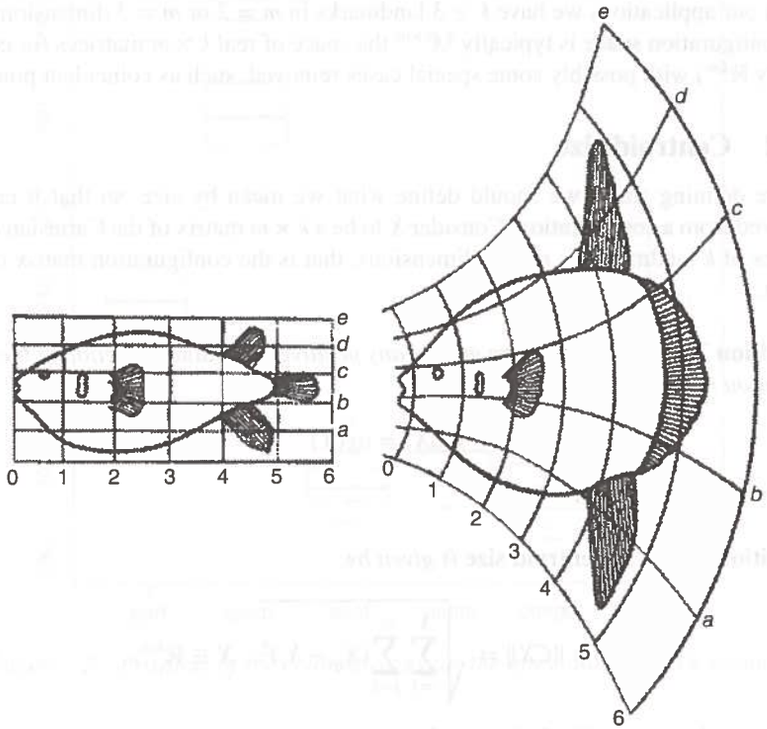


Figure 2.1 Thompson (1917)'s example of a species of fish Diodon *being geometrically transformed into another species* Orthagoriscus. *Source: Thompson 1917. Reproduced with permission of Cambridge University Press.*

some historical remarks on his development of shape theory, and some of the contributions of his colleagues. The first articles on the subject were by Kendall (1977), Ziezold (1977) and Bookstein (1978). Some key early papers in the field include those by Kendall (1984, 1989), Bookstein (1986), Goodall (1991), Le and Kendall (1993) and Kent (1994). Also, development in non-i.i.d. (independent and identically distributed) distribution theory for shape started with Mardia and Dryden (1989a).

2.2 Size

2.2.1 Configuration space

Definition 2.1 *The* **configuration** *is the set of landmarks on a particular object. The* **configuration matrix** X *is the* $k \times m$ *matrix of Cartesian coordinates of the* k *landmarks in* m *dimensions. The* **configuration space** *is the space of all landmark coordinates.*

In our applications we have $k \geq 3$ landmarks in $m = 2$ or $m = 3$ dimensions and the configuration space is typically $\mathbb{M}^{k \times m}$ the space of real $k \times m$ matrices (or equivalently $\mathbb{R}^{km}$) with possibly some special cases removed, such as coincident points.

2.2.2 Centroid size

Before defining shape we should define what we mean by size, so that it can be removed from a configuration. Consider X to be a $k \times m$ matrix of the Cartesian coordinates of k landmarks in m real dimensions, that is the configuration matrix of the object.

Definition 2.2 *A* **size measure** $g(X)$ *is any positive real valued function of the configuration matrix such that*

$$g(aX) = ag(X) \tag{2.1}$$

for any positive scalar $a > 0$.

Definition 2.3 *The* **centroid size** *is given by:*

$$S(X) = \|CX\| = \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{m}(X_{ij} - \bar{X}_j)^2}, \quad X \in \mathbb{R}^{km}, \tag{2.2}$$

where X_{ij} *is the* (i,j)*th entry of* X*, the arithmetic mean in the* j*th dimension is* $\bar{X}_j = \frac{1}{k}\sum_{i=1}^{k} X_{ij}$,

$$C = I_k - \frac{1}{k}1_k 1_k^{\mathrm{T}} \tag{2.3}$$

is the centring matrix,

$$\|X\| = \sqrt{\mathrm{trace}(X^{\mathrm{T}}X)}$$

is the Euclidean norm, I_k *is the* $k \times k$ *identity matrix (diagonal matrix with ones on the diagonal), and* 1_k *is the* $k \times 1$ *vector of ones.*

Obviously $S(aX) = aS(X)$ for $a > 0$, thus satisfying Equation (2.1). The centroid size $S(X)$ is the square root of the sum of squared Euclidean distances from each landmark to the centroid, namely

$$S(X) = \sqrt{\sum_{i=1}^{k} \|(X)_i - \bar{X}\|^2},$$

where $(X)_i$ is the ith row of $X (i = 1, \ldots, k)$ and $\bar{X} = (\bar{X}_1, \ldots, \bar{X}_m)$ is the centroid. This measure of size will be used throughout the book. In fact, the centroid size is the most commonly used size measure in geometrical shape analysis (e.g. Kendall 1984;

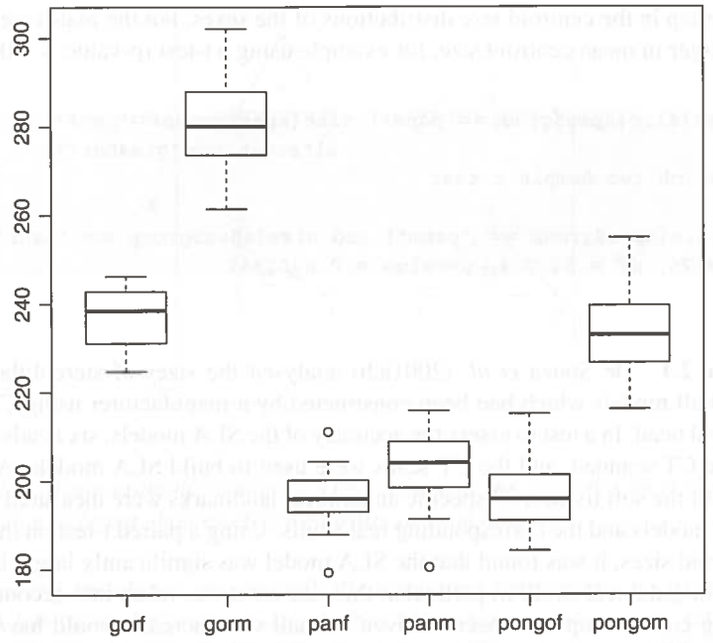


Figure 2.2 Boxplots of the centroid sizes for the apes' data by each group.

Bookstein 1986; Goodall 1991; Dryden and Mardia 1992). The centroid size could also be used in a normalized form, for example $S(X)/\sqrt{k}$ or $S(X)/(km)^{1/2}$, and this would be particularly appropriate when comparing configurations with different numbers of landmarks. The squared centroid size can also be interpreted as $(2k)^{-1}$ times the sum of the squared inter-landmark distances since

$$\sum_{j=1}^{m}\sum_{i=1}^{k}\sum_{l=1}^{k}(X_{ij} - X_{lj})^2 = \sum_{j=1}^{m}\sum_{i=1}^{k}\sum_{l=1}^{k}(X_{ij} - \bar{X}_j + \bar{X}_j - X_{lj})^2$$
$$= 2k\sum_{i=1}^{k}\sum_{j=1}^{m}(X_{ij} - \bar{X}_j)^2 = 2kS(X)^2.$$

In Figure 2.2 we display boxplots of the centroid size of the six groups of great apes which were described in Section 1.4.8. The plot is obtained using the R commands:

```
data(apes)
size<-centroid.size(apes$x)
plot(apes$group,size)
```

It is clear that there are sex differences in centroid size for gorillas and orangutans (*pongo*), with the males all being larger than the females. For chimpanzees (*pan*) there

is an overlap in the centroid size distributions of the sexes, but the males are significantly larger in mean centroid.size, for example using a t-test (p-value = 0.0002):

```
> t.test(size[apes$group=="panm"],size[apes$group=="panf"],
                                    alternative="greater")
      Welch Two Sample t-test

data: size[apes$group == "panm"] and size[apes$group == "panf"]
t = 3.8176, df = 50.704, p-value = 0.0001841
```

