Statistical methods for the analysis of tooth shape

David L. Robinson

VOL. 1 (of 2)

Submitted for the degree of Doctor of Philosophy

November 2005

Dept. of Probability and Statistics, School of Mathematics and Statistics
& Dept. of Oral Health and Development, School of Clinical Dentistry

University of Sheffield

Summary

The study of tooth shape has traditionally involved analysing distances or angles between established points of correspondence, known as landmarks. Digital imaging has aided this process, yet improved statistical techniques, which offer advantages by retaining information on the geometry of objects throughout the analysis, have so far received little attention. Since methods must be suitable for use on un-extracted teeth, a key difficulty is that unwanted variation in recorded shape results from differences in the position of patients' gingival (gum) tissue.

Here we present new methodology for addressing this problem and for use in more general applications, where objects are analysed as configurations of landmarks and one would wish to account for lack of precise correspondence between certain points, in a better way than is possible using existing techniques.

After introducing the ideas of Procrustes analysis to this field, we use newly proposed methods of reliability assessment to show how, in addition to failing to allow for gum variation, implementation of this technique in its standard form is problematic, due to the poor reproducibility of particular landmarks. Use of Bookstein's (1996a,d,e) semi-landmark method, which aims to overcome lack of precise correspondence along certain directions by allowing landmarks to move iteratively along chords during Procrustes registration, is investigated but found to produce unrealistic results in certain situations.

Novel modifications of this method are then proposed and evaluated in terms of addressing the issues noted above. Alternatives to minimising the 'bending energy' of a pair of splines mapping from the mean shape, in order to determine new semi-landmark positions are explored and two new methods, using a 'nearest point' or 'full Procrustes' criterion, identified as most promising. Further investigation, by application to tooth shape problems (including a simulation study of gingival tissue variation) and use on distorted configurations generated from simple geometric shapes, show that these methods offer improvements over existing techniques in terms of filtering out unwanted variation.

Acknowledgements

Supervisors:

Dr. Eleanor Stillman & Prof. Paul Blackwell

Dept. of Probability and Statistics

University of Sheffield

Prof. Alan Brook

Dept. of Oral Health and Development

University of Sheffield

Post-graduate researchers who provided reliability data for Chapter 4:

Anwar Shah, Khaled Khalaf, Mayada Karmo, Mohammad Al-Sharood

(Formerly of the Dept. of Child Dental Health, University of Sheffield)

Technical advice and assistance:

Jim Rohlf, David Sheets, Fred Bookstein, Claire Elcock, Andrew Rawlinson

Nick Fieller, Ian Dryden

Other help and moral support:

Charlotte Birch, Mum & Dad, All three supervisors, David Patrick, Helen Owen,

Richard Smith, Jason and Samantha Poole, Ed Casson.

Publications & presentations

Published refereed papers:

Robinson, D. L., Blackwell, P. G., Stillman, E. C. and Brook, A. H. (2001). Planar Procrustes Analysis of Tooth Shape. *Archives of Oral Biology,* 46(3):191-199.
(Material from Chapter 3)

Robinson, D. L., Blackwell, P. G., Stillman, E. C. and Brook, A. H. (2001). Development of a coherent methodology for the Analysis of Tooth Shape. In Brook, A.H., editor (2001). *Dental Morphology 2001: 12th International Symposium on Dental Morphology, Sheffield, UK.* Sheffield Academic Press.
(Material from Chapters 3, 4 & 5)

Robinson, D. L., Blackwell, P. G., Stillman, E. C. and Brook, A. H. (2002). Impact of landmark reliability on the planar Procrustes analysis of tooth shape. *Archives of Oral Biology,* 47(1):545-554.
(Material from Chapter 4)

Conference presentations:

Talk: Landmark approaches to shape analysis of teeth (and likely problems). *Statistics Research Students Conference, Bristol, 22nd-24th March 1999.*
(Material from Chapters 1 & 3).

Talk: Procrustes matching and application to tooth images. *COST Dental Research Meeting, Berlin, 19th - 20th November 1999.*
(Material from Chapters 1 & 3).

Poster: Methods for describing and comparing variation in tooth shape. *British Society of Dental Research, Lancaster, April 2000.*
(Material from Chapters 1, 2 & 3).

Poster: Planar Procrustes analysis of tooth shape. *Young Statisticians Meeting, London School of Hygiene and Tropical Medicine, April 2000.* Winner of conference poster prize.
(Material from Chapters 3 & 4).

Talk: Problems in describing and comparing variation in tooth shape. *Statistics of Directions, Shapes and Images: Leeds Annual Statistics Research Workshop, July 2000.*
(Material from Chapters 4, 5, 6 & 8).

Poster: Developing a coherent methodology for the analysis of tooth shape. *12th International Symposium on Dental Morphology, Sheffield, August 2001.*
(Material from Chapters 3, 4 & 6)

Talk: Developing a coherent methodology for the analysis of tooth shape. *International conference of the Royal Statistical Society, Glasgow, July 2001.*
(Material from Chapters 5, 6 & 8).

Talk (as invited speaker): Impact of unreliable landmarks on Procrustes analysis of shape: Identification of unwanted directions of variation and possible workarounds. *Young Statisticians meeting, Edinburgh, March 2002.*
(Material from Chapters 4, 6, 8 & 9).

Internal presentations:

Talk: Problems in the analysis of tooth shape.
Seminar, Department of Child Dental Health, May 1999.
(Material from Chapters 1, 2 & 3).

Talk: Analysis of tooth shape.
Seminar, Department of Probability and Statistics, December 2000.
(Material from Chapters 1 & 3).

Talk: Reliability of landmark data from tooth images: Impact on the Procrustes analysis of shape. *Cluster group meeting, Dept. Probability & Statistics, December 2001.*
(Material from Chapters 4 & 6).

Poster & Talk: Procrustes analysis and the use of new semi-landmark methods in the study of tooth shape. *4th year part time and 2nd year full time research student assessment, School of Dentistry, April 2002.* Winner of prize for best presentation.
(Material from Chapters 4, 6, 8 & 9).

Contents

Chapter 1

Introduction and background

1.1 Motivation

For many years, dental researchers have attempted to measure and describe variations in tooth shape. Identifying variation between individuals is important in population and family studies of dental development so that we may understand the genetic and environmental influences on the shape of a patient's teeth.

Abnormal shape affects the contact relationships of teeth, their function (primarily the comminution of food), physiological adaptation (ability to sustain themselves and assist in the development and protection of the tissues that support them) and the dento-facial aesthetics of a patient. Consequently, early diagnosis of non-normal appearance is important in order to enable improved planning of treatment and control of development, through surgical, orthodontic and restorative procedures.

Studies of tooth shape have typically involved taking traditional, linear orthodontic measurements and performing an analysis using univariate statistical methods on quantities derived from them. In recent years, advances in technology have aided the process of obtaining these measurements and have also allowed a greater variety of more detailed information to be collected. At the same time, improved techniques for the analysis of shape have emerged in the fields of mathematics and statistics, allowing analysis to be based directly on information such as 'landmark' locations or outline coordinates which may be easily obtained from digital images. In contrast to the approaches described above, these more powerful methods allow work with the full geometry of the objects involved, which is otherwise ignored or lost.

However, so far little attention has so far been given to these ideas for the study of human tooth shape.

1.2 Preliminary definitions

We first describe some of the terms and definitions used throughout this project when referring to the anatomy of the permanent dentition. The different crown surfaces are described and definitions given of standard dental measurements. We also give formal definitions of terms such as 'landmarks' and, of course, 'shape'.

1.2.1 Tooth types, features and surrounding tissue

Teeth fall into one of 8 possible types and 4 possible classes. In each of the maxillary (upper) or mandibular (lower) arches there are two (left and right) central and lateral 'incisors', 'canines', first and second 'pre-molars' and first, second and third 'molars'. Incisors are cutting teeth, which have an 'incisal edge'. Canines have a single, pointed 'cusp', a pointed elevation or mound, for tearing and incision, whereas pre-molars and molars have two or more cusps, to break up and grind food. As with most studies of tooth morphology, we are concerned only with the adult permanent teeth, excluding the third molars, rather than the deciduous teeth (so 28 teeth in a full compliment). However, all of the definitions and methods of analysis presented here may also be used on children's teeth.

The pairs of teeth are arranged symmetrically on either side of the mouth and occur in the same order within each dental arch, with the 'median line' passing between the two central incisors. The left-right pairs within each arch have long been established to be roughly symmetrical in size (e.g. Moorrees & Reed, 1964) and shape (e.g. Garn *et al.*, 1966b), however between the upper and lower arches, the shape of each tooth type is different. Consequently there are 16 different tooth shapes (the 8 on one side of the upper arch and the 8 on one side of the lower). In any study of size or shape teeth are only compared with those at the corresponding position within the dentition. It makes no sense to compare molars, canines and incisors or to make comparisons within classes (e.g. central vs. lateral incisors). Fig 1.1, based on a diagram in Wheeler (1962), shows the maxillary jaw, with the tooth types labelled. (It also illustrates the standard MD, LAAC, LACC and OG measurements, to be described in section 1.2.3). In the lower arch the arrangement of the teeth is identical (although their appearance will be different

from those shown here). The incisors and canines are often referred to as 'anterior' teeth whereas pre-molars and molars are called 'posterior' teeth.

By the shape of a tooth we refer to the shape of the 'crown' only and not the 'root', which fixes the tooth into the bony process of the jaw. The crown is the portion of the tooth covered with enamel and joins the cementum covered root at the 'cemento-enamel junction' (CEJ). After full eruption soft tissue known as gingiva or gum tissue covers part of the tooth crown and fills the interproximal spaces between teeth, where it is referred to as the interdental papilla. The 'gingival margin' is the part of the gingival tissue that follows the curvature, but not necessarily the level, of the CEJ. While the CEJ is a stable demarcation, the gingival margin which (usually) covers it may vary and merely represents the gingival level on the tooth at any one point in time. (See fig. 1.2(a)). 'Grooves' are lines between parts of the crown and on the occlusal surface they are known as 'fissures'. 'Pits' are small pin-point depressions located at the junctions or terminals of grooves or fissures.

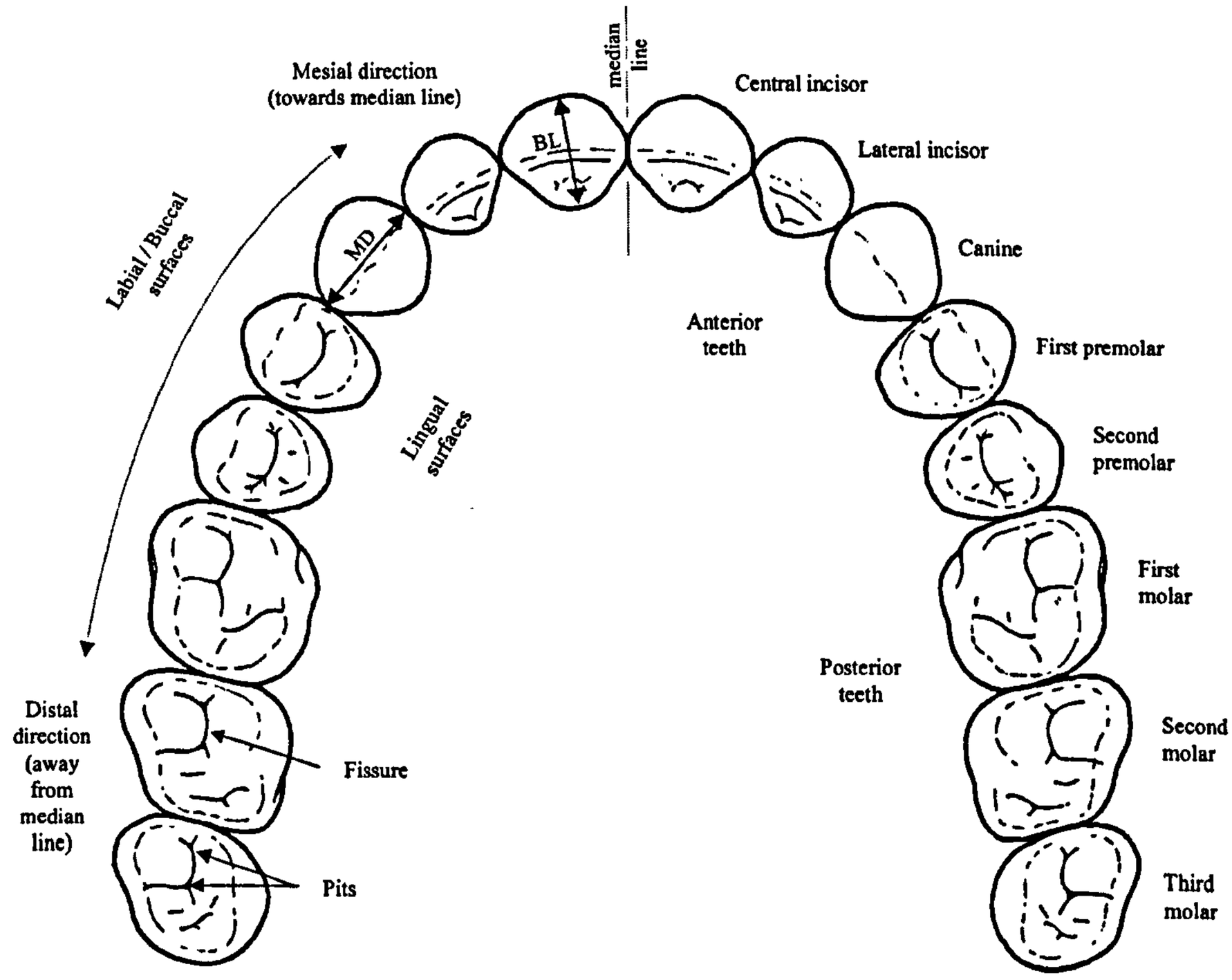


Fig. 1.1 Occlusal view of maxillary arch

1.2.2 Crown surfaces

The crown surfaces facing the lips or cheek are known collectively as the buccal surfaces, as shown in figs.1.1 and 1.2 (a and b). For the incisors and canines these are also referred to as the labial (facing the lips) surface. Those facing the tongue are known as the lingual surfaces (see fig. 1.1). The surfaces that come into contact with those in the opposite jaw when the mouth is closed, are known collectively as the occlusal surfaces, as shown in fig. 1.1. For incisors and canines these are also referred to as incisal surfaces. The proximal surfaces are those facing the adjoining teeth in the same arch and for any particular tooth are referred to as either the mesial (facing toward the median line) or distal (facing away from the median line) proximal surface. The proximal surfaces of neighbouring teeth touch each other in their 'contact areas'.

As with the majority of the studies, here we consider only the buccal and occlusal surfaces. On the occlusal surfaces, most of the characteristics of the different teeth types are visible (edge, cusps, fissure patterns etc.) and so this is the surface most often used for analysis. The buccal surfaces are also important, particularly from an aesthetic point of view, since this is the view of the teeth seen by the patient and others. Although there have been many studies considering the proximal surfaces of teeth, obtaining information from them is either intrusive (requiring extracted teeth) or time consuming (if casts of teeth are taken and sectioned). The lingual surface however is hardly ever studied since this does not contain any features or other information that is not already represented by the examining the buccal and occlusal surfaces.

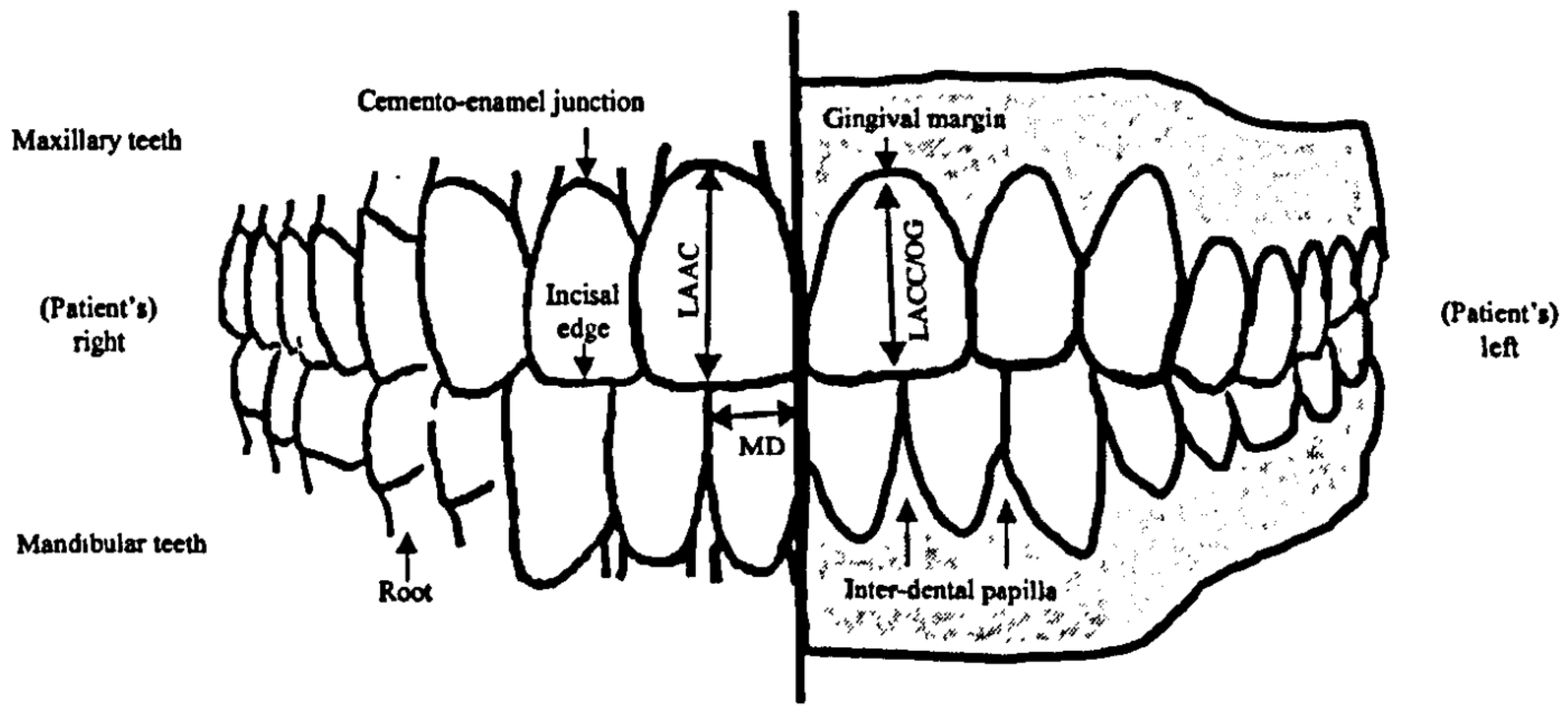


Fig. 1.2(a): Buccal (labial) view of the anterior dentition, with and without gingival tissue

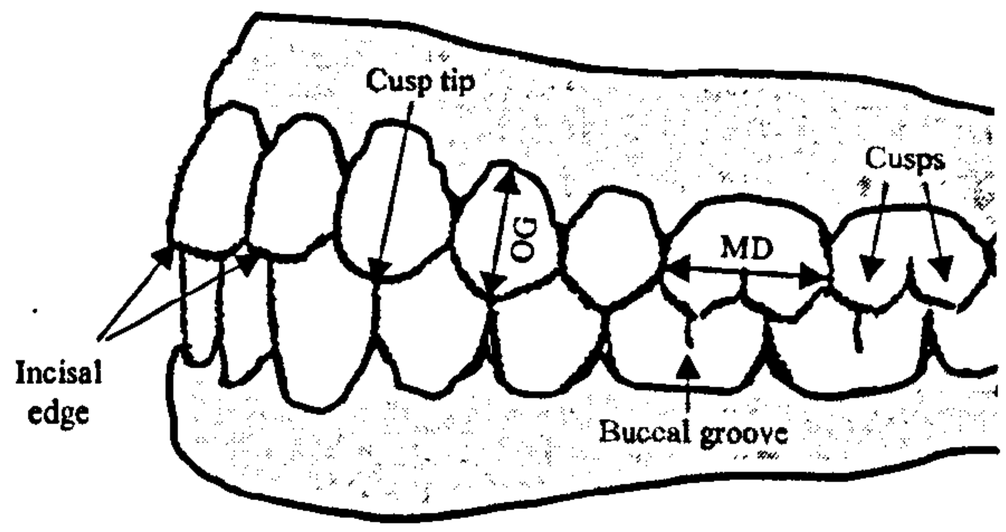


Fig. 1.2(b): Buccal view of the posterior dentition, with gingival tissue

1.2.3 Standard dental measurements

Many dental measurements have been defined over the years (see for example, Remane, 1930), but those which are routinely taken in almost any study of tooth size and shape are the maximum dimensions of the teeth.

1.2.3.1 Mesiodistal diameter (MD)

The most frequently used definition of the mesiodistal diameter is that of Moorrees *et al.* (1957) which is the maximum distance between the tooth contact areas, taken parallel to the buccal and occlusal surfaces. From the buccal view this means approximately perpendicular to the long axis of the clinical crown (LACC) or occluso-gingival dimension (see 1.2.3.3) and from the occlusal view, roughly perpendicular to the bucco-lingual dimension (see 1.2.3.2). Note that it is not defined as the distance between the actual contact points made with neighbouring teeth, since often there may be teeth rotated or displaced within the arch, no contact or an adjacent tooth may be missing. To allow for this and to ensure that the endpoints of MD are in homologous positions, Goose (1963) and Wolpoff (1971) stated that the MD diameter should be taken between where the points of contact would be in perfect occlusion. See figs. 1.1 and 1.2.

1.2.3.2 Bucco-lingual diameter (BL)

The most frequently used definition is the maximum distance between the buccal and lingual surfaces of the crown, perpendicular to the MD dimension (Moorrees *et al.*, 1957). For canines, molars and pre-molars the curved (bulbous) nature of the both these

surfaces, means that both endpoints are on the most prominent parts of each part of the crown. For the incisor teeth, the buccal endpoint will also be at the most prominent part of the buccal crown surface, but the lingual endpoint will be on the gingival margin at the back (lingual side) of the tooth. See fig. 1.1.

1.2.3.3 Occluso-gingival dimension (OG) / Long axis of the clinical crown (LACC)

For molars, Moorrees *et al.* (1957) defined the occluso-gingival dimension or 'crown height' as the distance between the tip of the mesiobuccal cusp and the most apical point of the gingival margin, taken in a direction perpendicular to the MD dimension. For incisors, canines and pre-molars, Lavelle (1968) defined the OG dimension as the greatest distance between the highest point on the occlusal surface (cusp tip or position along the incisal edge) and the lowest point on the gingival margin, again perpendicular to the MD width. A related, often identical, measure for canines and incisors is the 'long axis of the clinical crown' (LACC). This passes from the canine cusp tip or centre of the incisal edge to the centre of the gingival margin, roughly separating the buccal surface into mesial and distal halves. (There is no requirement that OG dimension is down the centre of the buccal surface). The 'long axis of the actual crown' (LAAC) passes along the same direction as the LACC, but to the centre of the cemento-enamel junction. See fig. 1.2.

1.2.4 Shape and similarity groups

Shape is defined by Kendall (1977) as "all the geometrical information that remains when location, scale and rotational effects are filtered out from an object". Similarly, Bookstein (1998) noted that "In ordinary language, the shape of an object is described by words or quantities that do not vary when the object is moved, rotated, enlarged or reduced". Dryden & Mardia (1998) added that the shape of an object consists of all its geometrical properties that are unchanged when it is rotated, translated or re-scaled in an arbitrary coordinate system and that two objects have the same shape if they can be rotated, translated or rescaled so that they match each other. Any measure of shape must therefore be invariant to the three 'similarity transformations' of rotation, translation or scale, or in other words, unaffected by any 'registration differences'.

1.2.5 Landmarks

Landmarks are key points of biologically correspondence, located and defined in the same way on each object of a similar type (here tooth types), so that they match both between and within populations. Examples of landmarks on the buccal surface of an upper central incisor and occlusal surface of an upper first pre-molar are presented in fig. 1.3 below. Landmarks may be 'anatomical', homologously identifiable points that correspond between objects in some meaningful way (e.g. the tip of a cusp or corner of an incisal edge) or 'mathematical', located according to some geometrical or mathematical property (e.g. points of maximum curvature or diameter, such as the endpoints of the mesio-distal dimension). In addition, 'pseudo landmarks' may also be constructed either around the outline of an object (e.g. at equally spaced points) or in between anatomical or mathematical landmarks.

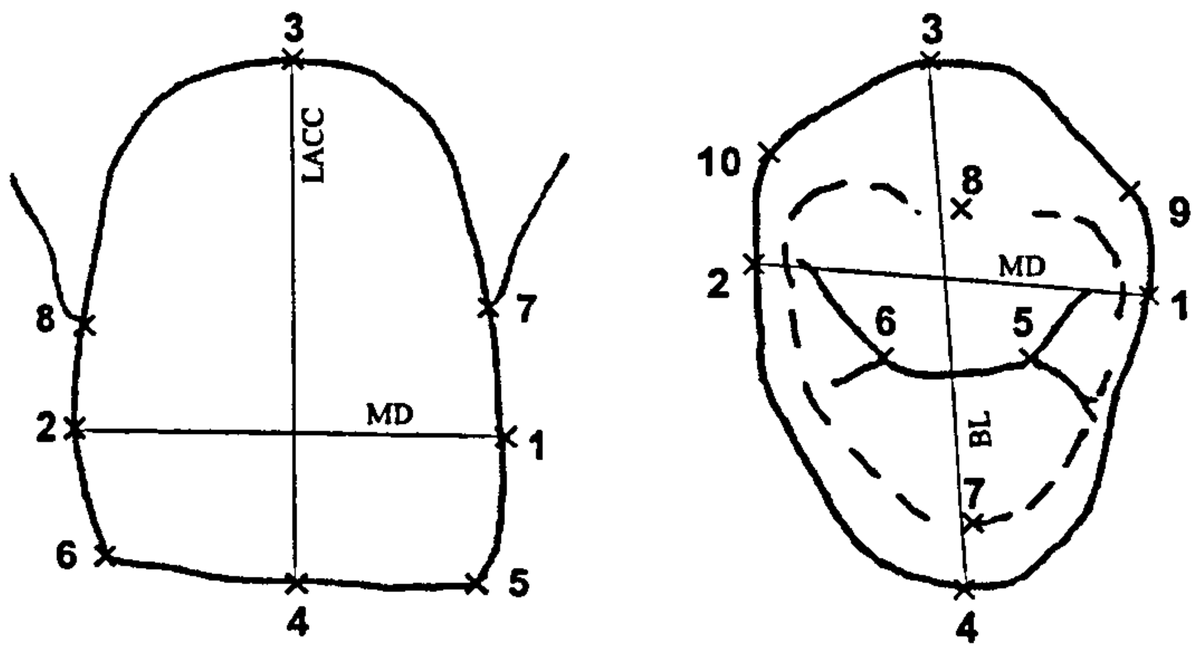


Fig 1.3: (Left) Possible landmarks on buccal surface of upper central incisor. (1) & (2) mesial & distal endpoints of MD, (3) & (4) gingival & incisal edge endpoints of LACC, (5) & (6) corners of mesial & distal sides and incisal edge, (7) & (8) ends of mesial & distal papilla. (Right) Possible landmarks on occlusal surface of upper first molar: (1) & (2) mesial & distal endpoints of MD, (3) & (4) buccal & lingual endpoints of BL, (5) & (6) mesial & distal pits/fissure junctions, (7) & (8) lingual and labial cusp tips, (9) & (10) mesial & distal endpoints of maximum labial cusp width.

1.3 Variations in size and shape

The subject of variation in size and shape has generated a larger literature than any other aspect of dental anthropology (Hillson, 1996). Recent interest in this field has inspired a series of international symposia on dental morphology, which began in Fredensorg,

1965 (Pedersen *et al.*, 1967) and more recently has included meetings in Berlin, 1995 (Radlinski & Renz, 1995), Oulu, 1998 (Mayhall & Heikkinen, 1999) and Sheffield, 2001 (Brook, 2001). Tooth crowns are formed to their permanent size and shape during childhood and so can be studied in mixed collections of individuals of varying ages (post-eruptive effects permitting). In addition, by taking dental impressions, living people (with known biological affinities) can be compared directly with ancient dental remains, due to the durability of the teeth. However, while there is extensive work on the size of teeth, for shape the literature is very limited.

Several authors have described how the shape of the different tooth types may vary within and between populations. Lamb (1988) described how the buccal outline of upper central incisors may be square, oval, or tapered (widening towards incisal edge). Hillson (1996) and Alt & Turp (1998) described how incisors may also be barrel-shaped, peg-shaped (narrower at the incisal edge) or shovel-shaped and that canines may be 'blunted' or 'nippled'. In addition, for molars and premolars, variation in the shape of the cusps may occur in terms of their 'pointedness', 'height' or 'bulge/ roundness' (Hillson, 1996).

Factors which play a role in controlling both the size and shape of tooth crowns are essentially either pre-eruptive (genetic or environmental) or post-eruptive. It is the former that has generated the vast majority of research in the field, whereas most post-developmental changes are often seen as a nuisance to investigating pre-eruptive causes.

1.3.1 Pre-eruptive variations

1.3.1.1 Crown dimensions

Variations in tooth size (as measured by the MD and/or BL dimensions) have long been established as being under the control of genetic factors (e.g. Moorrees & Reed, 1964, Garn *et al.*, 1965, Dempsey *et al.*, 1995). Heritability studies have included Alvesalo & Tigerstedt (1974), who described how approximately 60% of variation in size can attributed to parental differences. Osbourne *et al.* (1958) demonstrated strong degrees of similarity in the tooth dimensions of Caucasian twins and Townsend (1980) made a similar observation in Australian aboriginals. The determination of crown dimensions

by inheritance has also been investigated in like-sexed and unlike-sexed siblings, see for example, Alvesalo (1971), Townsend & Alvesalo (1985). Of particular interest in Sheffield has been alterations in size with tooth number in genetic conditions such as hypodontia (congenital absence of one of more teeth). Lavelle *et al.* (1970), Brook (1984) and others have shown that these patients are more likely to have smaller teeth. In terms of gender differences, males have long been shown to have larger tooth dimensions than females (e.g. Moorrees *et al.*, 1957, Lavelle, 1970, Miethke, 1972) with the canines exhibiting the largest differences and pre-molars the least (Garn *et al.*, 1967a, Rieger, 1993). Differences between ethnic groups have also been reported. For example, Lavelle (1972) noted differences in the MD dimensions of Negroes, Mongoloids and Caucasoids (anthropological classifications), with the largest dimensions observed in the former. Similarly, Bailit (1975) compared the tooth dimensions of Australian Aborigines with Norwegian Lapps. Environmental factors determining size may be related to the persistence of ameloblast activity at various stages of the crown formation. Health during pregnancy, birth weight, maternal age and birth order have all been said to have an influence on tooth size (see for example, Garn *et al.*, 1979, Bailitt, 1975 or Fearne & Brook, 1993). In addition diet or disease may also affect the process of development (Keene, 1966).

1.3.1.2 Crown shape

While many studies have demonstrated genetic and environment influences on particular tooth dimensions, investigations of influences on tooth shape are harder to find. In part this may be due to the fact that over the years, very few techniques for measuring and describing tooth shape dissimilarity have actually been considered, rather than the fact that differences in shape do not exist and have therefore not been reported. For example, Dahlberg (1945) observed that patients with hypodonia often had 'peg shaped' incisors, but did not quantify this in any way. Many authors have used descriptions (such as those at the start of section 1.3) to categorise teeth into various groups, which may then be compared in terms of frequency. Carbonell (1963) reported how shovel-shaped incisors were more often seen in Asians and Native Americans than Europeans. Moskona *et al.* (1997) analysed presence/absence of 'shovel', 'peg shaped' or 'curved' incisors when studying the dental traits of Sinai Bedouin tribes. Usually, classifications are based on those of Dahlberg (1949) who also

defined different types of fissure pattern for molars (as used by Garn *et al.*, 1966a) and additional classifications such as 'barrel shaped'. Some findings on shape have also been reported by way of findings from the MD and BL dimensions. Garn *et al.* (1967b) claimed that tooth shape was different in males and females since women had a larger difference in BL dimensions (compared to the men) than in the MD dimension. In contrast, Townsend (1983) found that in individuals with Down's syndrome, the gender difference was larger for the MD rather than the BL dimension. The only quantitative measures that have been used to describe tooth 'shape' (in terms of values invariant to the effects of rotation, location and scale, as described in 1.2.4), are angles and ratios. Peck & Peck (1972) proposed the 'crown shape index', the ratio of MD/BL, originally to investigate the relationship of tooth shape and crowding. This has been used by many other authors since as a comparative measure of tooth shape between populations, including, for example, Wood & Engleman (1988) and Bodner *et al.* (2001). More recently, Shah *et al.* (2003) used the ratio of the MD dimension and the width of the tooth at 50% of the way along the LACC (parallel to MD) to investigate the relationship between tooth shape and crowding of lower incisors. Angles between measurements have also been used. For example, Peretz *et al.* (1998), who analysed angles between inter-cusp measurements in examining molar teeth of individuals with Down's syndrome. Note that many studies report the occurrence of an accessory (Carabelli) cusp on the labial surface maxillary molars as an instance of abnormal tooth shape, although others prefer to regard this as the presence of an additional rare feature, rather than a difference in shape (see, for example, Alt & Turp, 1998).

1.3.2 Post-eruptive changes

Although teeth do not grow once formed, there are still factors which affect their dimensions and shape after eruption. Size and/or shape may be changed due to bacterial action (caries), trauma or mechanical intervention by a dentist, gradual loss of material due to attrition (tooth on tooth contact), abrasion (physical wear by objects other than other teeth), or erosion (loss of tooth substance by chemical processes, e.g. diet containing acids or due to gastrointestinal disorders). Many of these are regarded as unwanted or nuisance effects when trying to study actually genetic or environmental variations in tooth form. For example it may be that the degree of attrition is related to the populations under investigation, due to differences in diet or methods of food

preparation (Molnar, 1971). Reinhardt (1983) documented how attrition also increases with age. As recommended by Keiser (1990), carious or restored teeth or those with marked attrition are almost always excluded from any analysis of size or shape.

1.4 Data collection

Because of the difficulty in obtaining measurements directly from a patient's mouth, particularly from the posterior teeth, dental casts are nearly always used. Ballard (1944) and Moorrees *et al.* (1957) each documented similar basic principles for acquiring stone dental casts, by taking alginate impressions and using plaster of Paris. Barrett *et al.* (1963) reported how standard MD and BL measurements obtained from casts were more reliable than those taken directly in the mouth and Doris *et al.* (1981) noted similar findings, particularly for posterior teeth.

For the current study, alginate impressions (Alginoplast, Bayer) cast in hard dental stone (Kaffir D stone, British Gypsum) were available from patients of the Charles Clifford Dental Hospital, Sheffield. These had already been used for previous research projects, for which ethics committee approval had been obtained.

Traditionally measurements have been obtained using sliding callipers or engineering dividers. A limitation however, is the amount of information they can provide, particularly when compared to methods which allow the exact locations of different features or 'landmarks' on a tooth surface to be recorded, from which a greater variety of more detailed measurements can then be studied (see, for example, Khalaf *et al.*, 2001 or Shah *et al.*, 2001). Methods that have been used to obtain two-dimensional (2D) landmark locations from teeth have included the Optocom system (see for example, Van Der Linden *et al.*, 1972), which consists of a microscope mounted over a table, moveable in two dimensions. More popular, though, has been the use of photographs of the buccal or occlusal surfaces of individual teeth (orientated in some standardised way) from which landmarks locations (and distances between them) can be recorded in two-dimensions, using the perpendicular sides of the resulting image as a horizontal and a vertical axis. In recent years, advances in digital imaging had aided this process considerably, allowing landmarks to be located on a screen and recorded

with greater accuracy. Biggerstaff (1969) first digitised photographic negatives and converted the resulting (x,y) coordinates into a measure of surface area. Brook *et al.* (1983) developed a cathode ray tube based image analysis system for the measurement of tooth dimensions from dental casts, mounted on a calibrated stage, adjustable in 3 dimensions. Peretz & Smith (1993) used image analysis involving a video camera and monitor to measure distances between cusps in the occlusal view.

For the current study, images of buccal and occlusal surfaces were obtained using the system described in Brook *et al.* (1998). This comprises a 32-bit digital camera (Kodak/Nikon DCS 410 with 90mm Elicar macro lens), mounted above a platform (which is moveable in 3 planes), which is connected to a PC. Once a tooth surface has been orientated to a pre-specified criterion, for example so that the buccal surface is perpendicular to the direction of the lens or so that the maximum surface area is visible, images are acquired by Adobe Photoshop v4.0 (Adobe systems) and processed using Image Pro-Plus v3.01 (Media Cybernetics, USA). For calibration of images, a small section of steel rule is placed in the plane of the surface being imaged, which is then used to specify the scale of the x-axis and perpendicular y-axis of the 2D image (running parallel to the bottom and sides of the image). The software then allows information such as the (image) coordinates of landmarks and distances between landmarks to be recorded. In addition, the binary outline image of the surface may be obtained using the edge detection trace facility in Image Pro Plus v3.01. See fig. 1.4 (a and b).

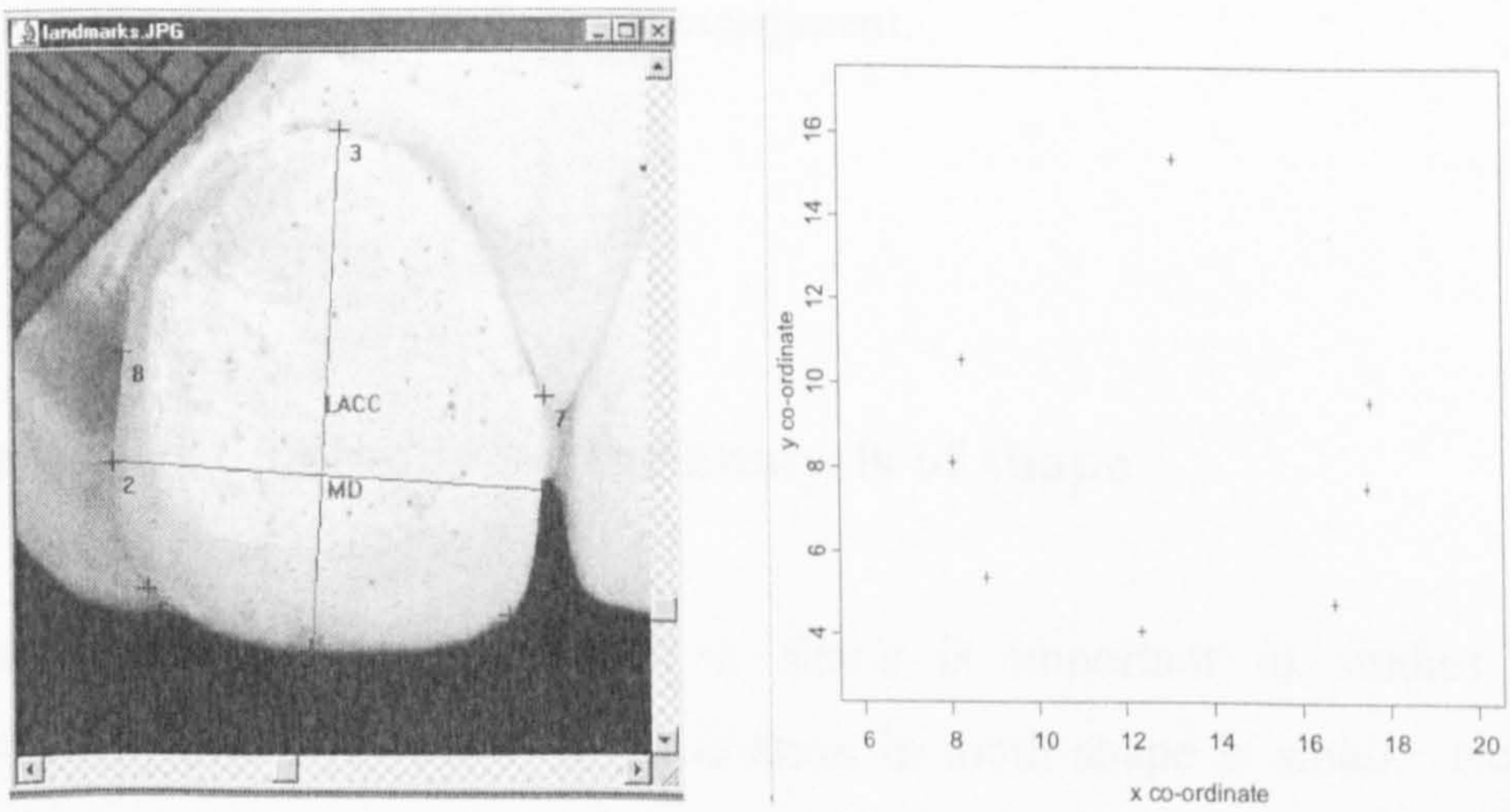


Fig 1.4(a): (Left) Image of buccal surface of an upper central incisor on a study cast, with MD and LACC dimensions displayed. Landmarks identified as in fig 1.3 (left). (Right) Plot of coordinates in S-Plus.

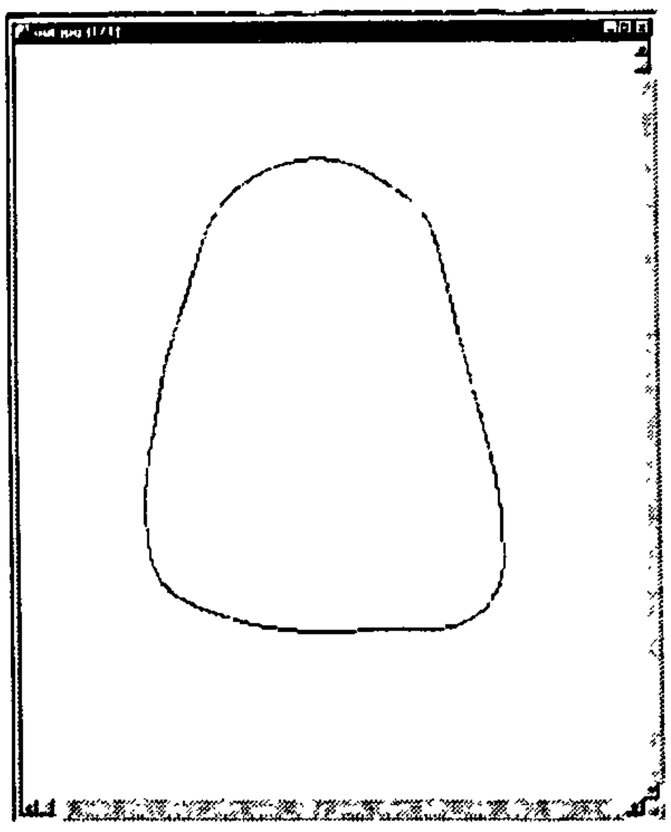


Fig 1.4(b): Binary outline of the buccal surface of an upper central incisor from Image Pro-Plus

In attempts to obtain three-dimensional (3D) coordinates, methods have included the reflex metro graph (see for example, Richmond, 1987), where points are marked directly onto an object and 3D coordinate locations recorded using a moveable, semi-reflecting mirror and the travelling microscope moveable along the frame of a glass box, with the object of interest placed inside (see, for example, Bahti & Harrison, 1987). Taverne *et al.* (1979) described how a 3D representation of a tooth crown could be constructed by sectioning the cast into different slices and obtaining a series of 2D photographs. Of course, the collection of 3D data is now quick and easy using laser scanning, which typically comprises two CCD (Charge Coupled Device) video cameras which capture reflected images or a single CCD and two mirrors, with the object fitted on a rotary table Kuroda *et al.* (1996), describe in detail how tooth dimensions could be recorded using laser scanners. However, the main limitation of obtaining 3D coordinates at present is the expense of equipment.

1.5 Need for this study

1.5.1 Improved methods for the analysis of shape

Accurate quantification of variation in shape is important in studies of dental development, particularly since most variation in tooth shape is small. However, an examination of the literature reveals that most of the methods used have so far been limited in terms of how shapes are described and in terms of the analysis performed.

Most of these methods have involved either categorising teeth into various groups, recording the presence/absence of abnormal shape characteristics, which may then be compared in terms of frequency, or taking traditional orthodontic measurements and analysing quantities or 'indices' derived from them, such as angles or ratios between them (for example, Peck & Peck's 1972, crown shape index). Standard univariate statistical techniques have then been employed to form inferences based on these quantities.

Advances in technology have aided the process of obtaining measurement data and have allowed a greater variety of more detailed measurements to be collected. However, difficulties still exist with the use of such methods. The same set of distances can often be obtained from two different shapes because the locations of measurements, relative to one another, are not represented in the data. Because of this it not usually easy to generate graphical representations of shape from results based on measurements alone (Marcus & Corti, 1996).