Example 2.1 de Souza *et al.* (2001a,b) analysed the sizes of stereolithography (SLA) skull models which had been constructed by a manufacturer using CT scans from a real head. In a test to assess the accuracy of the SLA models, six heads (cadavers) were CT scanned, and the CT scans were used to build SLA models. After the removal of the soft tissues, 37 specific anatomical landmarks were measured on both the SLA models and the corresponding real skulls. Using a paired t-test on the log of the centroid sizes, it was found that the SLA model was significantly larger by 1.5% than the original real skull. In particular, this size error was taken into account when designing custom implants, because even a small size increase would have severe clinical consequences. □

2.2.3 Other size measures

An alternative size measure is the baseline size, that is the length between landmarks 1 and 2:

$$D_{12}(X) = \|(X)_2 - (X)_1\|. \tag{2.4}$$

The baseline size was used as early as 1907 by Galton (1907) for normalizing faces and its use came into prominence with Bookstein coordinates, described in Section 2.4. This size variable is also useful in calculating size-and-shape distributions (Dryden and Mardia 1992) which we outline in Section 11.5.

Other alternative size measures include the square root of the area of the convex hull for planar configurations or the cube root of the volume of the convex hull for configurations in three dimensions which intuitively describes size, but these measures have not become popular. For triangles with coordinates $(X_1, Y_1), (X_2, Y_2)$ and (X_3, Y_3) the area is simply given by:

$$A = \tfrac{1}{2}\left|(Y_3 - Y_1)(X_2 - X_1) - (X_3 - X_1)(Y_2 - Y_1)\right|.$$

Note that the volume of a parallelepiped in three dimensions formed by v_1, v_2, v_3 is given by the absolute value of the determinant of the matrix with columns v_1, v_2, v_3. Other size measures for triangles include the radius of the inscribed circle and the circumradius (e.g. Miles 1970), the latter being equal to $S(X)/\sqrt{3}$.

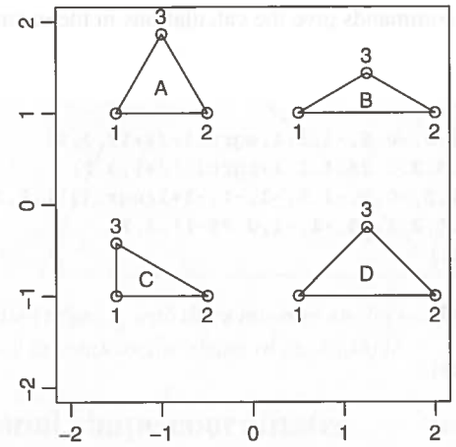


Figure 2.3 Four example triangles. The ranking of the triangles in terms of size differs when different choices of size measure are considered.

In general the choice of size will affect reported conclusions. In particular, if object A has twice the centroid size of object B it does not necessarily follow that object A is twice the size of object B when using a different measure. When reporting size differences one needs to state which size measure was used. The exception is when both objects have the same shape, in which case the ratio of sizes will be the same regardless of the choice of size variable. If comparing sets of collinear points, then very different conclusions about size are reached, with these pathological configurations all having zero area but non-zero centroid size for non-coincident collinear points.

Example 2.2 Consider the triangles in Figure 2.3. The centroid sizes of the triangles to three decimal places are:

$$\text{A} : 1.000,\ \text{B} : 1.118,\ \text{C} : 0.943,\ \text{D} : 1.225.$$

The baseline sizes are:

$$\text{A} : 1,\ \text{B} : 1.5,\ \text{C} : 1,\ \text{D} : 1.5.$$

The square root of area sizes to three decimal places are:

$$\text{A} : 0.658,\ \text{B} : 0.570,\ \text{C} : 0.537,\ \text{D} : 0.750.$$

The relative ranking of triangles in terms of size differs depending on the choice of size measure. In particular, in terms of area A is larger than B, but in terms of centroid size B is larger than A. This example demonstrates that different choices of size measure can lead to different conclusions about size. □

The following R commands give the calculations in the example:

```
> library(shapes)
> x1<-matrix(c(-1.5,-0.5,-1,1,1,sqrt(3)/2+1),3,2)
> x2<-matrix(c(0.5,2,1.25,1,1,1+sqrt(3)/4),3,2)
> x3<-matrix(c(-1.5,-0.5,-1.5,-1,-1,-1+1/sqrt(3)),3,2)
> x4<-matrix(c(0.5,2,1.25,-1,-1,0.75-1),3,2)
> centroid.size(x1)
[1] 1
> centroid.size(x2)
[1] 1.118034
> centroid.size(x3)
[1] 0.942809
> centroid.size(x4)
[1] 1.224745
> #baseline size
> sqrt( (x1[2,1]-x1[1,1])**2+(x1[2,2]-x1[1,2])**2 )
[1] 1
> sqrt( (x2[2,1]-x2[1,1])**2+(x2[2,2]-x2[1,2])**2 )
[1] 1.5
> sqrt( (x3[2,1]-x3[1,1])**2+(x3[2,2]-x3[1,2])**2 )
[1] 1
> sqrt( (x4[2,1]-x4[1,1])**2+(x4[2,2]-x4[1,2])**2 )
[1] 1.5
> #areas
> sqrt(abs((x1[3,2]-x1[1,2])*(x1[2,1]-x1[1,1])-
                  (x1[3,1]-x1[1,1])*(x1[2,2]-x1[1,2]))/2)
[1] 0.658037
> sqrt(abs((x2[3,2]-x2[1,2])*(x2[2,1]-x2[1,1])-
                  (x2[3,1]-x2[1,1])*(x2[2,2]-x2[1,2]))/2)
[1] 0.5698768
> sqrt(abs((x3[3,2]-x3[1,2])*(x3[2,1]-x3[1,1])-
                  (x3[3,1]-x3[1,1])*(x3[2,2]-x3[1,2]))/2)
[1] 0.537285
> sqrt(abs((x4[3,2]-x4[1,2])*(x4[2,1]-x4[1,1])-
                  (x4[3,1]-x4[1,1])*(x4[2,2]-x4[1,2]))/2)
[1] 0.75
```

Mosimann (1970) considers a definition of size based on length measurements which satisfies Equation (2.1), but with vectors of positive measurements (length, width, etc.) in place of the configuration matrix X.

In order to describe an object's shape it is useful to specify a coordinate system. We initially consider some of the most straightforward coordinate systems, which helps to provide an elementary introduction to aspects of shape. A suitable choice of coordinate system for shape is invariant under translation, scaling and rotation of the configuration. Further coordinate systems are discussed later in Section 4.4.

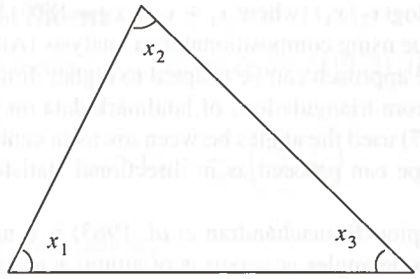


Figure 2.4 A labelled triangle with three internal angles marked. Two of the internal angles could be used to measure the shape of the triangle.

2.3 Traditional shape coordinates

2.3.1 Angles

For $k = 3$ points in $m = 2$ dimensions two internal angles are an obvious choice of coordinates that are invariant under the similarity transformations. For example, x_1 and x_2 could measure the shape of the triangle in Figure 2.4, where $x_1 + x_2 + x_3 = 180°$.

However, it soon becomes apparent that using angles to describe shape can be problematic. For cases such as very flat triangles (three points in a straight line) there are many different arrangements of three points. For example, see Figure 2.5, where the triangles all have $x_1 = 0°, x_3 = 0°, x_2 = 180°$, and yet the configurations have different shapes.