In recent years however, improved techniques for the analysis of shape have emerged in the fields of mathematics and statistics. During the late 1980's and early 1990's a shift occurred in the way morphological structures were quantified and how the data were analysed, using methods which captured the geometry of the objects and preserved this information throughout the analysis (for example, by representing shape as a configuration of landmark locations or a mathematical description of the object's outline). A short summary of the various approaches to shape analysis, in a historical perspective, was presented by Bookstein (1998) in a paper titled 'A hundred years of morphometrics'. This compared and described the development of the various ideas and the extent to which they have now been standardised. In particular, for objects represented as configurations of landmark locations, numerous methodological approaches evolving from the ideas of Bookstein (1986) and Kendall (1984) have been consolidated into what Bookstein (1998) refers to as the 'morphometric synthesis'; an established framework based around the idea of 'Procrustes analysis' by which an investigation of shape, based on landmark data should proceed.

Much of this development has taken place through a series of workshops since the late 1980's onwards, which have brought researchers together to discuss various ideas and

theoretical developments in statistical shape analysis and brought many of the methodologies to the attention of other researchers. The edited volumes of these workshops contain many of the papers which have had a key impact on this subject. See for example, Rohlf & Bookstein (1990), Marcus *et al.* (1993), Mardia & Gill (1995), Marcus *et al.* (1996) and Mardia *et al.* (1996a).

As these methods become more established, their application to various biological disciplines has becoming increasingly widespread. For applications when landmark data are available, the use of Procrustes analysis in particular has become particularly popular, due to the advantages it offers, in terms of developed theory, statistical properties such as consistency and power and the variety of descriptive and inferential techniques, which may be utilised. A key benefit is that it allows interpretation of results as pictures in the original space of the objects being considered, since the geometry of the objects is always retained throughout the analysis. Examples of studies making use of this and other methods can be found from many fields, from zoology to medicine. See for example, Kingenburg & Bookstein (eds.) (1998). However, so far little attention has so far been given to these ideas for the study of tooth shape.

1.5.2 Difficulties in the study of tooth shape and the need for statistical development

Successful implementation of any method of analysis depends on the accuracy and reliability of the data involved, i.e. that the data is reproducible and has good correspondence between cases. In the study of tooth shape, however, there are several difficulties in meeting these requirements that need to be addressed.

As with any investigation, inconsistencies between or within operators in recording data will carry through into the inferential procedures, inflating residual variance and diluting 'real' differences between individuals and so the effect of this on any method of shape analysis must always be considered.

However, an important difficulty that has yet to be addressed when considering an analysis of tooth shape (particularly of buccal surfaces) is the fact that part of a patient's

crown is usually obscured by gingival (gum) tissue. Variation between patients in the position of the gingival margin and inter-dental papilla will affect data collected in these regions representing the shape of a tooth, producing variations in recorded shape which are of no interest.

The level and positions of these gingival features vary throughout an individual's life once teeth are fully erupted and at any one time can depend on many factors. These include gum health, underlying bone morphology, tooth brushing technique, age and the contact relationships with neighbouring teeth. For example, with increasing age or bad gum health (periodontal disease), gingival recession becomes more common (Wilson & Kornman, 1996). The soft tissue profile of the interdental area is also determined by the contact relationships with the neighbouring teeth (Linde, 1990) as well as the nature of the interproximal spaces (Wheeler, 1974). For the gingival margin, Wilson & Kornman (1996) state that the normal range of variation after complete tooth eruption is approximately 0.5-2.0 mm from the centre of the cemento-enamel junction. Linde (1983) provided a similar range, whereas Heasman (1997) suggested 0.5-3.0mm.

Consequently, teeth may have the same true shape but differences in how this is represented or recorded due to the extent to which the visible part of the tooth crown is obscured the position of the gingival tissue. For example, fig 1.5 shows how different LACC dimensions are recorded from the same underlying tooth shape, but differing position of the gingival margin. The diagram also illustrates how any data around the crown outline will also be affected by the interdental papilla.

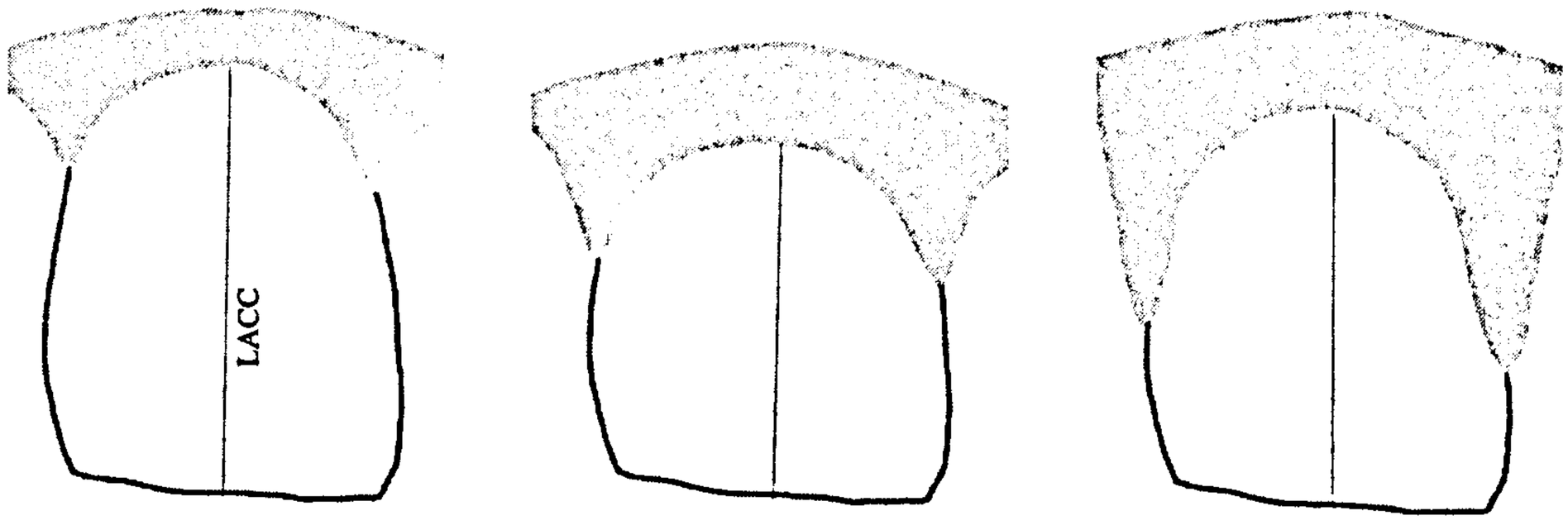


Fig. 1.5: Three teeth with identical shape but different coverage by gingival tissue

Obviously, investigations based on extracted teeth would not suffer this problem. However, this sort of data is typically difficult to acquire, as permanent teeth are not usually extracted unless they are unsound. The shape of the actual (rather than clinical) crown (including the cemento enamel junction) could be made visible with the use of X-rays. However, these are also not typically used for such investigations, for ethical reasons, as X-rays are only taken when clinically necessary. (So one could not set up a large population study for example). In addition, they have poor dimensional stability compared with study models and if they are used (retrospectively), curving or other distortions are often evident on the resulting images, due to the way they have been obtained.

Consequently, since we require methods suitable for use on teeth in the mouth, the effect of gingival variation between different individuals must also be considered in any method of analysis. whilst using only information from the visible parts of the surfaces of a tooth. Techniques must be able to recognise that data obtained from teeth which are identical in shape, but with the gingival margin and/or interdental papilla in different positions, are in fact still the same shape. Therefore in addition to registration differences (described in 1.2.4), we must also remove, filter out or account for gum variation in some way, before appropriate comparisons of shape can be made.

Variants of existing methods of shape analysis have been proposed for situations where lack of homology of features occurs around the outline of an object, for example, Bookstein (1996a,d,e), but these methods have not been thoroughly tested or applied to situations where unwanted variation in shape does not occur in directions around an object's outline (as we have here and as may occur in may other biological applications). In the proceedings of recent meetings, Rohlf (1996) and Bookstein (1998) note current areas for development and each highlight the need for progress in the analysis of data lacking precise landmark correspondence.

Since actual variations in tooth shape are small, development or modification of the existing statistical techniques may be required in order to address the problems described above. Different surfaces each present their own particular difficulties and so a variety of solutions may be required.

1.6 Aims and thesis outline

1.6.1 Statement of aims

Our aim is to produce a coherent statistical methodology for the analysis of tooth shape, allowing future investigation into the dependency of shape on explanatory variables. The techniques must be suitable for removing unwanted variation in shape in directions other than around the outline, in order to address such problems as gingival tissue variation. In addition, any methodology to be used also needs to be well suited for user interaction, which is important in clinical situations.

1.6.2 Thesis outline

In chapter 2 we review and evaluate the most popular methods of shape analysis, in terms of their suitability as a starting point for our investigations. The techniques are separated into landmark (distance based and geometric) and outline approaches as described above. We provide reasons as to why the Procrustes technique (along with the other tools of the 'morphometric synthesis') has, in recent years, become the standard protocol for investigations of shape when 'landmark' data is available and why this is also the most logical place to begin our investigations.

Chapter 3 introduces the ideas of Procrustes analysis for the first time to the study of tooth shape, where teeth are represented and analysed as configurations of 'landmarks' from digital images. Using buccal images of central incisors from patients with hypodontia and a corresponding control group, we illustrate how it is possible to estimate mean shapes and investigate shape variability. We also demonstrate how conventional 2D inferential techniques may be adapted to address hypotheses concerning shape and provide full mathematical details throughout, including the ideas of shape space and tangent space.

The techniques assume that the data is reproducible and has good correspondence between cases. Therefore in chapter 4 we set out to establish how well the Procrustes method can be expected to perform (in its standard form) when used on surfaces from a

variety of tooth types and, in particular, how much impact inconsistencies in the positioning of landmarks may have on investigations of shape. From buccal and occlusal views of a variety of tooth types, we establish the level of variation due to operator inconsistency in obtaining landmarks, relative to actual variation in tooth shape.

In chapter 5 we turn our attention to addressing some of the issues highlighted in 1.5.2 and investigate the use of 'semi-landmarks' (Bookstein, 1996a,d,e), an extension of the standard Procrustes methodology which recognises that landmarks may be known to lie along particular lines or curves, such as around the outline of an object, but are difficult to locate precisely. The technique allows these landmarks to move in pre-specified directions to overcome the lack of precise correspondence by introducing an additional standardisation step in the Procrustes procedure and so may offer a way of removing unwanted variations in shape due to differences in the position of a patient's gingival tissue.

In chapter 6, we consider other ways in which the semi-landmark and/or Procrustes method can be modified to overcome this problem. In particular we examine other possible mappings and criteria that may be optimised when determining the new positions of the semi-landmarks as well as the use of alternative existing techniques, which aim to allow for lack of landmark correspondence. Chapter 7 then describes a specifically written routine for investigating the selected methods in S-plus v.4.5 (Insightful Corporation). We give details of the various options for specifying how the positions of semi-landmarks are determined (including the importance of different scaling and convergence choices) and for displaying the initial, final and individual steps of the different procedures.

In chapter 8 we investigate how well each of the newly proposed methods performs in addressing the problem of unwanted variation in shape due differences in the position of a patient's gingival tissue. As well as removing this variation between individuals, a successful method must also be able to identify that representations of the same individual tooth, but with the gingival margin and/or interdental papilla in different positions, are in fact, still the same shape. Since we have no way of investigating this using study casts (as the gum is always in one position), images of patients' teeth where

the entire actual crown is visible are used and different possible clinical crown shapes generated by simulating different possible positions of a patient's gingival tissue. Use of probability models is made with parameters chosen to reflect what is known about gingival variation from both the dental literature and experienced periodontologists. The choice of summary measures for assessing and comparing the results of the different methods, is also carefully considered. The investigation centres on the buccal surface of the upper central incisor, although the methods described are intended to be suitable for buccal surfaces of all tooth types.

In the next two chapters we consider other applications and uses of the new methods, both in the study of tooth shape and when applied to configurations generated from simple geometric forms. In chapter 9, we consider how the different methods perform in addressing some of the other problems and difficulties associated in the study of tooth shape on a variety of different tooth types. In particular, we return to the data and results of the reliability study in chapter 4 and investigate use of the new methods in addressing issues of operator inconsistency. In chapter 10 we consider use of the new methods in filtering out unwanted patterns of variation generated from basic known shapes, so that we may extend our investigations of the new methods to applications beyond those encountered in the study of tooth shape. Simulating variation in specific controlled ways, also allows us to monitor and assess how the different methods perform when trying to remove certain patterns of variation.

Finally, chapter 11 summarises and discusses the main points and findings from this study. We offer recommendations and advice as to when use of the new methods presented here may be more beneficial than a standard Procrustes analysis of shape or use of Bookstein's original semi-landmark method, both for the study of tooth shape and in wider applications. Ideas for future development of this work are also suggested and considered.

The descriptions of many of the techniques involved, throughout this thesis, do require some expertise in mathematical notation, which may be unfamiliar to most dental researchers. However, we hope that readers will be able to appreciate the main points of the different methods involved from the accompanying information in the text.

Chapter 2

Review of main methods of shape analysis

2.1 Introduction

The field of shape analysis involves methods for the study of the shape of objects where location, rotation and scale are removed or accounted for in some way. The main aims are to estimate average shapes and the structure of shape variability and to carry out inferences on population quantities.

Methods of shape analysis may be categorised by their associated data source, as either 'landmark' or 'outline' based (see for example, Marcus *et al.*, 1996, Bookstein, 1998 or Adams *et al.*, 2004). Landmark techniques may be further separated into 'distance' methods, based on inter-landmark measurements or 'geometric/superimposition' techniques, where information on the relative locations of landmarks is always retained (which is generally not usually possible with 'distance' methods). In the current study we have the option of using either type of data. As described in section 1.4, digital imaging enables the shape of an object to be represented as a series of Cartesian coordinates of landmark positions. Alternatively, the outline of an object may be recorded as a bivariate string of x and y coordinates or as a single list of data representing the value of some function at equally spaced positions along an outline, from some arbitrary starting point.

The chapter begins with a review of the key publications and other sources of information on the subject, including details of discussion forums, websites and mailing lists. It then continues with a detailed review of the most popular methods of shape analysis, which we describe, evaluate and compare, in terms of their suitability as a starting point for our investigations. Methods included are those considered well

enough known to feature in the ongoing 'glossary of morphometrics', originally compiled by Slice *et al.* (1996) and continually updated at the State University of New York (SUNY) Stony Brook website (see section 2.1.2).

The techniques are separated into landmark (distance and geometric) and outline approaches, as described above, in sections 2.2 and 2.3 respectively. Where relevant, examples of their use with applications involving human tooth shape are noted, although for all but the 'traditional' methods of analysis, described in section 2.2.2.1, these are rare. Note that methods that are considered in later sections of this thesis are given only a brief description here, since a more detailed review will be presented in the relevant chapters.

At the end of section 2.2 we discuss the advantages and disadvantages of distance versus geometric landmark methods, noting the arguments of previous authors as well as issues likely to arise when used for the study of tooth shape. In particular, we consider reasons as to why, in recent years, Procrustes analysis (along with the other tools of the morphometric synthesis) has become the standard approach for investigation of shape when landmark data is available. In section 2.4, the benefits and drawbacks of landmark and outline techniques are then compared. Since we have both types of data available for this study, we consider which is the more useful approach to take with the current application and which techniques offer the best scope for possible development and/or modification, in order to address some of the likely issues arising with the study of tooth shape, described in section 1.5.2. We highlight how, in addition to the benefits offered over other landmark methods, the Procrustes technique also has advantages over outline methods, in terms of developed theory, variety of descriptive and inferential techniques which may be utilised and in terms of ease of interpretation of results as pictures in the original space of the objects being considered.

Following the discussion of 2.4 and identification of an appropriate place to start our investigation, the chapter ends with a brief overview of the software currently available for performing the various landmark and outline methods described.

2.1.1 Main texts

A short summary of the various approaches to shape analysis, in a historical perspective, is Bookstein's (1998) paper, titled 'A hundred years of morphometrics', which compares and describes the development of the various ideas we describe in this chapter and the extent to which they have now been standardised. Adams *et al.* (2004) also present a broad and up-to-date review of the field of morphometrics, emphasising the developments of the last 10 years and identifying future possible directions.

'Statistical shape analysis' by Dryden & Mardia (1998) is the most comprehensive recent book, dealing, in particular, with the analysis of landmark data. Other introductory texts on geometric landmark methods include Bookstein (1991) and more recently the primer by Zelditch *et al.* (2004). A comprehensive (but very technical) account of shape theory for landmark data is that of Kendall *et al.* (1999). Other recent texts dealing with the geometric analysis of inter-landmark measurements and outline data respectively are those by Lele & Richtsmeier (2001) and Lestrel (1997).

Prior to these texts, various proceedings from workshops and meetings provided most of the basic literature on shape. A series of workshops during the late 1980s to mid 1990s brought many of the methodologies to the attention of researchers. Publications from these workshops are available as books, known by their colour:

- 'Red book': Morphometrics in evolutionary biology, Bookstein *et al.* (1985).
- 'Blue book': Proceedings of the Michigan workshop in morphometrics, Rohlf & Bookstein (1990).
- 'Black book': Contributions to morphometrics, Marcus *et al.* (1993).
- 'White book': Advances in morphometrics, Marcus *et al.* (1996).

Additionally, over the past 10 years, the Leeds Annual Statistics research workshop has brought researchers together to discuss various ideas and theoretical developments in statistical shape analysis, with participants from a wide variety of fields. The edited volumes of these workshops contain many papers which have had an impact on the subject, namely:

- Current issues in statistical shape analysis, Mardia & Gill (1995)
- Image fusion and shape variability techniques, Mardia *et al.* (1996a)

Accompanying the growth in literature and the increased profile of these methods has been their application to various biological disciplines. Workshops are now offered around the world drawing in a wide variety of researchers. Examples of studies making use of shape methods can be found in all the proceedings listed above as well as more recently, a special issue of 'Acta Zoologica Academiae Scientiarum Hungaricae', Klingenburg and Bookstein (1998).

2.1.2 Other sources of information

An excellent source of information on shape analysis is the Stony Brook website at the State University of New York:

http://life.bio.sunysb.edu/morph/

Here researchers can find information on past and forthcoming meetings, workshops and courses, new and recent texts on shape analysis and details of currently active workers in the field of morphometrics. The website also has FTP links to download various software routines which are described in section 2.5. While demand clearly exists, there are as yet no mainstream off the shelf packages available that perform shape analyses, just a series of downloadable routines from various authors which still require a good level of understanding of the methods involved and of how to interpret the results. The site also contains an extensive and continually updated 'glossary for morphometrics', maintained by F.J. Rohlf, F.L. Bookstein and D.E. Slice. The original version (Slice *et al.*, 1996) can be found in the 'white book', 'Advances in morphometrics' by Marcus *et al.* (1996) mentioned above.

There is also an active mailing list/discussion group MORPHMET, moderated by D. E. Slice (formerly F. J. Rohlf and L. Marcus) and accessible at the address:

http://morphometrics.org/morphmet.html

In addition, an extensive bibliography on morphometrics, frequently updated by J. Lynch, is available at:

http://www.public.asu.edu/~jmlynch-geometric-index.html.

2.2 Landmark methods

2.2.1 Preliminary definitions and notation

The shape of an object may be represented by a number of points called 'landmarks'. As described in section 1.2.5, these are key points of correspondence, selected to be informative about particular characteristics of interest and are defined in the same way on each object of a given type, so that they match both between and within populations. Examples of 'anatomical' and 'mathematical' buccal and occlusal tooth surface landmarks were presented in fig. 1.3.

For most methods, the configuration of landmarks of an object X are represented as a $k \times m$ matrix of coordinates where k is the number of landmarks and m the dimensionality of the space within which the objects were digitised. For example, in m=2 dimensions:

$$X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{k1} & x_{k2} \end{pmatrix} \text{ or } \begin{pmatrix} x_{1x} & x_{1y} \\ x_{2x} & x_{2y} \\ \vdots & \vdots \\ x_{kx} & x_{ky} \end{pmatrix}. \qquad (2.1)$$

The order of the landmarks is arbitrary, but this order must correspond between the objects being compared.

Occasionally, a configuration is denoted in vectorised form, as $vec(X)$. For a $k \times m$ matrix X, the *vec* operator stacks the m columns of X (in left to right order) to give a vector of length km. For example in m=2 dimensions, $vec(X) = (x_{1x}, \ldots, x_{kx}, x_{1y}, \ldots, x_{ky})^T$.

The reverse operation, vec_m^{-1}, is used for converting a vector of length km into a matrix of m columns and k rows, with $vec_m^{-1}(vec(X))=X$.

Before considering methods for shape, we also define what is meant by the size of an object for landmark data. The choice of size measure used is important since it will usually affect the reported conclusions of any analysis and so much attention has been given to what the most useful measure of size actually is. The most popular, and now almost standard choice, is the 'centroid size', calculated as the square root of the sum of squared distances of each landmark from the configuration's centre. For configuration X, this corresponds to:

$$\|X\|=\|CX\|=\sqrt{\sum_{d=1}^{m}\sum_{j=1}^{k}(x_{jd}-\frac{1}{k}\sum_{j=1}^{k}x_{jd})^2} \qquad (2.2)$$

where C is the centring matrix $I_k-\frac{1}{k}1_k1_k^T$, with I_k the $k\times k$ identity matrix, 1_k a $k\times 1$ vector of ones and Euclidean norm $\|A\|=\sqrt{trace(A^TA)}$. The centre of the configuration is given by $\bar{x}=(\frac{1}{k}\sum_{i=1}^{k}x_{i1},...,\frac{1}{k}\sum_{i=1}^{k}x_{im})^T$. CX translates X by $-\bar{x}$ so that the centre of CX is at the origin. So, for example, for m=2, the centroid size of configuration X is given by:

$$\sqrt{\sum_{j=1}^{k}(x_{jx}-\frac{1}{k}\sum_{j=1}^{k}x_{jx})^2+\sum_{j=1}^{k}(x_{iy}-\frac{1}{k}\sum_{j=1}^{k}x_{jy})^2}\ .$$

This measure is used (particularly in geometric methods) because is not linearly dependent on any of the 'shape variables' on which subsequent analysis is based. See for example, Bookstein (1986), Kendall (1984) or Goodall, (1991). It could also be used in a normalised form (e.g. by dividing by $\sqrt{k}$), which would be particularly appropriate when comparing configurations with a different number of landmarks (Dryden & Mardia, 1998). For Procrustes methods, scaling by centroid size is especially important since this results in configurations that can be considered as points in Kendall's shape space (see section 2.2.3.2).

2.2.2 Distance based methods

The earliest methods of shape analysis to appear were multivariate morphometrics (or 'traditional morphometrics'), which are still used in many research investigations today. Measurements are typically multiple inter landmark distances from which basic shape features can be calculated, ignoring the geometry of the objects involved, e.g. which distances share end-points. More recent 'distance' based methods have used manipulations of matrices of measurements between all pair of landmarks (Euclidean distance matrix analysis).

2.2.2.1 Traditional (multivariate) morphometrics

Traditional methods of shape analysis typically involve measuring distances between landmarks (lengths or widths) and then calculating angles or ratios of these distances. Ratios and angles are used since they are invariant to differences in the location, size and the rotation of objects. Standard multivariate techniques are then applied to these variables, e.g. t-tests, ANOVA, and MANOVA to investigate differences in shape.

A general review of 'multivariate morphometrics' is given by Reyment *et al.* (1984). Despite the growing popularity of the newer 'geometric methods' (see section 2.2.3), traditional morphometrics is still widely and successfully used today and examples can be found in many different fields.

Investigations using 'traditional' morphometric methods are abundant in the Dental literature. In earlier studies authors used calliper measurements to obtain distances between subjectively identified positions. See for example, Peck & Peck (1972) and Wood & Engleman (1988) who used the ratio of MD and BL measurements as a measure of crown shape. Today, image analysis has aided the process of taking such measurements and allowed new, more detailed features to be quantified and investigated. For example, Peretz *et al.* (1998) used digital imaging to measure the angles between inter-cusp measurements on molar teeth. More recently, Shah *et al.* (2003) utilised the imaging system used in the current study, to obtain measurements from of casts of incisors sectioned at the midpoint of the LACC, to calculate a ratio of mesio-distal width measurements, in order to describe tooth shape.

Having collected a series of measurements on each object, an alternative approach to using ratios and angles is to use principal components analysis on the length and width measurements and then interpret the resulting components as measures of different aspects of size or shape. For example, Khalaf *et al.* (2001) use this approach to quantify and analyse shape characteristics of teeth in patients with supernumerary teeth.

In most dental and other biological applications, summaries of the findings are usually presented as a table of means and standard deviations of the measurements or derived variables involved. However, pictorial representation of variation in shape is often *ad hoc* and usually limited to simply displaying series of scatter plots of pairs of the different measurements.

This lack of a simple means by which to visualise resulting shapes is a commonly cited drawback of these methods. While distances may be constructed from landmark coordinates, the reverse is generally not true unless as extensive series of measurements is recorded that includes the same landmarks as endpoints of several distances (Marcus & Corti, 1996). Consequently, visual representation of shape variation and shape differences is not generally straightforward. Marcus (1990) reviews the numerous diagrammatic displays attempted from traditional morphometric methods. A recent attempt to address this difficulty is given by Carpenter *et al.* (1996). They detail procedures to transform distances to landmarks using iterative multidimensional scaling methods (see section 2.2.2.3) and introduce otherwise redundant measurements, such as the 'truss protocol' to allow a reasonably accurate conversion of inter-landmark distance data into coordinate data, which may then be plotted and visualised.

2.2.2.2 Allometry

Introduced by Huxley (1924, 1932), allometry is a technique using inter-landmark distances for the study of size and shape. Differences or changes in shape are often associated with size, for example, smaller potatoes are more spherical (Glaseby *et al.*, 1988). It is often found that the sizes of different parts of organisms are linearly related on a logarithmic scale. For example, length (L) and width (W) may obey the relationship:

$$LogL = \alpha + \beta \log W \text{ or equivalently } L = e^{\alpha} W^{\beta}.$$

Keiser (1990) devotes an entire chapter of his book on adult odontometrics to the study of allometric relations of tooth size, although emphasis is placed on investigating regression relationships with the size or mass of other parts of the body, such as height or the length of bones, rather than between measurements on the same tooth. Mosimann (1970) gives a rigorous treatment of the subject of allometry and provides theorems for the independence of population size and shape.

However, as we note in section 2.2.4, considering just distance and angular measurements can often be inferior to methods where the geometry of an object is retained. One reason for this is that interpretation of the important linear combinations of measurements of the form above can often be difficult, particularly when more than just two measurements are involved.

2.2.2.3 Euclidean distance matrix analysis (EDMA)

Another method utilising inter-landmark distances is EDMA, first proposed by Lele & Richtsmeier (1991). EDMA involves computation of matrices of all inter-landmark distances for each configuration. These are called 'form' matrices (form being size and shape information) and contain Euclidean distances between every possible pair of the k landmarks (k landmarks generate $n(n\text{-}1)/2$ distances). For a sample of n form matrices it is possible to produce an estimate of mean size and shape using multi-dimensional scaling. However, this leads to biased estimates, so instead Lele (1993) suggested using a corrected method of moments. The ideas rely on those used in multidimensional scaling, as we explain below.

2.2.2.3.1 Multidimensional scaling and EDMA

For a $k\times m$ configuration X, consider the $k\times k$ squared Euclidean distance matrix $D(X)$, with entry r,s ($r,s=1,\ldots,k$) given by the squared Euclidean distance between points r and s. For example, for $m=2$, $D(X)$ has entries:

$$D(X)_{rs}=(x_{rx}-x_{sx})^2+(x_{ry}-x_{sy})^2.$$

Let W be the (component-wise) average squared Euclidean distance matrix over n configurations in a sample. Using multidimensional scaling (MDS) it is then possible to construct a configuration of k points in Euclidean space corresponding to the information in W (or any Euclidean distance matrix D) which also has squared Euclidean matrix approximately equal to W (or D). See for example Mardia *et al.*, 1979. W can therefore be used to provide an estimate of the mean size and shape configuration by calculating:

$$\hat{\mu}_{MDS(W)} = MDS(W) = [f_1, \ldots, f_m] \tag{2.3}$$

where $f_1, \ldots, f_m$ are the first m eigenvectors (of length k) of $-\frac{1}{2}CWC$, scaled so that $f_l^T f_l = a_l$, where the a_l are the corresponding eigenvalues. C is the $k \times k$ centring matrix, as defined in section 2.2.1. The approximation is reasonable providing the first m eigenvalues are large compared with the rest (Dryden & Mardia, 1998). Any rotation (and/or reflection) of $\hat{\mu}_{MDS(W)_m}$ will also be an equivalent solution and so this is often called the estimated mean reflection size and shape. Of course, it is then possible to plot $\hat{\mu}_{MDS(W)_m}$ for visualisation.

However, the classical MDS solution leads to a biased estimate of mean size and shape under normal errors, so EDMA corrects for this by estimating population distances using a method of moments, under certain modelling assumptions.

Let $F(X)$ be the $k \times k$ 'form distance matrix' of configuration X, with entries $F(X)_{rs}$ ($r,s=1,\ldots,k$) being the Euclidean distances between landmarks r and s. e.g. for $m=2$:

$$F(X)_{rs} = \sqrt{(x_{rx} - x_{sx})^2 + (x_{ry} - x_{sy})^2}\,. \tag{2.4}$$

For a sample of configurations, the idea is then to obtain an estimate of the population form distance matrix $F(\mu)$. Assuming, for $m=2$, that the coordinates of the jth landmark (t_{jx}, t_{jy}) in T are independently, multivariate-normally distributed, with mean (μ_{jx}, μ_{jy}) and covariance $\sigma^2 I_2$, $j=1,\ldots,k$, an unbiased moment estimate of the population squared Euclidean distance between landmarks r and s,

$$\delta_{rs}=(\mu_{rx}-\mu_{sx})^2+(\mu_{ry}-\mu_{sy})^2, \tag{2.5}$$

is given by:

$$\hat{\delta}_{rs}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}d_{rs_i}^2-\frac{1}{n}\left\{\sum_{i=1}^{n}(d_{rs_i}^2-\frac{1}{n}\sum_{i=1}^{n}d_{rs_i}^2)^2\right\}}$$

where the $d_{rs_i}^2$, i=1,…,n are a random sample of squared distances of the form of (2.5) between landmarks r and s in each of the n configurations. It is then easy to construct a $k\times k$ matrix δ with elements $(\delta)_{rs}=\hat{\delta}_{rs}$, representing the population form matrix and by (2.3), to obtain an estimate of mean (reflection) size and shape:

$$\hat{\mu}_{MDS(\delta)}=MDS(\delta)\ .$$

If variations are small then the EDMA reconstructed mean shape of a sample will be very similar to that of MDS, which in turn is very similar the full Procrustes mean in section 2.2.3.2.1 (Dryden & Mardia, 1998). Note that as with any size and shape study, all objects have to be commensurate in scale.

2.2.2.3.2 Examining differences in shape

A statistical test to compare mean shapes in two independent populations has been developed by Lele & Richtsmeier (1991).

To compare two shapes (which may be two means or two individuals), a form difference matrix comprising element-wise ratios of the corresponding distances in the two form matrices could be computed. For example for two shapes X and Y, we could calculate:

$$\frac{F_{rs}(X)}{F_{rs}(Y)}.$$

If the ratio of all the distances is constant, then the forms can be said to have the same shape, the common ratio being the difference in size. The amount of shape difference could be given by the ratio of the largest and smallest of the elements of the form

difference matrix. Unfortunately, one must then examine long lists of distance ratios to interpret how exactly the shapes differ.

Instead, in order to test for a difference in mean reflection size and shape between two samples, Lele & Richtsmeier (1991) propose the statistic:

$$\max_{r,s} \frac{F_{rs}(\hat{\mu}_1)}{F_{rs}(\hat{\mu}_2)} \Big/ \min_{r,s} \frac{F_{rs}(\hat{\mu}_1)}{F_{rs}(\hat{\mu}_2)}$$

where $\hat{\mu}_1$ and $\hat{\mu}_2$ are the EDMA estimators of mean size and shape in each sample. Bootstrap procedures must be used to estimate the null distribution of the test statistic. A more powerful, although less frequently used test, EDMA-II, was also later proposed by Lele & Cole (1995).

For a more comprehensive description and discussion of EDMA, see the recent text by Lele & Richtsmeier (2001).

2.2.2.3.3 A related approach using logarithms of distances

A related approach to shape analysis based on inter-landmark distances is that of Rao & Suryawanshi (1996), who proposed comparing samples of shapes using matrices of average log distances and again multidimensional scaling. See also Dryden & Mardia, (1998) for a brief description.

2.2.3 Superimposition (geometric) methods

Over time, analysis of linear distances gave way to analyses based directly on landmark configurations (or the entire outlines of objects, as described in section 2.3), so that instead of working with a few quantities derived from objects such as angles, distances and ratios, one works with the complete geometry of them.

In the last two decades there have been many key developments in the statistical analysis of shape, in particular in terms of geometric methods for landmark data. As recently as the late 1980s/early 1990s several 'competing' approaches existed, one of

which was the calculation of 'two-point shape coordinates' by edge superimposition (see section 2.2.3.1 below). Since then a standard set of procedures for analysing shape has emerged, based around the ideas of Procrustes superimposition and Kendall's shape space and now referred to by Bookstein (1998) as the 'morphometric synthesis'.

In any study of shape, a particular difficulty is that Euclidean methods of analysis cannot be used to analyse differences in the locations of points on objects directly, because registration differences due to size, location and orientation are still present. Superimposition methods eliminate this non-shape variation from the configurations of landmarks by overlaying them according to different protocols and optimisation criteria. The shape of a configuration may then be represented as the resulting coordinates of the points after the effects of these 'similarity transformations' are removed and differences in shape described by differences in the coordinates of corresponding landmarks. It is then possible to compare samples using standard multivariate inferential techniques and to produce graphical displays to visualise variations in shape.

For a comprehensive survey of early uses of superimposition methods to study differences in shapes, see Cole (1996).

2.2.3.1 Edge superimposition (Bookstein coordinates)

For planar (m=2) data, Bookstein (1984, 1986) suggested removing the similarity transformations by translating, rotating and re-scaling the configurations so that two chosen landmarks (say landmarks 1 and 2) are sent to a fixed position or baseline, e.g. (-1/2,0) and (1/2,0). The coordinates of the remaining k-2 landmarks after these transformations may then be used to represent variation in shape.

For landmarks j=1,…,k the new coordinates of each configuration X are given by:

$$\begin{pmatrix} -0.5 & 0 \\ 0.5 & 0 \\ \vdots & \vdots \\ \dfrac{(x_{21}-x_{11})(x_{j1}-x_{11})+(x_{22}-x_{12})(x_{j2}-x_{12})}{(x_{21}-x_{11})^2+(x_{22}-x_{12})-0.5} & \dfrac{(x_{21}-x_{11})(x_{j2}-x_{12})-(x_{22}-x_{12})(x_{j1}-x_{11})}{(x_{21}-x_{11})^2+(x_{22}-x_{12})} \\ \vdots & \vdots \end{pmatrix}.$$

The choice of baseline is in fact arbitrary and below we present formulae to obtain edge superimposition coordinates when landmarks 1 and 2 may be sent to any specified baseline with coordinates $p_1=(p_{1x}, p_{1y})$ and $p_2=(p_{2x}, p_{2y})$. The transformed configuration is calculated as:

$$X^B = h(X - 1_k b)\Gamma + 1_k p \tag{2.6}$$

where 1_k is a vector of 1's of length k and

$$b=\begin{pmatrix} \frac{x_{1x}+x_{2x}}{2} & \frac{x_{1y}+x_{2y}}{2} \end{pmatrix}^T, p=\begin{pmatrix} \frac{p_{1x}+p_{2x}}{2} & \frac{p_{1y}+p_{2y}}{2} \end{pmatrix}^T, \Gamma=\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix},$$

$$h=\frac{\sqrt{(p_{2x}-p_{1x})^2+(p_{2y}-p_{1y})^2}}{\sqrt{(x_{2x}-x_{1x})^2+(x_{2y}-x_{1y})^2}}, \theta=\tan^{-1}\left(-\frac{(p_{1x}-p_{2x})(x_{1y}-x_{2y})-(x_{1x}-x_{2x})(p_{1y}-p_{2y})}{(x_{1x}-x_{2x})(p_{1x}-p_{2x})-(x_{1y}-x_{2y})(p_{1y}-p_{2y})}\right).$$

A benefit of using Bookstein or 'two-point' shape coordinates is that the graphical displays and coordinates are easy to interpret. Fig 1.1 shows a plot of the Bookstein coordinates of a small sample of upper central incisors, with landmarks defined as in 1.2.5, aligned so that the mesio-distal endpoints are sent to positions (-0.5,0) and (0.5,0). Extensions to three dimensions have also been developed (see e.g. Dryden & Mardia, 1998).

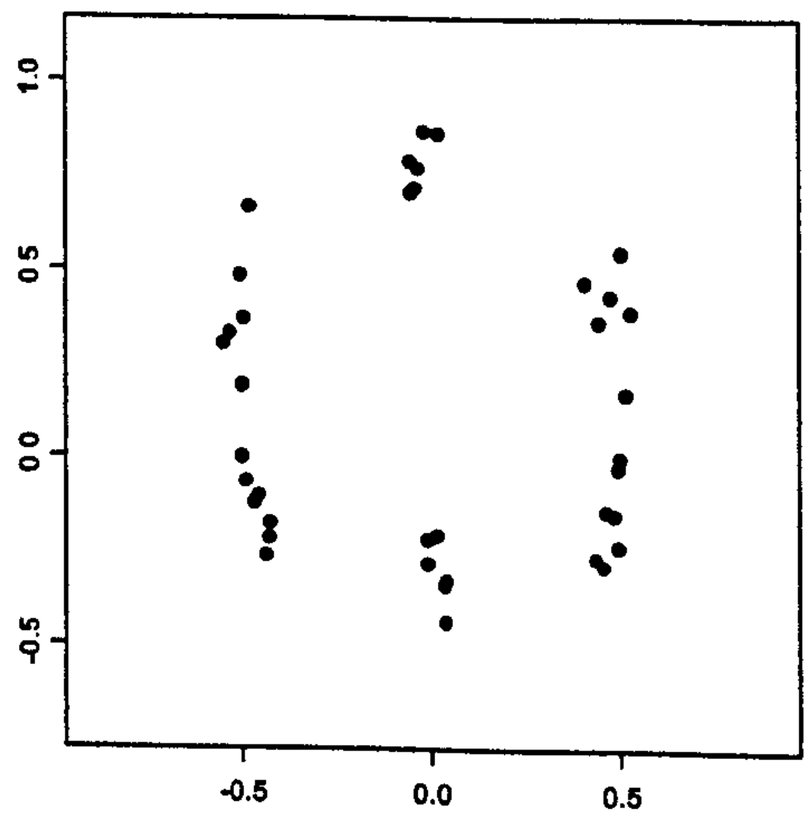


Fig 1.1: Bookstein coordinates of upper central incisor configurations with MD endpoints registered to baseline (-0.5,0) and (0.5,0). Gingival landmarks at top, incisal edge landmarks at bottom.

One approach to analysis is to use standard multivariate methods on the k-2 Bookstein coordinates directly, ignoring the 'non-Euclidean' nature of the shape space. Euclidean

distances between configurations in some regions of the space may be larger/smaller when the apparent difference in shape is smaller/larger than in other regions. For example, for triangles, registered to a common baseline, e.g. (-0.5,0) and (0.5,0), a Euclidean distance of 1 between configurations whose third vertices are nearer to the origin results from a larger difference in shape than when configurations have third vertices further away from the origin. However, providing variations in the data are small and the points chosen as the baseline are not close together, the method is adequate for mean estimation and hypothesis testing, using (say) Hotelling's T^2 tests on the k-2 non-fixed coordinates (Dryden & Mardia, 1998).

The main disadvantage, however, is that since the configurations are registered to a common edge, this induces correlations into the k-2 other coordinates, leading to spurious relationships, and so the method should not be used to interpret the structure of shape variability. In particular, examination of the sample covariance matrix (by e.g. PCA) can be very misleading and should not be used. An artefact of using this type of coordinate system is that variability in points away from the baseline often appears larger than at the nearer points.

The approach is also often criticised because unlike the Procrustes methods, it does not summarise covariation across all landmarks. In addition, different choices of baseline can often produce very different results. For these reasons the use of Procrustes superimposition is now nearly always preferred.

2.2.3.2 Procrustes analysis and the morphometric synthesis

During the last two decades, a rigorous theory for shape analysis had been developed, that makes possible the combined use of multivariate statistical techniques and methods for the direct visualisation of shape. By way of the pioneering articles of Kendall (1984) and Bookstein (1986), the two major advances have been the formulation of Kendall's space (with Procrustes metric measuring dissimilarities in shape) and the adaptation of Procrustes analysis to the analysis of shape (Goodall, 1991).

These articles then provided the theoretical setting for the development of inferential and descriptive statistical methods that followed, where shapes are defined as

equivalence classes of landmark sets under the operations of the 'similarity group' of transformations described in the introduction to 2.2.3, comprising rotation, translations and scalings. These included the development of specific distributional methods of shape analysis (starting with Mardia & Dryden, 1989a,b), and also Kent (1994), who introduced the Procrustes tangent space, enabling conventional methods of statistical analysis to be used to for descriptive and inferential investigations of variation in shape.

2.2.3.2.1 Ordinary and generalised Procrustes analysis

Consider two k landmark configurations in m dimensions, expressed as $k \times m$ matrices X and Y as in equation (2.1). We wish to compare the configurations and obtain a measure of dissimilarity or distance between the shapes.

Ordinary Procrustes superimposition of (say) configuration X to configuration Y involves matching X to Y as closely as possible, by least squares, over the similarity transformations of scale, location and rotation. For this we seek to minimise the sum of squared Euclidean distances between the landmarks of Y and corresponding points of $\beta X\Gamma + 1_k \gamma^T$, i.e. we seek

$$\min_{\gamma,\beta,\Gamma} \left\| Y - (\beta X\Gamma + 1_k \gamma^T) \right\|^2, \tag{2.7}$$

where γ is a translation vector of length m, β a scale parameter >0 and $\Gamma(\theta)$, an $m \times m$ special orthogonal rotation matrix. The minimised value of (2.7) and corresponding parameters estimates $\hat{\gamma}$, $\hat{\beta}$ and $\Gamma(\hat{\theta})$ are obtained by first translating the centre of each configuration to (0,0), and then rotating (and resizing) X so that the sum of squared distances between corresponding landmarks is minimised. $\hat{\beta} X \hat{\Gamma} + 1_k \hat{\gamma}^T$ is the full Procrustes fit of X to Y. Details of how to obtain $\hat{\gamma}$, $\hat{\beta}$ and $\Gamma(\hat{\theta})$ algebraically are given in an expanded version of this description in section 3.2.1.

However, note that if we had instead superimposed Y to X then:

$$\min_{\gamma,\beta,\Gamma} \left\| Y - (\beta X\Gamma + 1_k \gamma^T) \right\|^2 \neq \min_{\gamma,\beta,\Gamma} \left\| X - (\beta Y\Gamma + 1_k \gamma^T) \right\|^2.$$

However, a symmetric measure of dissimilarity in shape can be easily obtained if both configurations are first scaled to unit centroid size. We then have:

$$\min_{\gamma,\beta,\Gamma}\left\|\frac{Y}{\|Y\|}-(\beta\frac{X}{\|X\|}\Gamma+1_k\gamma^T)\right\|^2=\min_{\gamma,\beta,\Gamma}\left\|\frac{X}{\|X\|}-(\beta\frac{Y}{\|Y\|}\Gamma+1_k\gamma^T)\right\|^2$$

$=d_F^2(X,Y)$, the squared full Procrustes distance between X and Y.

For a sample of configurations X_i, i=1,...,n, the full Procrustes estimate of mean shape $\hat{\mu}$ is obtained as the shape which has least sum of squared full Procrustes distances to each configuration in the sample. i.e.

$$\hat{\mu} = \arg\min \sum_{i=1}^{n} d_F^2(X_i,\mu)= \arg \min_{\gamma_i,\beta_i,\Gamma_i} \sum_{i=1}^{n}\left\|\hat{\mu}-(\beta_i X\Gamma_i+1_k\gamma_i^T)\right\|^2, \qquad (2.8)$$

subject to some constraint on the size of $\hat{\mu}$ or of the $\beta_i X\Gamma_i+1_k\gamma_i^T$ and choice of an arbitrary orientation for $\hat{\mu}$. The 'full Procrustes registered fits' are then given by $X_i^P=\hat{\beta}_i X_i\hat{\Gamma}_i+1_k\hat{\gamma}_i^T$ where $\hat{\beta}_i,\hat{\Gamma}_i$ and $\hat{\gamma}_i^T$ are the similarity transformations minimising (2.8) above. Several different methods have been proposed for obtaining $\hat{\mu}$ and the X_i^P. For m=2 an explicit solution is available using complex vector notation (see for example, Bookstein, 1991 or Kent, 1994). For $m\geq 2$ however, the iterative method of Gower (1975) and Ten Berge (1977), also described by Rohlf & Slice (1990) and Goodall (1991), must be used. Full descriptions are given in section 3.2.2.