For larger numbers of points ($k > 3$) one could subdivide the configuration into triangles and so $2k - 4$ angles would be needed. For the $m = 3$ dimensional case angles could also be used, but again these suffer from problems in pathological cases. Also, probability distributions of the angles themselves are not easy to work with (see Mardia *et al.* 1977). If the angles of the triangle are x_1, x_2 and x_3, then the

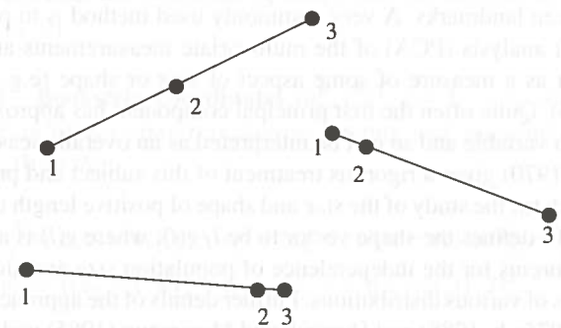


Figure 2.5 Examples of pathological cases where angles are an inappropriate shape measure. The landmarks are at the centre of the discs.

use of $\log(x_1/x_3)$ and $\log(x_2/x_3)$ (where $x_1 + x_2 + x_3 = 180°$) has some potential for analysing triangle shape using compositional data analysis (Aitchison 1986; Pukkila and Rao 1988) and the approach can be adapted to higher dimensions. Mardia *et al.* (1996b) used angles from triangulations of landmark data on photographs of faces, and Mardia *et al.* (1977) used the angles between towns in central place data. In such cases, analysis of shape can proceed as in directional statistics (Mardia and Jupp 2000).

A Ramachandran plot (Ramachandran *et al.* 1963) is a method for visualizing backbone dihedral torsion angles ψ versus ϕ of amino acid residues in protein secondary structures. The two angles describe the torsion angles either side of each α-carbon atom, and provide very useful shape information. This plot contains important angular shape information about the protein backbone and has proved very popular in the proteomics literature. The torsion angles are helpful in the prediction of a protein's 3D folding. Note that each angle is on the circle, S^1, and so the plots are on a torus $S^1 \times S^1$. The plot is also known as a $[\phi, \psi]$ plot or a Ramachandran diagram. Another geometrical description of a protein backbone was given by Zacharias and Knapp (2013) called a (d,θ) plot.

2.3.2 Ratios of lengths

Another traditional approach to representing shape is through ratios of length measurements, or lengths divided by size. A considerable amount of work has been carried out in multivariate morphometrics using distances, ratios, angles, and so on and it is still very commonly used in the biological literature (Reyment *et al.* 1984). There are many situations, such as classification problems, where the techniques can be very powerful. However, sometimes the interpretation of the important linear combinations of ratios of lengths and angles can be difficult. It is often easier to interpret pictures in the original space of the specimens than in some derived multivariate space, so we consider methods where the geometry of the object is retained.

Typical applications of the multivariate morphometrics approach include the classification of species (taxonomy) or the sexing of skulls using lengths, angles or ratios of lengths between landmarks. A very commonly used method is to perform a principal component analysis (PCA) of the multivariate measurements and to interpret each component as a measure of some aspect of size or shape (e.g. Jolicoeur and Mosimann 1960). Quite often the first principal component has approximately equal loadings on each variable and so can be interpreted as an overall measure of size.

Mosimann (1970) gives a rigorous treatment of this subject and provides a theoretical framework for the study of the size and shape of positive length measurements. Mosimann (1970) defines the shape vector to be $l/g(l)$, where $g(l)$ is a size measure. He provides theorems for the independence of population size and shape, including characterizations of various distributions. Further details of the approach can be found in Mosimann (1975a,b, 1988) and Darroch and Mosimann (1985) and the references therein.

2.3.3 Penrose coefficent

Given a set of measurements $d_1, \ldots, d_m$ the Penrose (1952) size coefficient is:

$$P_{Size} = C_Q^2 = \frac{1}{m^2}\left(\sum_{i=1}^{m} d_i\right)^2,$$

and the Penrose shape coefficient is:

$$P_{Shape} = \sum_{i=1}^{m} \frac{d_i^2}{m} - C_Q^2.$$

The coefficients are related to a coefficient of racial likeness (Penrose, 1952; after Pearson, 1926) given by:

$$CRL = P_{Size} + P_{Shape}.$$

These very traditional shape measures usually only capture part of the shape or size-and-shape information, so we prefer a newer approach where the geometry of the objects under study is retained.

2.4 Bookstein shape coordinates

2.4.1 Planar landmarks

Let $(x_j, y_j), j = 1, \ldots, k$, be $k \geq 3$ landmarks in a plane ($m = 2$ dimensions). Bookstein (1986, 1984) suggests removing the similarity transformations by translating, rotating and rescaling such that landmarks 1 and 2 are sent to a fixed position. If landmark 1 is sent to $(0,0)$ and landmark 2 is sent to $(1,0)$, then suitable shape variables are the coordinates of the remaining $k-2$ coordinates after these operations. To preserve symmetry, we consider the coordinate system where the baseline landmarks are sent to $(-\frac{1}{2}, 0)$ and $(\frac{1}{2}, 0)$.

Definition 2.4 **Bookstein coordinates** $(u_j^B, v_j^B)^{\mathrm{T}}, j = 3, \ldots, k$, *are the remaining coordinates of an object after translating, rotating and rescaling the baseline to* $(-\frac{1}{2}, 0)$ *and* $(\frac{1}{2}, 0)$ *so that*

$$u_j^B = \{(x_2 - x_1)(x_j - x_1) + (y_2 - y_1)(y_j - y_1)\}/D_{12}^2 - \tfrac{1}{2}, \tag{2.5a}$$

$$v_j^B = \{(x_2 - x_1)(y_j - y_1) - (y_2 - y_1)(x_j - x_1)\}/D_{12}^2, \tag{2.5b}$$

where $j = 3, \ldots, k$, $D_{12}^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 > 0$ *and* $-\infty < u_j^B, v_j^B < \infty$.

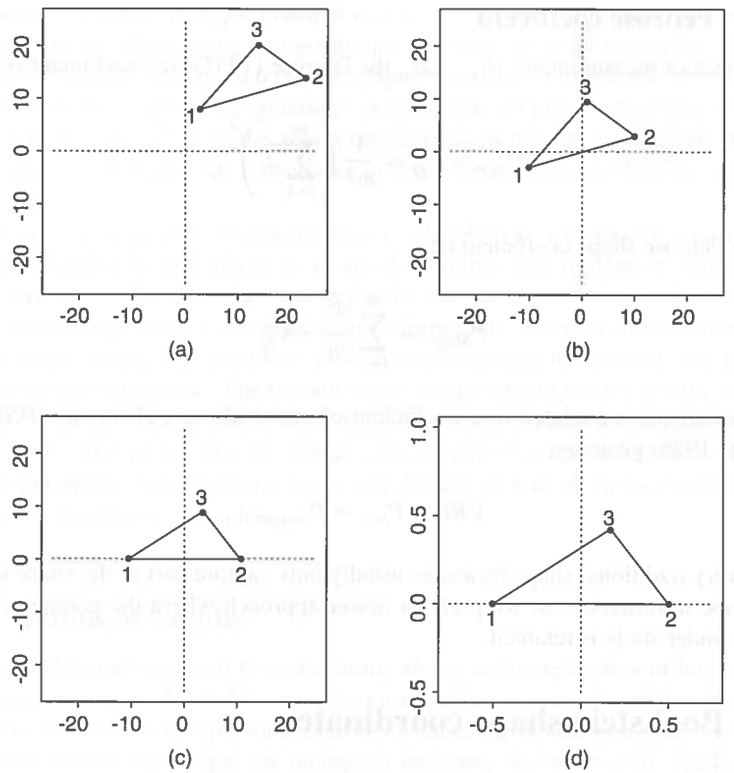


Figure 2.6 The geometrical interpretation of Bookstein coordinates. The original triangle in (a) is transformed by (b) translation, (c) rotation and finally (d) rescaling to give the Bookstein coordinates as the coordinates of the point labelled 3 in plot (d).