The procedure is known as Generalised Procrustes analysis (GPA) and in removing the variation in location, rotation and size, the n configurations objects are now registered to a common (though arbitrary) coordinate system. The sample variation in shape may then be visualised by plotting the Procrustes fits, as for example in fig 2.2 below, which shows the fits of the same sample of upper central incisors used in fig. 2.1.

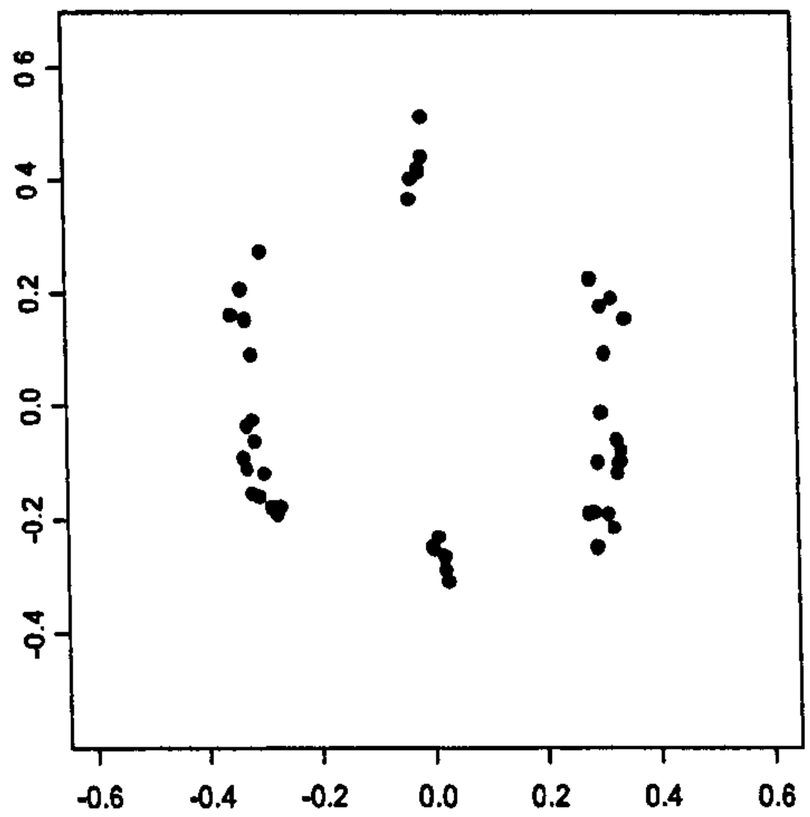


Fig 2.2: Procrustes fits of a small sample of upper central incisor configurations. Gingival landmarks at top, incisal edge landmarks at bottom.

2.2.3.2.2 Tangent space analysis

After removing the unwanted registration differences using GPA, the landmark configurations can be regarded as points in Kendall's shape space (Kendall, 1984). Each point in this space represents the shape of a configuration of points irrespective of size, location and orientation (points correspond to entire landmark configurations, not just single landmarks). Underlying this space is the Procrustes metric ρ for measuring dissimilarity between points (shapes). Minimisation of distances between points in shape space corresponds directly with the minimisation of sums of squared Euclidean distances in configuration space using $d_F^2(X,Y)$, as described above, with $d_F(X,Y) \approx \rho(X,Y)$. Consequently, providing variation in shape is small, it can be shown (see section 3.3.2) that the Procrustes fits approximate coordinates in a tangent space to Kendall's shape space; a linearised version of shape space in the vicinity of the Procrustes mean which has Euclidean metric approximating Procrustes distances ρ in the shape space. This result forms the fundamental basis for further analytical investigations, allowing most standard descriptive and inferential techniques (which rely on the Euclidean metric) to be adapted and applied successfully.

For instance, sample shape variation may be investigated by principal components analysis of the Procrustes fits and visualised by displaying hypothetical shapes at different extremities of variation. Scores for each configuration on each component may be used to investigate associations of certain patterns of variation in shape with other covariates, e.g. centroid size (see section 3.4.2.2).

As Adams *et al.* (2004) note, today nearly all landmark based morphometric studies analyse shape with procedures based on Kendall's shape space, Procrustes distances or their tangent space approximations. However, despite this rise in popularity, the profile of the Procrustes technique and its related methods is still relatively low in the mainstream scientific literature and almost non-existent for studies of tooth shape.

2.2.3.3 Post superimposition techniques

Following superimposition, transformation techniques can be used to map one configuration to another, so that differences in shape can then be visually described. For example, mappings may be made from a mean or reference configuration to an individual object or between two means of different samples. We describe some of these methods below.

2.2.3.3.1 Thin plate splines and deformation grids

Visualisation of differences in shape may be obtained using a pair of thin plate spline (PTPS) transformations (Bookstein, 1991). Here these are used to represent shape differences between two configurations in a manner analogous to D. W. Thompson's (1917) deformation grids. The technique is described in detail in chapter 5 and so we give only a brief description of the ideas involved here.

Imagine one configuration drawn on a piece of squared graph paper, the other on plain paper. If we 'deform' the graph paper so that the corresponding landmarks of the two configurations can be placed directly over each other, the resulting deformed grid tells us where and how the configurations differ.

A pair of thin plate splines is used to describe the mapping of the 2D space in which a set of landmarks lie to designated locations in another by minimising the 'bending energy' (integral quadratic variation) over the entire picture plane, constrained so that corresponding landmarks match exactly. The square grid referred to above thereby 'bends' as little as possible and so the deformation is 'optimal' in this sense. An illustration is given in fig. 2.3, which shows the difference in shape between a normal upper central incisor, from a patient with a full number of teeth and the same tooth from

a patient with hypodontia. Landmarks correspond to those described in section 1.2.5, with the exception of the mesio-distal endpoints which were omitted due to the difficulty in identifying these positions in the hypodontia population. Both configurations have been centred and scaled to unit size and the hypodontia incisor matched to the normal incisor by Procrustes superimposition.

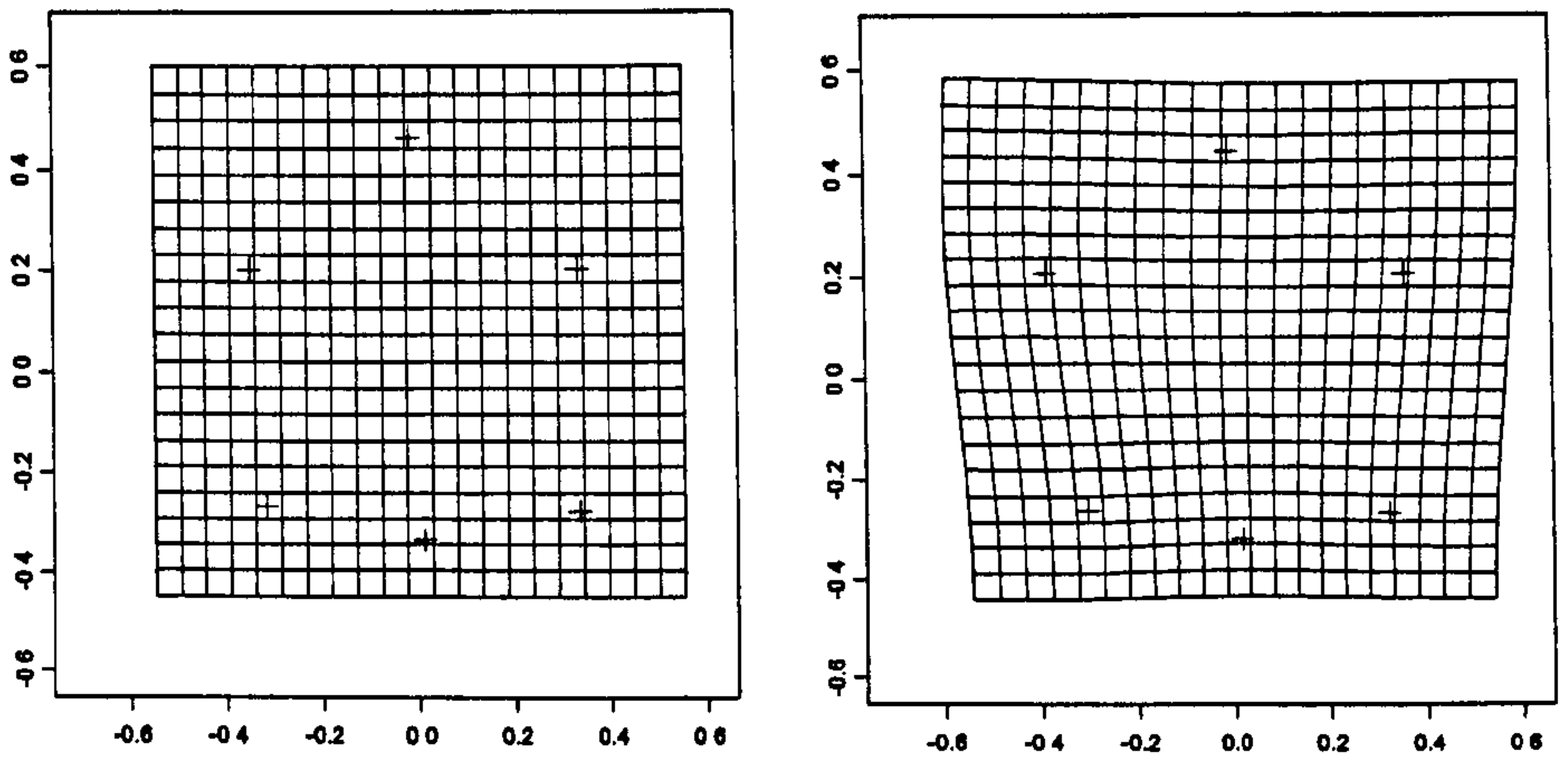


Fig 2.3: (Left) Square grid superimposed onto normal central incisor configuration. (Right) Square grid deformed to lie on hypodontia incisor.

Parameters describing the deformations ('partial' or 'relative' warp scores) of the estimated mean shape to each configuration in a sample, can be used as shape variables for subsequent statistical comparisons, in a manner similar to that described for principal components above.

2.2.3.3.2 Relative warps analysis

For relative warp analysis (Bookstein 1991), the parameters of the PTPS transformation from the mean $\hat{\mu}$ to the Procrustes fit of each configuration in a sample, can be used as alternative variables in a principal components analysis (PCA) (or any multivariate method) and a special scaling parameter used to emphasise small or larger scale deformations from the mean shape. Rather than using the covariance matrix of the vectorised Procrustes fits (with Euclidean metric), PCA is performed with respect to the power of the 'bending energy (or inverse bending energy) matrix' of $\hat{\mu}$ (with respect to the 'bending energy' metric from $\hat{\mu}$ to each X_i^P). Further details of can be found in section 5.3.6.3.

2.2.3.3.3 Finite element scaling

An alternative method for pairs of objects is finite element scaling. In this technique, the set of landmarks on each configuration serve as vertices of r small closed regions, usually triangles in two dimensions. These are the finite elements. Two configurations X and Y may be compared by computing a separate 'affine' transformation, of the form below, between each pair of elements, Y_{FE_r} and X_{FE_r} with:

$$Y_{FE_r} = X_{FE_r} A_r + 1_k c_r^T + E_r .$$

The affine transformation of X_{FE_r} is given by multiplication by a $m \times m$ matrix A_r and translation by a vector c_r, of length m. If we were to consider points on square grid, such an affine transformation would deform the grid uniformly into a parallelogram grid. E_r is a matrix of residual errors, added so that each transformation maps the vertices of each small region in one configuration to their corresponding positions in the other. (For triangular elements in two dimensions the transformation is necessarily affine and maps vertices exactly with E_r=0).

The shape differences are usually displayed in terms of strained crosses inside each element (triangle), representing the principal (major and minor) axes of the deformation of a unit circle inscribed in each element, which is transformed into an ellipse by the affine mapping. These are then compared in terms of their magnitudes and directions.

The estimated strains are said to represent hypothetical forces that deform one object to another. The technique can also be used with non-homogenous elements, e.g. elements not necessarily all triangles, see for example, Cheverud & Richtsmeir (1986).

One problem however, is that the selection of landmarks and finite elements to be used is arbitrary. A different partitioning of the configuration into different elements can often produce very different results and can change the whole perception of the shape difference between two objects (see Zienkiewicz, 1971 for more details). Bookstein (1991) also criticises the use of the technique on this basis.

2.2.3.4 Resistant superimposition

Least squares matching of configurations can be greatly influenced by landmarks that are unusually located on different objects. Siegel & Benson (1982) made the observation that if the two objects being compared are identical except for the position of a few landmarks, a global ordinary least squares criterion (as in (2.7)) usually results in a general lack of fit at most points. This can make the interpretation of Procrustes fits misleading, masking the nature of actual differences between configurations. They instead proposed a resistant fit method for matching X to Y over the similarity transformations, that is better able to reveal such localised differences.

Here we denote the superimposed version of X to Y as:

$$X^{RF} = \beta_R X\Gamma(\theta_R) - 1_k \gamma_R^T . \tag{2.9}$$

However, unlike with Procrustes superimposition, there is no quantity to be optimised between Y and X^{RF}. Instead, estimation of the scale β_R, rotation angle θ_R and translation vector γ_R involves sequential updating by repeated median estimators.

Following an initial approximate alignment of configuration X to Y to facilitate the estimation (Rohlf & Slice (1990) suggest using the Procrustes fit), the scale factor is computed as the repeated median of the ratios of corresponding distances between each pair of landmarks j and l in Y and X;

$$\beta_R = \underset{j}{median}\left(\underset{l \neq j}{median} \sqrt{\frac{(y_{lx} - y_{jx})^2 + (y_{ly} - y_{jy})^2}{(x_{lx} - x_{jx})^2 + (x_{ly} - x_{jy})^2}} \right). \tag{2.10}$$

A similar procedure is then used to estimate the rotation angle θ_R, by evaluating the median angle between pairs of vectors connecting pairs of points;

$$\theta_R = \underset{j}{median}\left(\underset{l \neq j}{median}(\theta_{jl}) \right) \tag{2.11}$$

where θ_{jl} is the rotation required to point a vector connecting points j and l in Y in the same direction as the corresponding vector in X. Siegel & Benson (1982) recommend using a range of $-\pi$ to π for θ_{jl}, since one would expect their average to be close to zero following the initial fitting (this then saves having to consider the fact that $2\pi-\varepsilon$ is the same as $-\varepsilon$ in the calculation of the median angle and ensures it is properly defined). This can be achieved, for example, by converting each pair of vectors to complex numbers and following the details of section 3.2.2.3. γ_R is then computed using ordinary medians, with:

$$\gamma_R = \underset{j}{median}(Y - \beta_R X\Gamma(\theta_R)). \tag{2.12}$$

Siegel and Benson (1982) showed that the procedure has a 'breakdown value' of almost 50%, this being the number of points that can be perturbed before the fit of the unperturbed points is affected. In other words, whenever $(k+1)/2$ of the points can be made to fit closely, the resistant fit method will obtain the correct solution. By contrast the least squares solution can be affected by a change in just a single point. It is therefore hoped that the excessively larger or smaller, scale and shift parameters associated with outlying landmarks are ignored and the fit allocates most of the variation to the outlying landmarks.

To extend these ideas for the matching of multiple objects (and obtaining an average shape, $\hat{\mu}_R$), Rohlf & Slice (1990) modified Gower's (1975) iterative approach to generalised Procrustes analysis to use resistant fitting. Firstly, each configuration X_i is centred and scaled to a common size, so that for each X_i the median squared inter-landmark distance equals unity, i.e.

$$\underset{j\neq l}{median}((x_{jx} - x_{lx})^2 + (x_{jy} - x_{ly})^2) = 1,\ j, l = 1,\ldots,k. \tag{2.13}$$

An initial estimate of the mean to use in the iterations is obtained as follows.

(i) Firstly choose $\hat{\mu}_R$ as one of the scaled centre configurations.

(ii) Rotate each X_i to $\hat{\mu}_R$ using least squares rotation.

(iii) Compute a new $\hat{\mu}_R$ as the median of the rotated X_i.

(iv) For convenience in testing convergence, scale $\hat{\mu}_R$ so that (2.13) holds.

An average shape and corresponding set of registered 'fits' are then obtained using the following iteration steps.

1. Rotate each current X_i to fit $\hat{\mu}_R$ using the resistant fit procedure (equations (2.9) to (2.12)), to obtain the X_i^{RF}.
2. Compute new mean $\hat{\mu}_{R_{new}}$ as the arithmetic mean of the positions in the X_i^{RF}.
3. Test convergence by examining the median difference between $\hat{\mu}_{R_{new}}$ and $\hat{\mu}_R$. (Rohlf & Slice (1990) suggest this is more in line with the goals of the robust procedure than testing convergence using least squares).
4. If the median absolute difference between $\hat{\mu}_{R_{new}}$ and $\hat{\mu}_R$ is not less than some pre-specified tolerance, then rescale the new $\hat{\mu}_{R_{new}}$ to have unit squared inter-landmark distance, set $\hat{\mu}_R = \hat{\mu}_{R_{new}}$ and $X_i = X_i^{RF}$ and repeat from 1.

Bookstein & Sampson (1990) showed how one can use the technique to detect the effect of outliers by matching objects with various combinations of landmarks left out of the fitting process. Having identified outliers as those with large residuals, one course of action suggested is then to ignore a suspect point(s) and proceed with a conventional, more efficient analysis (e.g. Procrustes least squares) on the rest of the data.

An alternative to the repeated median method using 'least median of squares (LMS) estimators', was introduced by Rousseeuw (1984) who noted that the residual discrepancy measures from Siegel & Benson's (1982) method were still somewhat dependent on the original registration of the two objects. If E denotes the $k \times m$ matrix of residuals of the fit of $X^{RF} = \beta_R X\Gamma(\theta_R) - 1_k \gamma_R^T$ to Y, with jth row entries (e_{jx}, e_{jy}), $j=1,\ldots,k$, then estimates of the parameters, β_R, θ_R and γ_R are obtained to optimise:

$$\text{median}\left((e_{1x}^2 + e_{1y}^2), \ldots, (e_{kx}^2 + e_{ky}^2)\right).$$

Many other 'robust' regression procedures have been proposed for superimposition of configurations such as GS-estimators, M-estimators and least absolute deviations. For example, Dryden & Walker (1997, 1999) adapt the S-estimator of Rousseeuw and Yohai (1984) to minimise the residuals of the fit of X to Y.

Note, however, that unlike with the least squares approach (and some other resistant techniques, e.g. Dryden & Walker, 1999), the parameter estimates determined by repeated medians are obtained using procedures that are not based on any explicit optimisation principle. Because there is no objective function to be minimised, Rohlf & Slice (1990) argue that this makes it unclear how to judge the quality of the final results. Frequently, the view is that subsequent analysis of the X_i^{RF} should not be performed as the resistant fit procedures lack the well developed theory associated with the least squares Procrustes methods (Rohlf & Slice, 1990, Adams *et al.*, 2004). By Kendall's (1984) definition of shape, configurations aligned by any other method than OLS or GLS, still contain 'non-shape' variation due to location, orientation and size (Bookstein, 1996c). Others however, for example Walker (2000), argue that providing the larger/smaller variance is not only limited to a few (<25%) of landmarks, analysis should proceed in the same way as when having obtained Procrustes fits, using ordinary multivariate techniques, as the estimated covariance structure of the X_i^{RF} will be a better reflection of the true variation in shape. Goodall (1991) also suggested using a robust superimposition when the variance at the different landmarks is known to be 'non-isotropic' (which may involve different weighting of landmarks in different configurations) and to base inference on the covariance estimated from the resulting fits. However, as Walker (2000) also notes, because of the acceptance of Kendall's shape space, resistant fit methods have been largely ignored.

2.2.4 Geometric methods vs Distance methods

The preceding sections of 2.2 have presented an overview of the main methods of shape analysis using landmarks. The various approaches were separated into distance methods (traditional morphometrics, allometry, EDMA) and superimposition/geometric methods (Bookstein, resistant methods and in particular Procrustes superimposition, including various post superimposition techniques).

When only a few specific length measurements on an object are of interest then clearly the distance methods of section 2.2.2 will be entirely adequate. However, when the complete geometry of an object is to be considered (as we hope to do in this study), then the various advantages and disadvantages of the different methods need to be thought out in order to identify the most appropriate place to start with our investigations.

Basing an analysis of shape on just distances, ratios and angles can be inferior to using the actual landmark coordinates of the objects themselves for several reasons. In general a set of linear distances will usually be insufficient to capture the geometry of objects, since they carry no information about the spatial relationships among the measured distances. As Marcus (1996) notes, use of actual landmark configurations archives the size and shape of an object, an accomplishment that no 'simple' set of distances can duplicate. Ratios of distances and angles can be easily calculated from landmark coordinates whereas the converse is not generally true. In addition, the same set of measurements, angles and/or ratios can often result from two objects of different shape but since information on the location of where the distances were made, relative to each another is not included in the data, we would have no way of knowing this.

With traditional morphometrics, interpretation of the important linear combinations of ratios and angles resulting from statistical analyses can be difficult and for EDMA, the long list of numbers in a resulting form difference matrix can be especially hard to understand in terms of the shape that is being described. Another disadvantage of these methods is they are invariant under reflections, which may not be desirable for certain applications (although it may be for others).

A major strength of the newer geometric methods is that graphical representations of results are possible in terms of configurations of landmark points, whereas distance methods usually only allow one to visualise statistical relationships as scatter plots of pairs of distances and not as representations of the shapes themselves. As noted in section 2.2.2.1, reconstruction of configurations is possible when using distance methods, but this requires enough measurements are taken with enough common endpoints. However, in general, generation of graphical representations of shapes from

linear distances will be difficult because the geometric relationships among the variables have not been preserved.

In this respect the methods of section 2.2.3 have a clear advantage, since researchers can then easily examine and visualise results in the original space of the objects. The use of principal components to visualise shape variation is a particular benefit of these methods and an equivalent technique for distance methods is not obvious.

However, while the newer geometric approaches, in particular the Procrustes methods, offer clear advantages in terms of visualisation and presentation of geometrical information, distance based methods do have some benefits. In fact, Kent (1994) notes that if variations are small then registration methods and distance-based methods will give very similar conclusions about shape, because the coordinates used in the different methods are effectively linear transformations of each other.

An advantage of the methods in 2.2.2 is that they can be applied to distances that do not require the precise location of landmarks. For example the exact positions of the mesio-distal widths of teeth are often difficult to locate, but the actual measurements are less prone to error and so for features such as these, superimposition methods may be more difficult to apply.

Localisation of shape analysis to distinct subsets of landmarks is also straightforward using distance based methods, whereas geometrical shape methods use information at all the landmarks. Edge superimposition however, does allow a straightforward localisation by choosing baseline points within the desired subset of landmarks (Dryden & Mardia, 1998).

In order to justify the choice of one landmark method over another, the different techniques have in recent years been investigated in terms of various statistical properties. These include 'consistency' (i.e. does a method give the correct estimate of mean shape as sample sizes become infinite), the size of any bias in estimating the mean shape (i.e. how close estimates are to the true mean shape, on average) and statistical power in detecting differences between mean shapes. Lele (1993) showed that EDMA estimates of mean shape are consistent and Kent (1994) showed that the Procrustes

estimates of average shape were also consistent under the same model of independent isotropic variation at each landmark. Several recent studies have used both analytical and simulation approaches to compare different landmark techniques, each demonstrating that Procrustes superimposition methods are the preferred methods for comparing shape statistically. Rohlf (2000a) found that in many instances, the statistical power for methods based on inter-landmark distances and angles was much lower (and never higher) than for tests based Procrustes methods, and also reported a programming error in the simulations of Lele & Cole (1996), who had previously reported rather different results. Another criterion for assessment is that the methods should not impose constraints on the patterns of variation displayed in the results. Rohlf (1999, 2000b) found that the methods based on inter-landmark distances and angles gave distinct patterns of covariation that depend on their mean shapes. However, no such problems were found when examining patterns of variation in Procrustes fits (or any other choice of tangent coordinates).

In their recent review article, Adams *et al.* (2004) discuss these articles further and in view of these findings, advocate the use of the Procrustes method as the basis for any statistical analysis of shape based on landmarks.

2.3 Outline methods

In many cases there may be insufficient landmarks, or not enough landmarks in the right places, to capture the variation in shape of an object adequately (e.g. there may be large regions of an object where no biologically meaningful landmarks can be identified). Points may be sampled around an object's outline but often these will lack the correspondence required by most of the landmark methods in section 2.2. Instead therefore, 'outline' based methods may be used.

The first step in any such analysis is to digitise points around the two-dimensional outline of an object into Cartesian pixel coordinates. The curve passing through these points can then be represented in a variety of ways, producing shape variables and measures of dissimilarity for comparison of different objects.

Below we review the most popular methods for analysing shape based on outline data. We start by considering simple univariate shape descriptors, which can be easily derived from the pixel outlines of objects providing measures of 'compactness', 'elongation' and 'roundness' as well as other moment-based quantities. Fourier analysis represents each outline as a weighted sum of wave functions, the resulting coefficients of which may then be used as 'shape variables' and standard multivariate analysis applied. A benefit of these methods is that resulting points in the parameter space can be transformed back into the space of the original data and visualised as outlines. Alternatively, an empirical distance function may be derived to describe differences between outlines at different points and the method of 'eigenshapes' used based on the principal components of these distances.

Most of these methods assume that the complete outline of an object is available, yet for the study of tooth shape this may not always be possible. While complete outlines may be easily obtainable from around occlusal surfaces, only open, incomplete curves can be obtained the buccal surfaces of teeth, where part of the outline will nearly always be obscured by the gingival margin. The ability of these methods to cope with incomplete outlines must therefore be considered when assessing their suitability for the analysis of tooth shape.

2.3.1. Shape descriptors based on outlines

Glasbey & Horgan (1995), summarise several simple shape descriptors for use on strings of 2D pixel outline coordinates, each invariant to location, orientation and size. Many of these statistics can be found as standard summary descriptors for outlines in most image analysis packages (e.g. Image ProPlus, v. 2.0 and above, Media Cybernetics).

Probably the most commonly used of these shape statistics is the measure of 'compactness', which is defined as the ratio of the area of an object to the area of a circle with the same perimeter. A circle is used for comparisons as it is has the most 'compact' shape. 'Convexity' is an alternative measure that is sensitive to only certain departures from circularity. This is obtained as the ratio of an object's convex hull perimeter to the perimeter of the object itself. The convex hull of an object is defined as

the smallest convex shape, which contains the object, where a convex shape is one where if any two points within the shape are joined by a straight line, all points along the line are also within the shape. A measure of 'roundness' can be obtained as the ratio of the area of an object to the area of a circle with the same convex hull perimeter. Finally, 'elongation' is defined as the ratio of the maximum length of the object to some measure of width. Width could be the sum of the largest perpendicular distances to the outline at any point either side of the object's length or simply the maximum width at any fixed position, perpendicular to the maximum length. Length and width could also be based on the major and minor axes of the object. The relevant formulae and associated ranges of values are summarised in table 2.1.

Measure	Formula	Max and min values
Compactness	$4\pi \frac{area}{perimeter^2}$	Max value 1 (when object a circle) <1 for elliptical shapes or shape with an irregular border (rather than smooth)
Convexity	$\frac{convex\ hull\ perimeter}{perimeter}$	Max value 1 (when object convex) <1 when object has irregular borders
Roundness	$4\pi \frac{area}{convex\ perimeter^2}$	Max value 1 (when object a circle) <1 for departures from circularity, but relatively insensitive to irregular borders.
Elongation	$\frac{length}{breadth}$	Max value is infinite (an object may be infinitely long compared to its width). Min value 0

Table 2.1: Commonly used shape descriptors for outline data

Alternative methods are based on moments, which may be used to describe the spatial distribution of an object represented as a binary image (e.g. object entirely black on a white background). For pixels with coordinates (x,y) within the area of an object, the (k,l)th order moment is defined as:

$$\mu_{kl} = \sum_x \sum_y x^k y^l \text{ for } k,l=0,1,2,\ldots$$

For example, if $k=l=0$ then μ_{00} is the area and $(\mu_{10}/\mu_{00}, \mu_{01}/\mu_{00})$ the centroid (centre of gravity). Central moments are defined as:

$$\mu'_{kl} = \sum_x \sum_y \left(x - \frac{\mu_{10}}{\mu_{00}}\right)^k \left(y - \frac{\mu_{01}}{\mu_{00}}\right)^l \text{ for } k+l>1.$$

e.g. μ'_{20} and μ'_{02} measure dispersion in the x and y direction, respectively and μ'_{11} in a diagonal direction. These measures are not rotationally invariant although Hu (1962) derived ones which are, namely the moment of inertia and the isotropic/directional measure of dispersion given by:

$$\mu'_{20} + \mu'_{02}$$

and

$$(\mu'_{20} + \mu'_{02})^2 + 4\mu'_{11}$$

respectively. Alternatively one could specify the direction in which the object has maximum variation (the major axis of the object) as well as the direction perpendicular (the minor axis) and define second order moments λ_1 and λ_2, measuring dispersion in these directions;

$$\lambda_1 = \mu'_{20}\sin^2\phi + \mu'_{02}\cos^2\phi + 2\mu'_{11}\sin\phi\cos\phi\,,$$

$$\lambda_2 = \mu'_{20}\cos^2\phi + \mu'_{02}\sin^2\phi + 2\mu'_{11}\sin\phi\cos\phi\,.$$

These will then also be rotationally invariant. Shape measures for compactness and elongation based on moments could then be:

$$\lambda_1/\lambda_2 \text{ or } \frac{\mu'_{20} + \mu'_{02}}{\mu_{00}}$$

respectively. The former measure quantifies how dispersed pixels in an object are from their centroid, in comparison with the most compact arrangement of the pixels.

Marshall (1989) notes however that while the measures may be useful for describing simple shape, they often tend to fail for more complicated ones. Another major drawback is that many of the descriptors above can yield the same values for objects widely differing in shape. Moreover, we have absolutely no way of reconstructing what an object of (say) mean compactness (or any other measure) may look like.

In many studies of tooth shape, subjective classifications of outline shape are made by visually comparing the buccal surface outline to pictures of several possible hypothetical outline shapes. For example, Lamb (1998) categorised the buccal surface

shapes of upper central incisors as one of three possibilities (square, oval or taped), Carbonell (1963) classified incisor teeth as either normal or 'shovel' shaped whereas Moskana *et al.* (2001) analysed frequencies of teeth that were either normal or visually identified as 'peg' shaped.

2.3.2 Fourier analysis

Devised by Jean Baptiste Fourier (1768-1830) and first published at the beginning of the 19th century, Fourier analysis decomposes a continuous curve or outline (which may be incomplete – see section 2.3.2.5) into a weighted sum of sine and cosine functions, of increasing frequency and varying amplitude. If $\phi(t)$ denotes some value on an outline of length L and t provides some arbitrary correspondence between specimens, with $t \in [0, L]$ (for example t could be a proportional distance around the outline from some arbitrary starting point), the (discrete) Fourier series decomposition of $\phi(t)$ is given by:

$$\phi(t) = a_0 + \sum_{n=1}^{\infty} \left\{ a_n \cos\left(\frac{2\pi n t}{L}\right) + b_n \sin\left(\frac{2\pi n t}{L}\right) \right\} \tag{2.14}$$

where $a_0 = \frac{1}{L}\sum_0^L \phi(t)$, $a_n = \frac{2}{L}\sum_0^L \phi(t)\cos\left(\frac{2\pi n t}{L}\right)$, $b_n = \frac{2}{L}\sum_0^L \phi(t)\sin\left(\frac{2\pi n t}{L}\right)$.

Typically t is pre-standardised to have range $[0,2\pi]$ (by multiplication by $2\pi/L$), so (2.14) becomes:

$$\phi(t) = a_0 + \sum_{n=1}^{\infty} \{ a_n \cos(nt) + b_n \sin(nt) \} \tag{2.15}$$

where $a_0 = \frac{1}{2\pi}\sum_0^{2\pi} \phi(t)$, $a_n = \frac{1}{\pi}\sum_0^{2\pi} \phi(t)\cos(nt)$, $b_n = \frac{1}{\pi}\sum_0^{2\pi} \phi(t)\sin(nt)$.

The coefficient (or amplitude) a_n or b_n of each harmonic term quantifies its relative contribution to the description of the curve or outline. Because the sine and cosine components are $\pi/2$ radians out of phase, the Fourier series can describe highly irregular curves. In general, the higher frequency (larger n) sine and cosine terms have smaller coefficients and describe the finer aspects of the curve, contributing less to the

description of the outline. Simpler shapes will be adequately described by the earlier harmonics, whereas more complex shapes will require more components. In practice the 'finer detail' may represent noise and so may be disregarded without any detriment to shape analysis. Consequently, the Fourier series is usually 'truncated' after a finite number N of these harmonics, where N depends on the irregularity of the original outline curve, and the smaller terms ignored. The sum of the retained components then give an approximate mathematical reconstruction of the original outline.

The coefficients can be computed by a variety of algorithms e.g. Ralston (1965) for equally spaced data points. Most often, the Fourier series is thought of as a multiple regression analysis in which the best fitting values of the coefficients a_n and b_n n=1,...,N (in a least squares sense) are obtained by regressing the observed set of values of $\phi(t)$ onto the sine and cosine terms. Alternatively, fast Fourier transform algorithms may be used to compute the coefficients, providing N is either a power of 2 or a product of small prime factors (see, for example, Cooley & Tukey, 1965 or Press *et al.*, 1992). If necessary, extra boundary points can be introduced by duplicating or interpolating between pixels to produce such a value for N.

Two curves will have the same shape if they differ only by translation, rotation and change in size. By careful pre- or post-normalization of the outline of the resulting Fourier descriptors, new coefficients may be obtained which are invariant to geometric changes, so that the same shapes occurring at different positions, scales and orientations would all yield the same set of descriptors. Several different approaches have been considered, which we describe below. For example, many authors note that changes in the scale and orientation of the outlines result in simple transformations of the coefficients which can be accounted for in the choice of $\phi(t)$. Alternatively, the outline may be translated and rotated to a standard reference frame and normalized for size in some way before the calculation of the Fourier series. Note also that dependency on the arbitrary choice of starting position (t=0) must also be considered.

Following standardisation, the Fourier coefficients, a_n and b_n may then be regarded as a new dataset of 'shape variables' and standard multivariate analysis applied to compare the outlines of different objects. One of the advantages of the methods is that values in

the Fourier coefficient space resulting from such analyses can be transformed back into outlines. For example, outlines may be reconstructed corresponding to mean values of the coefficients or to points along principal component axes.

Examples of the use of Fourier series are widespread. Kaesler & Water's (1972) analysis was one of the first applications of Fourier descriptors to study morphological shapes, along with Lestrel (1974). Rohlf & Archie (1984) compared different Fourier methods for the description of wing shape in mosquitoes. Other applications from various fields can also be found in the first five chapters of the recent (1997) text by Lestrel.

The only example that can be found of the application of Fourier analysis for investigations of tooth shape is that of Ferrario *et al.* (1999), who used the technique to examine the shape of closed occlusal molar outlines. However, their reliability assessment of the method appears questionable.

Different methods depend on the different choices of $\phi(t)$ and t used to describe the outline and we consider the three most frequently used of these below. If t defines some position on the outline, with associated value $\phi(t)$, commonly used representations for $\phi(t)$ are :

a) the radial distance from some reference point at angular intervals t;
b) the tangent angle (local boundary slope) or change in tangent angle from some reference direction to position t along the outline,
c) the actual x or y coordinates (or changes in them) on the outline itself (or complex number representation of the coordinates as a function of distance t along the curve).

Each of these options is illustrated in fig 2.4 below, with further details given in the following three sections. After briefly describing each of these methods, including the various standardisation steps for size, orientation, location and starting point, section 2.3.2.5 then considers in how each these methods handle open curves (incomplete outlines). One of the underlying assumptions of the fitted Fourier series is that the outline is continuous over [0, L], with no endpoints and $\phi(0)=\phi(L)$, i.e. that the outline

curve is 'closed'. As mentioned at the start of this section, only incomplete (open) curves can be obtained from the buccal surfaces of teeth.

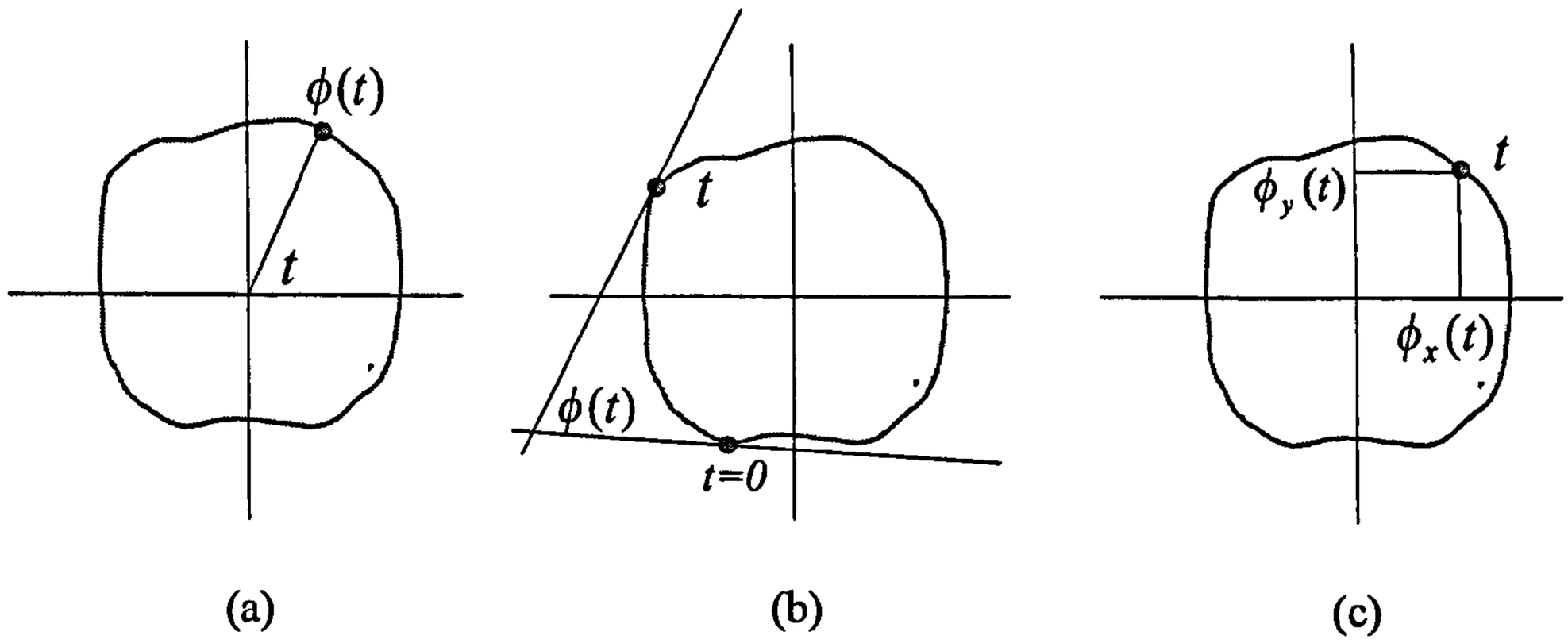


Fig. 2.4: Fourier series representation of function value $\phi(t)$ at position t around the outline: (a) polar or radii representation, (b) tangent angle, (c) Cartesian coordinates.

2.3.2.1 Polar (radial) representation

Often known as 'classical Fourier series', early uses of Fourier series used lengths of equally spaced radii from the centroid of the outline coordinates or a chosen reference landmark to describe the outline of an object. If t ($0 \leq t < 2\pi$) represents the angle from some reference axis then $\phi(t)$ is the length of a line joining a point on the outline to the reference point at angle t. Interpolation may be used whenever the sampling results in an outline position between recorded coordinates. See fig 2.4 (panel a).

Ferrario *et al.* (1999) sampled 360 distances to the occlusal outline of molars, using the centroid as the branching point, at every 1 degree. In order for the Fourier coefficients to become shape variables, the outlines were normalised prior to polar representation by setting the object area to an arbitrary size and orientating the outline and starting angle using two landmarks on the lingual grooves.

Johnson *et al.* (1985) also used polar representation with centroid as the reference point when investigating the shape of mouse vertebrae. Size was again standardised to an

arbitrary value and the outlines iteratively rotated (about their centre) to minimise the sum of squared differences along each of 128 polar radii to a reference outline template. Problems arise with the polar method where the function $\phi(t)$ becomes multivalued; the technique is limited only to simple outlines in which the radius is a single value. Although this may be adequate for many outline shapes, it was for this reason that the alternative methods below were developed, so that more general outline shapes may be considered.

2.3.2.2 Tangent angle representation

Zahn & Roskies (1972) suggested that the most general way to represent an outline is by the cumulative change in the angle of a tangent vector to the outline, again as a function of distance $t \in [0, L]$ along the outline from some starting point. See fig 2.4 (panel b). If the outline is first scaled from [0,L] to [0, 2π], we can define a cumulative angle function ϕ between the starting point and t as:

$$\phi(t) = \theta(t) - \theta(0) - t$$

where $\theta(t)$ is the angle of a tangent vector at a distance t from the starting point (t=0). Typically the function is computed for 100 equally spaced values of t where $\theta(t)$ is calculated by averaging over several boundary points (Rohlf & Archie, 1984). Since the domain of the domain of the outline has been scaled from [0,L] to [0, 2π], we can also define the 'normalised cumulative angular function' as:

$$\phi^*(t) = \phi\left(\frac{Lt}{2\pi}\right) + t, \text{ with } \phi^*(0) = \phi^*(2\pi) = 0. \qquad (2.16)$$

Zahn & Roskies (1972) show that $\phi^*(t)$ is invariant under translations, rotations and changes in the location of the outline in the 2D plane so that all closed curves identical in shape and starting point give the same Fourier coefficients. To remove the remaining invariance due to differences in starting point, a well defined landmark, with high correspondence between objects, is recommended for t=0.

The authors show how to obtain Fourier coefficients from a polygonal outline of length L (with straight lines, rather than a smooth curve between vertices), so that $\phi^*(t)$ may be represented as in (2.14), with coefficients given by:

$$a_0=\frac{1}{L}\sum_0^L \phi(\lambda),$$

$$a_n=\frac{2}{L}\sum_0^L\left(\phi(\lambda)+\frac{2\pi\lambda}{L}\right)\cos\left(\frac{2\pi n\lambda}{L}\right),\quad b_n=\frac{2}{L}\sum_0^L\left(\phi(\lambda)+\frac{2\pi\lambda}{L}\right)\sin\left(\frac{2\pi n\lambda}{L}\right),$$

where $\lambda=\frac{Lt}{2\pi}$. For a curve of m vertices (points) labelled 0 to m-1, the length of the edge between vertices i-1 and i is denoted as Δt_i (with $L=\sum_{i-1}^{m}\Delta t_i$) and the change in angular direction at vertex i is denoted as $\Delta\phi_i$. If t is the distance along the curve from some starting point, define:

$$\begin{aligned}\phi(t) &= \sum_{i-1}^{k}\Delta\phi_i && \text{for } \sum_{i-1}^{k}\Delta t_i \le t < \sum_{i-1}^{k+1}\Delta t_i\\ &= 0 && \text{for } 0\le t<\Delta t_1 .\end{aligned}$$

Using (2.16) and setting $t_k=\sum_{i=1}^{k}\Delta t_i$, the coefficients of the Fourier series for $\phi^*(t)$ are therefore given by:

$$a_0=-\pi-\frac{1}{L}\sum_{k=1}^{m}l_k\Delta\phi_k\ ,a_n=\frac{-1}{n\pi}\sum_{k=1}^{m}\Delta\phi_k\sin\left(\frac{2\pi n l_k}{L}\right),b_n=\frac{1}{n\pi}\sum_{k=1}^{m}\Delta\phi_k\cos\left(\frac{2\pi n l_k}{L}\right).$$

For the tangent angle approach, one of the main problems, first pointed out by Zahn & Roskies themselves, is that reconstructions from relatively few harmonics do not usually result in figures with closed contours, which may often be undesirable and may not look very realistic. Rohlf & Archie (1984) also note that estimation of the coefficients seems to be more sensitive to noise in the image than other methods.

2.3.2.3 Cartesian representation and elliptic Fourier analysis

Granlund (1972) proposed using two Fourier series, one to represent the *x*-coordinates and one to represent the *y*-coordinates of pixels around the outline directly. See fig 2.4 (panel c). If $\phi_y(t)$ denotes (say) the y coordinate of pixel t, where t=1,...L is the position (or path length) along a string of equally spaced outline coordinates from some chosen starting point, equation (2.14) written for $\phi_y(t)$ is:

$$\phi_y(t) = a_0 + \sum_{n=1}^{N}\left\{a_{ny}\cos\left(\frac{2\pi nt}{L}\right) + b_{ny}\sin\left(\frac{2\pi nt}{L}\right)\right\}.$$

If $N \geq (1/2)L$ then the curve passes through all the boundary pixels, whereas if $N < (1/2)L$ then the curve provides a smooth approximation. By (2.14), the Fourier coefficients are given by:

$$a_{ny} = \frac{2}{L}\sum_{t=1}^{L} y_n \cos\left(\frac{2\pi nt}{L}\right),\ b_{ny} = \frac{2}{L}\sum_{t=1}^{L} y_n \sin\left(\frac{2\pi nt}{L}\right).$$

Analogous equations are then also obtained for the *x*-coordinates.

A more general approach was proposed by Kuhl & Giardina (1982), which does which not require the points to be equally spaced and can fit any closed outline, given enough harmonics. Here the Fourier coefficients for (say) the *y* pixel coordinates of the outline are given by:

$$a_{ny} = \frac{V}{2n^2\pi^2}\sum_{t=2}^{L}\frac{\Delta y_t}{\Delta v_t}(\cos(2n\pi v_t / V) - \cos(2n\pi v_t - 1/V)), \qquad (2.17)$$

$$b_{ny} = \frac{V}{2n^2\pi^2}\sum_{t=2}^{L}\frac{\Delta y_t}{\Delta v_t}(\sin(2n\pi v_t / V) - \sin(2n\pi v_t - 1/V)),$$

where

L is the number of steps in the trace around the outline (indexed as above by t),

Δy_t is the displacement along the y-axis of the curve between steps t-1 and t ,

Δv_t is the length of the linear segment between these steps,

v_t is the accumulated length of such measurements, and

V=v_L is the total length of the contour as approximated by the polygonal outline.

The coefficients a_{nx} and b_{nx} for $\phi_x(t)$ (the x-coordinates), are found in the same way using the incremental changes in the x-direction.

The technique is known as 'elliptic Fourier analysis' and is the method used in most available software programs today. It is so called because it transforms the outline into the sum of a series of ellipses (i.e. each nth pair of harmonic terms, for x and y describes an ellipse).

Ferson *et al.* (1985) suggest performing simple transformations of the outlines beforehand so that the resulting coefficients will be invariant to size, location, rotation and starting point. These are used in the software program EFA, by F.J. Rohlf (see section 2.5). Size is standardised by measuring the area of the ellipse defined by the first harmonic in (2.17) and then dividing all the x and y coordinates by its square root. Invariance to location is achieved by calculating the position of the centroid of the outline and then subtracting this from each of the x and y-coordinates. Orientation is standardised by rotating the outline so that the major axis of the first harmonic ellipse is parallel to the x-axis. Finally, invariance to starting position is achieved by setting t=0 at a point at the end of this ellipse in the positive direction of the x-axis.

Conversely, Rohlf & Archie (1984) chose not to standardise the outlines beforehand but instead to calculate the Fourier series first and then standardise the coefficients in a corresponding way. If the zeroth harmonics in (2.17) are ignored, the rest of the coefficients may be normalised to be invariant to size, rotation and starting position of the outline trace, by calculating the following transformations of the coefficients. (Location is standardised by using the centre of the first elliptical harmonic).

$$\begin{pmatrix} a'_{nx} & b'_{nx} \\ a'_{ny} & b'_{ny} \end{pmatrix} = \frac{1}{E^*} \begin{pmatrix} \cos\psi & \sin\psi \\ \sin\psi & \cos\psi \end{pmatrix} \begin{pmatrix} a_{nx} & b_{nx} \\ a_{ny} & b_{ny} \end{pmatrix} \begin{pmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{pmatrix},$$

where

$$E^* = \sqrt{(a_{1x}\cos\theta + b_{1x}\sin\theta)^2 + (a_{1y}\cos\theta + b_{1y}\sin\theta)^2},$$

$$\psi = \arctan\left(\frac{a_{1y}\cos\theta + b_{1y}\sin\theta}{a_{1x}\cos\theta + b_{1x}\sin\theta}\right), \quad \theta = 0.5 \times \arctan\left(\frac{2a_{1x}b_{1x} + a_{1y}b_{1y}}{(a_{1x}^2 + a_{1y}^2) - (b_{1x}^2 + b_{1y}^2)}\right).$$

E^* is the magnitude of the semi-major axis if the first harmonic ellipse, ψ is the angle of rotation of this ellipse and θ is the angle of rotation of the starting point at the end of the ellipse (standardised arbitrarily to one end).

Note however that in both instances, the standardisation of rotation and starting position, actually depends on the outline shape. Instead, it may therefore be preferable to standardise orientation and starting position by aligning the outlines (beforehand) using well defined landmarks and setting t=0 at one of them.

Finally, Marshall (1989) describes how a closed curve of Cartesian pixel coordinates may instead be coded as a complex Fourier series with $\phi(t)=(x(t), y(t))$, where t is again the path length from an arbitrary starting point, from 0 to L-1, with

$$\phi(t)= x(t)+iy(t)+\sum_{n}^{N} c_n \exp(i2\pi tn / L)$$

and c_n are the Fourier coefficients. When the shape undergoes a translation this only results in a change in c_0. Variations in other parameters are shown to result in a set of Fourier coefficients c'_n related to the original set by

$$c'_n = Sc_n \exp(i(R + Pn)), \text{ for } n\neq 0,$$

where S results from a change in scale, R from a change in rotation and P from a change in starting position along the contour.

2.3.2.4 Other approaches

Renaud (1995) considered the Fourier series of the inverse of curvature radius as a solution to avoiding problems of standardisation, when analysing the shape of upper molar of mio-pliocene rodents. $\phi(t)$ is defined as $1/R = d\theta/dt$, where R is the curvature radius, t the cumulative length of the outline θ the tangent angle at point t. The x-axis of each tooth was taken as long axis of each tooth's major ellipse (the first harmonic

from a separate elliptic Fourier series), and the starting point as the intersection of the outline and the x-axis. A discrete Fourier transform of the form:

$$\phi(t)=\sum_{n}^{N} c_n \exp(-i\omega t)$$

was used and the modulus of each of the resulting Fourier coefficients taken to eliminate information concerning starting point. The c_n are then invariant to rotation and translation and are divided by the length L of the outline to eliminate size.

It is also possible to use polygons (Pavlidis, 1977), splines (Hill & Taylor, 1992) and conic sections (Bookstein, 1978). Crimmins (1982) also showed other ways in which Fourier series can be transformed to provide measures that are invariant to the object's registration.

2.3.2.5 General implementation issues and use on incomplete outlines

The fitted Fourier series of each of the above methods assumes a continuous outline over [0,L], with no endpoints and $\phi(0)=\phi(L)$. For the buccal surfaces of teeth however, only part of the tooth outline will be available from which to compute the series, due to the presence of the gingival margin and inter-dental papilla.

The fitted Fourier representations always result in a closed curve, estimating the missing portion(s) of an outline, based on the parts of the outline for which values of $\phi(t)$ (or $\phi_x(t)$ and $\phi_y(t)$) are available. In one sense, this means that the missing parts of the outline will be handled in a standardised way. On the other hand, what this does enforce is an element of symmetry and bias in the fitted outline (and hence the coefficients) by assuming that the shape of the unobserved outline can be described by the same Fourier representation as the observed outline.

Additionally, with different proportions of each outline missing (due to variation in gum position), standardisation by use of centroid positions and area measurements will clearly be unwise. Homologous landmarks located on the observed part of the outline would need to be used, if available, to standardise location and starting points and could

also be used to orientate the partial outline. Standardisation for size would be more difficult, although some measure based on suitable inter-landmark distance(s), could be worth considering.

One proposal as to how to handle partial outlines, using the complex Fourier approach in 2.3.2.3, was made by Lin & Chapellappa (1987), who found that when up to 30% outline points are missing, estimates of the coefficients may be obtained by minimising:

$$\sum_{t=0}^{L-1}\left|\phi(t)-\sum_{k}c_n\exp(i2\pi tn/N')\right|^2+WT(perimeter^2/area)_{N'},$$

over c_n and N', where WT is a weighting factor and N' is the estimated length of the complete outline. The second term corresponds to another measure of 'compactness' (see section 2.3.1) and so the partial outline is closed, subject to a constraint of minimising its compactness of the outline.

2.3.2.6 Use of other functions (wavelets)

In a recent review article of Lestrel (1997), F.J. Rohlf questioned as to why, when outlines could be fitted by a variety of functions, is Fourier series usually used and does it matter? An alternative approach could be to consider the method of wavelets to approximate the outlines (see for example, Nason & Silverman, 2000). Like Fourier series, wavelets are orthogonal functions that separate a function $\phi(t)$ into components of different frequency. However, wavelets allow a very wide class of smooth functions to be considered, instead of just sine and cosine harmonics.

A function $\phi(t)$ is represented as a double sum, with coefficients c_{jk}:

$$\phi(t)=\sum_j\sum_k c_{jk}\psi_{jk}(t).$$

Instead of sine and cosine terms, here the basis functions are given by:

$$\psi_{jk}(t)=2^{j/2}\psi(2^j t-k).$$

Each ψ_{jk} is obtained from the same 'mother wavelet (function)' $\psi(t)$, by scaling it by a factor of 2^{-j} and/or shifting to position $2^{-j}k$. e.g. $\psi_{10}(t)$ makes $\psi(t)$ oscillate twice as fast, $\psi_{01}(t)$ shifts $\psi(t)$ one unit to the right. The mother wavelet is chosen so that the ψ_{jk} are mutually orthogonal, the classic mother wavelet being the Haar function (Haar, 1910), given by:

$$\psi(t)=\begin{cases}1, & 0\le t\le 1/2\\ -1, & 1/2\le t\le 1/2\\ 0, & otherwise\end{cases}$$

Each coefficient c_{jk} depends only locally on $\phi(t)$ because:

$$c_{jk}=\int_{-\infty}^{\infty}\phi(t)\psi_{jk}(x)dx\,.$$

Each c_{jk} gives information on scale 2^{-j} (i.e. at frequencies around 2^{j}) near position $2^{-j}k$, whereas with Fourier series each coefficient depends on all values of $\phi(t)$.

Whether or not wavelet representation offers a significant advantage over Fourier series depends on the curve (outline) being approximated. If $\phi(t)$ contains discontinuities either side of a given t, or sharp spikes and displays a varying frequency behaviour then wavelets will represent the function more efficiently (in terms of the number of coefficients needed to represent $\phi(t)$). However, if $\phi(t)$ is reasonably smooth, the Fourier series is more efficient (see for example, Strang, 1993, for further details). With most tooth outline data we would not expect to find such discontinuities in $\phi(t)$ and would expect the outline representations of most tooth types to be smooth enough not to require the use of functions other than sine or cosines to describe the outline.

2.3.3 Eigenshape analysis

Eigenshape analysis (Lohmann, 1983) is essentially principal components for outline data. An eigenshape analysis begins with the selection of method to assign homologous positions on pairs of outlines and a distance function to measure differences in shape.

The former is usually the tangent angle function of Zahn & Roskies (1972) as described in 2.3.2.2, which assigns points by way of distance *t* around the outline from some starting point. Distances between outlines at corresponding points *t* are then given by the sum of squared distances between tangent angles. Alternatively, Bookstein (1991) suggested use of radii distances from the outlines centroid at angular intervals *t* and Rohlf (1996) proposed using elliptic Fourier analysis values, as another choice of distance metric.

Having obtained a mean outline shape in some way (see for example, Sampson *et al.* (1996), described below), the *n* outlines of a sample are represented by *n* vectors of distance values from points on the mean to homologous points on each sample outline. A principal components analysis of the sample matrix of these distance vectors is performed and the principal modes of variation are the 'eigenshapes'. They have the usual properties of PC vectors in that they are uncorrelated and describe the sample in decreasing order of variance and are outline shapes in themselves.

For each object one can then compute a 'score' on each PC. A subset of these scores could then be combined to produce an outline that approximately reconstructs the original. Statistical analysis such as tests of group differences or correlations with other variables can then be performed by using the 'scores' on each of the eigenshapes as values of new variables for each of the outlines (in the same way as relative warps are used for landmark configurations).

A limitation of the method however, is the arbitrary choice of metric used to measure the difference in shape between two outlines and in the way homologies are computed. (A similar problem was noted in the arbitrary choice of $\phi(t)$ in Fourier series).

For this reason, Sampson *et al* (1996) presented a modification of eigenshape analysis that is based on a Procrustes metric (sum of squared distances between least squares aligned points), providing the potential for combining outline and landmark information a non-arbitrary way. After aligning the outlines by Procrustes superimposition (using a set of chosen landmarks around the outline), a mean outline is obtained using point wise averages of the sample of outlines at different positions using the iterative closest point algorithm of Besl & McKay (1992). Individual outlines are then represented by vectors

of deviations normal to the mean outline and variation in shape analysed in terms of a principal components analysis of the sample matrix of these deviations as described above. However, Rohlf (1996) points out that a limitation of this modification is that the outlines must be fairly smooth and the shapes near convex, otherwise the method can imply rather counterintuitive assignments of homology between points around the outlines (or curves) being compared.

2.3.4 Other outline methods

Alternative outline methods can be found in a review by Marshall (1989). In this review of shape coding techniques, the methods in 2.3.1 and 2.3.2, as well as several other lesser-known techniques, such as the use of skeletons, distributions of chord-length (lines joining any pair of points on the outlines), and an approach based on circular autoregressive models, are separated into various groups. These characterise the different approaches as either external or internal (depending on whether they represent the outline itself or the area within it) and also according to whether or not the shape descriptors (shape variables) allow reconstruction of outlines.

2.4 Outline vs. geometric landmark methods

In contrast to the landmark techniques of section 2.3 and in particular the Procrustes approach, a review of the literature reveals that there is far less certainty about the best method of analysis for 2D outlines.

The results of the two best-known methods, elliptic Fourier series and eigenshape analysis, do not generally agree and can result in different variations of the sample of outlines being considered. Both methods are characterised by arbitrary choices of homology and distance functions, which influence results in ways that are difficult to describe (Sampson *et al.* 1996). Additionally, as Rohlf (1996) notes, analyses based on the different Fourier methods of 2.3.2.1 to 2.3.2.3 give different results and there is no agreed theory that would enable the researcher to select the best approach.

Authors often claim that Fourier methods are preferable to landmark methods because they do not require knowledge of homologous features, yet in many of their publications, they then use biological landmarks to define reference points as starting points along the outline and/or to orientate the outline (e.g. Ferrario *et al.*, 1999).

Despite the benefits of Procrustes methods in terms of developed theory and post superimposition methods of analysis, a limitation of landmark methods is that a sufficient number may not be available to capture the shape of a structure. Important shape differences may be located in the regions between landmarks and so one would expect the outline methods to offer a clear advantage in such situations. However, as Adams *et al.* (2004) note, methods such as Fourier series go to the other extreme by ignoring differences in the relative positions of landmarks, which may be available and so ignore potentially important information on differences in shape.

In deciding whether to use landmark or outline methods a key issue is not whether it is possible to locate a sufficient number of landmarks but whether one should ignore homology information when one has it. Ideally both types of information should be analysed jointly. One promising solution to this is the method of semi-landmarks Bookstein (1996a,d,e). This lesser known technique extends the standard generalised Procrustes superimposition procedure of 2.2.3.2 to allow both landmarks and outlines to be combined in one analysis. In addition to optimally translating, rotating and scaling the landmark data, an extra step is introduced to the Procrustes matching which allows the semi-landmarks of a configuration to move away from their outline positions along specific directions, approximating the outline of an object. These directions are typically along escribed chords (a line through the original position parallel to the line joining the two neighbouring points). The new locations along these lines are those which minimise the bending energy of the thin plate spline from the Procrustes mean shape. The whole process is repeated for each configuration with new mean shapes and directions computed until convergence. Once the optimal positions of the fixed and semi-landmarks are determined, they can all be treated in the same way as fixed landmarks in subsequent statistical analyses.

This method could have been included under either 'landmark methods' (section 2.2) or 'outline methods' (section 2.3), but was not since it is not yet regarded as an established

method of analysis. Only papers with Bookstein as first author have used the method. In addition, the extension of existing of morphometric software to include semi-landmarks, has not yet taken place (Adams *et al.*, 2004), although during the time this article was in press, an update was made to the TPS series of software (see section 2.5) to allow inclusion of semi-landmarks. Other specific and application-driven variants of the methods in sections 2.2 and 2.3 are considered in chapter 6 and the semi-landmark technique itself is described in much greater detail in chapter 5.

However, on the basis of the discussion of above and that of 2.2.4, we begin our investigations in chapter 3 by considering the Procustes approach to the analysis of shape. Along with the other tools of the 'morphometric synthesis' this has, in recent years, become the standard protocol for investigations of shape when landmark data is available and as discussed in 2.2.4, offers clear advantages over other landmark and outlines methods, in terms of developed theory, statistical properties such as consistency and power and variety of descriptive and inferential techniques which may be utilised. Since clinical interpretation will be required from the resulting analysis, a key benefit is that this also allows interpretation of results as pictures in the original space of the objects being considered, since the geometry of the objects is always retained throughout the analysis.

2.5 Summary of available software

Methods of shape analysis are not yet built into the standard statistical packages such as SPPS, SAS, MINITAB, S-PLUS, MATLAB and others. However, a number of specific programs for use with landmark or outline data are available online.

The best and most extensive source of Windows/Mac/Unix/Linux/DOS based software and routines is available through the State University of New York (SUNY) at Stony Brook website:

http://life.bio.sunysb.edu/morph/

The collection is maintained and updated by F.J. Rohlf, who until recently, also maintained the discussion list MORPHMET (now D.E. Slice). It includes cross

platform programs for computation of thin plate splines, superimposition methods, Fourier series and eigenshape analysis as well as many other support routines for image extraction and manipulation, generation of output files of landmark coordinates or outlines and multivariate analysis. A brief summary of the software is given in tables 2.2 and 2.3. Outdated software, for which more advanced versions are now available from the same authors, has been omitted. Most are designed for DOS, Windows and UNIX operating systems and a few are for Apple Mac. All accept landmark data in the form of text files in ASCII codes.

The largest series of programmes is the TPS series by F.J. Rohlf. Earlier versions were designed for DOS but more recent Windows applications are now available. They allow work with 2D objects only. Many have a built-in screen graphic editor to manipulate colours, labelling etc. and save images in graphic file formats. All have detailed, help files. These programmes are continually updated with new/alternative features so the summary descriptions in tables 2.2 and 2.3 represent the latest versions at the time of writing.

I.L. Dryden has also made available his own R routines for use with landmark data, at his University of Nottingham website:

http://www.maths.nottingham.ac.uk/personal/ild//shapes/

The routines also work in older versions of S-Plus (at least up to v.4.5, which was used throughout this study).

The use of a programming language package such as S-plus or R is perhaps more desirable than the software above, in view of the fact that methods will most likely need modification and/or development to incorporate specific tasks and features particular to the investigation of tooth shape.

In this study all routines have been purposely written by the author, although some parts of the calculations have been adapted from Dryden's functions.

Software for landmark data	Platform	Details
The shapes package Dryden, 2000-04	Windows, Unix & Linux	Set of routines for performing shape analysis of landmark data in R. Bookstein baseline registration for 2D data. GPA. Plots of principal components and thin plate spline transformation grids. Goodall's F-test and Hotelling's T^2 test for difference in mean shape. See http://www.maths.nottingham.ac.uk/personal/ild//shapes/
IMP (Integrated Morphometrics Package, comprising 6 programs) Sheets, 2001-02	Windows	Generates different types of shape coordinates. Principal components analysis (PACA) for landmark data based on the thin plate spline (Partial warps). Canonical variate analysis of the partial warp scores. Regression analysis of partial warp scores on a dependent variable. Generates a set of lengths from landmark data. Performs Goodall's F-test and Hotelling's T^2 test for mean shape difference as well as bootstrap versions. Allows files to be transferred to and from the TPS file format (used by software in table 2.3). See also http://www.canisius.edu/~sheets/morphsoft.html.
Morphologika O'Higgins & Jones, 1998-2004	Windows	GPA from 3D data and PCA of Procrustes fits. Displays scatter plots of configurations and transformation grids representing relative warps.
Planar Kunkel, 1997	DOS	Least squares and resistant-fit superimposition of landmarks. Can assign configurations to populations using linear discriminant functions.
MOGwin Dujardin, 2003	Windows	Displays visual steps of a Procrustes registration. Calculates centroid sizes and partial warps.
Procrustes bin / Procrustes hqx David & Laurin, 1994-2004	Mac	Pairwise as well as multi-specimen comparisons of landmark configurations
Edgewarp 2 / Edgewarp 3 Green & Bookstein, 1994-2001/05	Linux & others / Linux, Mac	Reads 2D and 3D images. Allows landmark identification and subsequent manipulation by image warping and averaging, using thin-plate splines.

Software for outline data	Platform	Details
EFA / EFAV / EFA3D Rohlf, 1993 / 1995 / 2003	DOS Windows	Elliptic Fourier analysis for 2 and 3-dimensional contours. EFA3D can also estimate coordinates of points from a set of Fourier coefficients.
Eshape MacLeod, 2000	MAC & Windows	Eigenshape analysis on 2D and 3D coordinate data from object outlines. See also: http://www.ucmp.berkeley.edu/Paleonet/ftp/ftp.html
Shape Iwata, 2001-02	Windows	Extracts outline from a bitmap image. Calculates elliptic Fourier descriptors and performs PCA. See also: http://cse.naro.affrc.go.jp/iwatah/shape/

Multi-format software	Platform	Details
Morpheus *et al.* Slice, 1994-2000	Windows, DOS, Linux, Unix & Mac	Comprehensive, cross-platform environment for the acquisition, processing, and analysis of landmark or outline data. Generalised Procrustes analysis (GPA), transformation grids, Fourier series. Replaces the programs GRF (1992) & GFR-ND (1993-99) by Rohlf & Slice, for GPA. See also http://life.bio.sunysb.edu/morph/morpheus/
PAST Hammer, 2002-03	Windows	Palaeontological multivariate Statistics package with thin plate splines, PCA, Procrustes superimposition, partial warps analysis, elliptic Fourier analysis and eigenshape analysis. See also: http://folk.uio.no/ohammer/past
Morphometrika Walker	Mac	See http://www.usm.maine.edu/%7Ewalker/software.html.

Table 2.2: Shape analysis software available at or linked to by the SUNY Stony Brook website

TPSsplin 2000-04	Displays D'Arcy Thompson style deformation grids, based on the thin plate spline. Recently updated (2003) to include option for specifying semi-landmarks, as in TPSrelw below.
TPSsuper 1998-2004	Generalised Procrustes superimposition, image un-warping of individual configurations to the concensus (using the thin-plate spline) then averages the un-warped images. Can also unwrap and average to an arbitrary reference. Helps user decide whether he/she has sufficient (or too many) landmarks.
TPSpower 1998-	Given estimates of the means of two populations, the expected amount of variability in the landmarks and a sample size, computes the statistical power expected using various statistical methods. Has recently been removed or is currently unavailable.
TPSregr 1998-2004	Multiple regression of partial/relative warps using continuous or group variables. Displays shape changes predicted by the regression model in the form of transformation grids. Provides statistical estimates of associations between explanatory and shape variables using standard or permutation methods. Displays empirical distribution of specimens along regression line. Can also be used to study allometry and perform MANOVA/MANCOVA.
TPSrelw 1998-2004	Computes principal, partial and relative warps. Displays shape changes as deformation grids and scatter plots of specimens in the space of the partial and relative warps. All images able to be saved as graphic files. Recent (2003) update to allow semi-landmarks to slide along chords when computing consensus (mean shape), with directions defined interactively by selecting pairs of landmarks.
TPSsmall 1998-2003	Test to determine whether the amount of variation in shape is small enough to permit statistical analysis in the tangent space.
TPStree 2000-04	Makes it possible to trace shape changes on a user defined hierarchical tree, displaying tree and transformation grid plots, which represent hypothetical configuration changes according to position on the tree.

Table 2.3: Components of the TPS software series for landmark data by F.J. Rohlf

Chapter 3

Procrustes analysis

3.1 Introduction

Having adopted Procrustes analysis as the most logical starting point for our investigations, this chapter now provides a more comprehensive description, demonstration and consideration of the ideas briefly outlined in the preceding chapter.

With the shape of our objects (here the tooth surfaces) represented as configurations of landmark coordinates (as shown in section 1.4), we describe how after optimally matching these configurations to account for the unwanted effects of location, scale and rotation, using orthogonal least squares, we can then define ways of measuring dissimilarity (or distance) between the shapes, estimate mean shapes, and examine shape variation.

After removing the unwanted registration effects, the landmark configurations can be regarded as single points in Kendall's shape space, with metric defined by the 'Procrustes' distance between shapes. It is this representation and the calculation of an appropriate tangent space, a linearised version of shape space in the vicinity of a particular point (here the estimated average shape), that forms the fundamental geometric setting for further analytical investigations, allowing most standard descriptive and inferential techniques to be adapted and applied successfully.

Demonstration of the application of techniques uses a sample of buccal surface images of central incisors from patients with hypodontia and a corresponding control group, introducing these ideas for the first time to the study of tooth shape.

Methods are described for $m \geq 2$ dimensions, even though the study data are two-dimensional, to illustrate how the techniques can readily cope with three dimensional

tooth landmark data, when this becomes available in the future. Much of the notation and definitions introduced here will also form the basis for use in the rest of this work.

The final part of this chapter raises issues relating to the general suitability of these methods for comparing and describing variation in tooth shape, highlighting some of the possible caveats of using Procrustes methods that are to be investigated and considered in subsequent chapters of this thesis.

3.2 Procrustes registration

In Chapter 2 we noted how the 'shape' of a configuration of landmark data may be represented as the coordinates of the points after the effects of differences in translation, rotation and scale are removed. The Procrustes approach is to filter out these similarity transformations by matching configurations using least squares.

We start by describing Ordinary Procrustes analysis where one configuration is matched to another and give solutions for determining the parameters of the optimal fit. When there is a sample of configurations we wish to match and obtain an estimate of average shape the technique is referred to as 'Generalised' Procrustes analysis, for which iterative algorithms must generally be used. However, in the case of two dimensions, the use of complex notation leads to an explicit solution.

The term 'Procrustes' is used because the matching operations are identical to those of Procrustes analysis in multivariate statistics for comparing matrices (see for example, Mardia *et al.*, 1979). Dryden & Mardia (1998) trace the technique back to Mosier (1939), although Cole (1996) reports that Boas (1905) suggested a 'method of least differences' (essentially ordinary Procrustes analysis) as a means of comparing homologous point sets. Later references, using the idea of a 'Procrustes distance' to measure the dissimilarity between two configurations of points include Hurley & Cattell (1962), who developed the method for factor analysis and Sneath (1967) who first used the measure in the context of two biological shapes. The idea of 'Generalised' Procrustes analysis was originally proposed by Kristof & Wingersky (1971) and later adapted by Gower (1975).

3.2.1 Ordinary (full) Procrustes superimposition

Consider two k landmark configurations in m dimensions, expressed as $k\times m$ matrices X and Y, e.g. in m=2 dimensions:

$$X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{k1} & x_{k2} \end{pmatrix} \text{or} \begin{pmatrix} x_{1x} & x_{1y} \\ x_{2x} & x_{2y} \\ \vdots & \vdots \\ x_{kx} & x_{ky} \end{pmatrix} \text{ and } Y = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ \vdots & \vdots \\ y_{k1} & y_{k2} \end{pmatrix} \text{or} \begin{pmatrix} y_{1x} & y_{1y} \\ y_{2x} & y_{2y} \\ \vdots & \vdots \\ y_{kx} & y_{ky} \end{pmatrix}.$$

We shall assume throughout this chapter that $k>m$ and $m\geq 2$.

In order to compare the configurations and obtain a measure of dissimilarity or distance between the shapes, Ordinary Procrustes analysis (OPA) involves matching configuration X to Y as closely as possible, by ordinary least squares (OLS), over the similarity transformations of scale, location and rotation. The differences between the fitted X and Y then indicate the size of the difference in shape.

The (Euclidean) similarity parameters, which may be applied to a configuration matrix such as X without changing its shape, are denoted as γ, a translation vector of length m (used as a $k\times m$ matrix $1_k\gamma^T$), $\beta>0$, a scale parameter and $\Gamma=\Gamma(\theta)$, an $m\times m$ orthogonal rotation matrix, satisfying $\Gamma^T\Gamma = \Gamma\Gamma^T = I_m$ and $|\Gamma| = +1$. For example, in 2 dimensions:

$$\Gamma = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \text{ where } \theta \text{ is the angle of rotation (about (0,0)).}$$

In matching X to Y by OPA we therefore seek to minimise the sum of squared Euclidean distances between corresponding landmarks of Y and $\beta X\Gamma + 1_k\gamma^T$, i.e.

$$\left\| Y - (\beta X\Gamma + 1_k\gamma^T) \right\|^2 = tr\{(Y - (\beta X\Gamma + 1_k\gamma^T))^T (Y - (\beta X\Gamma + 1_k\gamma^T))\}. \qquad (3.1)$$

The estimates of γ, Γ and β are obtained by the following sequence of steps.

Let X_C and Y_C denote the 'centred' versions of X and Y, obtained by calculating the average coordinate values (centroid) of each of X and Y and subtracting this from each set of landmarks. (Alternatively X_C and Y_C may be obtained by pre-multiplying X and Y by the centring matrix C defined in (2.2)). Following Goodall (1991), $\hat{\gamma}$ is given by the difference:

$$1_k \hat{\gamma}^T = (Y - Y_C) - \hat{\beta}(X - X_C)\hat{\Gamma}.$$

Rotation matrix $\hat{\Gamma}$ and scale $\hat{\beta}$ are then determined as those which minimise:

$$\|Y_C - \beta X_C \Gamma\|^2 = tr(Y_C^T Y_C) + \beta^2 tr(X_C^T X_C) + 2\beta tr(Y_C^T X_C \Gamma).$$

Following Goodall (1991) and Dryden & Mardia (1998) to obtain $\hat{\Gamma}$, if the single value decomposition of $Y_C^T X_C$ in the above is $V\Delta U^T$, then for any scale β, the minimising Γ is UV^T and so

$$\hat{\Gamma} = (X_C^T Y_C Y_C^T X_C)^{-1/2} X_C^T Y_C. \qquad (3.2)$$

For a given $\hat{\Gamma}$, the scale parameter is then:

$$\hat{\beta} = \frac{\sqrt{tr(X_C^T Y_C Y_C^T X_C)}}{tr(X_C^T X_C)} = \frac{\|Y_C\|}{\|X_C\|} tr\{(Y_C/\|Y_C\|)^T (X_C \hat{\Gamma}/\|X_C\|)\} = \frac{tr(Y_C^T X_C \hat{\Gamma})}{\|X_C\|^2} \qquad (3.3)$$

where $\|X_C\|$ and $\|Y_C\|$ are the centroid sizes of X_C and Y_C, as defined in (2.2).

Note that the order in which $\hat{\Gamma}$ and $\hat{\beta}$ are determined is important. $\hat{\Gamma}$ does not depend on the scale of the two centred configurations, but $\hat{\beta}$ depends on $\hat{\Gamma}$. Note also that $\|\hat{\beta} X_C\| \leq \|Y\|$ since $\hat{\beta} X_C = (\|Y_C\|/\|X_C\|) tr\{(Y_C/\|Y_C\|)^T (X_C \hat{\Gamma}/\|X_C\|)\} X_C$ scales X_C to unit size and then multiplies by the result by $\|Y_C\| A$ where $A = tr\{(Y_C/\|Y_C\|)^T (X_C \hat{\Gamma}/\|X_C\|)\} \leq 1$.

The configuration $\hat{\beta} X \hat{\Gamma} + 1_k \hat{\gamma}^T$ is then the 'full Procrustes fit' of X to Y. The term 'full' is used since the full set of similarity transformations have been considered in the matching.

Having determined $\hat{\gamma}$, the full Procrustes fit of X to Y is given by $(Y-Y_C)+\hat{\beta}X_C\hat{\Gamma}$. The rotation $\hat{\Gamma}$ (of X_C about (0,0)) and scale $\hat{\beta}$ are then determined to minimise the OLS distance between X_C and the centred version of Y. The resulting configuration, $\hat{\beta}X_C\hat{\Gamma}$ is then superimposed to Y by translation by $(Y-Y_C)$.

However, the more popular convention (e.g. Dryden & Mardia, 1998, Bookstein, 1991, Rohlf & Slice, 1990, Rohlf, 1999) is to have X and Y already centred to start with. The OLS superimposition of $X=X_C$ to $Y=Y_C$ is still given by $\hat{\beta}X\hat{\Gamma}+1_k\hat{\gamma}^T$, with parameter estimates as above and the same minimised value of (3.1) results. Since $X=X_C$ and $Y=Y_C$, $\hat{\Gamma}$ and $\hat{\beta}$ will be exactly the same as in (3.2) and (3.3), but now $\hat{\gamma}=0$.

The minimised value of (3.1) is denoted as OSS(X,Y), the ordinary (Procrustes) sum of squares, or in the following notation as:

$$\mathrm{OSS}(X,Y)=\min_{\beta,\gamma,\theta}\left\|\beta X\Gamma(\theta)+1_k\gamma^T-Y\right\|^2=\min_{SG}\left\|SG(X)-Y\right\|^2$$

where SG denotes the equivalence class or 'shape set' of configurations with the same shape as X which may be obtained by using the 'similarity group' of transformations of X, i.e. $\beta X\Gamma+1_k\gamma^T$. In the literature the word 'icon' is used to denote a representative member of a shape set and so both X and $\hat{\beta}X\hat{\Gamma}+1_k\hat{\gamma}^T$ are 'icons' of shape set of X.

Note that if the roles of X and Y are reversed, the estimated rotation is $\hat{\Gamma}^T$ but the scale is not $1/\hat{\beta}$. In addition, OSS(X,Y) $\neq$ OSS(Y,X), i.e.

$$\min_{SG}\left\|SG(X)-Y\right\|^2\neq\min_{SG}\left\|SG(Y)-X\right\|^2,$$

so this is still not a suitable choice of shape distance. However, if both configurations are standardised to be unit centroid size, as well as centred, before determining $\hat{\beta}$ then this does produce a suitable measure, with

$$OSS(X/\|X\|, Y/\|Y\|) = \min_{SG} \|SG(X/\|X\|) - Y/\|Y\|\|^2 = \min_{SG} \|SG(X) - Y/\|Y\|\|^2$$

$$= OSS(Y/\|Y\|, X/\|X\|) = \min_{SG} \|SG(Y/\|Y\|) - X/\|X\|\|^2 \quad \min_{SG} \|SG(Y) - X/\|X\|\|^2$$

$$= d_F^2(X, Y), \tag{3.4}$$

the 'squared full Procrustes distance' between X and Y. Further properties of the 'full Procrustes distance' are considered in section 3.3.2.

The matching/superimposition procedure is illustrated in fig. 3.1 on the following page, where two landmarks sets have been obtained from buccal images of two lower left canine teeth. The landmarks used were the MD and LACC endpoints and the positions along the sides of the tooth where the inter-dental papilla begins to cover the buccal surface. Both configurations have been centred prior to carrying out OPA. In the third row, the configurations are scaled to unit centroid size so that scaling step determining $\hat{\beta}$ in the final row obtains the full Procrustes distance between the two configurations.

3.2.2 Generalised (full) Procrustes analysis

Now consider the case where there are $n \geq 2$ configuration matrices, $X_1, ..., X_n$, sampled from some population and we wish to obtain an estimate of the population average shape μ and investigate the variation around it.

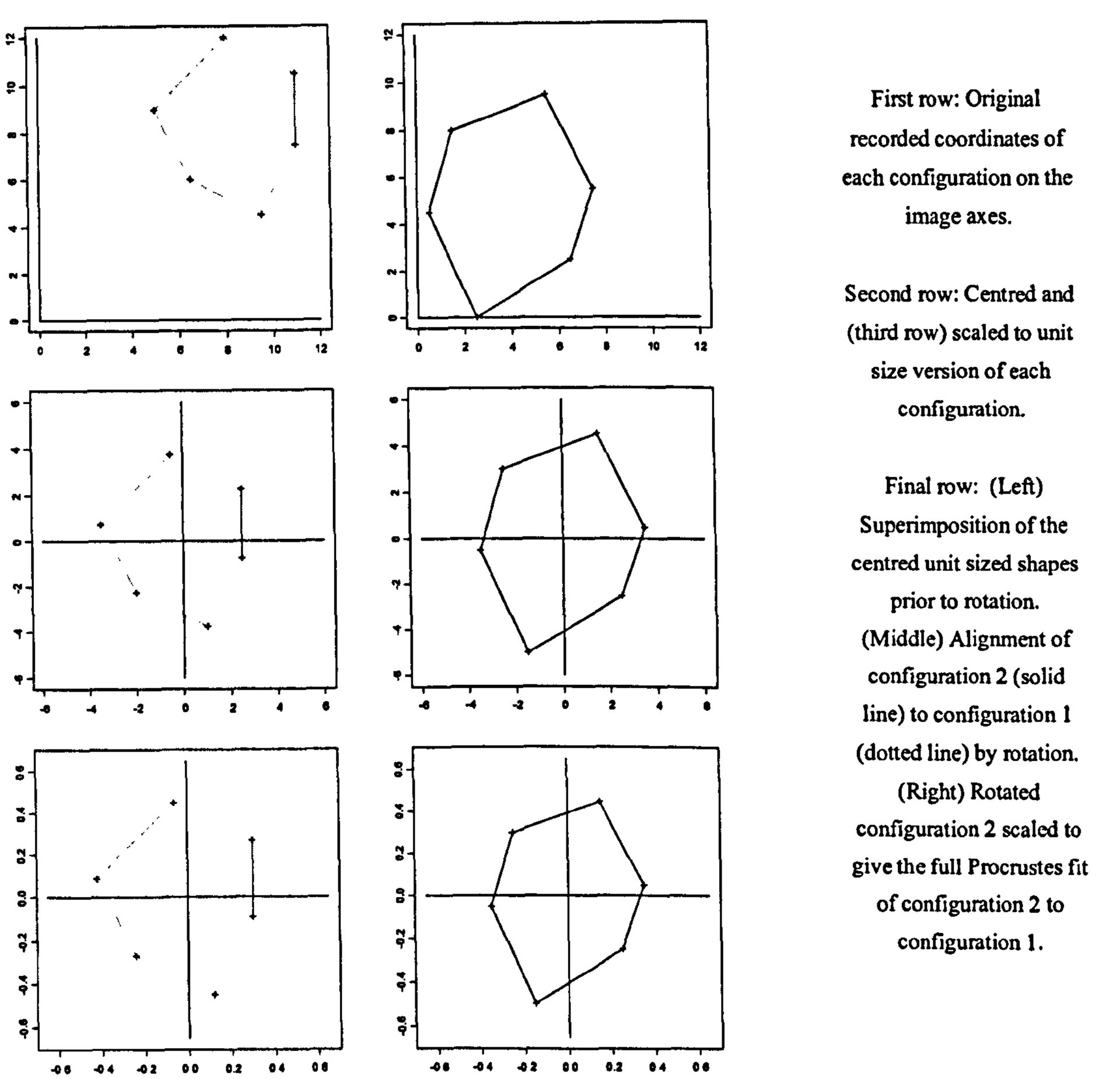
A model for the 'population' of shapes is given by:

$$X_i = \beta_i' (\mu + E_i) \Gamma_i' + 1_k \gamma_i^T. \tag{3.5}$$

Configuration $(\mu + E_i)$ is a perturbation from the mean, which is then translated, rotated, re-scaled (by γ_i', Γ_i' and β_i' respectively), and observed in some new coordinate system as X_i. The E_i are zero mean $k \times m$ independent random error matrices. As well as estimating μ, generalised Procrustes analysis (GPA) also involves estimating the reverse transformations β_i, Γ_i and $1_k \gamma_i^T$ of β_i', Γ_i' and $1_k \gamma_i'^T$ in (3.5)

respectively, to minimise a total sums of squares relative to μ, to obtain the full Procrustes registered fits of the X_i (relative to $\hat{\mu}$) given by:

$$X_i^P = \hat{\beta}_i X_i \hat{\Gamma}_i + 1_k \hat{\gamma}_i^T .$$



First row: Original recorded coordinates of each configuration on the image axes.

Second row: Centred and (third row) scaled to unit size version of each configuration.

Final row: (Left) Superimposition of the centred unit sized shapes prior to rotation. (Middle) Alignment of configuration 2 (solid line) to configuration 1 (dotted line) by rotation. (Right) Rotated configuration 2 scaled to give the full Procrustes fit of configuration 2 to configuration 1.

Fig. 3.1: Procrustes superimposition of two landmark configurations.

3.2.2.1 Mean shape and Procrustes fits

The full Procrustes estimate of mean shape is a configuration $\hat{\mu}$ which minimises the total sum of squared full Procrustes distances between each of the configurations and μ, given by

$$\sum_{i=1}^{n} d_F^2(X_i, \mu). \tag{3.6}$$

Note that μ is identified up to a similarity transformation and so the shape of μ is actually estimated by an 'icon' $\hat{\mu}$, of the shape set $SG(\hat{\mu})$.

Generalised Procrustes analysis is a generalisation of ordinary Procrustes analysis, originally developed by Gower (1975), with subsequent work by Ten Berge (1977), and Goodall (1991), which involves estimating parameters β_i, Γ_i and γ_i and μ so as to minimise:

$$\sum_{i=1}^{n} \|SG(X_i) - \mu\|^2 = \sum_{i=1}^{n} \|(\beta_i X_i \Gamma_i + 1_k \gamma_i^T) - \mu\|^2 \tag{3.7}$$

or alternatively,

$$\sum_{i=1}^{n} \sum_{i'=i+1}^{n} \|(\beta_i X_i \Gamma_i + 1_k \gamma_i^T) - (\beta_{i'} X_{i'} \Gamma_{i'} + 1_k \gamma_{i'}^T)\|^2 = \sum_{i=1}^{n} \left\| \beta_i X_i \Gamma_i + 1_k \gamma_i^T - \frac{1}{n} \sum_{i=1}^{n} \beta_i X_i \Gamma_i + 1_k \gamma_i^T \right\|^2 \tag{3.8}$$

subject to one of the following constraints on the size of μ or on the $\beta_i X_i \Gamma_i + 1_k \gamma_i^T$, in order to avoid all β_j becoming close to zero. The most popular choices of constraints and associated methods of analysis are:

(i) $\|\mu\| = 1$, as in Bookstein (1991), Dryden & Mardia (1998) for m=2 or as in Rohlf (1999) or Slice (2001) for $m \geq 2$.

(ii) $\sum_{i=1}^{n} \|\beta_i X_i \Gamma_i + 1_k \gamma_i^T\|^2 = n$, as in Gower (1975), Rohlf & Slice (1990).

(iii) $\sum_{i=1}^{n} \|\beta_i X_i \Gamma_i + 1_k \gamma_i^T\|^2 = \sum_{i=1}^{n} \|X_i\|^2$, i.e. sum of squared centroid sizes unchanged, as in Ten Berge (1977), Goodall (1991), Dryden & Mardia (1998).

For (ii) and (iii) GPA involves estimating the $\hat{\beta}_i$, $\hat{\Gamma}_i$ and $\hat{\gamma}$, to minimise (3.8), by iterative superimposition to successive estimates of $\hat{\mu}$. (The formula for $\hat{\beta}_i$ must adjusted so as to satisfy the constraints on the $\beta_i X_i \Gamma_i + 1_k \gamma_i^T$). The resulting $\hat{\mu}$ and X_i^P then also minimise (3.7), but subject to the constraints on the $\beta_i X_i \Gamma_i + 1_k \gamma_i^T$.

For (i) and $m \geq 2$, $\hat{\mu}$ is obtained by one of methods (ii) or (iii) and scaled to unit size. For $m=2$, an explicit complex eigenvector solution for $\hat{\mu}$ is available using complex variable notation. $\hat{\beta}_i$, $\hat{\Gamma}_i$ and $\hat{\gamma}$ are then calculated using the OLS formulae of section 3.2.1 to minimise (3.7), where $\|\mu\|=1$.

We describe the different possible approaches in detail in 3.2.2.2 and 3.2.2.3.