A geometrical illustration of these transformations is given in Figure 2.6. If the baseline is taken as $(0,0)$ and $(1,0)$, then there is no $-\frac{1}{2}$ in the equation for u_j^B, as originally proposed by Bookstein (1986). We subtract the $\frac{1}{2}$ in u_j^B to also simplify the transformation to Kendall coordinates in Section 2.5, although precisely where we send the baseline is an arbitrary choice. These coordinates have been used widely in shape analysis for planar data. Bookstein coordinates are the most straightforward to use for a newcomer to shape analysis. However, because of the lack of symmetry in choosing a particular baseline and the fact that correlations are induced into the coordinates, many practitioners often prefer to use the Procrustes tangent coordinates (see Section 4.4).

Galton (1907) defined precisely the same coordinates [with baseline (0,0), (1,0)] at the beginning of the last century, but the statistical details of using this approach needed to be worked out.

The construction of Bookstein coordinates is particularly simple if using complex arithmetic. Bookstein coordinates are obtained from the original complex coordinates $z_1^o, \ldots, z_k^o$, where $z_j^o \in \mathbb{C}, j = 1, \ldots, k$:

$$u_j^B + iv_j^B = \frac{z_j^o - z_1^o}{z_2^o - z_1^o} - \frac{1}{2} = \frac{2z_j^o - z_1^o - z_2^o}{2(z_2^o - z_1^o)}, \quad j = 3, \ldots, k, \tag{2.6}$$

where $i = \sqrt{-1}$. Consider how the formulae in Equation (2.5a) are obtained (Mardia 1991). To find (u_j^B, v_j^B) for a fixed $j = 3, \ldots, k$, we have to find the scale $c > 0$, the rotation A, and the translation $b = (b_1, b_2)^{\mathrm{T}}$ such that

$$U = cA(X - b), \tag{2.7}$$

where $X = (x_j, y_j)^{\mathrm{T}}$ and $U = (u_j^B, v_j^B)^{\mathrm{T}}$ are the coordinates of the jth point before and after the transformation, $j = 3, \ldots, k$, and A is a 2×2 rotation matrix, namely

$$A = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}, \quad |A| = 1, \quad A^{\mathrm{T}}A = AA^{\mathrm{T}} = I_2,$$

where we rotate clockwise by θ radians. Applying the transformation to landmarks 1 and 2 gives four equations in four unknowns (c, θ, b_1, b_2)

$$cA\left(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} - b\right) = \begin{bmatrix} -\frac{1}{2} \\ 0 \end{bmatrix}, \quad cA\left(\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} - b\right) = \begin{bmatrix} \frac{1}{2} \\ 0 \end{bmatrix}.$$

Now, we can solve these equations and see that the translation is

$$b = (\tfrac{1}{2}(x_1 + x_2), \tfrac{1}{2}(y_1 + y_2))^{\mathrm{T}},$$

the rotation (in the appropriate quadrant) is

$$\theta = \arctan\{(y_2 - y_1)/(x_2 - x_1)\}$$

and the rescaling is $c = \{(x_2 - x_1)^2 + (y_2 - y_1)^2\}^{-1/2}$. So,

$$A = c \begin{bmatrix} x_2 - x_1 & y_2 - y_1 \\ -(y_2 - y_1) & x_2 - x_1 \end{bmatrix}. \tag{2.8}$$

Substituting $X = (x_j, y_j)^{\mathrm{T}} (j = 3, \ldots, k)$ and c, A, b into Equation (2.7) we obtain the shape variables of Equation (2.5a). This solution can also be seen from Figure 2.6 using geometry.

Example 2.3 In Figure 2.7 we see scatter plots of the Bookstein coordinates for the T2 Small vertebrae of Section 1.4.1. We take landmarks 1 and 2 as the baseline.

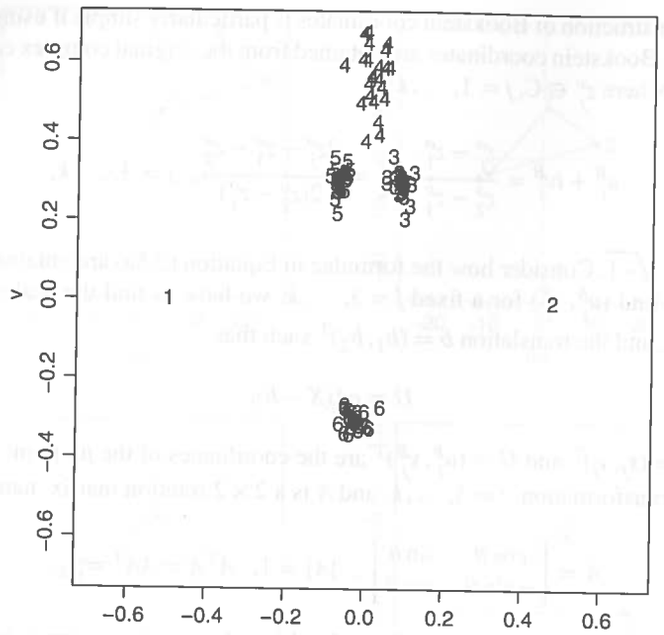


Figure 2.7 Scatter plots of the Bookstein coordinates of the T2 Small vertebrae registered on the baseline 1, 2.

Note that the marginal scatter plots at landmarks 4 and possibly 3 and 5 are elliptical in nature. The plot can be obtained using:

```
data(mice)
qset2<-mice$x[,,mice$group=="s"]
u<-bookstein2d(qset2)$bshpv
plotshapes(u,symbol=as.character(1:6))
```

For example, the first mouse has Bookstein shape variables given by the last four rows of:

```
> u[,,1]
, , 1
          [,1]  [,2]
[1,] -0.50000000 0.0000000
[2,]  0.50000000 0.0000000
[3,]  0.09121051 0.2826022
[4,]  0.04065577 0.5089238
[5,] -0.07367575 0.2735386
[6,] -0.02672182 -0.3054334
```


and this could also be obtained directly using the command:

```
bookstein.shpv(qset2[,,1])
```

For practical data analysis it is sensible to choose the baseline as landmarks that are not too close together, as in this example. As often happens with using Bookstein coordinates, the variability in the points away from the origin appears larger than the points nearer to the origin. This is an artifact of this coordinate system, and will be explored further in Sections 9.4 and 11.1.4. The centroid sizes of the bones are obtained using:

```
sz<-centroid.size(qset2)
```

A pairwise scatter plot of Bookstein's shape variance and centroid size is obtained using:

```
x<-rbind(u[3:6,1,],u[3:6,2,])
x<-rbind(matrix(sz,1,23),x)
pairs(t(x),labels=c("s","u3","u4","u5","u6","v3","v4","v5","v6"))
```

In Figure 2.8 we see pairwise scatter plots of Bookstein coordinates. There are strong positive correlations in the v_3^B, v_4^B and v_5^B coordinates. There are also positive correlations between v_3^B and v_6^B, between v_5^B and v_6^B, and between u_3^B and u_4^B. However, it is also an artifact of the coordinate system that correlations are induced into the shape variables, even when the landmark coordinates are uncorrelated (see Section 11.1.4), and so correlations can be difficult to interpret.

We can also examine the joint relationship between size and shape, which will be considered in more detail in Chapter 5. Scatter plots of the centroid size S versus each of the shape coordinates are also given in Figure 2.8. We see that there are quite strong positive correlations between S and each of v_3^B, v_4^B and v_5^B. □

Throughout the text we shall often refer to the real $(2k-4)$-vector of Bookstein coordinates $u^B = (u_3^B, \ldots, u_k^B, v_3^B, \ldots, v_k^B)^T$, stacking the coordinates in this particular order.