Note that with any method, once the X_i^P have been obtained and the configurations have been matched into optimal positions with respect to each other, the full Procrustes mean shape may be given by calculating the arithmetic mean configuration; the coordinate-wise average of the Procrustes registered configurations, given by:

$$\overline{X}^P = \frac{1}{n} \sum_{i=1}^{n} X_i^P .$$

(For a proof, see for example, Dryden & Mardia, 1998). With methods (ii) and (iii), $\hat{\mu} = \overline{X}^P$ is used iteratively to update the estimates of the β_i, Γ_i and γ_i as described above (see 3.2.2.2 for more detail) with the final coordinate-wise average then the resulting estimate of mean shape. When method (i) is used $\overline{X}^P$ will have the same shape as the estimated $\hat{\mu}$, but will differ in size. The average configuration $\overline{X}^P$ is often used as the mean shape in subsequent 'tangent space' analysis (see section 3.4.2). Based on $\overline{X}^P$ we can also calculate the Procrustes residuals, given by:

$$R_i^P = X_i^P - \overline{X}^P .$$

Having obtained $\hat{\mu}$ and the full Procrustes fits, we can compute various measures of shape variability, either about $\hat{\mu}$ or $\overline{X}^P$. One option would be the optimised value of (3.6), the sum of squared full Procrustes distances of each X_i to $\hat{\mu}$, given by $\sum_{i=1}^{n} d_F^2(X_i, \hat{\mu})$. Another could be the optimised value of (3.7), given by:

$$RSS(\hat{\mu}) = \sum_{i=1}^{n} \left\| X_i^P - \hat{\mu} \right\|^2,$$

the residual (Euclidean) sum of squares of the X_i^P about $\hat{\mu}$, or alternatively, the residual sums of squares, of the R_i^P, written as:

$$RSS = \sum_{i=1}^{n} \left\| X_i^P - \overline{X}^P \right\|^2.$$

Note that depending on the GPA method used, the sizes of $\hat{\mu}$ and the X_i^P will differ and so differences in the values of the various measures of shape variability suggested above may also result. In addition, the various summary measures above may or may not give the same values for any particular method. We discuss differences between the results of the GPA methods in detail in section 3.3.2.5.

Note that the orientation of the estimated mean $\hat{\mu}$ and the associated X_i^P is arbitrary. Typically, in order to aid subsequent visualisation, $\hat{\mu}$ (and the X_i^P) are rotated so that two particular chosen landmarks in $\hat{\mu}$ are horizontal (or vertical) or so that one of the coordinate axes is aligned with the principal axis of $\hat{\mu}$.

3.2.2.2 Algorithms for GPA

3.2.2.2.1 Gower's method with modification by Rohlf & Slice

We first outline the method described by Gower (1975), but using the subsequent modification to the initial scaling step by Rohlf & Slice (1990). The process involves iteratively updating the X_i^P by translating, rotating and scaling each configuration to a

current estimate of μ, whilst ensuring that constraint (ii) in 3.2.2.1 always holds, until (3.8) can no longer be improved.

Preliminary steps:

i. Centre each configuration X_i to (0,0) to remove differences in location once and for all. Then scale each configuration to have centroid size 1 by dividing by its original centroid size, i.e. compute $X_i' = X_i / \|X_i\|$, $i=1,\ldots,n$. Originally, Gower (1975) multiplied each centred X_i by $\sqrt{n / \sum_{i=1}^{n} \|X_i\|^2}$, where $\|X_i\|^2$ is the sum of squared distances of each landmark to the objects centre, but Rohlf & Slice (1990) argue that initial estimates then tend to be dominated by the largest objects. However in both cases the constraint $\sum_{i=1}^{n} tr(X_i' X_i'^T) = n$ holds.

ii. Set the initial estimate of mean shape $\hat{\mu}$ as X_1'.

iii. Rotate each X_i' to fit $\hat{\mu}$ by the least squares rotation in (3.2) to give X_i''

iv. Set all the starting individual scale factors $\beta_i = 1$.

v. Compute a new mean configuration $\hat{\mu}$ as the arithmetic mean of the X_i''.

Having obtained an initial mean and aligned the configurations approximately to it, calculate RSS_1 the initial residual sum of squared Euclidean distances of all configurations about $\hat{\mu}$. Setting $X_i = X_i''$, the following sequence of steps is then iterated.

1. Rotate each X_i to fit $\hat{\mu}$ by the least squares rotation in (3.2) to obtain $X_i \Gamma_i$.
2. Scale each $X_i \Gamma_i$ by β_i to obtain $X_i^* = \beta_i X_i \Gamma_i$.
3. Compute a new estimate of the mean $\hat{\mu}^*$ as arithmetic mean of X_i^*.
4. Scale to $X_i^{**} = \frac{\beta_i^*}{\beta_i} X_i^* = \beta_i^* X_i \Gamma_i$, where $\frac{\beta_i^*}{\beta_i} = \frac{\sqrt{tr(X_i^* \hat{\mu}^{*T})}}{\sqrt{tr(X_i^* X_i^{*T})}\sqrt{tr(\hat{\mu}^* \hat{\mu}^{*T})}}$.

   (This ensures that $\sum_{i=1}^{n} tr(X_i^{**} X_i^{**T}) = n$).

5. Compute new estimate of the mean $\hat{\mu}^{**}$ as the arithmetic mean of the X_i^{**}.
6. Compute RSS_r, the residual sum of squared Euclidean distances about $\hat{\mu}^{**}$ for iteration *r*.
7. If $RSS_r - RSS_{r-1}$ is less than some pre-specified tolerance, iteration is complete and $X_i^P = X_i^{**}$, $\hat{\mu} = \hat{\mu}^{**}$ (with $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i^P$). Otherwise set $\hat{\mu} = \hat{\mu}^{**}$, $X_i = X_i^{**}$, $\beta_i = \beta_i^*$ and repeat. Gower (1975) and Rohlf & Slice (1990) suggest a tolerance of 0.01 or 0.001 to be satisfactory.

As described above, for convenience the final $\hat{\mu}$ and X_i^P may then be rotated for easier visualisation.

3.2.2.2.2 Ten Berge's modification

Ten Berge (1977) suggested modifications of Gower's original method, also described in Goodall (1991) and Dryden & Mardia (1998). Instead of constraint (ii) in 3.2.2.1, here constraint (iii) must be satisfied.

Having first centred the X_i (i.e. setting $X_i = CX_i$) to remove location, once and for all, the iterative steps are as follows:

1. For each configuration *i*, let $\bar{X}_{(i)} = \frac{1}{n-1}\sum_{j \neq i} X_j$, and calculate $X_i^* = X_i \Gamma_i$, where Γ_i rotates X_i onto $\bar{X}_{(i)}$ by the OLS formula (3.3).
2. Repeat step 1 until (3.5) stops decreasing
3. Let Φ be the correlation matrix of $vec(X_i^*)$, the vectorised coordinates of configuration X_i^*, with principal eigenvector ϕ corresponding to the largest eigenvalue. Then take:

$$\hat{\beta}_i = \left(\frac{\sum_{i=1}^{n} \|X_i^*\|^2}{\|X_i^*\|^2} \right)^{1/2} \phi_i \text{ and calculate } X_i^{**} = \hat{\beta}_i X_i^* \text{ for each configuration } i.$$

(This ensures $\sum_{i=1}^{n}\|X_i^{**}\|^2=\sum_{i=1}^{n}\|X_i\|^2$)

4. Compute RSS_r, the residual sum of squared Euclidean distances of the X_i^{**} about $\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}X_i^{**}$, for iteration r.

5. If $RSS_r - RSS_{r-1}$ is less than some pre-specified tolerance, iteration is complete with $X_i^P=X_i^{**}$ and $\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}X_i^P$. Otherwise set $X_i=X_i^{**}$ and repeat from 1.

3.2.2.2.3 Comparison of algorithms

The main difference between the two iterative GPA methods is that the $\hat{\mu}$ and X_i^P obtained using the second algorithm (method (iii) in 3.2.2.1) are on the scale of the original configurations, X_i, whereas, with the first algorithm (method (ii) in 3.2.2.1), the X_i^P are on a scale consistent with those obtained with use of the constraint $\|\mu\|=1$ (see section 3.2.2.5 and below). However, note that if the X_i are pre-scaled to unit size when using the second approach then constraint (iii) equals constraint (ii) (since then $\sum_{i=1}^{n}\|X_i\|^2=n$) and identical configurations are produced for the $\hat{\mu}$ and X_i^P.

The approach of Gower/Rohlf & Slice (GPA method (ii)), is used by Rohlf's TPS series of programs (see section 2.5) to obtain $\hat{\mu}$ and the X_i^P. In some components of the software, $\hat{\mu}$ is then scaled to size 1, to satisfy constraint (i) in 3.2.2.1 and the X_i^P re-superimposed to $\hat{\mu}$ using the OLS formulae of section 3.2.1 (GPA method (i)). For 2D data, this then gives identical results to the explicit complex eigenvector approach to GPA method (i) for m=2, described in 3.2.2.3 below.

Ten Berge's modification (method (iii)) is used by default for m>2 in Dryden's shape package (see section 2.5) and is as one of two options for m=2 (the other being the explicit eigenvector solution).

3.2.2.3 Explicit eigenvector solution using complex notation (m=2)

The use of complex notation for OPA and GPA (method (i)) in m=2 dimensions leads to some useful simplifications, as described by Dryden & Mardia (1998) and Bookstein (1991). Writing centred configurations X_C and Y_C as (centred) complex k-vectors $x_C=(x_{C_{1x}}+ix_{C_{1y}},...,x_{C_{kx}}+ix_{C_{ky}})$ and $y_C=(y_{C_{1x}}+iy_{C_{1y}},...,y_{C_{kx}}+iy_{C_{ky}})$, the Ordinary full Procrustes fit of X_C onto Y_C may be obtained by complex linear regression, with equation:

$$y_C=(a+ib)1_k+\beta e^{i\theta}x_C+\varepsilon \tag{3.9}$$

where $(a+ib)1_k$ is the translation vector, β is the scaling factor and $e^{i\theta}=\cos\theta+i\sin\theta$, $0<\theta<2\pi$, rotates each entry in complex vector x_C. ε is a complex error vector. Dryden & Mardia (1998) show that by minimising the least squares objective function,

$$\varepsilon^*\varepsilon=(y_C-(a+ib)1_k-\beta e^{i\theta}x_C)^*(y_C-(a+ib)1_k-\beta e^{i\theta}x_C)) \tag{3.10}$$

where a^* denotes the transpose of the complex conjugate of a, the full Procrustes superimposition of x onto y has parameters:

$(\hat{a}+i\hat{b})=0$ (since x_C and y_C centred), $\hat{\theta}=\arg(x_C^*y_C)$ and $\hat{\beta}=\dfrac{(x_C^*y_Cy_C^*x_C)^{1/2}}{(x_C^*x_C)}$.

The minimised value of (3.10) obtained by these estimates is equal to $OSS(X,Y)$ as in section 3.2.1 and is symmetric (gives the same value as when Y is superimposed to X) if $x_C^*x_C=y_C^*y_C$. As before, pre-scaling X and Y (and hence x_C and y_C) to be unit size ensures that $OSS(X,Y)=OSS(Y,X)=d_F^2(X,Y)$.

To obtain the estimate of mean shape $\hat{\mu}$ and corresponding Procrustes 'fits', from a sample of centred complex vector configurations $x_{C_1},...,x_{C_n}$, we first find $\hat{\mu}$ optimising (3.5) by using Kent (1994), who noted that $\hat{\mu}$ is given explicitly as the complex

eigenvector corresponding to the largest eigenvalue of the sum of squares and cross product matrix:

$$S = \sum_{i=1}^{n} \frac{x_i x_i^*}{x_i^* x_i} = \sum_{i=1}^{n} \frac{x_i}{\|x_i\|} \left(\frac{x_i}{\|x_i\|} \right)^*$$

with:

$$\hat{\mu} = \arg \sup_{\|\mu\|=1} \mu^* S \mu .$$

The corresponding Procrustes fits are then obtained by calculating the full Procrustes superimposition of each x_{C_i} to $\hat{\mu}$ (of unit size) using the formulae above.

3.2.2.4 Example

Fig 3.2 displays the full Procrustes fits and estimated mean shape of 20 upper right central incisors, with landmarks defined as in fig 1.3 (left). The configurations are (arbitrarily) orientated so that the mesio-distal landmarks of the mean shape are horizontal, with gingival landmarks at the top. The algorithm described in 3.2.2.2.1 (GPA method (ii)) was used to obtain $\hat{\mu}$ and the X_i^P. However, similar plots obtained using either of the other GPA methods would produce practically identical displays, the only difference being that the displays would be on the scale of the X_i, if the second algorithm (method (ii)) from 3.2.2.2 had been used.

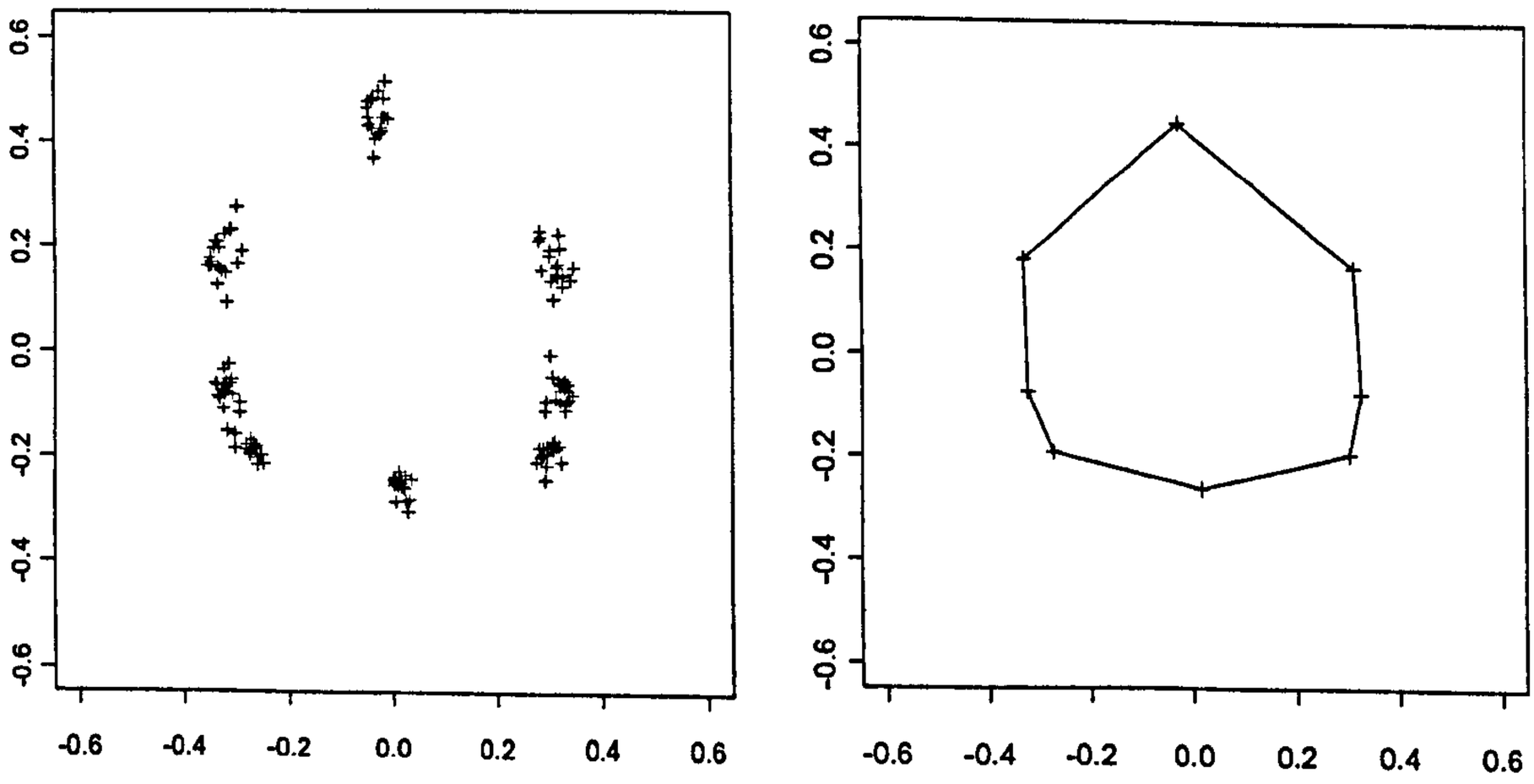


Fig. 3.2: (left) Procrustes fits and (right) mean shape of 20 upper central incisors

3.2.2.5 Differences between results of the GPA methods

Having rotated the resulting centred $\hat{\mu}$ from the different GPA methods to some common arbitrary orientation, differences in the sizes of $\hat{\mu}$ will still exist, depending on the method and constraint used. However, each estimate $\hat{\mu}$ produced by the GPA methods always has the same 'shape' (each estimate is a member of the same shape set, differing only in size), producing the same minimised sum of squared full Procrustes distances, $\sum_{i=1}^{n} d_F^2(X_i, \hat{\mu})$ in (3.6). (The same optimised value results, since $d_F^2(X_i, \hat{\mu})$ is invariant to differences in the sizes of the $\hat{\mu}$).

The full Procrustes fits, $X_i^P = \hat{\beta}_i X_i \hat{\Gamma}_i + 1_k \hat{\gamma}_i^T$ to each $\hat{\mu}$ are obtained as the transformed versions of each configuration which minimise (3.7), subject to any constraints on the sizes of the $\beta_i X_i \Gamma_i + 1_k \gamma_i^T$. Assuming the (centred) $\hat{\mu}$ have been rotated to some common arbitrary orientation, the associated X_i^P (also centred) always have the same optimal rotation $\hat{\Gamma}_i$ to $\hat{\mu}$, given by its OLS estimator ($\hat{\Gamma}_i$ is not dependent on the sizes of $\hat{\mu}$ or X_i), but the scaling $\hat{\beta}_i$, minimising (3.7), results in different sizes of the X_i^P, depending on the constraint used.

Using method (i) in 3.2.2.1 (with constraint $\|\mu\|=1$), the optimal superimpositions of each X_i to $\hat{\mu}$, are given by the OLS estimates of β_i, Γ_i and γ_i from 3.2.1 and since $\hat{\mu}$ is of unit size, $\left\|X_i^P - \hat{\mu}\right\|^2 = d_F^2(X_i, \hat{\mu})$. Therefore,

$$\min_{\substack{\beta_i, \Gamma_i, \gamma_i \\ \mu: \|\mu\|=1}} \sum_{i=1}^{n} \left\|\beta_i X_i \Gamma_i + 1_k \gamma_i^T - \mu\right\|^2 = \min_{\mu} \sum_{i=1}^{n} d_F^2(X_i, \mu)$$

and so in terms of the summary measures:

$$\sum_{i=1}^{n} \left\|X_i^P - \hat{\mu}\right\|^2 = \sum_{i=1}^{n} d_F^2(X_i, \hat{\mu}) = \min_{\mu} \sum_{i=1}^{n} d_F^2(X_i, \mu).$$

Note also that each X_i^P obtained using the OLS estimator for β_i, results in $\|X_i^P\| < \|\hat{\mu}\| = 1$, as described in 3.2.1 (and 3.3.2). Although $\hat{\mu}$ and $\overline{X}^P$ have the same shape, $\hat{\mu} \neq \overline{X}^P$ and so

$$RSS = \sum_{i=1}^{n} \|X_i^P - \overline{X}^P\|^2 \neq \sum_{i=1}^{n} \|X_i^P - \hat{\mu}\|^2 = \min_{\mu} \sum_{i=1}^{n} d_F^2(X_i, \mu),$$

as the X_i^P were obtained to optimise the OLS superimposition to $\hat{\mu}$, not $\overline{X}^P$. However, for small variations in shape, $RSS \approx \sum_{i=1}^{n} \|X_i^P - \hat{\mu}\|^2$. (See section 3.3).

Using methods (ii) and (iii) (with $\sum_{i=1}^{n} \|\beta_i X_i \Gamma_i + 1_k \gamma_i^T\|^2 = n$ or $\sum_{i=1}^{n} \|X_i\|^2$), $\hat{\mu} = \overline{X}^P$ but unlike above, is not necessarily of unit centroid size. The X_i^P optimising (3.7) and (3.8) are not the OLS superimpositions of the X_i to $\hat{\mu}$, i.e. each X_i^P does not minimise $\|\beta_i X_i \Gamma_i + 1_k \gamma_i^T - \hat{\mu}\|^2$. Regardless of the constraint used, the rotations $\hat{\Gamma}_i$ and translations $\hat{\gamma}_i$ will be OLS estimators but the $\hat{\beta}_i$ are not, due to the constraints on the sum of squared centroid sizes of the superimposed configurations. For the same reason and because $\|\hat{\mu}\|$ is not necessarily equal to 1, $\|X_i^P - \hat{\mu}\|^2 \neq d_F^2(X_i, \hat{\mu})$.

Despite this however, constraint (ii) actually ensures that for the summation of these terms:

$$\min_{\substack{\beta_i, \Gamma_i, \gamma_i \in \Omega \\ \mu = \frac{1}{n}\sum_{i=1}^{n}(\beta_i X_i \Gamma_i + 1_k \gamma_i^T)}} \sum_{i=1}^{n} \|(\beta_i X_i \Gamma_i + 1_k \gamma_i^T) - \mu\|^2 = \min_{\mu} \sum_{i=1}^{n} d_F^2(X_i, \mu) \qquad (3.11)$$

where $\Omega = \left\{ \beta_i, \Gamma_i, \gamma_i : \sum_{i=1}^{n} \|\beta_i X_i \Gamma_i + 1_k \gamma_i^T\|^2 = n \right\}$, and so in terms of the summary measures:

$$\sum_{i=1}^{n} \|X_i^P - \hat{\mu}\|^2 = \sum_{i=1}^{n} d_F^2(X_i, \hat{\mu}) = \min_{\mu} \sum_{i=1}^{n} d_F^2(X_i, \mu).$$

Since $\hat{\mu} = \overline{X}^P$ we also have:

$$RSS = \sum_{i=1}^{n} \left\| X_i^P - \overline{X}^P \right\|^2 = \sum_{i=1}^{n} \left\| X_i^P - \hat{\mu} \right\|^2 = \min_{\mu} \sum_{i=1}^{n} d_F^2(X_i, \mu).$$

For method (iii),

$$\min_{\substack{\beta_i, \Gamma_i, \gamma_i \in \Omega' \\ \mu = \frac{1}{n}\sum_{i=1}^{n}(\beta_i X_i \Gamma_i + 1_k \gamma_i^T)}} \sum_{i=1}^{n} \left\| (\beta_i X_i \Gamma_i + 1_k \gamma_i^T) - \mu \right\|^2 \propto \min_{\substack{\beta_i, \Gamma_i, \gamma_i \in \Omega \\ \mu = \frac{1}{n}\sum_{i=1}^{n}(\beta_i X_i \Gamma_i + 1_k \gamma_i^T)}} \sum_{i=1}^{n} \left\| (\beta_i X_i \Gamma_i + 1_k \gamma_i^T) - \mu \right\|^2,$$

where $\Omega' = \left\{ \beta_i, \Gamma_i, \gamma_i : \sum_{i=1}^{n} \left\| \beta_i X_i \Gamma_i + 1_k \gamma_i^T \right\|^2 = \sum_{i=1}^{n} \left\| X_i \right\|^2 \right\}$ and Ω is as defined above, with equality if the X_i are pre-scaled to unit size, prior to carrying out the GPA method in 3.2.2.2.2 (so that the constraint $\sum_{i=1}^{n} \left\| \beta_i X_i \Gamma_i + 1_k \gamma_i^T \right\|^2 = \sum_{i=1}^{n} \left\| X_i \right\|^2 = n$, as in method (ii)), or when a constant of proportionality is introduced equal to the ratio of the squared centroid sizes of the estimated means of the two GPA methods. Without pre-scaling, the X_i^P, $\hat{\mu}$ and value of the LHS of the above are on the scale of the original X_i.

By way of (3.11) and the equation above, we also have:

$$\min_{\substack{\beta_i, \Gamma_i, \gamma_i \in \Omega' \\ \mu = \frac{1}{n}\sum_{i=1}^{n}(\beta_i X_i \Gamma_i + 1_k \gamma_i^T)}} \sum_{i=1}^{n} \left\| (\beta_i X_i \Gamma_i + 1_k \gamma_i^T) - \mu \right\|^2 \propto \sum_{i=1}^{n} d_F^2(X_i, \hat{\mu}) = \min_{\mu} \sum_{i=1}^{n} d_F^2(X_i, \mu)$$

and so in terms of the summary measures:

$$\sum_{i=1}^{n} \left\| X_i^P - \hat{\mu} \right\|^2 \propto \min_{\mu} \sum_{i=1}^{n} d_F^2(X_i, \mu)$$

and since $\hat{\mu} = \overline{X}^P$,

$$RSS = \sum_{i=1}^{n} \left\| X_i^P - \overline{X}^P \right\|^2 = \sum_{i=1}^{n} \left\| X_i^P - \hat{\mu} \right\|^2 \propto \min_{\mu} \sum_{i=1}^{n} d_F^2(X_i, \mu).$$

In practice it does not matter which method is used, unless one wishes the results to be on the scale of the original data. All provide the same estimate of mean shape and align the configurations in such a way that visualisation and measurement of the scatter of Procrustes fits is indistinguishable between methods, except for scale. Both methods (i) and (ii) provide an estimate of mean shape and corresponding set of Procrustes fits with:

$$\sum_{i=1}^{n}\left\|X_i^P-\hat{\mu}\right\|^2=\sum_{i=1}^{n}d_F^2(X_i,\hat{\mu})=\min_{\mu}\sum_{i=1}^{n}d_F^2(X_i,\mu) \tag{3.12}$$

(although $\sum_{i=1}^{n}\left\|X_i^P-\overline{X}_i^P\right\|^2\neq\sum_{i=1}^{n}d_F^2(X_i,\hat{\mu})$ if method (i) is used), with the results of method (iii) differing only by a factor of scale from those of method (ii).

3.2.3 Alternative Procrustes methods

3.2.3.1 Partial Procrustes analysis (Procrustes without scaling)

Ordinary partial Procrustes involves superimposition of $k\times m$ matrix configuration X onto Y by just translation and rotation. Here we seek to minimise the sum of squared Euclidean distances between corresponding landmarks of Y and $X\Gamma+1_k\gamma^T$, i.e.

$$\min_{\Gamma,\gamma}\left\|Y-(X\Gamma+1_k\gamma^T)\right\|^2. \tag{3.13}$$

The estimates of γ, and Γ are obtained by the same procedure as in the full Procrustes case. If, X and Y have already been 'centred' to the origin (0,0), so that $X=X_C$ and $Y=Y_C$, then $\hat{\gamma}=0$ and (3.2) is used to obtain the rotation $\hat{\Gamma}$ (recall that $\hat{\Gamma}$ is independent of the size of X or Y).

The minimum of (3.13) is denoted as $OSS_P(X,Y)$. If the two configurations are also pre-scaled to unit size then

$$OSS_P(X/\|X\|,Y/\|Y\|)=OSS_P(Y/\|Y\|,X/\|X\|)=d_P^2(X,Y), \tag{3.14}$$

the 'squared partial Procrustes distance' between X and Y. Further properties of the 'partial Procrustes distance' are considered in section 3.3.2.

Referring back to fig. 3.1, the middle bottom row image corresponds to the partial Procrustes fit of configuration 2 onto configuration 1 (after pre-scaling to unit size),

before scaling takes place in the subsequent bottom right image. Both unit size configurations have been matched over only rotations and location.

For generalised partial Procrustes superimposition of configurations $X_1,...,X_n$, $n\geq 2$, which must be commensurate in scale and may be a sample from a population with mean shape μ, we minimise:

$$\sum_{i=1}^{n}\left\|(X_i\Gamma_i + 1_k\gamma_i^T) - \mu\right\|^2 \qquad (3.15)$$

over rotations Γ_i and translations γ_i^T to an unknown configuration $\mu \in SG(\mu)$, to obtain the partial Procrustes fits $X_i^{PP} = X_i\hat{\Gamma}_i + 1_k\hat{\gamma}_i^T$ to the estimated mean shape $\hat{\mu}$.

This may be appropriate in joint studies of size and shape, when the model assumed for the 'population' of shapes is of the form:

$$X_i = (\mu + E_i)\,\Gamma_i' + 1_k\gamma_i'^T$$

where perturbation $(\mu + E_i)$ is rotated and translated (by Γ_i' and γ_i') and observed as X_i and the E_i are again zero mean $k\times m$ independent random error matrices, as in (3.5). If all the configurations have unit size then minimisation of (3.15) is equivalent to minimizing $\sum_{i=1}^{n} d_P^2(X_i,\mu)$ over μ.

Both sets of software by Dryden and Rohlf allow computation of 'partial Procrustes fits', after obtaining an estimate of mean shape, $\hat{\mu}$ with $\|\hat{\mu}\|=1$, by GPA. Each X_i is centered and scaled to unit size and then rotated to $\hat{\mu}$ using (3.2), so that (3.14) is true.

3.2.3.2 Procrustes superimposition with reflection

When using Procrustes matching to compare matrices in multivariate statistics, the possibility of reflecting the axes is also permitted in addition to translation, rotating and scaling, if an improvement to the fit can be made, since orientation of the axes of many

variables is usually arbitrary. This could be considered as another similarity transformation of shape in all the methods above by using orthogonal matrices for Γ in the equations above, with $|\Gamma| = \pm 1$. In this study however, it makes no sense to allow reflections to be permitted and for datasets with small variation in shape (as we expect), with full rank configurations, there will be no difference between the approaches anyway (Dryden & Mardia, 1998, Goodall, 1991).

3.3 Shape space and shape distances

This section describes the theoretical connections of the methods in 3.2 with shape theory, in particular the shape and pre-shape space of Kendall (1984) and its link with the removal of similarity transformations, as first suggested by Bookstein (1986). Having had their location, rotation and scale differences removed, the registered landmark configurations can be represented as points in shape space. Minimisation of distances between these points corresponds precisely with the minimisation of sums of squared Euclidean distances in configuration space, as described in the preceding section. It is this link, which provided the theoretical setting for the development and adaptation of inferential and descriptive statistical methods, that then followed in the literature (see section 3.4).

3.3.1 Pre-shape and shape space

The pre-shape Z of a configuration X is given by standardising the configuration for location and size. All information about location and scale may be removed by centring the configuration to (0,0), by pre-multiplication by centring matrix $C = I_k - (1/k)1_k 1_k^T$ and dividing it by its centroid size. The pre-shape of X may therefore be given by:

$$Z = \frac{CX}{\|CX\|}. \tag{3.16}$$

The term 'pre-shape' indicates that we are one step away from removing all the similarity transformations to obtain the 'shape' of X (rotation still has to be removed).

The pre-shape space is the space of all possible pre-shapes Z, i.e. the space of configurations that have been centred and scaled (to unit size), and is a hypersphere of radius 1, since $\|Z\|=1$. (A 'hypersphere' is the generalisation of a sphere into any number of dimensions). The pre-shape space has dimension *km*-*m*-1 since we initially have *km* coordinates but lose *m* for location and one for scale.

The equivalence class of all rotated versions of the pre-shape $Z\Gamma$ is then the 'shape' of X and the 'shape space' the set of all possible shapes. Kendall (1984) showed that shape space is a *km*-*m*-1-*m*(*m*-1)/2 dimensional 'manifold', a generalisation to arbitrary dimensions of a curved surface in three dimensions. We lose another *m*(*m*-1)/2 dimensions from that of the pre-shape space for rotation. Each 'point' in the shape space then represents the shape of a configuration in Euclidean space, irrespective of size, location and orientation.

Rohlf (1999) and Dryden & Mardia (1998) demonstrate that for $k=3$ landmark configurations in $m=2$ dimensions, the shape space can be visualised as a sphere with the objects as points on its surface and the corresponding pre-shape as a unit hemisphere. For $k>3$ the geometry of the space is substantially more complicated and cannot be so easily visualised, being both high dimensional and non-linear (Goodall, 1991, Kendall, 1984).

In the following section we describe how shape may also be represented on the pre-shape hypersphere. The rotation of Z on the pre-shape, traces out a 'fibre' on the unit hypersphere, which we can think of as representing the shape of a configuration. This fibre will not overlap with fibres of other pre-shapes.

As in 3.2 the term 'icon' is used to denote a representative member of a particular shape set, e.g. the centred pre-shape Z_X of configuration X is an 'icon' of the shape set of X.

3.3.2 Shape distances

A metric in shape space will allow assessment of differences among shapes. Since shapes can be represented as fibres on the pre-shape hyper sphere, we can also consider

distances on the pre-shape hyper sphere. As the pre-shapes are rotated along their fibres, the closest distance attainable between positions along the fibres can be considered as the distance between the shapes and measured in terms of familiar formulae for distances between points on a sphere (here of unit radius).

It can be shown that calculation of these distances between pre-shapes is equivalent to minimising the sum of squared differences between corresponding landmarks of the two centred unit size configurations in Euclidean configuration space, as described in section 3.2.1, to obtain the full and partial Procrustes distances, $d_F(X,Y)$ and $d_P(X,Y)$. This is illustrated by figures 3.3 and 3.4 below.

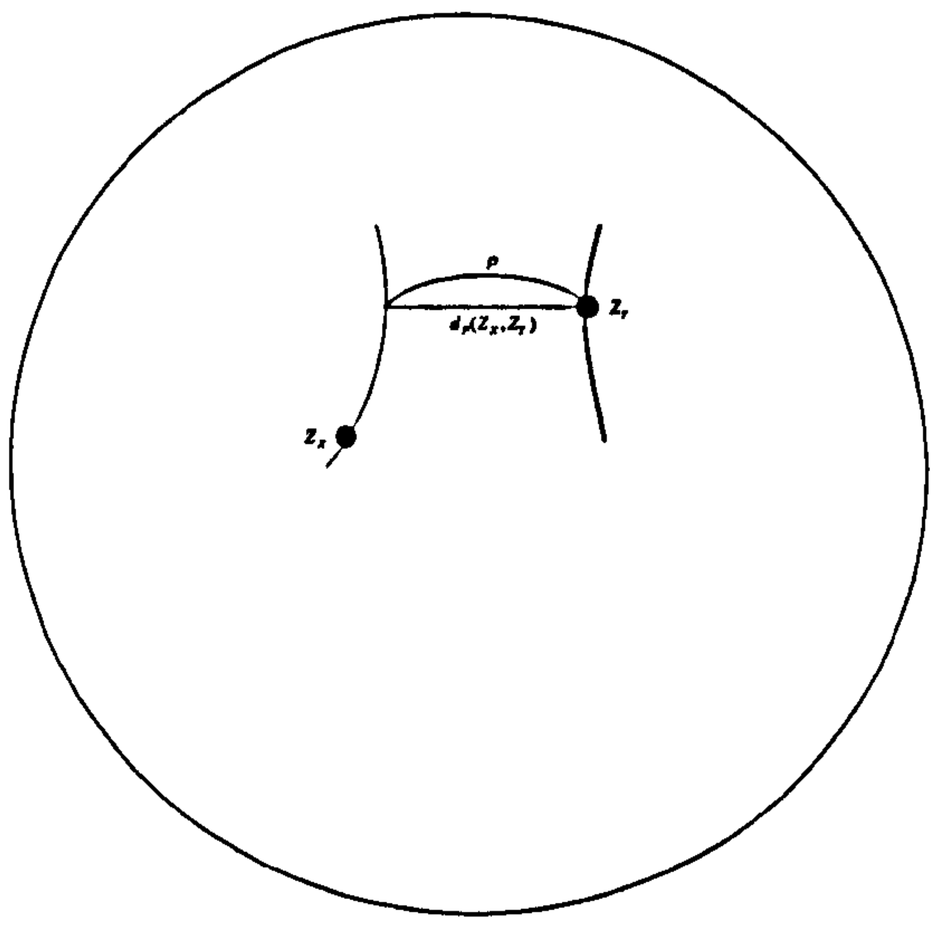


Fig 3.3: Fibre of pre-shape Z_X and Z_Y on the pre-shape sphere with closest great circle distance ρ and closest chordal distance $d_P(X,Y)$ between fibres

Fig. 3.3, reproduced from Dryden & Mardia (1998) shows a simplistic, illustrative view of two fibres on the pre-shape (hyper) sphere representing rotations of the pre-shapes Z_X and Z_Y corresponding to original configuration matrices X and Y. The closest great circle distance between rotations of Z_X and Z_Y on the pre-shape is denoted as ρ. The closest chordal distance is $2\sin(\rho/2)$ and is equal to the partial Procrustes distance $d_P(X,Y)$ in section 3.2.3.1.

Fig 3.4 shows an illustrative cross-section (hemisphere) of the pre-shape (hyper) sphere, with the measures ρ and d_P in more detail. In fact for k=3, m=2 this is the pre-shape space. The centred, unit size pre-shape of X is rotated to the pre-shape of Y until ρ or $d_P(X,Y)$ is minimised. If we allow further scaling of pre-shape Z_X in matching to Z_Y we minimise the full Procrustes distance $d_F(X,Y)$ of (3.4) as shown, with $d_F \leq d_P$.

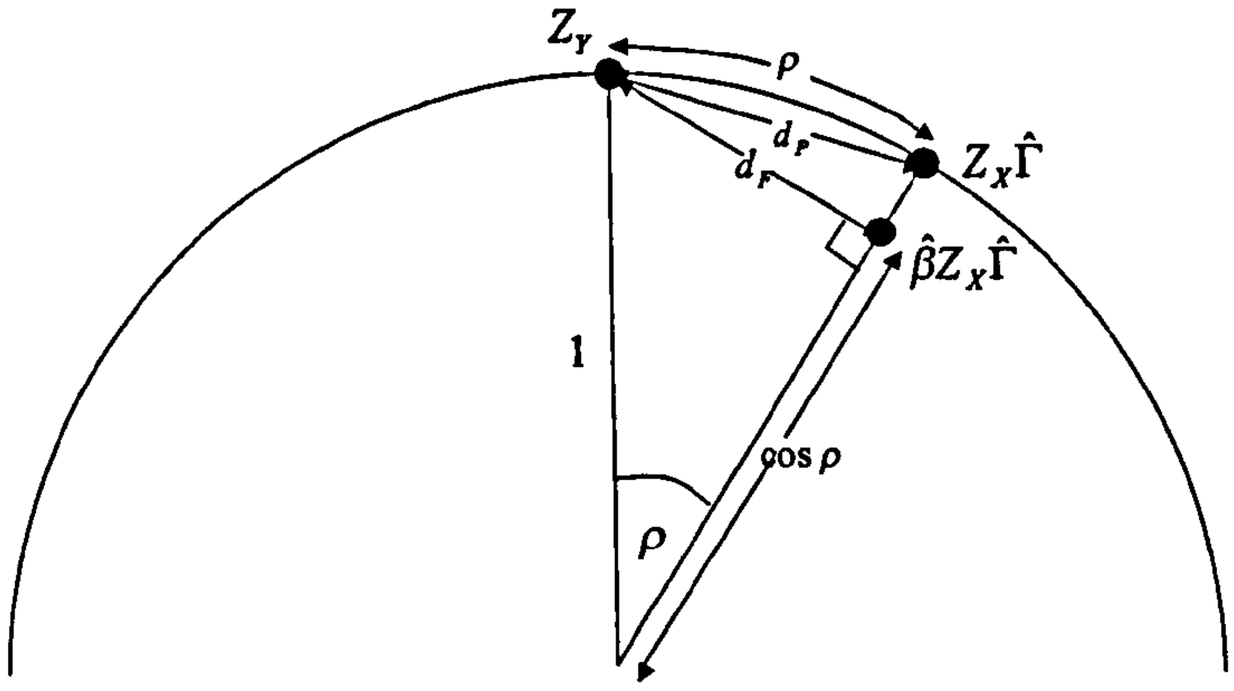


Fig 3.4: Simplistic cross section view of the pre-shape hyper sphere, illustrating the various Procrustes distances.

Algebraically;

$$\min d_F(X,Y) = \|Z_Y - \beta Z_X\Gamma\| = \left\|\frac{CY}{\|CY\|} - \beta\frac{CX}{\|CX\|}\Gamma\right\| = OSS(X/\|X\|, Y/\|Y\|)^{1/2} \quad (3.17)$$

and

$$\min d_P(X,Y) = \|Z_Y - Z_X\Gamma\| = \left\|\frac{CY}{\|CY\|} - \frac{CX}{\|CX\|}\Gamma\right\| = OSS_P(X/\|X\|, Y/\|Y\|)^{1/2}. \quad (3.18)$$

The optimal scaling of Z_X to Z_Y by $\hat{\beta}$ results in a fitted configuration of size $\cos\rho$ (and so the fitted configuration $\hat{\beta}Z_X\hat{\Gamma}$ will always have size $\leq \|Z_Y\|$=1). As described in 3.2.2, when the Procrustes estimate of mean shape $\hat{\mu}$ has unit centroid size, the full Procrustes fits of the configurations X_i to $\hat{\mu}$, each minimising (3.1), all have $\|X_i^P\|$<1.

Note that if we had set out to minimise ρ, d_P, or d_F then the optimal rotation to Z_Y on the pre-shape would be the same in each case, as seen is section 3.2.1. For shapes

which are close together there is usually very little difference between the three shape distances.

The steps of full and partial Procrustes matching were presented in fig. 3.1. In the current context, the third row shows the two pre-shapes Z_X and Z_Y of X and Y, after being centred and scaled to unit size. In the final row the Euclidean sums of squared distances between corresponding landmarks are minimised to give the partial (both scales fixed at 1) or full (allowing rescaling of Z_X) Procrustes distance between X to Y.

ρ is called the Procrustes angle or Procrustes distance between shapes and, as noted by Dryden & Mardia (1998) or Bookstein (1998) as well as others, is a natural distance inherited from the projection of fibres on the pre-shape sphere to points in the shape space. In geometric terms, this projection is termed a 'Riemannian submersion'; the metric is the same in both pre-shape space and shape space. In the following section we illustrate how most multivariate methods of analysis may be applied to shapes by considering linearisations of the distances ρ, d_P, or d_F in Kendall's shape space.

3.4 Shape variation: Description and Inference

In theory, the non-Euclidean geometry of shape space makes it difficult to use standard multivariate methods directly on the Procrustes fits as these generally rely on the use of Euclidean metrics as measures of differences and covariance. Specific models and probability distributions for use in shape space have been developed to take account of the non-Euclidean geometry and these are briefly summarised in section 3.4.3. However, under certain assumptions, it is possible to approximate the shape space by way of linearised 'tangent space' to the shape space, that has a Euclidean geometry and allows us to approximate distances in shape space by Euclidean distances. It is then possible to utilise any of the conventional multivariate statistical techniques to investigate variations in shape.

This section defines the 'tangent space' and discusses various choices of tangent coordinates, including how the Procrustes fits from a generalised full (or partial)

Procrustes analysis may be used as approximate tangent coordinates. We then illustrate how the tangent/approximate tangent coordinates may be used to address hypotheses concerning shape, by using a sample of buccal surface configurations from central incisors of patients with hypodontia and testing for a difference in mean buccal surface shape when compared to a corresponding control group.

3.4.1 Tangent space

Tangent space is a linearised version of the shape space at some particular point in shape space. This point is known as the 'pole' of projection and is usually chosen to be the average shape. The tangent space can be visualised as a sphere (the shape space) resting on a flat piece of paper (the tangent space). The scatter of points representing variation in shape on the (hyper) sphere, are projected onto the tangent plane in the same way as cartographer might project a map from a globe onto a flat piece of paper. Following Dryden & Mardia (1998), Rohlf (1999) or Slice (2001), we may equivalently consider a tangent projection from the pre-shape (hyper) sphere that does not depend on the original rotation of the configuration.

Consider configurations X_i with pre-shapes Z_i, i=1,...,n, and an estimate of average shape $\hat{\mu}$ with pre-shape $Z_{\hat{\mu}}$. Each Z_i is rotated on the pre-shape sphere and scaled to be as close as possible to $Z_{\hat{\mu}}$ as in fig 3.4 (minimising (3.17)) and then projected onto the tangent plane at pole $Z_{\hat{\mu}}$ to obtain a vector of full Procrustes tangent coordinates given by:

$$v_i^F = (I_{km-m} - vec(Z_{\hat{\mu}})vec(Z_{\hat{\mu}})^T)vec(\hat{\beta}Z\hat{\Gamma}) \qquad (3.19)$$

with $v_i^{F^T} Z_{\hat{\mu}} = 0$. If pre-shape Z_i is only rotated to $Z_{\hat{\mu}}$ and β =1 fixed when matching then the projection onto the tangent plane at $Z_{\hat{\mu}}$ gives a vector of partial Procrustes tangent coordinates (Kent, 1994):

$$v_i^P = (I_{km-m} - vec(Z_{\hat{\mu}})vec(Z_{\hat{\mu}})^T)vec(Z\hat{\Gamma}) \qquad (3.20)$$

with $v_i^{P^T} Z_{\hat{\mu}} = 0$. The matrix $(I_{km-m} - vec(Z_{\hat{\mu}})vec(Z_{\hat{\mu}})^T)$ in (3.19) and (3.20) is the matrix for projection onto the space orthogonal to $Z_{\hat{\mu}}$.

The coordinates of a configuration as the point on the hyper sphere have now become coordinates in the tangent plane. Both the full and partial Procrustes tangent coordinates can be visualised as in fig 3.5.