One approach to shape analysis is to use standard multivariate analysis of the Bookstein coordinates, ignoring the non-Euclidean nature of the space. Provided variations in the data are small, then the method is adequate for mean estimation and hypothesis testing. For example, we could obtain an estimate of the mean shape of the configuration by taking the arithmetic average of the Bookstein coordinates

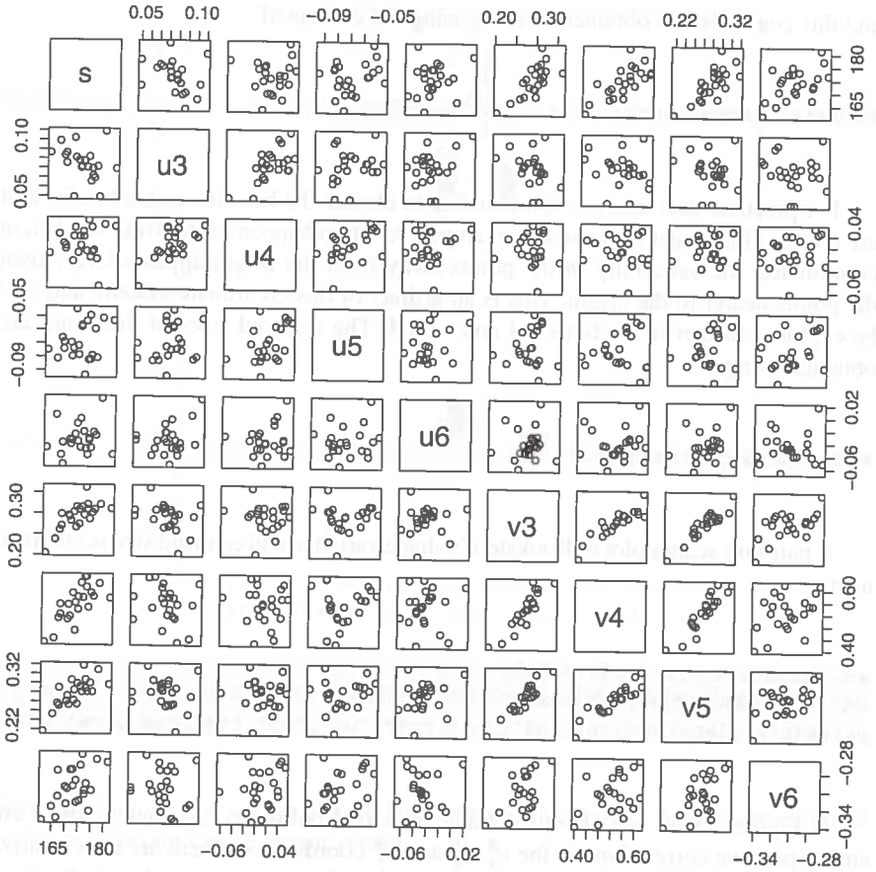


Figure 2.8 Pairwise scatter plots of centroid size S and Bookstein coordinates $u_3^B, \ldots, u_6^B, v_3^B, \ldots, v_6^B$ for the T2 Small vertebrae.

(the **Bookstein mean shape**). The Bookstein mean shape for the T2 Small mouse vertebrae of Example 2.7 is given by the last four rows of:

```
> bookstein2d(mice$x[,,mice$group=="s"])$mshape
          [,1]  [,2]
[1,] -0.50000000 0.000000e+00
[2,] 0.50000000 -6.634366e-21
[3,] 0.08469746 2.933430e-01
[4,] 0.01215768 5.613175e-01
[5,] -0.06874750 2.991278e-01
[6,] -0.02502185 -3.041418e-01
```

which when rounded to three decimal places is:

$$\bar{u}^B = (0.085, 0.012, -0.069, -0.025, 0.293, 0.561, 0.299, -0.304)^{\mathrm{T}}.$$

The variability of the shape variables is less straightforward to interpret. Transforming the objects (or registering) to a given edge induces correlations into the shape variables in general and this can lead to spurious correlations (Kent 1994), see Sections 9.4 and 11.1.4. So the method should not be used to interpret the structure of shape variability unless there is a good reason to believe that two landmarks are essentially fixed (see Section 9.4).

2.4.2 Bookstein-type coordinates for 3D data

Registration on a base to filter out the similarity transformations can also be carried out for 3D data, in a similar way to edge registration in Bookstein coordinates for 2D data. Consider k landmarks $X_j = (x_{1j}, x_{2j}, x_{3j})^{\mathrm{T}} \in \mathbb{R}^3, j = 1, \ldots, k$, in three dimensions. The number of shape variables is $3k - 7$ since we have $3k$ landmark coordinates but must remove 3 location, 1 scale and 3 rotation parameters. The Bookstein coordinates can be taken as

$$u_j = (u_{1j}, u_{2j}, u_{3j})^{\mathrm{T}} = \frac{1}{\|X_2 - X_1\|} A \left(X_j - \frac{(X_1 + X_2)}{2} \right), \quad j = 3, \ldots, k, \tag{2.9}$$

where A is a 3×3 rotation matrix (a function of (X_1, X_2, X_3)) and

$$X_1 \to (-1/2, 0, 0)^{\mathrm{T}}, \; X_2 \to (1/2, 0, 0)^{\mathrm{T}}, \; X_3 \to u_3 = (u_{13}, u_{23}, 0)^{\mathrm{T}},$$

where $u_{23} \geq 0, u_{33} = 0$ and $X_j \to u_j$ for $j = 4, \ldots, k$. We drop the superscript B denoting Bookstein coordinates in this section. Geometrically the figure has been translated, rotated and rescaled so that the baseline (landmarks 1 and 2) is of unit length, has midpoint at the origin and is rotated to lie along the x-axis. The figure is further rotated so that landmark 3 lies in the plane of the first two dimensions (X, Y), with $Y \geq 0$. The remaining non-fixed coordinates are the shape variables (cf. Mardia and Dryden 1997). The details are as follows: first translate the figure,

$$w_{li} = x_{li} - (x_{l1} + x_{l2})/2, \quad l = 1, 2, 3; \;\; i = 2, \ldots, k.$$

We then rescale by dividing by the length of the baseline between points 1 and 2: $D_{12} = 2(w_{12}^2 + w_{22}^2 + w_{32}^2)^{1/2}$, and rotate through clockwise angles of θ about the z-axis, ω about the y-axis and ϕ about the x-axis where

$$\theta = \arctan(w_{22}/w_{12}), \quad \omega = \arctan\left(w_{32}/(w_{12}^2 + w_{22}^2)^{1/2}\right),$$

$$\phi = \arctan\left(\frac{(w_{12}^2 + w_{22}^2)w_{33} - (w_{12}w_{13} + w_{22}w_{23})w_{32}}{(w_{12}^2 + w_{22}^2 + w_{32}^2)^{1/2}(w_{12}w_{23} - w_{13}w_{22})} \right).$$

So, using matrix notation if

$$R_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{bmatrix}, R_y = \begin{bmatrix} \cos\omega & 0 & \sin\omega \\ 0 & 1 & 0 \\ -\sin\omega & 0 & \cos\omega \end{bmatrix}, R_z = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

then the 3D Bookstein coordinates are:

$$u_j = R_x R_y R_z (w_{1j}, w_{2j}, w_{3j})^{\mathrm{T}} / D_{12}, \quad j = 3, \ldots, k.$$

An explicit expression for the shape coordinates is:

$$u_{13} = \left(w_{12}w_{13} + w_{22}w_{23} + w_{32}w_{33}\right)/a,$$
$$u_{23} = \left((w_{12}w_{23} - w_{13}w_{22})^2 + (w_{12}w_{33} - w_{32}w_{13})^2 + (w_{22}w_{33} - w_{23}w_{32})^2\right)^{1/2}/a,$$
$$u_{1j} = \left(w_{12}w_{1j} + w_{22}w_{2j} + w_{32}w_{3j}\right)/a,$$

$$u_{2j} = \frac{\begin{bmatrix} w_{12}^2 w_{23}w_{2j} + w_{12}^2 w_{33}w_{3j} - w_{12}w_{13}w_{22}w_{2j} - w_{12}w_{13}w_{32}w_{3j} \\ - w_{12}w_{22}w_{23}w_{1j} - w_{12}w_{32}w_{33}w_{1j} + w_{13}w_{22}^2 w_{1j} + w_{13}w_{32}^2 w_{1j} \\ + w_{22}^2 w_{33}w_{3j} - w_{22}w_{32}w_{23}w_{3j} - w_{22}w_{32}w_{33}w_{2j} + w_{32}^2 w_{23}w_{2j} \end{bmatrix}}{(ab)^{1/2}},$$