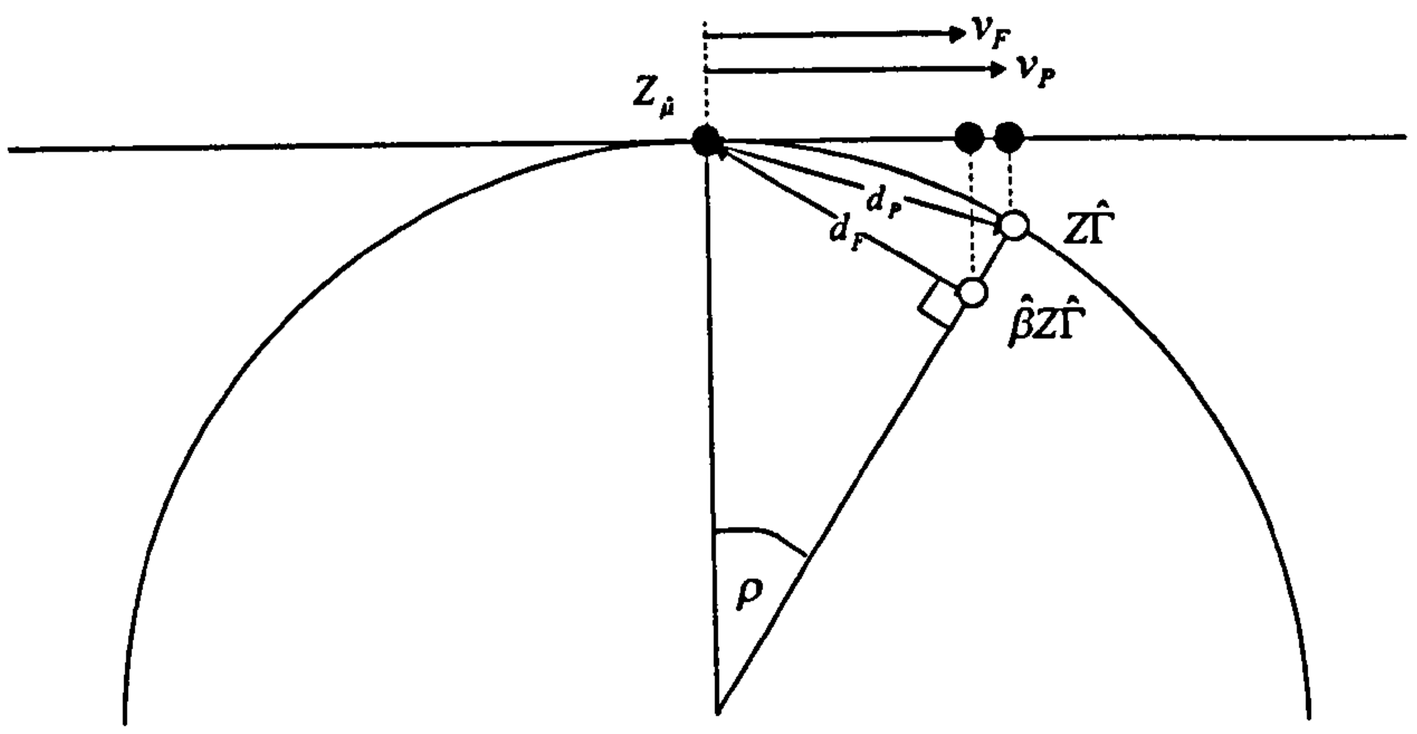


Fig 3.5: Simplistic view of the pre-shape (hyper) sphere showing the full and partial Procrustes tangent coordinates.

Note that when plotted, $vec_m^{-1}(v^F)$ or $vec_m^{-1}(v^P)$ will not have the same geometrical appearance as the shape of the original or pre-shape of the configuration. To visualise what tangent coordinates and results of operations on them look like, we need to reconstruct to give them the correct geometric appearance. It is possible to recover any shape from tangent space because each shape maps to a unique position in tangent space. The reverse transformation from coordinates $v = v^F$ or $v = v^P$ in tangent space to an icon configuration is given by:

$$vec_m^{-1}((1 - v^T v)^{1/2} vec(Z_{\hat{\mu}}) + v). \qquad (3.21)$$

The most important result regarding tangent space is as follows. If variation in shape is small, then Euclidean distances between points in tangent space can be used to

approximate non-Euclidean shape distances ρ, d_P or d_F in shape space. Dryden & Mardia (1998) show that with pole $Z_{\hat{\mu}}$,

$$\|v^P\| = d_F(X, \hat{\mu})$$

and close to $Z_{\hat{\mu}}$,

$$\|v^F\| = \cos\rho \sin\rho \approx d_F(X, \hat{\mu}) \approx \rho(X, \hat{\mu}) \approx d_P(X, \hat{\mu}). \qquad (3.22)$$

Consequently, for two configurations X_1 and X_2 that are close in shape, with (full or partial) tangent coordinates v_1 and v_2:

$$\|v_1 - v_2\| \approx d_F(X_1, X_2) \approx \rho(X_1, X_2) \approx d_P(X_1, X_2). \qquad (3.23)$$

This means that for post-superimposition analysis of shape, it is possible to use standard multivariate statistical approaches to exploration and inference based on Euclidean distances between coordinates in the tangent plane, as long as the projection has not encompassed a large proportion of the sphere. i.e., providing variation in shape about the pole is small.

An important approximation to tangent space when the pole is the pre-shape of the full Procrustes estimate of mean shape is as follows (Dryden & Mardia, 1998). After the full Procrustes coordinates $X_1^P, \ldots, X_n^P$ have been obtained by any of the GPA methods described in 3.2.2, the Procrustes residuals about $\hat{\mu}$, defined already in 3.2.2 as $R_i^P = X_i^P - \bar{X}_i^P$, i=1,…,n and here denoted as $r_i^P = vec(R_i)$, are approximate tangent coordinates.

If GPA method (i), with constraint $\|\mu\|$=1 was used, where the X_i^P are the OLS superimpositions to $\hat{\mu}$, using equations (3.1)-(3.3), then:

$$\|r_i^P\| \approx \|X_i^P - \hat{\mu}\| = d_F(X, \hat{\mu}).$$

Using GPA methods (ii) or (iii), the X_i^P are not the OLS superimpositions to $\hat{\mu}$ due to the constraint $\sum_{i=1}^{n}\left\|\beta_i X_i \Gamma_i + 1_k \gamma_i^T\right\|^2 = n$ or $\sum_{i=1}^{n}\|X_i\|^2$ (as described in 3.2.2.5) and since $\|\hat{\mu}\|$ is not necessarily equal to 1, $\left\|X_i^P - \hat{\mu}\right\| \neq d_F(X, \hat{\mu})$. However, for the X_i^P and $\hat{\mu}$ obtained using method (ii), with $\hat{\mu} = \overline{X}^P$,

$$\left\|r_i^P\right\| = \left\|X_i^P - \hat{\mu}\right\| \approx d_F(X, \hat{\mu})$$

and for those obtained using method (iii), for some scalar λ,

$$\left\|r_i^P\right\| = \left\|X_i^P - \hat{\mu}\right\| \approx \lambda\, d_F(X, \hat{\mu}).$$

Another alternative (Rohlf, 1999) could be to use the residuals of the partial Procrustes fits, obtained by minimising (3.15), with $\|\hat{\mu}\| = \|X_i\| = 1$, given by $r_i^{PP} = vec(X_i^{PP} - \overline{X}_i^{PP})$, for which:

$$\left\|r_i^{PP}\right\| \approx \left\|X_i^{PP} - \hat{\mu}\right\| = d_P(X, \hat{\mu}) \approx d_F(X, \hat{\mu}).$$

Consequently, standard multivariate methods can be carried out on the r_i^P (or r_i^{PP}) as well as the v_i^P or v_i^F. To visualise shapes to corresponding to operations on the r_i^P (or r_i^{PP}), we can simply use $vec_m^{-1}(r_i^P) + \hat{\mu}$ (or $vec_m^{-1}(r_i^{PP}) + \hat{\mu}$).

An important question is when are variations small, i.e. when can the curvature of the shape space be ignored and tangent space approximations used? Rohlf (1999) describes how in practical applications, plots of $\left\|v_i - v_j\right\|$ against $d_F(X_i, X_j)$, for all $i \neq j$ or of $\left\|v_i - \overline{v}\right\|$ against $d_F(X_i, \hat{\mu})$, $i=1,\ldots,n$, should have a linear relationship if the approximation is valid. This is performed by Rohlf's 'TPSsmall' routine (see table 2.3).

Dryden's shape package allows calculation of either partial tangent coordinates v_i^P or approximate tangent coordinates r_i^P following GPA using method (i) (m=2 only) or method (iii). Rohlf (1999) presents a comprehensive overview of these and other

possible choices of tangent coordinates, which are available as options in his TPS series of software. These include 'stereographic projections' as well as the use of the partial or full Procrustes fits ($vec(X_i^{PP})$ or $vec(X_i^{P})$) directly, which is equivalent to using the partial or full Procrustes residuals as described above. The stereographic tangent coordinates are constructed by rotating each shape in pre-shape space to the pole (pre-shape of full Procrustes mean) and then scaling each to configuration to have centroid size $1/\cos\rho$ (as opposed to $\cos\rho$ when obtaining the full Procrustes fits, with $\|\mu\|=1$).

Goodall (1991) notes that Bookstein coordinates (see section 2.2.3.1) are a special case of the stereographic projections. Rohlf (1999) suggests however that stereographic tangent coordinates represent a somewhat extreme non-linear projection of shape space unless the points are very close to the pole and that variance is inflated for shapes further from the reference. In light of these concerns we shall not consider stereographic projections further. See Slice (2001) and Small (1996) for an extended discussion of tangent coordinates.

Note however, that Rohlf (1999), Dryden & Mardia (1998), Slice (2001) and many others, all suggest that for the amount of variation in shape normally found in biological data, the different choices of tangent coordinates described in this section, will usually lead to very similar statistical conclusions.

3.4.2 Multivariate statistical methods in tangent space

For shapes that are concentrated in a small region of the shape space, we therefore can carry out most multivariate statistical analysis, such as principal components analysis or tests of group differences in mean shape, on either the tangent coordinates or on approximate tangent coordinates given by the full or partial Procrustes fits (or their residuals).

First note however that the tangent coordinates for the ith configuration v_i, $i=1,...,n$ are not linearly independent because the effects of translation, size and rotation to the reference have been removed. Multivariate statistical methods must take into account the fact that the dimension of the tangent space (and shape space) is $M=km-m-m(m-1)/2-1$ whereas the dimension of each tangent vector is greater

than M and so the covariance matrix of any choice of tangent coordinates will be singular. Consequently, generalised inverses must be used for methods where inversion of the covariance matrix S_v is necessary and adjustments to the degrees of freedom made to take account of the fact that there are fewer independent dimensions than there are shape variables. One option is to use the Moore-Penrose generalized inverse (which will be denoted as S_v^- below). See, for example, Mardia *et al.* (1979). Another solution is to perform a principal components analysis on the data and delete components corresponding to zero eigenvalues. Doing this will remove no information about shape variation, and will produce a new dataset of component scores with a non-singular covariance matrix (so no need for generalised inverses).

3.4.2.1 Example data

The data used in this section to demonstrate some of the different analyses which may be carried out on the tangent (or approximate tangent) coordinates, are two samples of 20 landmark configurations from images of buccal surfaces of upper left central incisors. The first set of configurations were obtained from images of study casts of 20 patients with moderate/severe hypodontia (three or more congenitally missing teeth). These patients' records were being investigated as part of a wider study of hypodontia for which ethics committee approval had been obtained. Teeth partially obscured by crowding or with evidence of attrition were excluded. Six landmarks were identified on each image, as defined in fig. 1.3 (left), with the exception of the mesio-distal endpoints since these are often difficult to locate in this population. The second sample comprises 20 corresponding control configurations, again obtained from images of study casts and following the same selection criteria, with the corresponding landmarks to those identified on the hypodontia images recorded.

3.4.2.2 Principal components analysis

Principal components analysis (PCA) in the tangent space was developed by Cootes *et al.* (1992) and Kent (1994). We can perform a PCA on any choice of tangent coordinates, e.g. v_i^F or v_i^P or approximate tangent coordinates such as r_i^P or $vec(X_i^P)$. The sample covariance of any set of tangent coordinates v_i is defined as:

$$S_v = \frac{1}{n}\sum_{i=1}^{n}(v_i - \bar{v})(v_i - \bar{v})^T \tag{3.24}$$

where $\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i$. (Note that $\bar{v}=0$ except for when $vec(X_i^P)$ or $vec(X_i^{PP})$ are used). The $min(n-1, M)$ non-zero orthonormal eigenvectors of S_v are the principal components, denoted as ψ_j, and have eigenvalues $\lambda_1 \geq ... \geq \lambda_{\min(n-1,M)}$, where M is the dimension of the shape space. We will assume that the rank of S_v is M, so that PCA of S_v results in M components. PCA in the tangent space/approximate tangent space decomposes the total variability in shape, measured by $\sum_{i=1}^{n}\|v_i - \bar{v}\|^2$, into orthogonal components with each PC successively explaining the largest remaining variability in the data, where for example,

$\sum_{i=1}^{n}\|v_i - \bar{v}\|^2 \approx \sum_{i=1}^{n} d_F^2(X_i, \hat{\mu})$, if $v_i = v_i^F$ or r_i^P or $vec(X_i^P)$, following GPA method (i),

$\sum_{i=1}^{n}\|v_i - \bar{v}\|^2 = \sum_{i=1}^{n} d_F^2(X_i, \hat{\mu})$, if $v_i = v_i^P$ or r_i^P or $vec(X_i^P)$, following GPA method (ii),

$\sum_{i=1}^{n}\|v_i - \bar{v}\|^2 \alpha \sum_{i=1}^{n} d_F^2(X_i, \hat{\mu})$, if $v_i = r_i^P$ or $vec(X_i^P)$, following GPA method (iii).

The proportion of variability captured by the jth PC is:

$$\lambda_j \Big/ \sum_{i=1}^{M}\lambda_j . \tag{3.25}$$

The PC score for the ith configuration on the jth component is then defined as:

$$s_{ij} = \psi_j^T (v_i - \bar{v}),\ i=1,\ldots,n;\ j=1,\ldots, M, \tag{3.26}$$

with standardised (mean=0, s.d.=1) PC score given by $s_{ij} / \sqrt{\lambda_j}$. We can visualise the effect of each PC by plotting icons corresponding to values of the standardised PC score given by $c\lambda_j^{1/2}\psi_j + \bar{v}$. Typically c=(-3,-2,-1,0,1,2,3) so that variation either side of the

mean shape can be visualized. Icons are obtained by projecting the scores in tangent space back into configuration space, e.g. by adding the Procrustes mean $\hat{\mu}$ if the v_i are Procrustes residuals, calculating $vec_m^{-1}(c\lambda_j^{1/2}\psi_j + \bar{v})$ if the v_i are vectorised Procrustes fits or by equation (3.2.1) if v_i^F or v_i^P are used. In practice the results will be almost identical for datasets with small variability.

The PC scores may then be correlated with other covariates to examine the relationships of shape with explanatory variables (e.g. shape with centroid size). A related approach, relative warps, provides a different orthogonal basis for the PCs and decomposes the variation in shape at a variety of scales. See for example, Bookstein (1996b) and later in section 5.3.6.3.

For the hypodontia example data described above, we use the full Procrustes fits from GPA method (ii) as approximate tangent coordinates and obtain the PCs using the sample covariance matrix:

$$\frac{1}{n}\sum_{i=1}^{n} vec(X_i^P - \bar{X})vec(X_i^P - \bar{X})^T . \qquad (3.27)$$

Fig. 3.6 (top) shows the mean and full Procrustes fits of the 20 upper central incisors from hypodontia patients. In fig. 3.6 (bottom), configurations at -3,-2,-1 and +1,+2,+3 standard deviations either side of the mean shape are plotted in the directions of the first and second principal components of variation. PC(1) in fig. 3.6 bottom (left), accounts for 62% of the variation in shape in the sample and contrasts the LACC length with the width of the tooth, particularly around the gingival margin. We see that relatively wider teeth are more 'tapered' in shape. PC(2) in fig. 3.6 bottom (right) accounts for just 15% of the variation in shape and reflects vertical variation in the central and distal gingival landmarks. (Identical results were found with the use of Procrustes fits from any of the GPA methods).

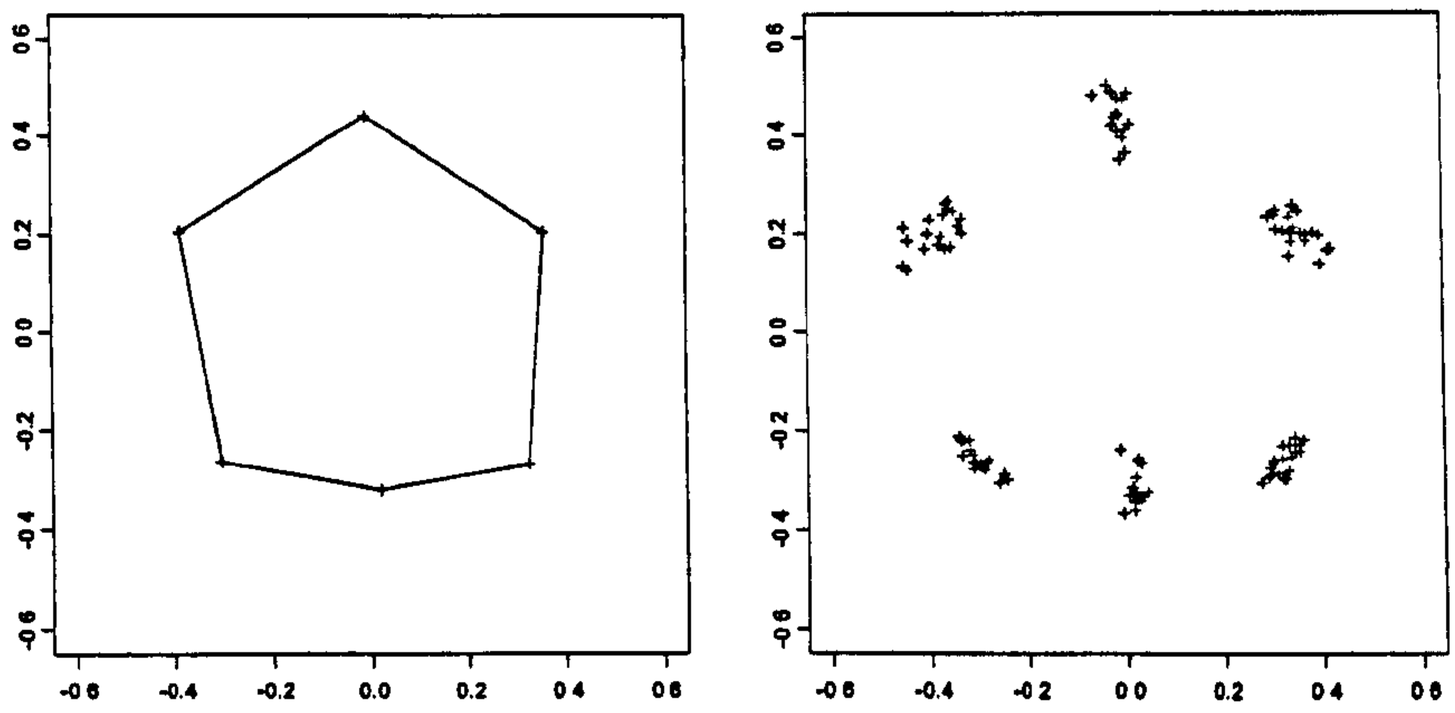


Fig 3.6 top: (Left) full Procrustes mean shape and (right) full Procrustes fits of Upper left central incisors in patients with hypodontia, gingival landmarks at the top.

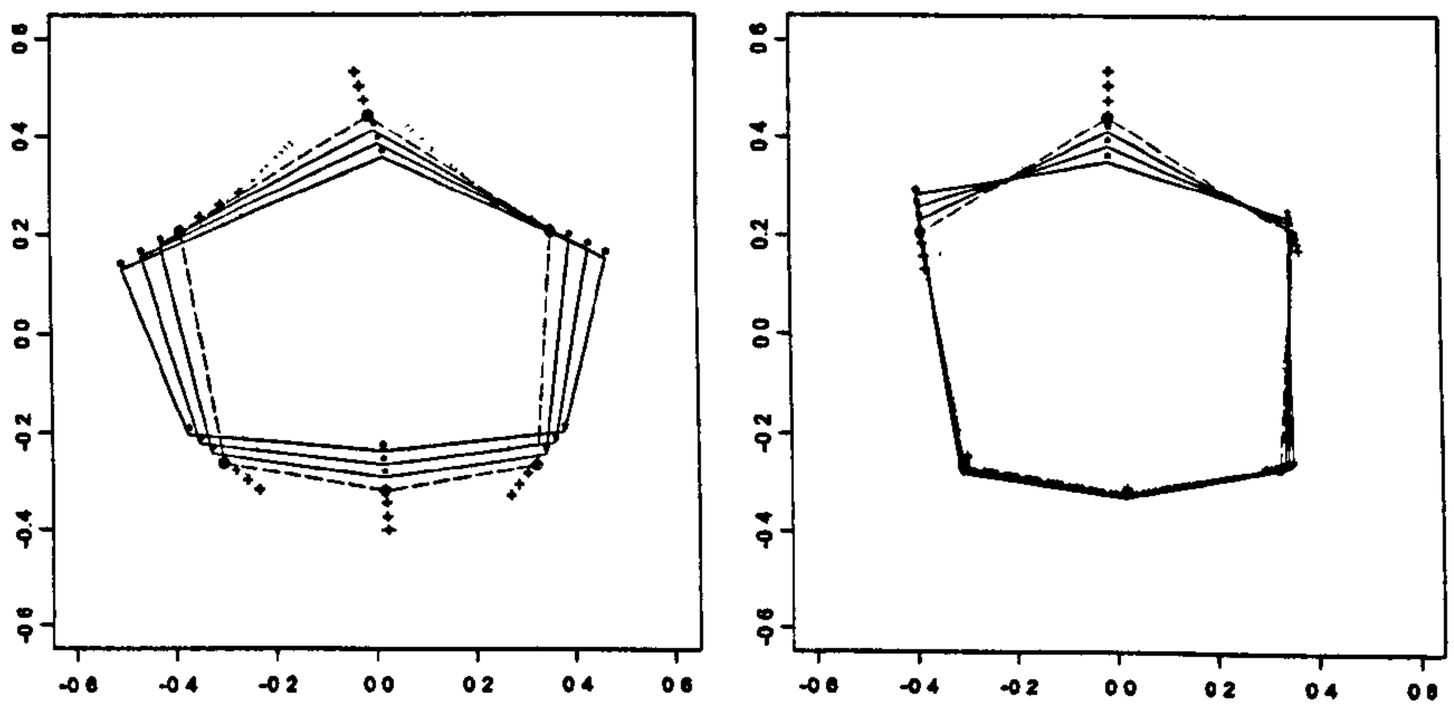


Fig. 3.6 bottom: (Left) *Dotted lines:* Shapes at -3,-2,-1 standard deviations along the 1st PC of variation. *Solid lines:* Shapes at +1,+2,+3 s.d.'s along the 1st PC. *Dashed line:* Procrustes mean shape. (Right) *Dotted lines:* Shapes at -3,-2,-1 standard deviations along the 2nd PC of variation. *Solid lines:* Shapes at +1,+2,+3 s.d's along the 2nd PC. *Dashed line:* Procrustes mean shape

3.4.2.3 Inference: Hotelling's T^2 tests

The most useful multivariate tests for comparison of group means are the well-known one and two sample Hotelling's T^2 tests. Consider a sample of configurations $X_1,...,X_n$, with tangent coordinates v_i. As with PCA we have several choices for the v_i. We can test whether or not the mean has a particular shape. i.e. test $H_0 : \mu = \mu_o$ vs. $H_1 : \mu \neq \mu_o$. Assuming each v_i is multivariate normal random variable with v_i independent of v_j ($i \neq j$) and letting v_{μ_0} be the tangent coordinates of μ_o, Hotelling's one sample T^2 statistic (under H_0) is taken as:

$$F = \frac{(n-M)}{M}(\bar{v} - v_{\mu_0})^T S_v^- (\bar{v} - v_{\mu_0}) \sim F_{M,n-M} \qquad (3.28)$$

where $\bar{v} = \frac{1}{n}\sum_{i=1}^{n} v_i$ and S_v^- is the generalised inverse of S_v. If the data are considered as being from two independent samples, $X_{11},...,X_{1n_1}$ and $X_{21},...,X_{2n_2}$ with mean shapes μ_1 and μ_2, and tangent coordinates $v_{11},...,v_{1n_1}$ and $v_{11},...,v_{2n_1}$, we can test $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$ using Hotelling's two sample T^2 statistic as below. The pole of projection for obtaining both sets of tangent coordinates is the pre-shape of the full Procrustes mean obtained using all $n_1 + n_2$ configurations. (The groups must be in one consistent tangent space, Bookstein, 1996c). The two sets of tangent coordinates are assumed mutually independent, multivariate normal with common covariance matrices. Providing that the variation in shape is reasonably similar in each group (a formal comparison can be made using e.g. Box's M test), then under H_0:

$$F = \frac{n_1 n_2 (n_1 + n_2 - M - 1)}{(n_1 + n_2)(n_1 + n_2 - 2)M}(\bar{v}_1 - \bar{v}_2)^T S_{v_1,v_2}^- (\bar{v}_1 - \bar{v}_2) \sim F_{M,n_1+n_2-M-1} \qquad (3.29)$$

where $S_{v_1,v_2} = (n_1 S_1 + n_2 S_2)/(n_1 + n_2 - 2)$, with generalised inverse S_{v_1,v_2}^- and $\bar{v}_1, \bar{v}_2$, S_{v_1} and S_{v_2} are the sample mean and covariance matrices in each group.

For the example data, fig. 3.7 (left) shows the full Procrustes mean from the n_1=20 control configurations matched to the estimated mean shape of the n_2=20 hypodontia cases. Fig. 3.7 (right) shows the scatter of Procrustes fits for each group around the pooled mean shape. Box's M gave a p-value of 0.3 and so we cannot reject the null hypothesis of equal covariance matrices. The observed value of the test statistic gave a p-value of 0.004, providing strong evidence of a difference between the hypodontia and control patients' mean incisor shapes. Fig. 3.7 (left) suggests that the mean shapes differ in the position of the incisal corners, hypodontia central incisors being more tapered in shape.

For testing the equality of mean shapes in several groups, the pre-shape of the overall full Procrustes mean is taken as the pole, as above and a multivariate analysis of

variance (MANOVA) carried out on the chosen tangent coordinates (again in one consistent tangent space). In fact, providing variations in shape are small, the entire complement of multivariate statistical methods could be used in the tangent space to analyse variations in shape.

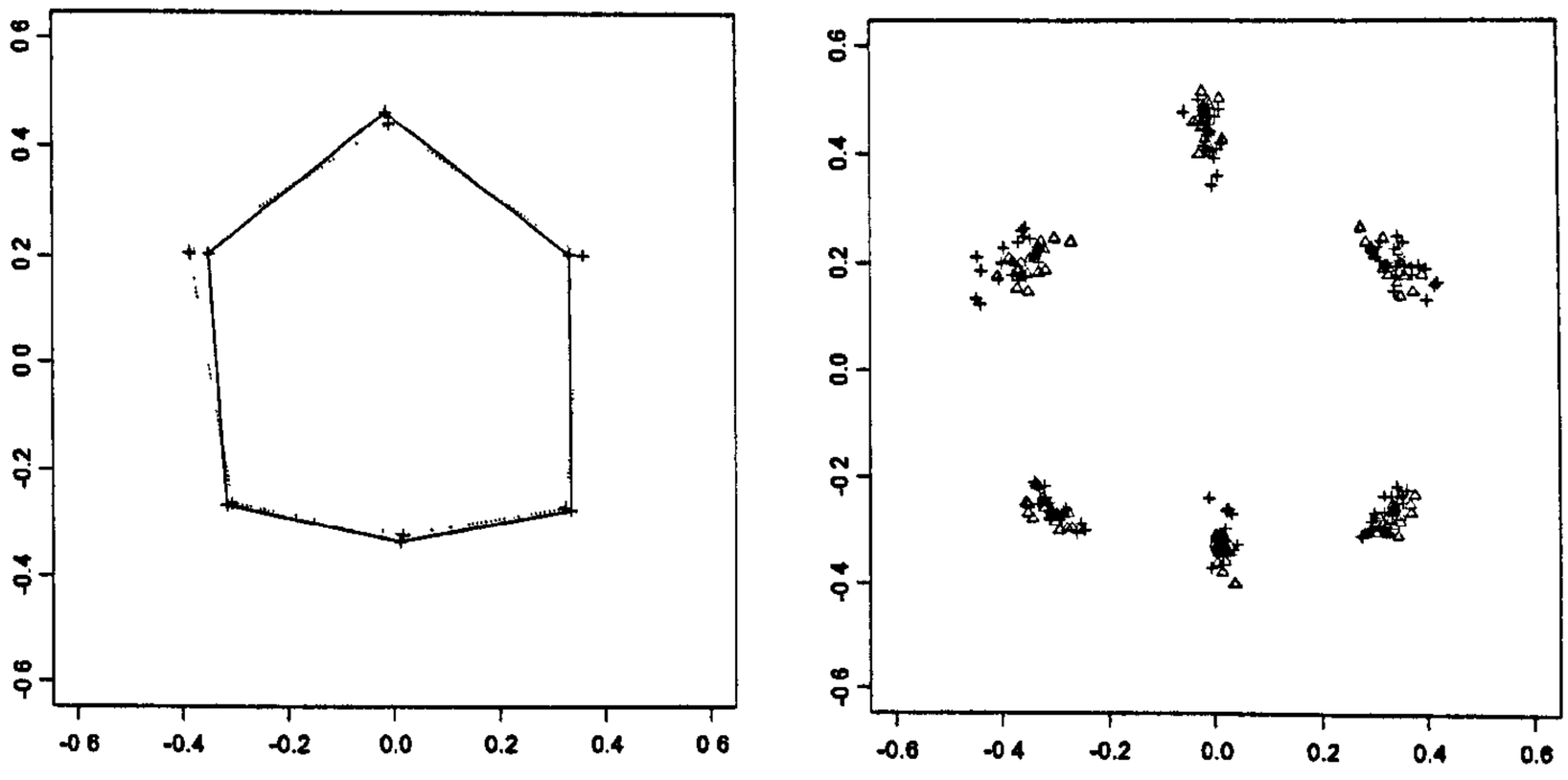


Fig. 3.7: (Left) Superimposed Procrustes mean shapes of control (solid) and hypodontia (dashed) upper left central incisors. (Right) Procrustes fits using a pooled sample (Δ = control, + = hypodontia). Gingival landmarks at the top of the plots.

3.4.3 Other inference methods

Aside from tangent space methods, comparisons of group means have been developed under the assumption that configurations are isotropic normal perturbations from mean configurations, which, when appropriate results in more powerful tests than those already described. These are outlined below. The second and final part of this section then briefly covers probability distributions for points in shape space, which may also be used to provide models for analysis. Although we do not make use of these methods in this thesis, they are included here for completeness.

3.4.3.1 Goodall's F-tests

An alternative approach to inference is to use statistics based on squared Procrustes distances (Goodall, 1991). Here the configurations are assumed to be perturbations from mean configurations, as defined in the model (3.5) with $vec(E_i) \sim N(0, \sigma^2 I_{km})$.

For testing $H_0 : \mu = \mu_o$ vs. $H_1 : \mu \neq \mu_o$. Goodall (1991) showed that, providing σ^2 is small and μ_o close to $\hat{\mu}$, then under H_0, we have the approximate result:

$$F = (n-1)n \frac{d_F^2(\mu_0, \hat{\mu})}{\sum_{i=1}^{n} d_F^2(X_i, \hat{\mu})} \sim F_{M,(n-1)M}.$$

For two independent samples, $X_{11},...,X_{1n_1}$ and $X_{21},...,X_{2n_2}$, each set of configurations is assumed to be from a population modelled by (3.5) with means μ_1 and μ_2 respectively and with a common variance σ^2 for each coordinate, so that $vec(E_{1i})$ and $vec(E_{2i}) \sim N(0, \sigma^2 I_{km})$. To test $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$, we use (under H_0 and for σ^2 small) the approximate result:

$$F = \frac{n_1 + n_2 + 2}{n_1^{-1} + n_2^{-1}} \frac{d_F^2(\hat{\mu}_{1,}, \hat{\mu}_{2,})}{\sum_{i=1}^{n_1} d_F^2(X_{1i}, \hat{\mu}_1) + \sum_{i=1}^{n_2} d_F^2(X_{2i}, \hat{\mu}_2)} \sim F_{M,(n_1+n_2-2)M}.$$

Dryden & Mardia (1998) show that Goodall's two sample test can be seen as a special case of the two sample Hotelling's T^2 under the isotropic model, with S_{ν_1,ν_2} in (3.29) replaced by $s_\nu^2 I_{2k-2}$, with s_ν^2 the unbiased estimate of variance. They also illustrate the lower power of Hotelling's two sample T^2 compared to Goodall's two sample F test when the isotropic normal model holds. Power is lost because of the many degrees of freedom used in estimating the covariance matrix in the T^2 test.

In the case of multiple independent samples, j=1,...,g, each of size n, Goodall (1991) suggested that under $H_0 : \mu_1 = ... = \mu_g$ we can use the following test statistic and approximate result:

$$F = n(n-1)g \frac{\sum_{j=1}^{n} d_F^2(\hat{\mu}_j, \hat{\mu})}{(g-1)\sum_{j=1}^{g}\sum_{i=1}^{n} d_F^2(X_{ji}, \hat{\mu})} \sim F_{(g-1)M,g(n-1)M}.$$

3.4.3.2 Shape distribution models for two dimensional data

By considering suitable probability distributions for shape, special statistical models have been developed for two-dimensional configurations that take account of the non-Euclidean geometry of shape space. Dryden & Mardia (1998) devote a chapter of their book in bringing the main techniques together, summarising the different approaches and considering the practical issues of using each method for inference.

The main distributions for shape include:

- Uniform distribution: Defined in pre-shape space to be invariant under rotations (see e.g. Kendall, 1984).
- Complex Bingham distribution: Analogous to the real Bingham distribution, defined in configuration space and conditioned on the similarity transformations (Kent, 1994).
- Complex Watson: A special case of the complex Bingham, again conditional on the similarity transformations. See e.g. Mardia & Dryden (1998)
- Offset Normal: a marginal distribution in configuration space (after the similarity transformations have been integrated out). See e.g. Kendal (1984) and further developments by Mardia & Dryden (1989a, b), Dryden & Mardia (1991).

More details of these methods can also be found in Dryden & Mardia (1993, 1998), Goodall (1991), Goodall & Mardia (1993), Mardia *et al.* (1995), Mardia & Dryden (1989a, b) and Small (1996).

Dryden & Mardia (1998) suggest that the most straightforward and preferred way to proceed is to use the complex Watson, which gives the full Procrustes mean as the maximum likelihood estimator of mean shape.

3.5 Discussion

The description and examples contained in this chapter reveal the potential value of this technique for the study of tooth shape, demonstrating a significant difference in mean buccal surface shape (Hotelling's two sample T^2 test; p=0.004), when comparing a hypodontia group to a corresponding control group. Details have also been published in the dental literature (see Robinson *et al.*, 2001).

For general application however, the methods described above assume that the landmarks on images are clearly located, reproducible and have good correspondence between cases. In addition, methods assume equal, circular (isotropic) variation at each of the landmarks during the superimposition and registration. For data collected from tooth images, such as the landmark sets suggested in fig 1.3, there are several violations of these assumptions.

In particular, success of the method depends on the reproducibility of the landmark data and its effects on the Procrustes analysis of shape. In the examples used here, certain landmarks, such as those at the corners of the incisal edge and the mesio-distal width, were found to be more difficult to identify than others, such as cusp tips or fissure junctions. This has implications in an analysis where each landmark carries equal 'importance'. Another potential source of such errors is in the subjective orientation of the tooth surface when its image is captured. Any landmark identification inconsistencies, both between occasions and different operators, will carry through into the inferential procedures, inflating residual variance, decreasing power and so diluting 'real' differences between patients.

In the following chapter we therefore set out to investigate the importance of these problems for landmark sets from both the buccal and occlusal surfaces of a variety of tooth types, quantifying the reliability of landmark representation and establishing how useful a technique Procrustes analysis is for the analysis of tooth shape.

Chapter 4

Reliability of landmark data: Impact on the planar Procrustes analysis of tooth shape

4.1 Introduction

In the preceding chapter we introduced the ideas and concepts of Procrustes analysis to the study of tooth shape, where teeth are represented and analysed as configurations of 'landmarks'. The techniques allow estimates of mean shapes and visualisations of shape variability to be obtained and conventional inferential techniques to be adapted to address hypotheses concerning shape. However, for these methods to perform well, landmarks must be reliably located. Otherwise this can lead to problems later in an analysis. Inconsistencies between operators in the positioning of landmarks will result in inconsistent representations of shape, as illustrated in fig. 4.1. 'Real' differences will become diluted by increased residual variance and statistical power will be reduced, lessening our ability to reject false hypotheses.

In this chapter we therefore aim to establish how useful a technique the Procrustes method is (i.e. how well it can be expected to perform), in its standard form, when used on buccal and occlusal surfaces from a variety of different tooth types. In particular we investigate how much impact inconsistencies in the locations of landmarks will have on investigations of shape.

In sections 4.2 and 4.3 we consider the sequence of steps involved in obtaining landmark data from study casts and identify the main possible sources of error. We also discuss how best to make use of the operators and study casts available for this investigation.

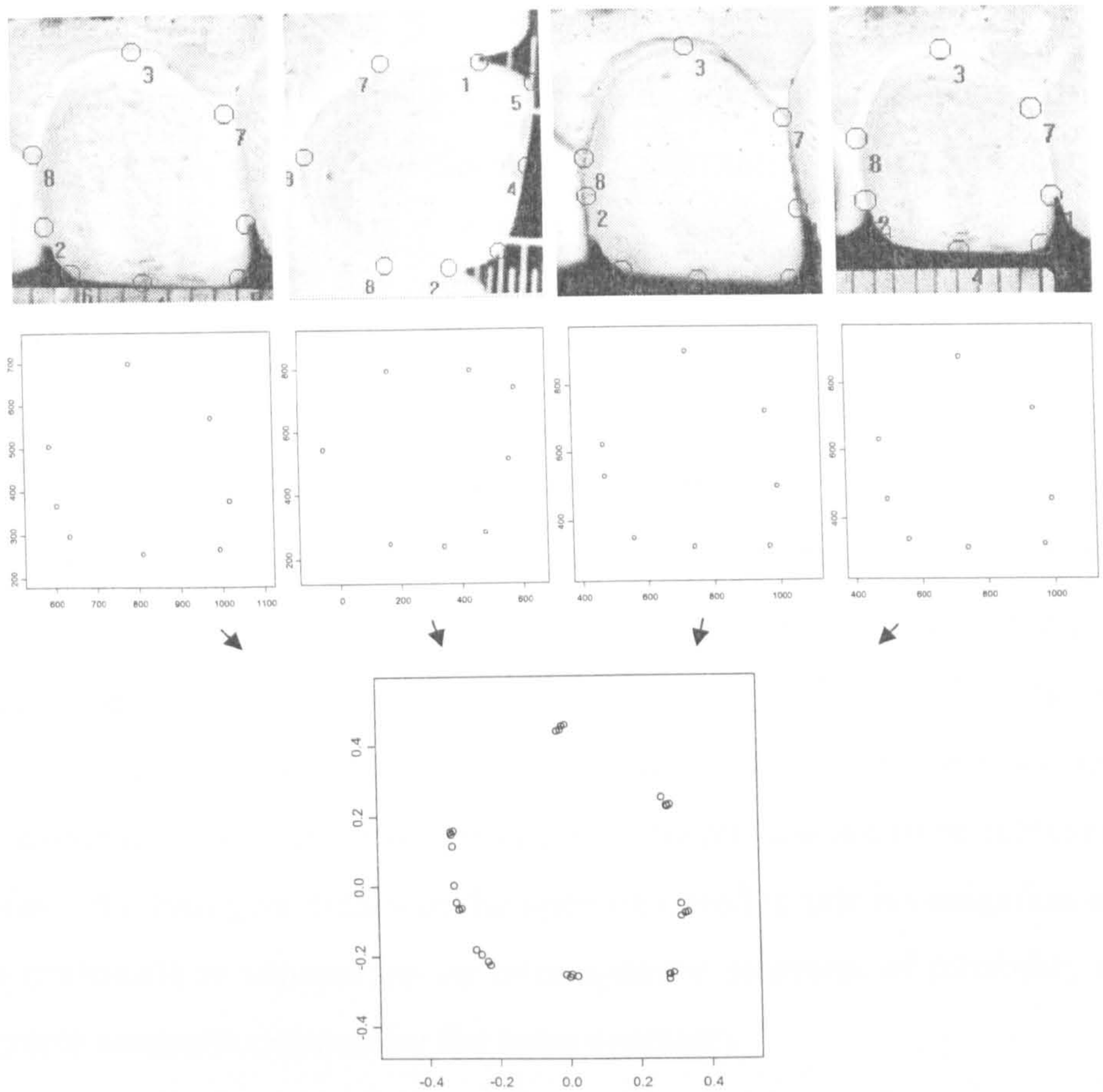


Fig 4.1: (Top and centre) Images and recorded landmarks of the same upper central incisor, obtained by four different operators. (Bottom) Procrustes matched configurations with shape dissimilarities due to inconsistent representation by the operators.

Section 4.4 then describes the methods of reliability assessment to be used, the results of which are presented in 4.5. When using images derived by different operators and landmarks of various surfaces from a variety of patients, the consequences of location inconsistency may be evaluated by calculating its effect on the recorded variation in Procrustes fits, obtained from each set of multiple representations. We establish the proportion of variation attributable to actual variation between patients relative to inter-operator inconsistency in obtaining landmarks and other sources of error and present a new measure of shape reliability, for use with Procrustes registered configurations.

Using principal components plots we can also visually investigate the different patterns of variation in the Procrustes fits based on the both the total and within patient covariance, represented by sums of squares and products matrices. The latter allows us to examine which particular directions of variation in shape have resulted from inconsistent representations of the same patient's tooth surface.

Having evaluated the importance of these problems, the final part of this chapter then considers how methods of analysis will need to accommodate such difficulties, if landmark data are to be used to represent and describe variations in tooth shape.

4.2 Sources of error

The precision of the recorded landmark data will depend on the methods used to obtain images as well as the subjective positioning of these points. In the first part of this section, we consider the sequence of steps involved in obtaining landmark configurations and identify and discuss possible sources of error. Our assessments of reliability will only encompass the parts of the data collection process that are repeated and so consideration is given to which stages of the process are to be represented in the final results. We then give details of the operators used in this investigation and discuss why it is preferable to concentrate on inter-operator measures of reliability rather than intra-operator measures (repeats by the same operator).

4.2.1 Data acquisition

To obtain landmark coordinates from dental study casts, the following sequence of steps is generally required and would be typical of any image capture and coordinate recording system.

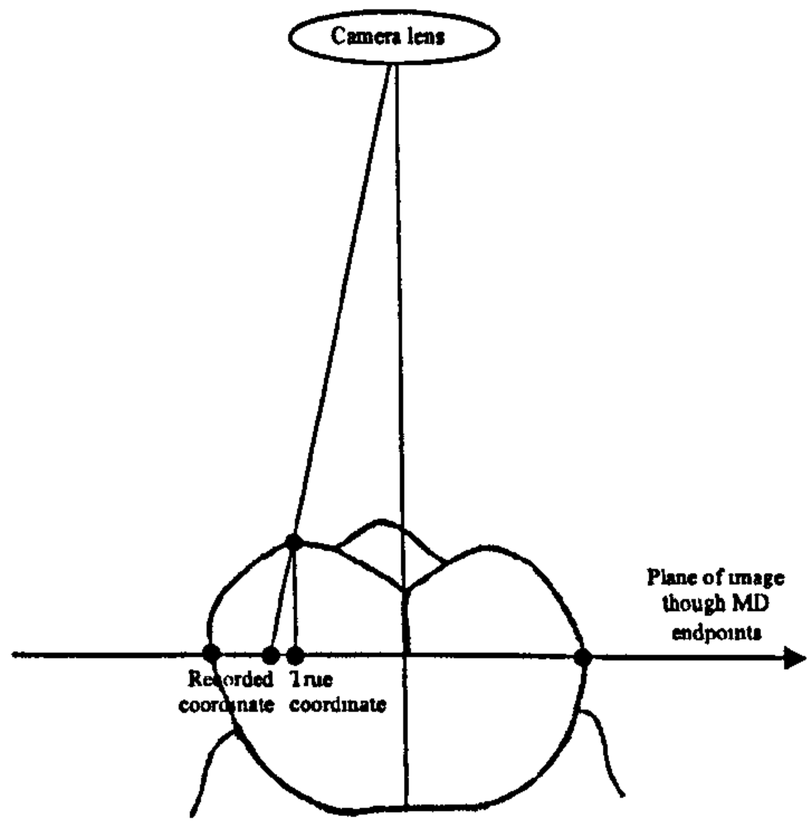
Each study cast must be subjectively orientated so that the buccal or occlusal tooth surface of interest may be imaged. The camera used to acquire the images must then be correctly focused and once the image is obtained, transferred to a computer screen for processing. As detailed in section 1.4, for the Sheffield system, this involves securely placing the study model on a platform adjustable in three planes beneath a mounted 32-bit digital camera. Standardised definitions for the orientation of each tooth surface at image capture will help aid correspondence between operators. The criterion used here was that each surface should be positioned parallel to the camera, so that the maximum surface area is visible.

Once displayed on screen, the locations of landmarks are then recorded as (x,y) pixel coordinates in the 2D plane of the image, using a suitable package (here Image ProPlus V3.01, Media Cybernetics). The x-axis of an image typically runs parallel to the bottom edge of the image window, with the perpendicular y-axis parallel to the left side of the image. Landmarks can be positioned using a mouse-controlled marker (typically accurate to a single pixel) and may be repositioned at any time during the landmark identification process, if required. In this study the images were typically 1000 × 1000 pixels, displayed on a 17" monitor.

As an additional step, the axes of the image may also be calibrated. This involves placing a rule in the plane of the tooth surface, prior to imaging, which is then used to specify the scale of the two axes, by drawing a line along the edge of the rule and specifying the number of units (in mm) represented. Note that the calibration is only an additional source of error when assessing the impact of landmark inconsistencies on measures of size. Recall that in an analysis of shape, size is irrelevant.

Note that for some imaging systems additional error in the landmark locations can arise from perspective effects if the camera is placed too close to the surface of an object (Arnqvist & Martensson, 1998). The generation of false landmark coordinates, which results from some points being closer to the lens of the camera than others, is illustrated in fig. 4.2 below. For the Sheffield system, the distance of the camera from the tooth surface was deemed sufficiently large for this to be inconsequential.

Fig 4.2: Perspective effects when camera too close to a 3D object



Arnqvist & Martensson (1998) present a comprehensive description of the components of error that may be encountered when identifying and recording landmark locations on 2D images of 3D objects by procedures such as those above. They suggest that 'total error' may be partitioned into 'methodological' (from specimen presentation), 'instrumental' (optical or digital distortion) and 'personal' errors (subjective decisions). Whilst there may be methodological errors involved in obtaining impressions and forming study casts from patients, this part of the process will not be repeated for this study. The reproducibility of study models is not in question here and our assessments of reliability are only required to represent the subsequent components of error that are repeated. It is also expected that with today's equipment, instrumental effects on the accuracy and precision of recorded data points due to the lens distortion or other characteristics of the camera, digital transfer and representation on a computer screen are negligible (see manufacturer's service manuals). Consequently, we expect that the primary source of landmark errors will be in the operators' subjective positioning of landmark points and/or due to subjective orientation of surfaces during image capture.

4.2.2 Operators

Four individuals were available to take part in this study, allowing variations between operators to be represented in the final results. Any single operator may be highly successful in recording similar representations of a configuration on multiple occasions, but these representations must also agree with those of other operators using the technique, so that results may validated or verified independently and to guard against systematic error. The operators were postgraduate students, each with over two years experience of using the imaging system and software. In the analysis which follows they are regarded as a typical sample from a larger population of trained operators so that reliability levels will be generalisable beyond just those operators who took part.

4.3 Data used: Patients, surfaces and landmarks

In order to provide a comprehensive representation of the different surfaces found within the dentition and to encompass a variety of different tooth features, up to twenty of each of the surfaces listed below were imaged and represented as landmark

configurations by each operator. Each tooth was required to be fully erupted, with no evidence of attrition and none of the surface obscured by crowding. The investigation made use of existing material currently under investigation for which ethics committee approval had been obtained. The twenty independent cases comprised 10 males and 10 females of a variety of different genetic and environmental backgrounds, such as may be of interest in some future study (as opposed to (say) using only Caucasian cases, which will not encompass the same range of actual variation in shape).

The surfaces considered are displayed, along with the definitions of landmarks used, in fig. 4.3. The buccal surfaces used were: upper right central incisor, upper right first molar, lower left central incisor, lower left canine and lower left first molar. The occlusal surfaces used were: upper right central incisor, upper right first pre-molar, lower left canine, lower left first molar and lower left second molar. The tooth outlines in the figure were created from diagrams in Wheeler (1962).

The landmarks we consider in fig 4.3 are primarily 'anatomical'; points assigned by dental experts of homologous, meaningful, biological correspondence. They consist of features such as cusp tips and fissure junctions in addition to positions corresponding to endpoints of commonly used clinical measurements such as the endpoints of mesio-distal width (MD), bucco-lingual width (BL) and the long axis of the clinical crown (LACC). MD was defined as the maximum diameter between the contact areas of a tooth. BL was defined as the maximum bucco-lingual diameter in the occlusal view, approximately perpendicular to MD. The LACC was to be placed to divide the buccal surface roughly in half, in the occlusal-gingival direction. The 'ends of the papilla' were defined as the last visible points on the sides of the teeth before they become obscured by the inter-dental papilla. Other points on the gingival margin were included for buccal views to provide some indication of the shape of teeth in these regions. In remaining regions with few characteristics of correspondence, points placed half way between other landmarks were used. These are known as 'pseudo-landmarks'. Other landmarks based on mathematical or geometric properties and identified by automatic processes were also considered. e.g. points of high curvature or extreme points. However, while these may be appealing in reducing variation in landmark positions, it proved difficult to obtain any meaningful points of correspondence between cases, using a variety of in-built and custom written routines for Image-Pro Plus.

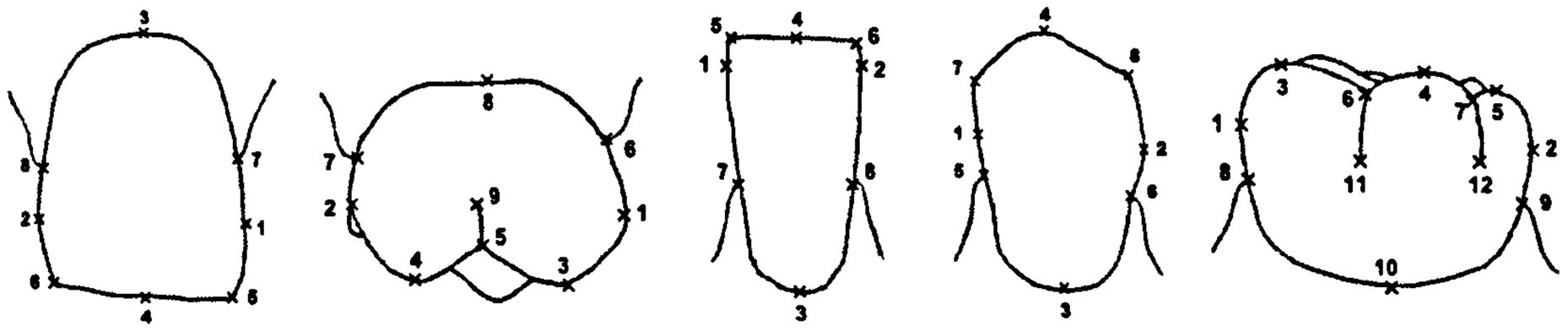


Fig. 4.3(a): Buccal surfaces and landmarks used. Left to right: *Upper right central incisor*: (1) & (2) mesial & distal endpoints of MD, (3) & (4) gingival & incisal endpoints of LACC, (5) & (6) corners of mesial & distal sides and incisal edge, (7) & (8) ends of mesial & distal papillae. *Upper right first molar*: (1) & (2) mesial & distal endpoint of MD, (3) & (4) mesial & distal labial cusp tips, (5) occlusal limit of buccal groove, (6) & (7) ends of mesial & distal papillae, (8) half way between 6 & 7 along gingival margin, (9) start of buccal groove. Lower *left central incisor:* (1) & (2) mesial & distal endpoints of MD, (3) & (4) gingival & incisal endpoints of LACC, (5) & (6) corners of mesial & distal sides of tooth & incisal edge, (7) & (8) ends of mesial & distal papillae. *Lower left canine:* (1) & (2) mesial & distal endpoints of MD, (3) & (4) gingival & cusp tip endpoints of LACC, (5) & (6) ends of mesial and distal papillae, (7) & (8) mesial & distal angles of cusp. *Lower left first molar:* (1) & (2) mesial & distal endpoints of MD, (3), (4) & (5) mesial, central & distal cusp tips, (6) & (7) occlusal limits of mesial & distal buccal grooves, (8) & (9) ends of mesial and distal papillae, (10) half way between 8 & 9 along gingival margin, (11) & (12) starts of mesial & distal buccal grooves.

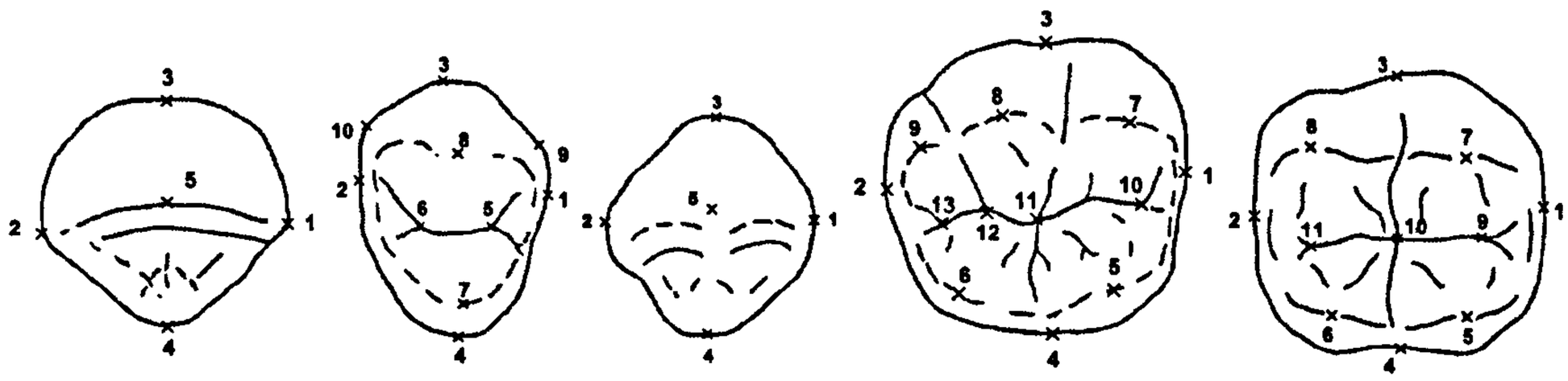


Fig. 4.3(b): Occlusal surfaces and landmarks used. Left to right: *Upper right central incisor:* (1) & (2) mesial & distal endpoints of MD, (3) & (4) buccal & lingual endpoints of BL, (5) point of bisection of incisal edge by BL. *Upper right first pre-molar:* (1) & (2) mesial & distal endpoints of MD, (3) & (4) buccal & lingual endpoints of BL, (5) & (6) mesial & distal pits/fissure junctions, (7) & (8) lingual and labial cusp tips, (9) & (10) mesial & distal endpoints of max labial cusp width. *Lower left canine:* (1) & (2) mesial & distal endpoints of MD, (3) & (4) buccal & lingual endpoints of BL, (5) cusp tip. *Lower left first molar:* (1) & (2) mesial & distal endpoints of MD, (3) & (4) buccal & lingual endpoints of BL, (5) & (6) mesial & distal lingual cusp tips, (7), (8) & (9) mesial, central & distal labial cusp tips (10), (11), (12) & (13) outer mesial, inner mesial, central and distal pits. *Lower left second molar:* (1) & (2) mesial & distal endpoints of MD, (3) & (4) buccal & lingual endpoints of BL, (5) & (6) mesial & distal lingual cusp tips, (7) & (8) mesial & distal labial cusp tips , (9), (10) & (11) mesial, central and distal pits.

4.4 Reliability assessment: Methods

In previous morphometric studies, reliability of shape representation has often been assessed in terms of the absolute precision of landmark locations, with data collected by repeated identification on the same images, e.g. Reig (1998). Other studies have compared various distances between pairs of landmarks on repeated images, see McKeown *et al.*, (2001) for an example using inter-landmark distances on teeth. The most popular measures of absolute precision, e.g. Bland & Altman (1986), typically involve assessing inconsistencies using measures of dispersion, such as standard deviations and it is then left to the investigator to deem the resulting values as either 'problematic' or 'negligible'. Because of this many authors such as Fleiss (1986) and Bailey & Byrnes (1990) stress the importance of relative measures of assessment. In contrast to absolute measures, which must be subjectively assessed, these relate the magnitude of measurement errors to the extent of true or actual variance between cases.

In the Procrustes analysis of shape, errors in the precision of landmark locations are carried through the superimposition and registration procedures to result in errors in the coordinates of the Procrustes fits. Since subsequent inference and analysis will be based on these coordinates, it seems logical that our assessments of reliability should also be based on them too and so for each of the ten surfaces, we measure the impact of operator inconsistency in landmark positioning by evaluating its effect on the variation in the Procrustes fits obtained from each set of multiple representations.

Since each set of Procrustes coordinates forms a multivariate dataset, a suitable approach to assessing reliability is not obvious. Following Arnqvist & Martensson (1998), the Procrustes fits for each set of repeated configurations may be parameterised by the calculation of principal components, describing the largest, second largest and so on patterns of variation in the fits. By separating the variation in shape into uncorrelated variables using PCA, we can quantify the reliability in each univariate dataset of component scores (i.e. in the direction of each PC) using well-established methods of assessment for univariate data. In particular, we use intraclass correlation coefficients, which have a long history in the study of errors (e.g. Fleiss & Strout, 1977). These involve identifying how much of recorded variation in shape is actual/true

(biological) variation and how much is variation in scores due to inconsistently located landmarks (and other errors), using estimates of components of variance.

We then extend Arnqvist & Martensson's approach, further utilising the variance estimates obtained for each PC. Since the PCs are uncorrelated (and further assuming independence) it is possible to combine the variance estimates from each PC to produce an 'overall' measure of reproducibility for the Procrustes fits of each tooth type. We also show how the same overall reliability figure may be obtained directly from sums of Euclidean distances between the Procrustes registered configurations.

4.4.1 Reliability along each principal component

Consider (say) the sample of upper central incisor (buccal surface) configurations. Let V denote the covariance matrix of the Procrustes fits of these configurations (to an estimated grand mean), with

$$V=\frac{1}{np-1}\sum_{i=1}^{n}\sum_{m=1}^{p}(vec(X_{im}^{P})-vec(\bar{X}^{P}))(vec(X_{im}^{P})-vec(\bar{X}^{P}))^{T} \qquad (4.1)$$

$$=\frac{1}{np-1}SSP(vec(X_{im}^{P})),$$

where X_{im}^{P} denotes the Procrustes fit of the representation of case (patient's surface) i, i=1,...,n, obtained by operator m, m=1,...,p, $\bar{X}^{P}=\frac{1}{np}\sum_{i=1}^{n}\sum_{m=1}^{p}X_{im}^{P}$ and $SSP(vec(X_{im}^{P}))$ denotes the sums of squares and products matrix of the vectorised Procrustes fits.

Now consider the principal components extracted from V or $SSP(vec(X_{im}^{P}))$. From 3.4.2.2, there will be $2k$-4 non-zero PCs (for the buccal surface of the upper central incisor there are k=8 landmarks). On each component there will be $p\times n$ univariate component scores, i.e. one for each of the p=4 operator's representations of each of n cases (patients' surfaces).

Considering any one particular component, PC_r, (r=1,...,2k-4), the component score for patient (or case) i, obtained by operator m, may be written as:

$$S_{r_{im}} = C_{r_i} + O_{r_m} + E_{r_{im}} . \tag{4.2}$$

Following Fleiss (1986), equation (4.2) defines a 2-way random effects ANOVA model; C_{r_i} denotes the mean of many such (replicated) observations on the randomly selected tooth i, and so may be thought of as the cases' (patients') 'true score' for this component. Each C_{r_i} is estimated as the mean score for each patient, averaged over the operators. O_{r_m} represents the randomly selected m^{th} operator's systematic (constant) additive effect on the scores (mean 0) and $E_{r_{im}}$ the random, non-constant part of the difference from the true score (mean 0) in this particular direction of variation. The terms are considered mutually independent with variances $\sigma^2_{C_r}$, $\sigma^2_{O_r}$ and $\sigma^2_{E_r}$ respectively and therefore:

$$\text{Var}(S_{r_{im}}) = \sigma^2_{C_r} + \sigma^2_{O_r} + \sigma^2_{E_r} .$$

The inter-operator coefficient of reliability (Fleiss & Strout, 1977, Bailey & Byres, 1990) for this PC, expressing the relative magnitude of the variation between patient means (true scores), is defined as:

$$R(C_r) = \frac{\sigma^2_{C_r}}{\sigma^2_{C_r} + \sigma^2_{O_r} + \sigma^2_{E_r}} . \tag{4.3}$$

i.e. actual variation between patients as a proportion of the total observed variation in the component scores.

If the variation due to inconsistency ($\sigma^2_{O_r} + \sigma^2_{E_r}$) is small relative to the variability between patients $\sigma^2_{C_r}$, reliability $R(C_r)$ in this direction of variation is high (with maximum value 1). Conversely if $\sigma^2_{O_r} + \sigma^2_{E_r}$ is large relative to $\sigma^2_{C_r}$, reliability is low (with minimum value 0). Note that $1-R(C_r)$ gives the proportion of observed variance attributable to systematic and random errors.

The estimates of the components of variance are based on expected mean squares from a 2-way analysis of variance. The partition of the total sum of squares, SSS_r into sums of squares between cases, SSC_r, operators, SSO_r, and errors, SSE_r, is given in the table below, along with the means squares and expected mean squares.

Source	d.f.	SS	MS	EMS
Between cases	n-1	$SSC_r = p\sum_{i=1}^{n}(\bar{s}_{r_i} - \bar{s}_r)^2$	$MSC_r = \frac{SSC_r}{(n-1)}$	$p\sigma^2_{C_r} + \sigma^2_{E_r}$
Between operators	p-1	$SSO_r = n\sum_{m=1}^{p}(\bar{s}_{r_m} - \bar{s}_r)^2$	$MSO_r = \frac{SSO_r}{(p-1)}$	$n\sigma^2_{O_r} + \sigma^2_{E_r}$
Other sources of error	$(n-1)\times(p-1)$	$SSE_r = \sum_{m=1}^{p}\sum_{i=1}^{n}(s_{r_{im}} - \bar{s}_{r_i} - \bar{s}_{r_m} + \bar{s}_r)^2$	$MSE_r = \frac{SSE_r}{(n-1)(p-1)}$	$\sigma^2_{E_r}$
Total	np-1	$SSS_r = \sum_{m=1}^{p}\sum_{i=1}^{n}(s_{r_{im}} - \bar{s}_r)^2$		

Table 4.1: ANOVA partition for 2-way random effects model, with $SSS_r = SSC_r + SSO_r + SSE_r$.

$\bar{s}_{r_i}$ = mean component score for case i, $\bar{s}_{r_m}$ = mean component score for operator m,

$\bar{s}_r$ = overall mean component score on PC_r.

Unbiased estimates of $\sigma^2_{C_r}, \sigma^2_{O_r}$ and $\sigma^2_{E_r}$ are then given by:

$$\hat{\sigma}^2_{C_r} = \frac{MSC_r - MSE_r}{p},\ \hat{\sigma}^2_{O_r} = \frac{MSO_r - MSE_r}{n},\ \hat{\sigma}^2_{E_r} = MSE_r. \qquad (4.4)$$

from which an estimate $\hat{R}(C_r)$ of $R(C_r)$ may then be obtained. (Note that Var($S_{r_{im}}$) is estimated by $\hat{\sigma}^2_{C_r} + \hat{\sigma}^2_{O_r} + \hat{\sigma}^2_{E_r}$ and not $\frac{SSS_r}{np-1}$, the sample variance in scores along PC_r, due to the assumptions of the random effects model). The estimated proportions of variance attributable to systematic operator differences or attributable to random errors may also be calculated as:

$$\hat{R}(O_r) = \frac{\hat{\sigma}^2_{O_r}}{\hat{\sigma}^2_{C_r} + \hat{\sigma}^2_{O_r} + \hat{\sigma}^2_{E_r}},\ \hat{R}(E_r) = \frac{\hat{\sigma}^2_{E_r}}{\hat{\sigma}^2_{C_r} + \hat{\sigma}^2_{C_r} + \hat{\sigma}^2_{C_r}}.$$

4.4.2 Overall reliability of Procrustes fits

4.4.2.1 Using variance components from PC scores

Since the principal components are uncorrelated (and if we further assume independence), then with same variance components model (4.2) assumed for each set of PC scores, an overall reliability figure may be produced for each of the ten tooth surfaces. The total variation in shape in the Procrustes fits for each set of repeated configurations is:

$$Var(S_1) + Var(S_2) + \ldots\ldots + Var(S_{2k-4}) = \sum_{r=1}^{2k-4}\sigma_{C_r}^2 + \sum_{r=1}^{2k-4}\sigma_{O_r}^2 + \sum_{r=1}^{2k-4}\sigma_{E_r}^2 .$$

So an overall reliability score for all variation in shape may be calculated as:

$$\mathrm{R}(C_{All}) = \frac{\sum_{r=1}^{2k-4}\sigma_{C_r}^2}{\sum_{r=1}^{2k-4}(\sigma_{C_r}^2 + \sigma_{O_r}^2 + \sigma_{E_r}^2)}. \qquad (4.5)$$

Using estimates, from (4.4):

$$\sum_{r=1}^{2k-4}\hat{\sigma}_{c_r}^2 = \frac{1}{p}\sum_{r=1}^{2k-4}(MSC_r - MSE_r),\ \sum_{r=1}^{2k-4}\hat{\sigma}_{o_r}^2 = \frac{1}{n}\sum_{r=1}^{2k-4}(MSO_r - MSE_r),\ \sum_{r=1}^{2k-4}\hat{\sigma}_{e_r}^2 = \sum_{r=1}^{2k-4}MSE_r \quad (4.6)$$

This extends the ideas of Arnqvist & Martensson (1998) and presents a new figure for calculating an overall reliability score for a set of multiple representations of different configurations. Overall measures of the proportion of variance attributable to systematic operator differences or random errors may similarly be defined in the same way as (4.5) as below and estimated using equations (4.6) above.

$$\mathrm{R}(O_{All}) = \frac{\sum_{r=1}^{2k-4}\sigma_{O_r}^2}{\sum_{r=1}^{2k-4}(\sigma_{C_r}^2 + \sigma_{O_r}^2 + \sigma_{E_r}^2)},\ \mathrm{R}(E_{All}) = \frac{\sum_{r=1}^{2k-4}\sigma_{E_r}^2}{\sum_{r=1}^{2k-4}(\sigma_{C_r}^2 + \sigma_{O_r}^2 + \sigma_{E_r}^2)}.$$

4.4.2.2 Using 2-way ANOVA identity for Procrustes fits

For any particular surface it is also possible to compute $\hat{R}(C_r)$ directly from the Procrustes fits, without the need to calculate PCs. For each set of PC scores we have:

$$SSS_r = SSC_r + SSO_r + SSE_r$$

and so:

$$\sum_{r=1}^{2k-4} SSS_r = \sum_{r=1}^{2k-4} SSC_r + \sum_{r=1}^{2k-4} SSO_r + \sum_{r=1}^{2k-4} SSE_r \,. \qquad (4.7)$$

Since the principal components of $SSP(vec(X_{im}^P))$ or V, defined in (4.1), are linear combinations of the coordinates of the Procrustes fits, which preserves the total variation, then (assuming PC scores were obtained from $SSP(vec(X_{im}^P))$)

$$\sum_{r=1}^{2k-4} SSS_r = \sum_{r=1}^{2k-4}\sum_{m=1}^{p}\sum_{i=1}^{n}(s_{r_{im}} - \bar{s}_r)^2$$

$$= \sum_{m=1}^{p}\sum_{i=1}^{n}(vec(X_{im}^P) - vec(\bar{X}^P))^T(vec(X_{im}^P) - vec(\bar{X}^P)) = \sum_{m=1}^{p}\sum_{i=1}^{n}\left\|X_{im}^P - \bar{X}^P\right\|^2 = RSS$$

(if the PC's were extracted from V, then $(np-1)\sum_{r=1}^{2k-4} SSS_r = \sum_{m=1}^{p}\sum_{i=1}^{n}\left\|X_{im}^P - \bar{X}^P\right\|^2$).

Furthermore,

$$\sum_{r=1}^{2k-4} SSC_r = \sum_{r=1}^{2k}\sum_{i=1}^{n}\sum_{m=1}^{p}(\bar{s}_{r_i} - \bar{s}_r)^2 = \sum_{i=1}^{n}\sum_{m=1}^{p}\left\|\bar{X}_i^P - \bar{X}^P\right\|^2,$$

$$\sum_{r=1}^{2k-4} SSO_r = \sum_{r=1}^{2k}\sum_{i=1}^{n}\sum_{m=1}^{p}(\bar{s}_{r_m} - \bar{s}_r)^2 = \sum_{i=1}^{n}\sum_{m=1}^{p}\left\|\bar{X}_m^P - \bar{X}^P\right\|^2,$$

$$\sum_{r=1}^{2k-4} SSE_r = \sum_{r=1}^{2k}\sum_{i=1}^{n}\sum_{m=1}^{p}(s_{r_{im}} - \bar{s}_{r_i} - \bar{s}_{r_m} + \bar{s}_r)^2 = \sum_{i=1}^{n}\sum_{m=1}^{p}\left\|X_{im}^P - \bar{X}_i^P - \bar{X}_m^P - \bar{X}^P\right\|^2, \qquad (4.8)$$

where $\bar{X}_i^P$ and $\bar{X}_m^P$ are the coordinate-wise averages of the Procrustes fits for case (patient) i, i=1,...,n, (averaged over operators) and for operator m, m=1,...,p (averaged over cases), respectively. Therefore from (4.7), we have:

$$\sum_{m=1}^{p}\sum_{i=1}^{n}\left\|X_{im}^{P}-\bar{X}^{P}\right\|^{2}=\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|\bar{X}_{i}^{P}-\bar{X}^{P}\right\|^{2}+\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|\bar{X}_{m}^{P}-\bar{X}^{P}\right\|^{2}+\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|X_{im}^{P}-\bar{X}_{i}^{P}-\bar{X}_{m}^{P}-\bar{X}^{P}\right\|^{2}. \tag{4.9}$$

The identity is analogous to a univariate 2-way partition of the total sum of squares into between cases (patients), operators and other sources of error. Each quantity in (4.9) is a summation of univariate values measuring the differences between observations and their means, with the Euclidean norm providing a univariate measure of difference between observations (configurations).

Note that this is not the 2-way MANOVA identity for the X_{im}^{P}, decomposing $SSP(vec(X_{im}^{P}))$ into sums of squares and products matrices between cases, operators and other sources, given by:

$$SSP(vec(X_{im}^{P}))=SSP(cases)+SSP(operators)+SSP(errors), \tag{4.10}$$

where

$$SSP(vec(X_{im}^{P}))=\sum_{i=1}^{n}\sum_{m=1}^{p}(vec(X_{im}^{P})-vec(\bar{X}^{P}))(vec(X_{im}^{P})-vec(\bar{X}^{P}))^{T},$$

$$SSP(cases)=\sum_{i=1}^{n}\sum_{m=1}^{p}(vec(\bar{X}_{i}^{P})-vec(\bar{X}^{P}))(vec(\bar{X}_{i}^{P})-vec(\bar{X}^{P}))^{T},$$

$$SSP(operators)=\sum_{i=1}^{n}\sum_{m=1}^{p}(vec(\bar{X}_{m}^{P})-vec(\bar{X}^{P}))(vec(\bar{X}_{m}^{P})-vec(\bar{X}^{P}))^{T},$$

$$SSP(errors)=\sum_{i=1}^{n}\sum_{m=1}^{p}\{(vec(\bar{X}_{im}^{P})-vec(\bar{X}_{i}^{P})-vec(\bar{X}_{m}^{P})-vec(\bar{X}^{P}))(vec(\bar{X}_{im}^{P})-vec(\bar{X}_{i}^{P})-vec(\bar{X}_{m}^{P})-vec(\bar{X}^{P}))^{T}\}.$$

Equation (4.9) holds regardless of the GPA method used to obtain the X_{im}^{P}, although recall from section 3.2.2.5 that $\sum_{m=1}^{p}\sum_{i=1}^{n}\left\|X_{im}^{P}-\bar{X}^{P}\right\|^{2}=\sum_{m=1}^{p}\sum_{i=1}^{n}d_{F}^{2}(X_{im},\bar{X}^{P})$, if the iterative method of Gower/Rohlf & Slice (method (ii) in 3.2.2) is used. Otherwise the sums of squared Euclidean norms and full Procrustes distances will only be approximately equal, although this approximation is commonly used by many tangent space methods. Use of the arithmetic means of the Procrustes fits for each case (patient) and operator is also in keeping with the tangent space methods of 3.4.2 (e.g. Hotelling's test), where the groups means and variance measures must be based on configurations registered in one

configurations registered in one consistent tangent space, rather than estimated by a GPA of each group. Therefore, for X_{im}^{P} obtained by any GPA method and with small variations in shape:

$$\sum_{m=1}^{p}\sum_{i=1}^{n} d_F^2(X_{im},\bar{X}^P) \approx \sum_{i=1}^{n}\sum_{m=1}^{p} d_F^2(\bar{X}_i,\bar{X}^P)+\sum_{i=1}^{n}\sum_{m=1}^{p} d_F^2(\bar{X}_m,\bar{X}^P)+\sum_{i=1}^{n}\sum_{m=1}^{p} d_F^2(X_{im},\bar{X}_i^P-\bar{X}_m^P+\bar{X}^P). \tag{4.11}$$

From (4.8), we also have:

$$\sum_{r=1}^{2k-4} MSC_r = \frac{1}{n-1}\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|\bar{X}_i^P-\bar{X}^P\right\|^2,\ \sum_{r=1}^{2k-4} MSO_r = \frac{1}{p-1}\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|\bar{X}_m^P-\bar{X}^P\right\|^2,$$

$$\sum_{r=1}^{2k-4} MSE_r = \frac{1}{(n-1)(p-1)}\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|X_{im}^P-\bar{X}_i^P-\bar{X}_m^P-\bar{X}^P\right\|^2$$

and so from (4.6):

$$\sum_{r=1}^{2k-4}\hat{\sigma}_{C_r}^2 = \frac{1}{p}\left(\frac{1}{n-1}\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|\bar{X}_i^P-\bar{X}^P\right\|^2-\frac{1}{(n-1)(p-1)}\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|X_{im}^P-\bar{X}_i^P-\bar{X}_m^P-\bar{X}^P\right\|^2\right),$$

$$\sum_{r=1}^{2k-4}\hat{\sigma}_{O_r}^2 = \frac{1}{n}\left(\frac{1}{p-1}\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|\bar{X}_m^P-\bar{X}^P\right\|^2-\frac{1}{(n-1)(p-1)}\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|X_{im}^P-\bar{X}_i^P-\bar{X}_m^P-\bar{X}^P\right\|^2\right),$$

$$\sum_{r=1}^{2k-4}\hat{\sigma}_{E_r}^2 = \frac{1}{(n-1)(p-1)}\sum_{i=1}^{n}\sum_{m=1}^{4}\left\|X_{im}^P-\bar{X}_i^P-\bar{X}_m^P-\bar{X}\right\|^2. \tag{4.12}$$

Consequently the same figure for $\hat{R}(C_{All})$ in (4.5) (or $\hat{R}(O_{All})$ or $\hat{R}(E_{All})$) may also be obtained using the calculations above. This would then remove the need to calculate PCs, although as we will see in 4.5, they do provide an informative pictorial summary of the patterns of variation.

4.4.2.3 1-way ANOVA model for Procrustes fits

Note that since

$$\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|X_{im}^P-\bar{X}_i^P\right\|^2 = \sum_{i=1}^{n}\sum_{m=1}^{p}\left\|\bar{X}_m^P-\bar{X}^P\right\|^2+\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|X_{im}^P-\bar{X}_i^P-\bar{X}_m^P+\bar{X}^P\right\|^2,$$

we may also write a 1-way ANOVA identity for the Procrustes fits as:

$$\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|X_{im}^{P}-\overline{X}^{P}\right\|^{2}=\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|\overline{X}_{i}^{P}-\overline{X}^{P}\right\|^{2}+\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|X_{im}^{P}-\overline{X}_{i}^{P}\right\|^{2}, \qquad (4.13)$$

where

$$\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|X_{im}^{P}-\overline{X}^{P}\right\|^{2}=\sum_{r=1}^{2k-4}SSS_{r}=\sum_{r=1}^{2k-4}\sum_{i=1}^{n}\sum_{m=1}^{p}(s_{r_{im}}-\overline{s}_{r})^{2},$$

$$\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|\overline{X}_{i}^{P}-\overline{X}^{P}\right\|^{2}=\sum_{r=1}^{2k-4}SSC_{r}=\sum_{r=1}^{2k}\sum_{i=1}^{n}\sum_{m=1}^{p}(\overline{s}_{r_{i}}-\overline{s}_{r})^{2},$$

$$\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|X_{im}^{P}-\overline{X}_{i}^{P}\right\|^{2}=\sum_{r=1}^{2k-4}SSW_{r}=\sum_{r=1}^{2k}\sum_{i=1}^{n}\sum_{m=1}^{p}(\overline{s}_{r_{im}}-\overline{s}_{r_{i}})^{2},$$

with

$$\sum_{i=1}^{n}\sum_{m=1}^{p}d_{F}^{2}(X_{im}^{P},\overline{X}^{P})\approx\sum_{i=1}^{n}\sum_{m=1}^{p}d_{F}^{2}(\overline{X}_{i}^{P},\overline{X}^{P})+\sum_{i=1}^{n}\sum_{m=1}^{p}d_{F}^{2}(X_{im}^{P},\overline{X}_{i}^{P}) \qquad (4.14)$$

for small variations in shape. SSW_r is as defined in table 4.2 below, the corresponding 1-way layout for r=1,...,2k PC scores of V or $SSP(vec(X_{im}^{P}))$.

Source	d.f.	SS	MS	EMS
Between cases	n-1	$SSC_r=p\sum_{i=1}^{n}(\overline{s}_{r_i}-\overline{s}_r)^2$	$MSC_r=\frac{SSC_r}{(n-1)}$	$p\sigma_{C_r}^2+\sigma_{W_r}^2$
Within cases	n (p-1)	$SSW_r=\sum_{i=1}^{n}\sum_{m=1}^{p}(\overline{s}_{r_{im}}-\overline{s}_{r_i})^2$	$MSW_r=\frac{SSW_r}{n(p-1)}$	$\sigma_{W_r}^2$
Total	np-1	$SSS_r=\sum_{m=1}^{p}\sum_{i=1}^{n}(s_{r_{im}}-\overline{s}_r)^2$		

Table 4.2: ANOVA partition for 2-way random effects model, with $SSS_r=SSC_r+SSW_r$ and $SSW_r=SSO_r+SSE_r$ from table 4.1.

The 1-way layout is used in simple replication reliability studies where data from each of the i=1,...,n cases, or patients is recorded m=1,...,p times but where the order in which the repeated representations were made is irrelevant. In the current study the 2-way model is appropriate, so that we may account for possible systematic operator differences, but in the investigations of later chapters, a 1-way model is required. Instead of (4.2), the score for case i, i=1,...,n on occasion m=1,...,p is given by $S_{r_{im}}=C_{r_i}+W_{r_{im}}$, where again, C_{r_i} denotes the cases' (patients') 'true score' for this

component and $W_{r_{im}}$ the random difference from the true score (mean 0). The terms are again considered mutually independent with variances $\sigma^2_{C_r}$ and $\sigma^2_{W_r}$ respectively and therefore $\text{Var}(S_{r_{im}}) = \sigma^2_{C_r} + \sigma^2_{W_r}$. The coefficient of reliability (Fleiss & Strout, 1977, Bailey & Byres, 1990) for each PC_r would then be:

$$\hat{R}_{1-way}(C_r) = \frac{\hat{\sigma}^2_{C_r}}{\hat{\sigma}^2_{C_r} + \hat{\sigma}^2_{W_r}} \neq \frac{\hat{\sigma}^2_{C_r}}{\hat{\sigma}^2_{C_r} + \hat{\sigma}^2_{O_r} + \hat{\sigma}^2_{E_r}} = \hat{R}_{2-way}(C_r). \qquad (4.15)$$

This is because although $SSW_r = SSO_r + SSE_r$, $\hat{\sigma}^2_{W_r} \neq \hat{\sigma}^2_{O_e} + \hat{\sigma}^2_{E_r}$ (since $MSW_r = (MSO_r + (n-1)MSE_r)/n \neq MSO_r + MSE_r$ in table 4.1) and the formula for $\hat{\sigma}^2_{C_r}$ is the not the same under both models. For the 1-way model:

$$\hat{\sigma}^2_{C_r} = \frac{MSC_r - MSW_r}{p},\ \hat{\sigma}^2_{W_r} = MSW_r\ .$$

Since the PCs are uncorrelated,

$$\sum_{r=1}^{2k-4} Var(S_t) = \sum_{r=1}^{2k-4} \sigma^2_{C_r} + \sum_{r=1}^{2k-4} \sigma^2_{W_r}$$

and so an overall reliability score may be calculated as:

$$\hat{R}_{1-way}(C_r) = \frac{\sum_{r=1}^{2k-4} \sigma^2_{C_r}}{\sum_{r=1}^{2k-4} (\sigma^2_{C_r} + \sigma^2_{W_r})}. \qquad (4.16)$$

From (4.13) and table 4.2:

$$\sum_{r=1}^{2k-4} MSC_r = \frac{1}{n-1} \sum_{i=1}^{n} \sum_{m=1}^{p} \left\| \bar{X}^P_i - \bar{X}^P \right\|^2,\ \sum_{r=1}^{2k-4} MSW_r = \frac{1}{pn-n} \sum_{i=1}^{n} \sum_{m=1}^{p} \left\| \bar{X}^P_{im} - \bar{X}^P_i \right\|^2$$

and so an overall measure of reliability, assuming a 1-way layout for the PC scores could be obtained by computing:

$$\sum_{r=1}^{2k-4}\hat{\sigma}^2_{C_r}=\frac{1}{p}\sum_{r=1}^{2k-4}\left(\frac{1}{n-1}\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|\overline{X}_i^P-\overline{X}^P\right\|^2-\frac{1}{pn-n}\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|\overline{X}_{im}^P-\overline{X}_i^P\right\|^2\right),$$

$$\sum_{r=1}^{2k-4}\hat{\sigma}^2_{W_r}=\frac{1}{pn-n}\sum_{r=1}^{2k-4}\sum_{i=1}^{n}\sum_{m=1}^{p}\left\|\overline{X}_{im}^P-\overline{X}_i^P\right\|^2. \qquad (4.17)$$

4.4.2.4 Corresponding covariance structure of the Procrustes fits

The variance assumptions in 4.4.1 and throughout the subsequent calculations are defined on the PC scores, rather than on the covariance matrix for the $vec(X_{im}^P)$. In the model of equation (3.5), Σ is the covariance matrix of the $vec(E_i)$, where the configurations $(\mu+E_i)$ are estimated by the X_{im}^P. The form of Σ corresponding to that of the variance components models assumed along each PC is given as follows.

If $S=(S_1,....S_{2k-4})^T$ denotes the vector of non-zero PC scores for a configuration, as in (4.2), then if the PCs are independent, we have:

$$Cov(S)= diag(\sigma_{C_1},...,\sigma_{C_{2k-4}})+diag(\sigma_{O_1},...,\sigma_{O_{2k-4}})+diag(\sigma_{E_1},...,\sigma_{E_{2k-4}})$$

$$=diag(\sigma_{C_1}+\sigma_{O_1}+\sigma_{E_1},...,\sigma_{C_{2k-4}}+\sigma_{O_{2k-4}}+\sigma_{E_{2k-4}}).$$

If $\psi_1,...,\psi_{2k-4}$ are the non-zero principal components of Σ (each of length $2k$) along which the S_r were obtained, with $S=U^T vec(E_i)$, where $U^T=(\psi_1^T,...,\psi_{2k-4}^T)^T$, then using the reverse transformation, $vec(E_i)=SU$, the model for the covariance matrix of the Procrustes fits would be:

$$\Sigma=Cov(vec(E_i))=U^T Cov(S)U$$

$$=U^T diag(\sigma_{C_1}+\sigma_{O_1}+\sigma_{E_1},...,\sigma_{C_{2k-4}}+\sigma_{O_{2k-4}}+\sigma_{E_{2k-4}})U=\sum_{r=1}^{2k-4}(\sigma^2_{C_r}+\sigma^2_{O_r}+\sigma^2_{E_r})\psi_r\psi_r^T.$$

The $\psi_1,...,\psi_{2k}$ are estimated by the PC's of V, the sample covariance matrix of the $vec(X_{im}^P)$ and $\sigma_{C_1},...,\sigma_{C_{2k-4}},\sigma_{O_1},...,\sigma_{O_{2k-4}},\sigma_{E_1},...,\sigma_{E_{2k-4}}$ as in.4.1.

An alternative reliability model for the X_{im}^{P} could also be based on the MANOVA partition of sums of squares and products (SSP) matrices in (4.10). In MANOVA applications, determinants of the SSP matrices are used to compare covariance between and within groups and so an overall measure of reliability could be based on these in some way. The main issue however, is how to obtain estimates of components of 'covariance' corresponding to a suitable multivariate random effects model.

4.5 Reliability assessment: Results

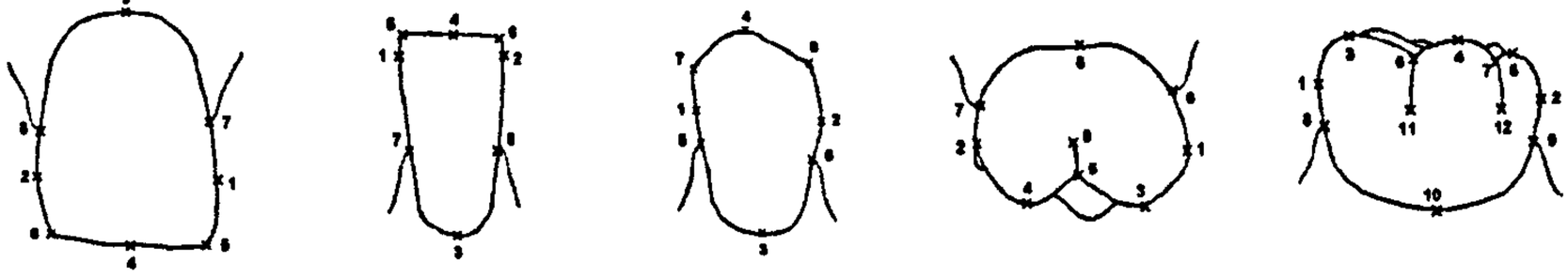
For several reasons, it was not always possible to obtain 4 representations of each of the 20 tooth surfaces of each type. In most instances this was because an operator had omitted a case accidentally. Other reasons for missing data were that the casts had become damaged between operators and, for the lower first molar, an unforeseen eventuality was that in some of the cases this tooth presented itself as a second molar (extraction of the first molar at an early age can mean that the second molar develops to occupy this position in the dental arch instead). Cases where a complete set of different operators' data was unavailable were omitted from the study. The reduced sample sizes for these surfaces are indicated in tables 4.4 and 4.5 below in parentheses.

Note also that using ANOVA to obtain estimates of the variance components along each PC sometimes produces negative values. In this event it is usual to set these estimates to zero. The estimates for the remaining components may then be taken from the ANOVA results or computed again using an iterative procedure such as maximum likelihood estimation, which re-estimates the remaining terms after setting the negative estimates to zero. The difference in estimates produced was found to be negligible (zero, to 2 decimal places) for the data considered here and made no difference to the reliability figures.

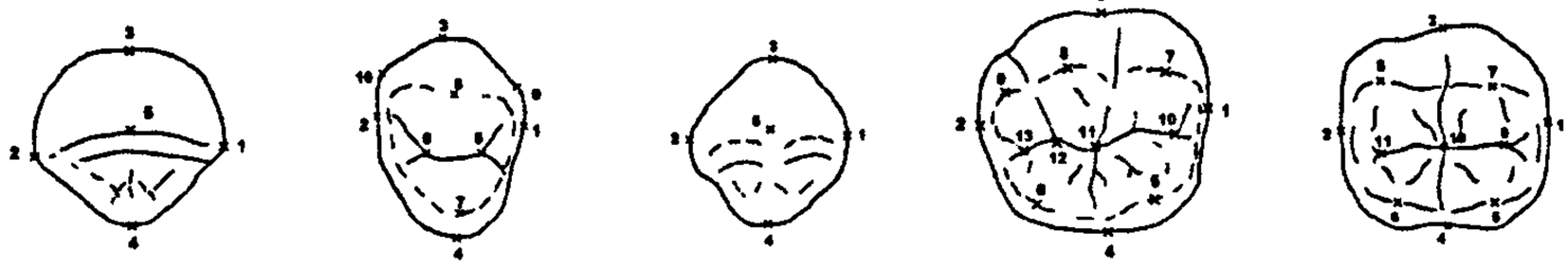
Observed values of reliability have previously been characterised by qualitative benchmarks, such as 'slight' or 'excellent' for values below 0.2 or above 0.8 respectively (Donner & Eliasziw, 1987). While these are frequently quoted, they are only arbitrary classifications and so will be avoided here.