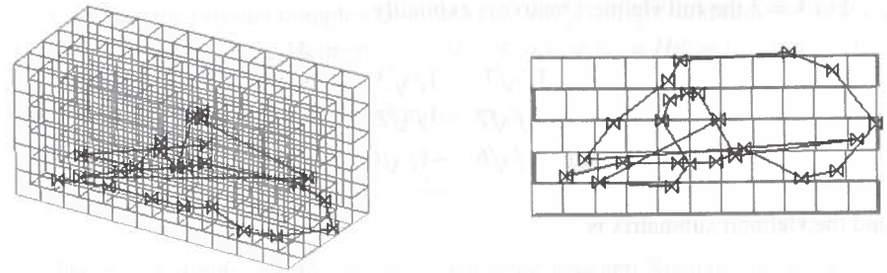
$$u_{3j} = \frac{\begin{bmatrix} w_{12}w_{23}w_{3j} - w_{12}w_{33}w_{2j} - w_{13}w_{22}w_{3j} \\ + w_{13}w_{32}w_{2j} + w_{22}w_{33}w_{1j} - w_{32}w_{23}w_{1j} \end{bmatrix}}{(2ab)^{1/2}},$$

where $a = 2(w_{12}^2 + w_{22}^2 + w_{32}^2)$, $j = 4, \ldots, k$, and

$$b = w_{12}^2 w_{23}^2 + w_{12}^2 w_{33}^2 - 2w_{12}w_{13}w_{22}w_{23} - 2w_{12}w_{13}w_{32}w_{33} + w_{13}^2 w_{22}^2 + w_{13}^2 w_{32}^2 + w_{22}^2 w_{33}^2 - 2w_{22}w_{32}w_{23}w_{33} + w_{32}^2 w_{23}^2.$$

As in the $m = 2$ dimensional case, care must be taken when interpreting correlations between the shape variables, because spurious correlations can occur with edge registration methods such as this (see Sections 9.4 and 11.1.4).

Example 2.4 In Figure 2.9(a) we see a male macaque skull from the dataset described in Section 1.4.3, with 24 landmarks located in three dimensions and a wire box regular grid drawn over the skull. Let us consider the page to be the x–y plane and the z-axis is perpendicular to the page. The 3D Bookstein coordinates involve selecting two landmarks (e.g. landmarks 1 and 6 in Figure 1.11) to translate, rotate and rescale to lie horizontal in the plane of the page at points $(\frac{1}{2}, 0, 0)$ and $(-\frac{1}{2}, 0, 0)$, and a third landmark (e.g. landmark 10) is chosen to lie in the x–y plane, as displayed in Figure 2.9(b). □



(a) (b)

Figure 2.9 The 24 landmarks located on a macaque monkey skull where the page is regarded as the x–y plane and into the page is the z-axis. The skull is translated, rotated and rescaled so that landmarks 1 and 6 (see Section 1.2.8) lie in the x–y plane of the page, at unit length apart, and landmark 10 is also rotated to be in the x–y plane.

Important point: For **affine shape** (where invariance is under affine transformations rather than Euclidean similarity transformations – see Section 12.2), there are 12 rather than 7 constraints. Hence all the coordinates of the first four landmarks are sufficient to form a base in the affine case and the expressions are simpler than for similarity shape.

2.5 Kendall's shape coordinates

Kendall coordinates (Kendall 1984) are related to Bookstein coordinates but location is removed in a different manner. We first need to define the Helmert submatrix which is used to remove location.

The Helmert submatrix H is the $(k-1) \times k$ Helmert matrix without the first row. The full Helmert matrix H^F, which is commonly used in Statistics, is a square $k \times k$ orthogonal matrix with its first row of elements equal to $1/\sqrt{k}$, and the remaining rows are orthogonal to the first row. We drop the first row of H^F so that the transformed HX does not depend on the original location of the configuration. Note $H^T H = C$, where C is the centring matrix of (2.3).

Definition 2.5 *The jth row of the* **Helmert submatrix** *H is given by:*

$$(h_j, \ldots, h_j, -jh_j, 0, \ldots, 0), \quad h_j = -\{j(j+1)\}^{-1/2}, \tag{2.10}$$

and so the jth row consists of h_j repeated j times, followed by $-jh_j$ and then $k-j-1$ zeros, $j = 1, \ldots, k-1$.

For $k = 3$ the full Helmert matrix is explicitly

$$H^F = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \end{bmatrix}$$

and the Helmert submatrix is

$$H = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \end{bmatrix}.$$

For $k = 4$ points the full Helmert matrix is

$$H^F = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} & 0 \\ -1/\sqrt{12} & -1/\sqrt{12} & -1/\sqrt{12} & 3/\sqrt{12} \end{bmatrix}$$

and the Helmert submatrix is

$$H = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} & 0 \\ -1/\sqrt{12} & -1/\sqrt{12} & -1/\sqrt{12} & 3/\sqrt{12} \end{bmatrix}.$$

We can obtain the Helmert submatrix using `defh(k-1)` where k is the number of landmarks, for example:

```
> defh(2)
        [,1] [,2] [,3]
[1,] -0.7071068 0.7071068 0.0000000
[2,] -0.4082483 -0.4082483 0.8164966
> defh(3)
        [,1] [,2] [,3] [,4]
[1,] -0.7071068 0.7071068 0.0000000 0.0000000
[2,] -0.4082483 -0.4082483 0.8164966 0.0000000
[3,] -0.2886751 -0.2886751 -0.2886751 0.8660254
> defh(4)
        [,1] [,2] [,3] [,4] [,5]
[1,] -0.7071068 0.7071068 0.0000000 0.0000000 0.0000000
[2,] -0.4082483 -0.4082483 0.8164966 0.0000000 0.0000000
[3,] -0.2886751 -0.2886751 -0.2886751 0.8660254 0.0000000
[4,] -0.2236068 -0.2236068 -0.2236068 -0.2236068 0.8944272
```


Consider the original complex landmarks $z^o = (z_1^o, \ldots, z_k^o)^{\mathrm{T}}$ and remove location by pre-multiplying by the Helmert submatrix H to give $z_H = Hz^o = (z_1, \ldots, z_{k-1})^{\mathrm{T}}$.

Definition 2.6 *The* **Kendall coordinates** *are given by:*

$$u_j^K + iv_j^K = \frac{z_{j-1}}{z_1}, \quad j = 3, \ldots, k. \tag{2.11}$$

There is a simple one to one correspondence between Kendall and Bookstein coordinates. If we write

$$w^B = (u_3^B + iv_3^B, \ldots, u_k^B + iv_k^B)^{\mathrm{T}}$$

for Bookstein coordinates and

$$w^K = (u_3^K + iv_3^K, \ldots, u_k^K + iv_k^K)^{\mathrm{T}}$$

for Kendall coordinates, then it follows that

$$w^K = \sqrt{2}H_1 w^B \tag{2.12}$$

where H_1 is the lower right $(k-2) \times (k-2)$ partition matrix of the Helmert submatrix H. Note that

$$H_1^{\mathrm{T}} H_1 = I_{k-2} - \frac{1}{k} 1_{k-2} 1_{k-2}^{\mathrm{T}}, \quad \|H_1\|^2 = 2/k, \quad (H_1^{\mathrm{T}} H_1)^{-1} = I_{k-2} + \frac{1}{2} 1_{k-2} 1_{k-2}^{\mathrm{T}}$$

so linear transformation from one coordinate system to the other is straightforward. The inverse transformation is:

$$w^B = (H_1^{\mathrm{T}} H_1)^{-1} H_1^{\mathrm{T}} w^K / \sqrt{2}.$$

For $k = 3$ we have the relationship:

$$u_3^B + iv_3^B = \frac{z_3^o - \frac{1}{2}(z_1^o + z_2^o)}{z_2^o - z_1^o} = \frac{\sqrt{3}}{2}(u_3^K + iv_3^K)$$

and so Kendall coordinates in this case are the coordinates of the third landmark after transforming landmarks 1 and 2 to $(-1/\sqrt{3}, 0)$ and $(1/\sqrt{3}, 0)$ by the similarity transformations. The transformation from (z_1^o, z_2^o, z_3^o) to Kendall coordinates is:

$$z_1^o \to -\frac{1}{\sqrt{3}}, \; z_2^o \to \frac{1}{\sqrt{3}}, \; z_3^o \to u_3^K + iv_3^K.$$

Throughout the text we shall often refer to the real $(2k-4)$-vector of Kendall coordinates $u^K = (u_3^K, \ldots, u_k^K, v_3^K, \ldots, v_k^K)^T$, stacking the coordinates in this particular order.

2.6 Triangle shape coordinates

2.6.1 Bookstein coordinates for triangles

For the case of a triangle of landmarks we have two shape coordinates.

Example 2.5 In Figure 2.3 we see the triangles with internal angles at points 1, 2, 3 given by:

A: 60°, 60°, 60° (equilateral),
B: 30°, 30°, 120° (isosceles),
C: 90°, 30°, 60° (right-angled),
D: 45°, 45°, 90° (right-angled and isosceles).

Bookstein coordinates for these three triangles with baseline points 1, 2 are:

A: $U_3^B = 0, V_3^B = \frac{\sqrt{3}}{2}$,
B: $U_3^B = 0, V_3^B = \frac{1}{2\sqrt{3}}$,
C: $U_3^B = -\frac{1}{2}, V_3^B = \frac{1}{\sqrt{3}}$,
D: $U_3^B = 0, V_3^B = \frac{1}{2}$. □

A plot of the shape space of triangles for Bookstein coordinates is given in Figure 2.10. Each triangle shape is drawn with its centroid in the position of the Bookstein coordinates (U^B, V^B) corresponding to its shape in the shape space. For example, the equilateral triangles are at $(0, \sqrt{3}/2)$ and $(0, -\sqrt{3}/2)$ (marked E and F). The right-angled triangles are located on the lines $U^B = -0.5, U^B = 0.5$ and on the circle $(U^B)^2 + (V^B)^2 = 0.5$; the isosceles triangles are located on the line $U^B = 0$ and on the circles $(U^B + 0.5)^2 + (V^B)^2 = 1$ and $(U^B - 0.5)^2 + (V^B)^2 = 1$; and the flat triangles (3 points in a straight line) are located on $V^B = 0$.

The plane can be partitioned into two half-planes – all triangles below $V^B = 0$ can be reflected to lie above $V^B = 0$. In addition, each half-plane can be partitioned into six regions where the triangles have their labels permuted. If a triangle has baseline 1, 2 and apex 3 and side lengths d_{12}, d_{13} and d_{23}, then the six regions are:

$$d_{23} \geq d_{13} \geq d_{12}; \quad d_{23} \geq d_{12} \geq d_{13}; \quad d_{12} \geq d_{23} \geq d_{13};$$
$$d_{12} \geq d_{13} \geq d_{23}; \quad d_{13} \geq d_{12} \geq d_{23}; \quad d_{13} \geq d_{23} \geq d_{12}.$$

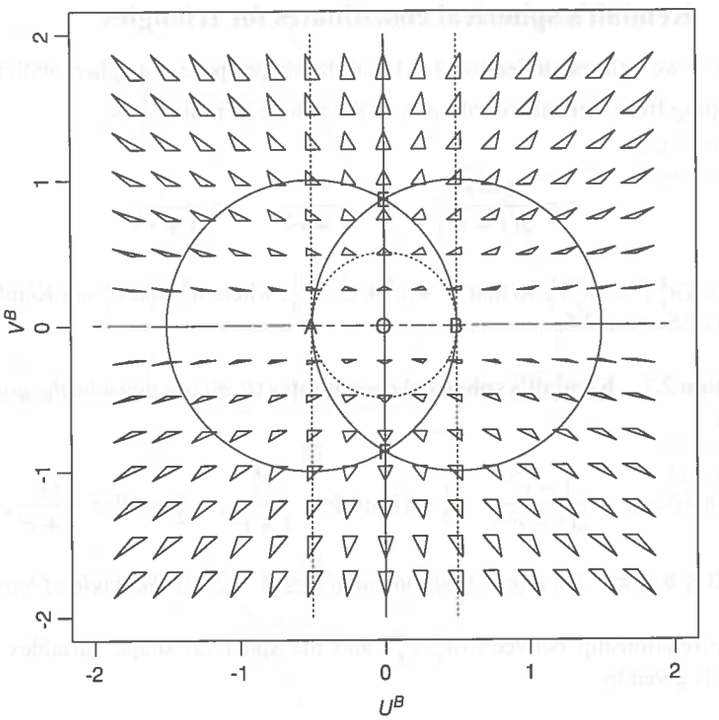


Figure 2.10 The shape space of triangles, using the Bookstein coordinates (U^B, V^B). Each triangle is plotted with its centre at the shape coordinates (U^B, V^B). The equilateral triangles are located at the points marked E and F. The isosceles triangles are located on the unbroken lines and circles (——) and the right-angled triangles are located on the broken lines and circles (- - - -). The flat (collinear) triangles are located on the $V^B = 0$ line (the U^B-axis). All triangles could be relabelled and reflected to lie in the region AOE, bounded by the arc of isosceles triangles AE, the line of isosceles triangles OE and the line of flat triangles AO.

Thus, if invariance under relabelling and reflection of the landmarks was required, then we would be restricted to one of the 12 regions, for example the region AOE, bounded by the arc of isosceles triangles AE, the line of isosceles triangles OE and the line of flat triangles AO in Figure 2.10.

It is quite apparent from Figure 2.10 that a non-Euclidean distance in (U^B, V^B) is most appropriate for the shape space. For example, two triangles near the origin that are a Euclidean distance of 1 apart are very different in shape, but two triangles away from the origin that are a Euclidean distance of 1 apart are quite similar in shape. Non-Euclidean shape metrics are described in Section 4.1.1. Bookstein (1986) also suggested using the hyperbolic Poincaré metric for triangle shapes, which is described in Section 12.2.4.

2.6.2 Kendall's spherical coordinates for triangles

For $k = 3$ we will see in Section 4.3.4 that the shape space is a sphere with radius $\frac{1}{2}$. A mapping from Kendall coordinates to the sphere of radius $\frac{1}{2}$ is:

$$x = \frac{1 - r^2}{2(1 + r^2)}, \quad y = \frac{u_3^K}{1 + r^2}, \quad z = \frac{v_3^K}{1 + r^2}, \tag{2.13}$$

and $r^2 = (u_3^K)^2 + (v_3^K)^2$, so that $x^2 + y^2 + z^2 = \frac{1}{4}$, where u_3^K and v_3^K are Kendall coordinates of Section 2.5.

Definition 2.7 **Kendall's spherical coordinates** (θ, ϕ) *are given by the polar coordinates*

$$\tfrac{1}{2}\sin\theta\cos\phi = \frac{1 - r^2}{2(1 + r^2)}, \quad \tfrac{1}{2}\sin\theta\sin\phi = \frac{u_3^K}{1 + r^2}, \quad \tfrac{1}{2}\cos\theta = \frac{v_3^K}{1 + r^2}, \tag{2.14}$$

where $0 \le \theta \le \pi$ *is the angle of latitude and* $0 \le \phi < 2\pi$ *is the angle of longitude.*

The relationship between (u_3^K, v_3^K) and the spherical shape variables (Mardia 1989b) is given by:

$$u_3^K = \frac{\sin\theta\sin\phi}{1 + \sin\theta\cos\phi},$$
$$v_3^K = \frac{\cos\theta}{1 + \sin\theta\cos\phi}. \tag{2.15}$$

The sphere can be partitioned into 6 lunes and 12 half-lunes. In order to make the terminology clear, let us note that one example full-lune is $0 \le \phi \le \pi/3, 0 \le \theta \le \pi$ and one example half-lune is $0 \le \phi \le \pi/3, 0 \le \theta \le \pi/2$.

In Figure 2.11 we see triangle shapes located on the spherical shape space. The equilateral triangle with anti-clockwise labelling corresponds to the 'North pole' ($\theta = 0$) and the reflected equilateral triangle (with clockwise labelling) is at the 'South pole' ($\theta = \pi$). The flat triangles (three collinear points) lie around the equator ($\theta = \pi/2$). The isosceles triangles lie on the meridians $\phi = 0, \pi/3, 2\pi/3, \pi, 4\pi/3, 5\pi/3$. The right-angled triangles lie on three small circles given by:

$$\sin\theta\cos\left(\phi - \frac{2k\pi}{3}\right) = \tfrac{1}{2}, \; k = 0, 1, 2,$$

and we see the arc of unlabelled right-angled triangles on the front half-lune in Figure 2.11.

Reflections of triangles in the upper hemisphere at (θ, ϕ) are located in the lower hemisphere at $(\pi - \theta, \phi)$. In addition, permuting the triangle labels gives rise to points

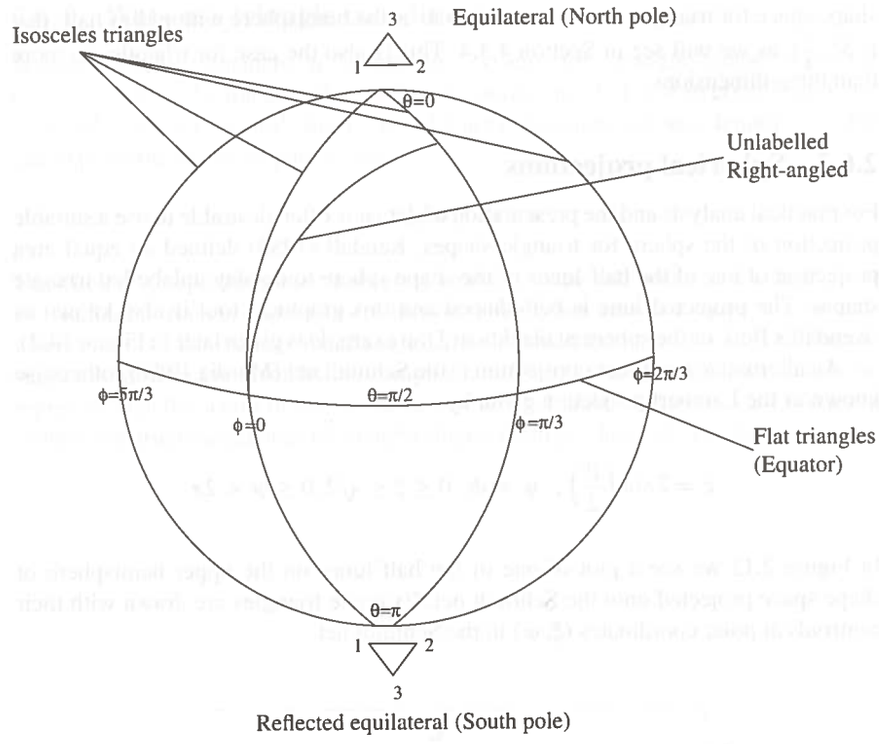


Figure 2.11 Kendall's spherical shape space for triangles in $m = 2$ dimensions. The shape coordinates are the latitude θ (with zero at the North pole) and the longitude ϕ.

in each of the six equal half-lunes in each hemisphere. Thus, if invariance under labelling and reflection was required, then we would be restricted to one of these half-lunes, for example the sphere surface defined by $0 \le \phi \le \pi/3, 0 \le \theta \le \pi/2$. Consider a triangle with labels A, B and C, and edge lengths AB, BC and AC. If the labelling and reflection of the points was unimportant, then we could relabel each triangle so that, for example, AB $\ge$ AC $\ge$ BC and point C is above the baseline AB.

If we have a triangle in three dimensions, then we see that we can translate, rotate and rescale so that

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \to \begin{bmatrix} -\frac{1}{2} \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} \to \begin{bmatrix} \frac{1}{2} \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} x_3 \\ y_3 \\ z_3 \end{bmatrix} \to \begin{bmatrix} u_{13} \\ u_{23} \\ 0 \end{bmatrix}$$

where $u_{23} > 0$. Hence, the triangle has two shape coordinates $(u_{13}, u_{23})^{\mathrm{T}}$ which lie in a half-plane. We could transfer to Kendall's spherical shape coordinates of Equation (2.14), although the range of the latitude angle θ will be $[0, \frac{\pi}{2}]$, as $u_{23} > 0$. Hence, the

shape space for triangles in three dimensions is the hemisphere with radius half, that is $S^2_+(\frac{1}{2})$, as we will see in Section 4.3.4. This is also the case for triangles in more than three dimensions.

2.6.3 Spherical projections

For practical analysis and the presentation of data it is often desirable to use a suitable projection of the sphere for triangle shapes. Kendall (1983) defined an equal area projection of one of the half-lunes of the shape sphere to display unlabelled triangle shapes. The projected lune is bell-shaped and this graphical tool is also known as 'Kendall's Bell' or the spherical blackboard (an example is given later in Figure 14.2).

An alternative equal-area projection is the Schmidt net (Mardia 1989b) otherwise known as the Lambert projection given by:

$$\xi = 2\sin\left(\frac{\theta}{2}\right), \quad \psi = \phi; \; 0 \le \xi \le \sqrt{2}, 0 \le \psi < 2\pi.$$

In Figure 2.12 we see a plot of one of the half-lunes on the upper hemisphere of shape space projected onto the Schmidt net. Example triangles are drawn with their centroids at polar coordinates (ξ, ψ) in the Schmidt net.

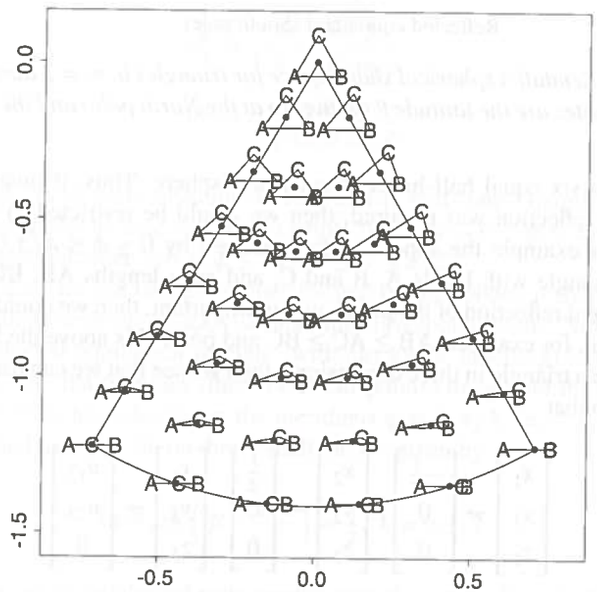


Figure 2.12 Part of the shape space of triangles projected onto the equal-area projection Schmidt net. If relabelling and reflection was not important, then all triangles could be projected into this sector.

2.6.4 Watson's triangle coordinates

Watson (1986) considers a coordinate system for triangle shape. Let $z = (z_1, z_2, z_3)^T \in \mathbb{C}^3$ be the complex landmark coordinates. Let $\omega = \exp(i\pi/3)$ and $u = (1, \omega, \omega^2)^T$, $\bar{u} = (1, \omega^2, \omega)^T$. So, 1_3, u and $\bar{u}$ are orthogonal and have length $\sqrt{3}$. One can express the triplet of points z as:

$$z = \alpha 1_3 + \beta u + \gamma \bar{u}.$$

The vector 1_3 represents the degenerate triangle shape (all points coincident), u is an equilateral triangle and $\bar{u}$ is its reflection. Since z has the same shape as $cz + d1_3$, where $c, d \in \mathbb{C}$, the shape of the triangle can be obtained from $z = u + b\bar{u}$. The Watson shape coordinates are given by the complex number b. All unlabelled triangles can be represented in the sector of the unit disc, with $|b| \le 1$ and $0 < \text{Arg}(b) < \pi/3$. Watson (1986) constructs sequences of triangle shapes from products of circulant matrices.