4.5.1 Main results: Overall reliability figures

Table 4.3 displays the overall reliability figures $\hat{R}(C_{All})$ for each of the five buccal and occlusal tooth surfaces.



Upper central incisor	Lower central incisor	Lower canine	Upper first molar	Lower first molar
0.52	**0.65**	**0.55**	**0.36**	**0.46**



Upper central incisor	Upper pre-molar	Lower canine	Lower first molar	Lower second molar
0.30	**0.18**	**0.14**	**0.15**	**0.20**

Table 4.3: Overall proportions of variance $\hat{R}(C_{All})$ attributable to actual between patient variation

For the buccal surfaces, the overall reliability figures, describing the proportion of variation in shape attributable to 'actual variation', ranged from 0.36 to 0.65. For the canine and two incisor surfaces, the variation observed in the Procrustes fits was mostly between patients, $\hat{R}(C_{All})$ ranging from 0.52 to 0.65. The two molar surfaces produced lower results with more than half of the observed variation in shape being attributable to systematic and random operator differences ($\hat{R}(C_{All})$=0.36 & 0.46).

For the occlusal surfaces, the overall reliability figures, $R(C_{All})$, are poor, all being less than the lowest calculated for the buccal surfaces. In all but one data set, more than 80% of occlusal variation in shape was found to be attributable to inconsistencies in the landmark representations ($\hat{R}(C_{All})$<0.20). For the upper central incisor, this figure was little better ($\hat{R}(C_{All})$=0.30).

Further investigation of these results follows in the next two sections. Visual plots of the largest PCs from the Procrustes fits of each set of repeated configurations and their components of variance are examined to see whether the largest patterns of variation contain high proportions of errors and inconsistencies or describe actual variation between cases. We also examine plots of PCs obtained from the 'within' patient covariance structure, which is described in 4.5.3, so that for each set of multiple configurations, we may directly identify the main patterns of error and inconsistency from representations of the same tooth surface.

4.5.2 Individual principal components

For each of the buccal and occlusal surfaces, tables 4.4 and 4.5 display the individual principal component reliability figures, $\hat{R}(C_r)$ obtained from the (total) covariance of each set of Procrustes fits, i.e. $SSP(vec(X_{im}^P))$, as described in section 4.4.1. To save space, only components describing at least 5% of variance in shape are given. The remaining proportions of variance attributable to systematic operator differences, $\hat{R}(O_r)$, and due to random error, $\hat{R}(E_r)$, on each component as well as overall ($\hat{R}(O_{All})$, $\hat{R}(E_{All})$) are also reported.

For both sets of surfaces, the proportion of variance on each PC attributable to random error ($\hat{R}(E_r)$) generally increases with component number, whereas reliability scores $\hat{R}(C_r)$ (between patients) and the proportion of variance due to constant differences between operators ($\hat{R}(O_r)$) generally decrease. This is expected since principal component analysis selectively recovers non-random structure in its early components (Arnqvist & Martensson 1998, Lougheed *et al.* 1991).

For each of the 5 buccal and 5 occlusal surfaces, figs. 4.4 and 4.5 display the first two of these PCs, representing the largest two patterns of variation in shape. Where only one of $\hat{R}(C_r)$ or $\hat{R}(O_r)$, r=1,2 is high, the diagrams provide an indication of likely biological differences or systematic irregularities in operators' data, respectively.

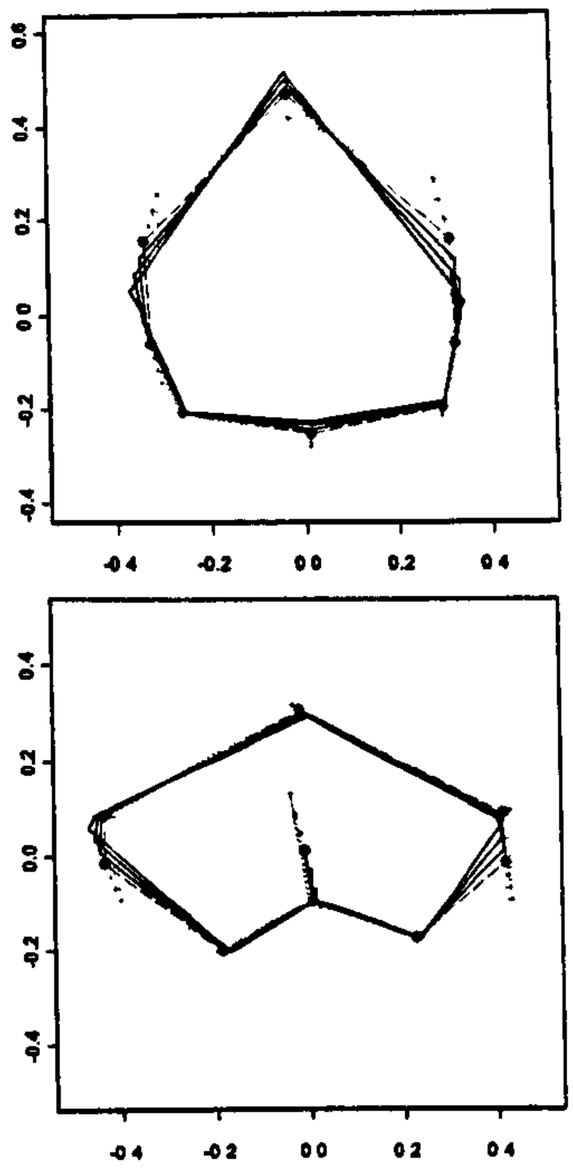
Table 4.4: Buccal surfaces: Proportions of variance attributable to different sources on principal components describing >5% of variance. No. of cases in brackets.

	Upper central incisor (19)				Upper first molar (20)				Lower central incisor (19)				Lower canine (20)				Lower first molar (15)			
	$\hat{R}(C_r)$	$\hat{R}(O_r)$	$\hat{R}(E_r)$	%var	$\hat{R}(C_r)$	$\hat{R}(O_r)$	$\hat{R}(E_r)$	%var	$\hat{R}(C_r)$	$\hat{R}(O_r)$	$\hat{R}(E_r)$	%var	$\hat{R}(C_r)$	$\hat{R}(O_r)$	$\hat{R}(E_r)$	%var	$\hat{R}(C_r)$	$\hat{R}(O_r)$	$\hat{R}(E_r)$	%var
PC_r																				
r=1	0.52	0.24	0.35	41.70	0.07	0.70	0.23	28.30	0.66	0.15	0.19	40.30	0.65	0.07	0.28	30.55	0.67	0.23	0.10	32.96
2	0.73	0.13	0.11	20.04	0.69	0.09	0.22	22.43	0.86	0.05	0.09	23.66	0.68	0.08	0.24	17.54	0.22	0.57	0.21	24.39
3	0.35	0.16	0.52	12.99	0.48	0.16	0.36	15.04	0.66	0.01	0.33	14.23	0.63	0.17	0.20	16.42	0.49	0.14	0.37	8.41
4	0.59	0.13	0.21	11.80	0.51	0.11	0.38	8.16	0.35	0.23	0.42	10.83	0.35	0.07	0.58	11.45	0.58	0.11	0.31	8.01
5					0.29	0.03	0.68	6.39					0.35	0.37	0.28	8.55	0.58	0.00	0.42	7.18
6													0.51	0.01	0.48	6.45				
All	0.52	0.19	0.29		0.36	0.28	0.36		0.65	0.11	0.25		0.55	0.11	0.34		0.46	0.25	0.29	

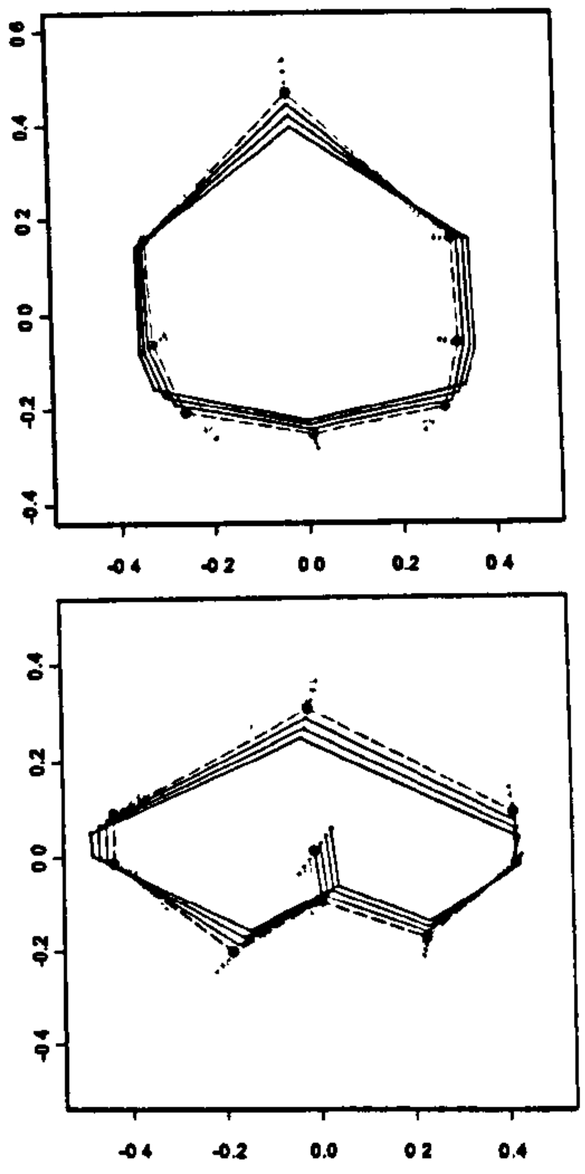
Table 4.5: Occlusal surfaces: Proportions of variance attributable to different sources on principal components describing >5% of variance. No. of cases in brackets.

	Upper central incisor (19)				Upper first pre-molar (18)				Lower canine (20)				Lower first molar (15)				Lower second molar (20)			
	$\hat{R}(C_r)$	$\hat{R}(O_r)$	$\hat{R}(E_r)$	%var	$\hat{R}(C_r)$	$\hat{R}(O_r)$	$\hat{R}(E_r)$	%var	$\hat{R}(C_r)$	$\hat{R}(O_r)$	$\hat{R}(E_r)$	%var	$\hat{R}(C_r)$	$\hat{R}(O_r)$	$\hat{R}(E_r)$	%var	$\hat{R}(C_r)$	$\hat{R}(O_r)$	$\hat{R}(E_r)$	%var
PC_r																				
r=1	0.23	0.55	0.22	59.75	0.06	0.75	0.19	30.03	0.06	0.47	0.47	46.99	0.01	0.92	0.07	26.03	0.02	0.90	0.08	26.54
2	0.65	0.18	0.17	22.04	0.43	0.14	0.43	15.79	0.25	0.36	0.39	23.71	0.18	0.34	0.48	15.61	0.28	0.36	0.36	13.80
3	0.10	0.01	0.89	7.18	0.16	0.25	0.59	13.14	0.21	0.03	0.76	9.93	0.18	0.42	0.40	10.43	0.38	0.08	0.54	10.44
4	0.13	0.00	0.87	6.34	0.17	0.02	0.81	8.12	0.07	0.00	0.93	8.16	0.33	0.05	0.62	7.85	0.26	0.14	0.60	8.93
5					0.30	0.06	0.64	6.18	0.29	0.03	0.68	5.99	0.18	0.00	0.82	6.33	0.29	0.04	0.67	7.36
6					0.12	0.00	0.88	5.28									0.37	0.21	0.42	6.93
All	0.30	0.38	0.32		0.18	0.34	0.48		0.14	0.32	0.54		0.15	0.38	0.48		0.20	0.37	0.44	

41.7%



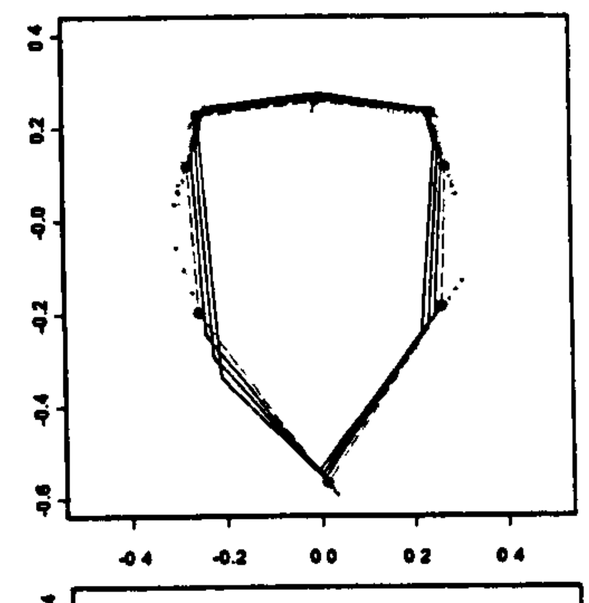
20.0%



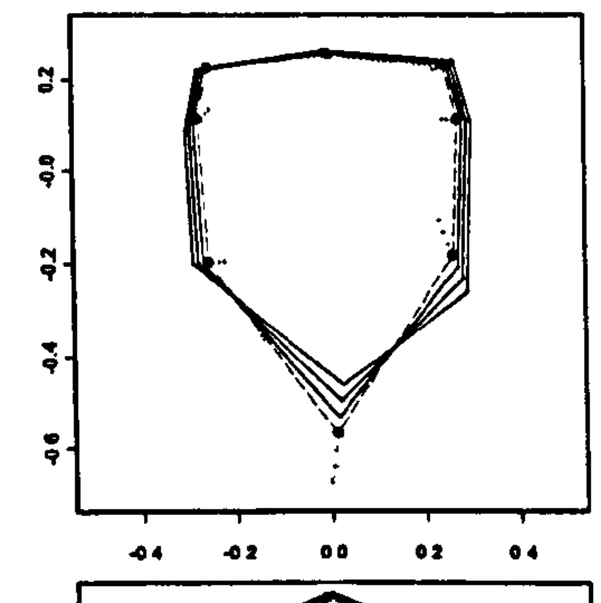
28.3%

22.4%

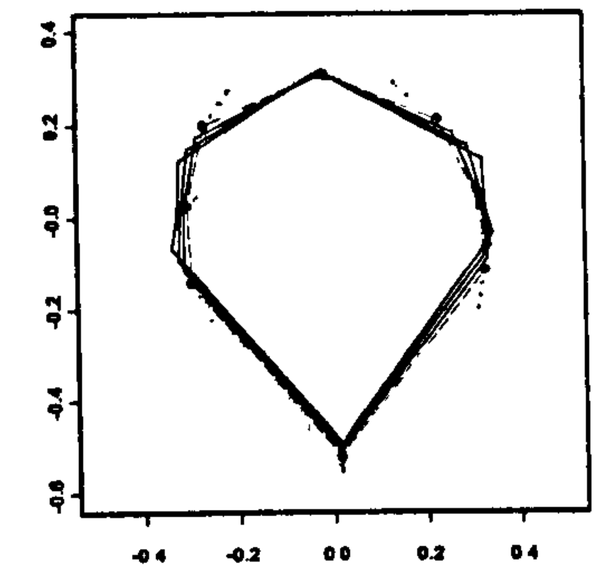
40.3%



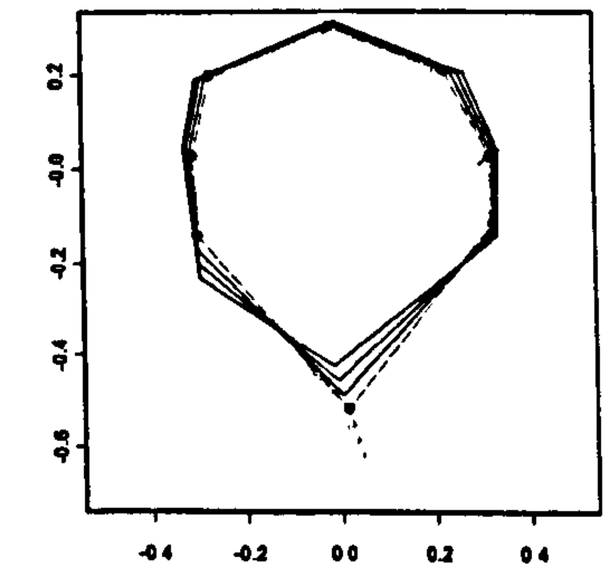
23.7%



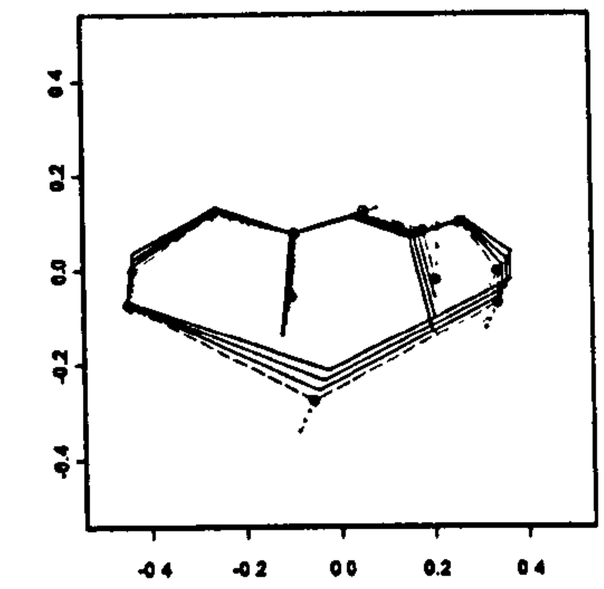
30.6%



17.5%



33.0%



24.4%

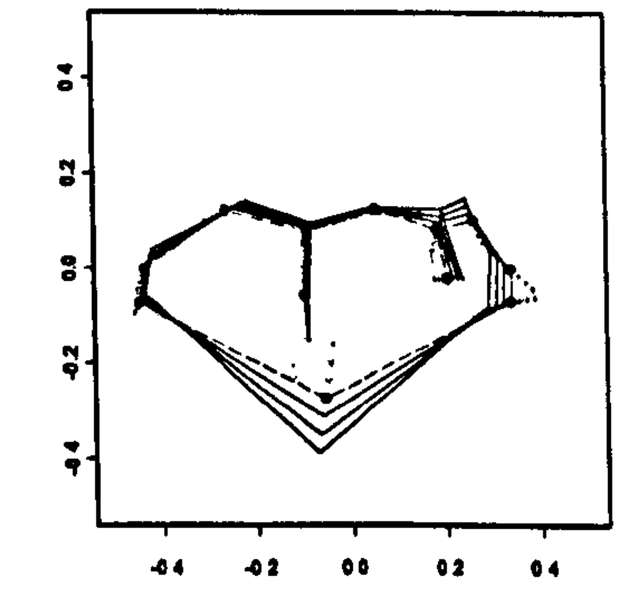


Fig 4.4: First (left column) and second (right column) principal components of variation from buccal surfaces. Lines represent configurations at -3,-2,-1 (dotted) and +1,+2,+3 (solid) standard deviations either side of the mean shape (dashed), along each PC. (Top row) Upper right central incisor. (Second row) Upper right first pre-molar. (Middle row) Lower left incisor. (Fourth row) Lower left canine. (Bottom row) Lower left second molar. (All rows) Labial landmarks at top, mesial landmarks on right. Percentage variation described by each PC indicated next to each plot.

4.5.2.1 Buccal surfaces

$\hat{R}(C_1)$ and $\hat{R}(C_2)$ from the first two principal components were found to be fairly high for each of the upper and lower central incisors and the lower canine surfaces (0.52, 0.66 and 0.65 respectively for PC_1, 0.73,0.86 and 0.68 for PC_2), and so the two largest components mainly describe actual variations in tooth shape between patients. For both incisors, PC_1 in fig. 4.4 describes the shape of the gum in relation to the taper of the sides of the teeth. If the papillae landmarks are located further towards the proximal edge, relative to the gingival endpoint of the long axis of the clinical crown, the teeth are more tapered in shape. The first PC for the lower canine describes how the cusp tip is more pronounced when the distance between the mesial and distal corners is similar to the mesio-distal width. The second component for each of the three surfaces mentioned so far indicates variation in the gingival end point of the long axis of the clinical crown (and hence the shape of the gum) relative to the width. These teeth appear to be proportionally narrower when more of the crown is exposed. $\hat{R}(C_1)$=0.67 for the lower first molar with PC_1 in fig. 4.4 describing the association between the relative height of the tooth and the lengths of both the mesial and distal buccal grooves, the distal cusp being relatively larger (wider) when the teeth are flatter in shape. For the upper first molar, $\hat{R}(C_2)$=0.69 and PC_2 contrasts the relative widths of the mesial and distal cusps with the height of the visible surface. This would suggest that for 'shorter' teeth (with smaller occlusal-gingival height), the distal cusp will be smaller than the mesial, but the converse will be true when height is proportionally larger relative to its width.

For the upper first molar, $\hat{R}(O_1)$=0.70 and so PC_1 in fig. 4.4 indicates that differences in the position of the mesio-distal width relative to the length of the buccal groove are largely attributable to consistent operator differences. Individuals who consistently placed the mesio-distal diameter of the tooth more occlusally/gingivally, also tended to record that the buccal groove was longer/shorter. In addition, $\hat{R}(O_2)$=0.57 for the lower first molar with PC_2 in fig. 4.4 describing variation in the relative height of the tooth and the length of the buccal groove. One possible explanation for both these observations would be that operators consistently orientate the surface differently in the

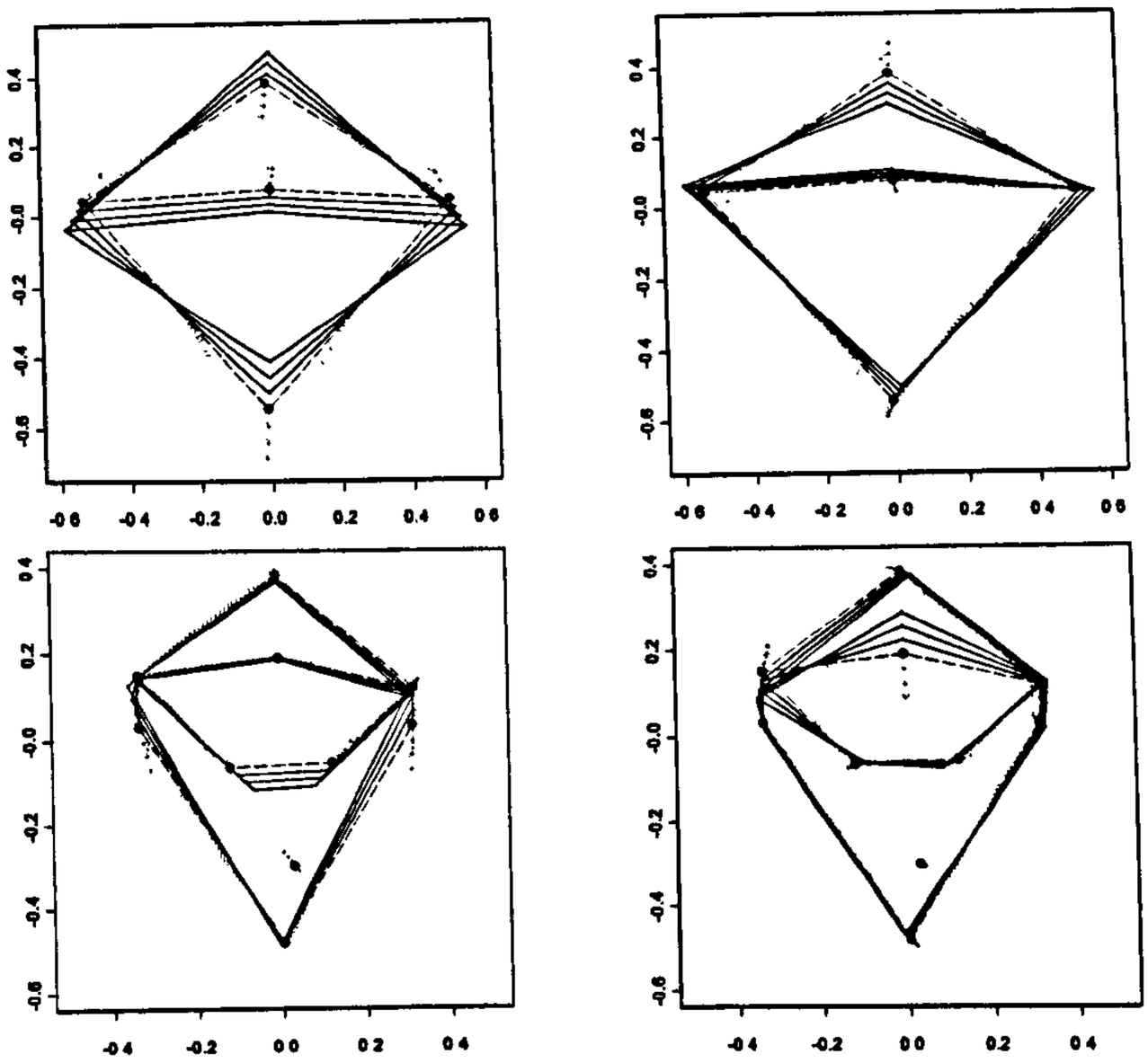
occlusal-gingival direction, before identifying the affected positions on screen. For several cases, an examination of the operators' images appeared to support this, however variation in these directions was also evident in the location of landmarks, once imaged had already been obtained.

4.5.2.2 Occlusal surfaces

The second principal component for the upper central incisor is the only direction of variation for any occlusal surface to offer reliable information on variation between patients ($\hat{R}(C_2)$=0.65). In fig. 4.5 this contrasts the bucco-lingual dimension with the mesio-distal width.

For all occlusal surfaces, a large proportion of variance on the first PC is attributable to operators' systematic errors. For the upper central incisor and lower canine PC_1 in fig. 4.5 shows differences in the position of the incisal edge or cusp edge (and hence the mesio-distal diameter) along the bucco-lingual width. For the lower first molar and lower second molar, $\hat{R}(O_2)$>90%. Each of the first components in fig. 4.5 contrasts the positions of all four cusp tips, relative to the endpoints of the mesio-distal width. The cusp tips are located more lingually when the mesio-distal width is placed more towards the outer buccal surface. For the upper first pre-molar, systematic operator differences exist along PC_1 in the position of the mesio-distal width and the position of the fissure, in the bucco-lingual direction. When operators place the mesio-distal diameter nearer towards the buccal cusp, they also locate the fissure back towards the lingual cusp. This and all the observations from the first PCs would most likely be due to individuals' consistent orientation differences in the imaging of this surface in the bucco-lingual direction, or where operators place the mesio-distal width having acquired the image. For the upper first pre-molar one would also expect that the position of the labial cusp tip, relative to the buccal endpoint of the bucco-lingual width would also be affected by orientation differences. This is represented in the second PC for this tooth in fig. 4, as actual and non-consistent operator variation ($\hat{R}(C_2)$=0.43, $\hat{R}(E_2)$=0.43), along with changes in the distal position of the maximum cusp width. Note that here, and on the other second components not discussed, the variations in scores are not solely attributable to just one source and so interpretation is not as straightforward.

59.6%

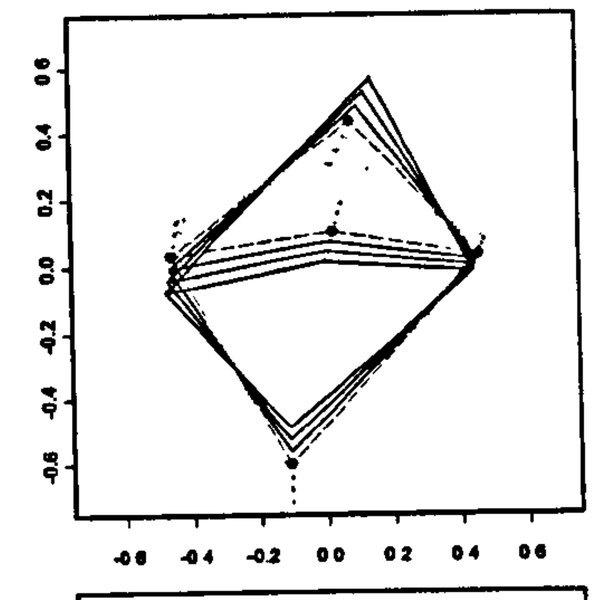


22.0%

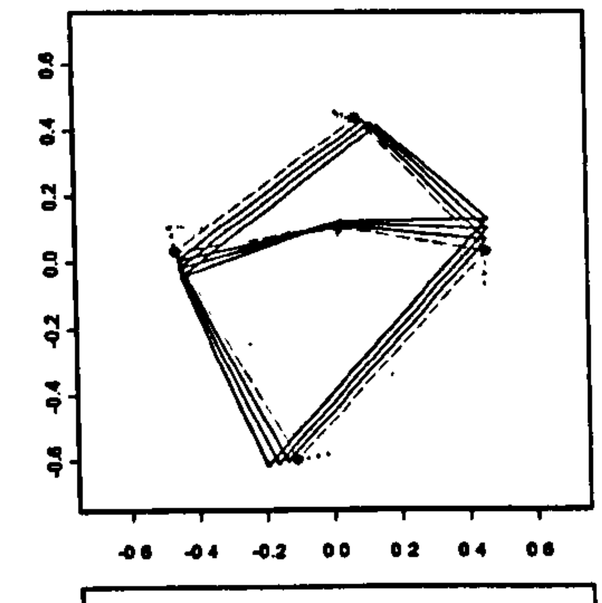
30.0%

15.8%

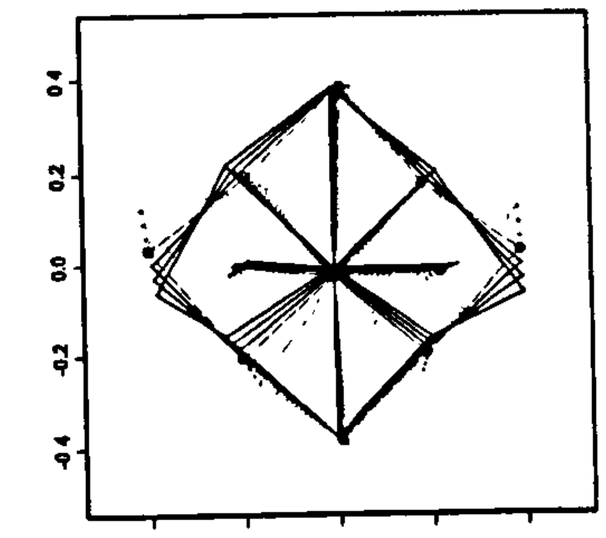
47.0%



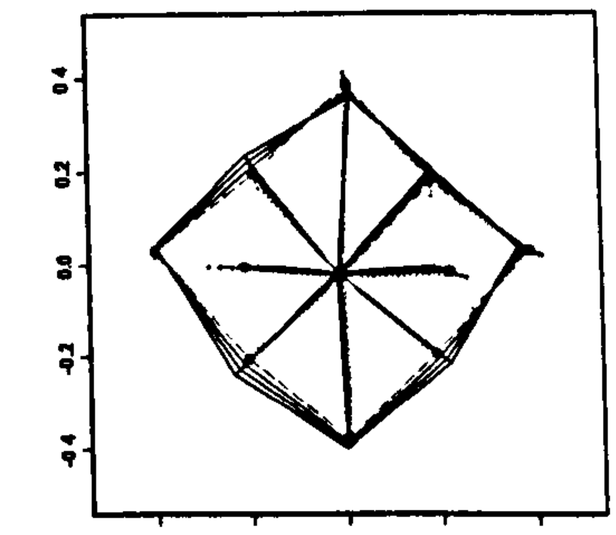
23.7%



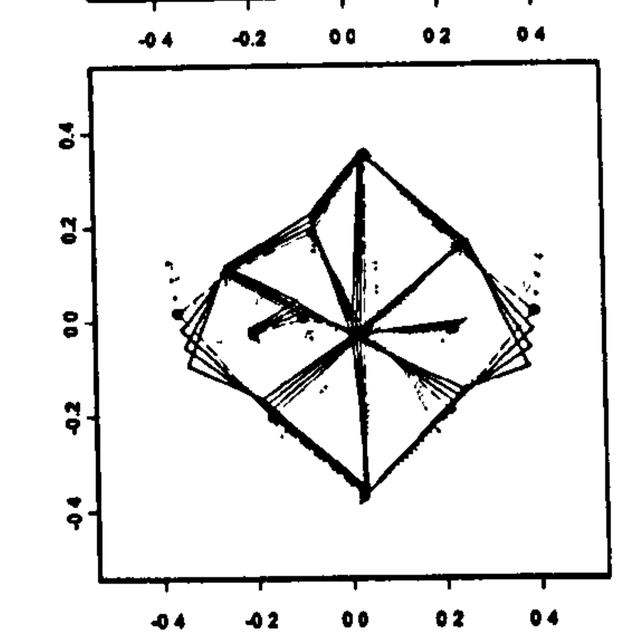
26.0%



15.6%



26.5%



13.8%

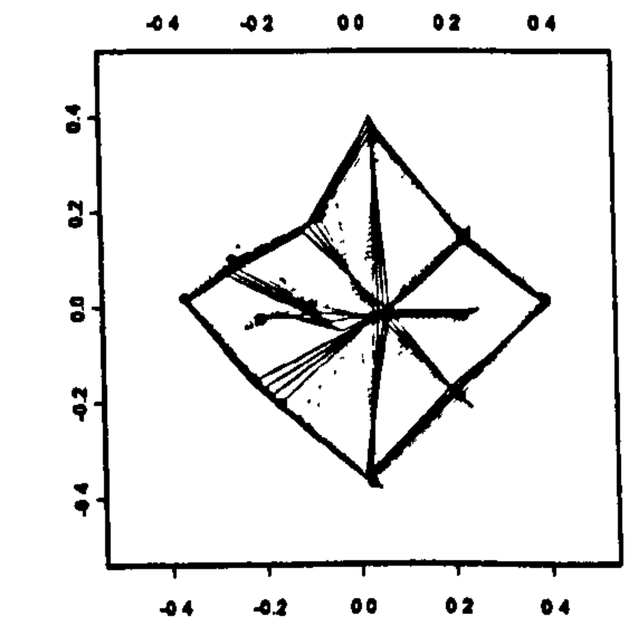


Fig 4.5: First (left column) and second (right column) principal components of variation from occlusal surfaces. Lines represent configs. at -3,-2,-1 (dotted) and +1,+2,+3 (solid) standard deviations either side of the grand mean shape (dashed), along each PC. (Top row) Upper right central incisor. (Second row) Upper right first molar. (Middle row) Lower left central incisor. (Fourth row) Lower left canine. (Bottom row) Lower left first molar. (Rows 1 and 2) Gingival landmarks at top, mesial on right. (Rows 3 to 5) Gingival landmarks at bottom, mesial on left. Percentage variation described by each PC indicated next to each plot.

4.5.3 Within case principal components

For each of the ten data sets, the principal components above are calculated from all variation in shape represented in the Procrustes fits, i.e. $SSP(vec(X_{im}^{P}))$ or V in (4.1). Consequently each set of component scores contains some 'real' variance between patients and some due to operator inconsistencies. The descriptions above suggest which patterns of variation contain large proportions of error, but do not directly indicate the actual areas of systematic and non-systematic operator inconsistency. Instead, an examination of the 'within case (patient)' covariance structures in the each set of Procrustes fits should help identify the patterns and directions of unwanted variation.

Since the 2-way MANOVA identity for $SSP(vec(X_{im}^{P}))$ in (4.10), is 'balanced', i.e. there are p=4 representations of each tooth by each of the same 4 operators, we can rewrite the partition of the total sum of squares and products in (4.10) as follows, corresponding to a 1-way MANOVA partition between patients and within patients. Writing

$$SSP(within\ cases) = SSP(operators) + SSP(errors), \qquad (4.18)$$

where $SSP(operators)$ and $SSP(errors)$ are as defined in 4.4.2.2, equation (4.10) becomes:

$$SSP(vec(X_{im}^{P})) = SSP(cases) + SSP(within\ cases),$$

where

$$SSP(vec(X_{im}^{P})) = \sum_{i=1}^{n}\sum_{m=1}^{p}(vec(X_{im}^{P}) - vec(\overline{X}^{P}))(vec(X_{im}^{P}) - vec(\overline{X}^{P}))^{T},$$

$$SSP(cases) = \sum_{i=1}^{n}\sum_{m=1}^{p}(vec(\overline{X}_{i}^{P}) - vec(\overline{X}^{P}))(vec(\overline{X}_{i}^{P}) - vec(\overline{X}^{P}))^{T},$$

$$SSP(within\ cases) = \sum_{i=1}^{n}\sum_{m=1}^{p}(vec(X_{im}^{P}) - vec(\overline{X}_{i}^{P}))(vec(X_{im}^{P}) - vec(\overline{X}_{i}^{P}))^{T}.$$

This collects all variation not attributable to actual variation between cases (patients) into a single covariance matrix, which may then be analysed by principal components. As defined above, $SSP(within\ cases)$ is calculated as a sum of the covariation within

each case about its mean. We can then plot the PCs to visualise and identify the main directions and patterns of inconsistency and error on each tooth surface.

Note that since (4.18) combines sums of squares and products, rather than mean squares, from a 2-way into a 1-way MANOVA partition, it is still appropriate. (Recall that in 4.4.2.2, a 1-way ANOVA partition of the PC scores gave different reliability figures from a 2-way partition, due to the fact that although $SSO_r + SSE_r = SSW_r$, $MSO_r + MSE_r \neq MSW_r$). Here we are only interested in examining patterns of variation other than that between cases (due to operators and errors), by considering PCA of the *SSP*(*within cases*) matrix, rather than estimating components of variance.

Having obtained the principal patterns of variation from each *SSP*(*within cases*), values of the standardised component scores are displayed as variation either side of the grand mean shape for each of the 10 datasets. Technically the components describe variation about each case mean but this would involve plotting PCs for each of the ≤20 cases. Instead the Procrustes mean is used as a 'typical' individual tooth shape and the within case variation, represented by scores at –3, -2, -1 and +1, +2, +3 standard deviations either side of the mean standardised score, displayed as configurations either side of the overall mean shape.

For each of the 5 buccal and 5 occlusal surfaces, figs. 4.6 and 4.7 display the first two of the within case PCs representing the largest two patterns of within case variation in shape, as variation about the grand mean. Again there will be $2k$-4 PCs for each set of k-landmark configurations and so to save space we will again examine only the first two PCs for each surface. However, this still meant that for every surface we consider every PC describing at least 12% of the within case variation.

Note that for each of the first PCs in figs. 4.3 and 4.4 of section 4.5.2 where $\hat{R}(C_1)$ was very low (say <0.3), the first PCs of within case variation in figs. 4.6 and 4.7, describe the same pattern of variation. This is expected since small values of $\hat{R}(C_1)$ indicated that the main pattern of variation in the Procrustes fits comprises mostly within case variation. The surfaces in question are: upper first molar (buccal surface) and all of the

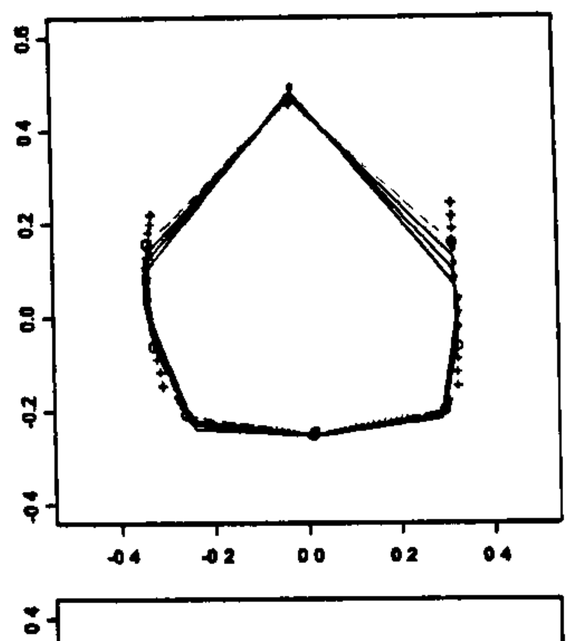
occlusal surfaces. In addition, for the lower canine, lower first molar, and lower second molar, the plots of the second PC from *SSP*(*within cases*) also have a similar appearance to PC_2 obtained from $SSP(vec(X_{im}^{P}))$. $\hat{R}(C_2)$ was again small (<0.3) for each of these surfaces.

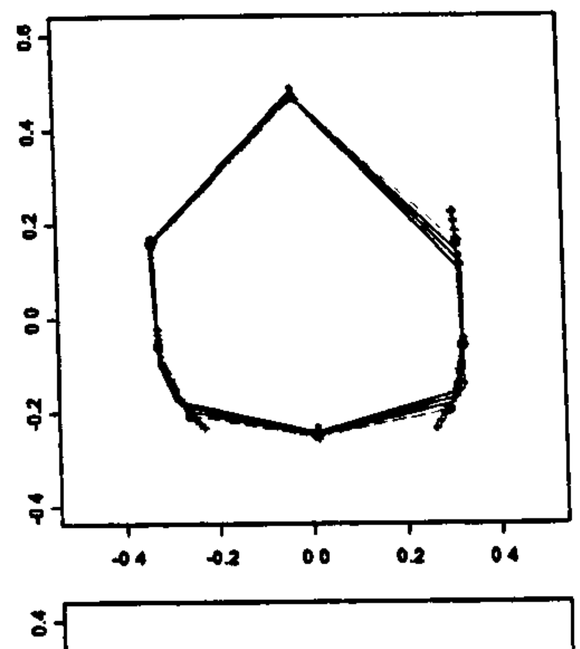
4.5.3.1 Buccal surfaces

For the upper and lower central incisors, as well as the lower canine, the largest components of variance from *SSP*(*within cases*) suggest that the main errors and inconsistencies are in the (vertical) position of the ends of the MD width and the location of the ends of the inter-dental papilla landmarks along the sides of the teeth. The definition used for the papilla landmarks was 'the highest/lowest position on the sides of the teeth before being obscured by the inter-dental tissue. Because the papilla curves round from the surface of the crown and tapers between the sides to the contact point with neighboring teeth, these locations can be difficult to identify consistently. Identification of the endpoints of the MD width along the sides of the teeth was also reported to be uncertain in practice, particularly on cases where the sides of the teeth are relatively straight in this region.

For the upper central incisors and lower canines an additional source of inconsistency represented in the second PCs appears to be in the landmarks at the mesial and distal corners of the incisal edge/cusp. Again, these positions were reported by the operators as being difficult to identify with confidence. This variation is regarded by the PCA as independent from the MD and papilla landmark inconsistencies on the first PC for these surfaces. This seems plausible since their identification does not depend on either of these other features in any way. For all three of these surfaces the patterns of variation on the first two PCs together account for over 70% (upper central and lower central) and 55% (lower canine) of observation variation within cases.

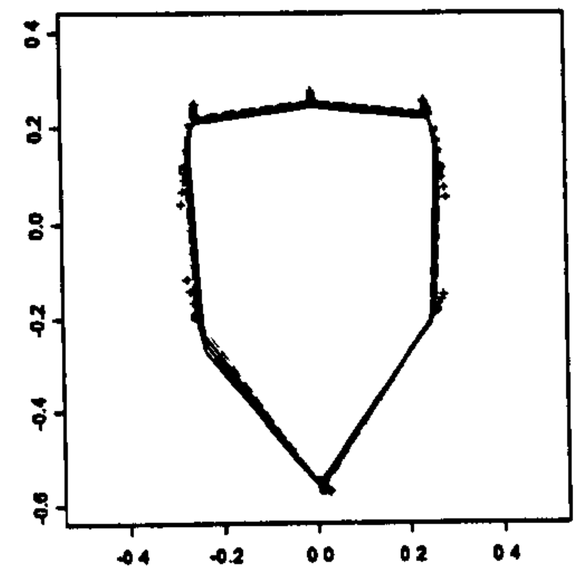
51.0%

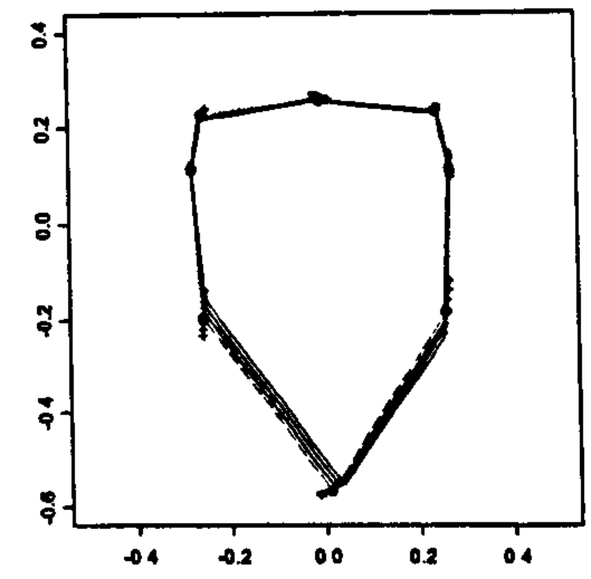




17.4%

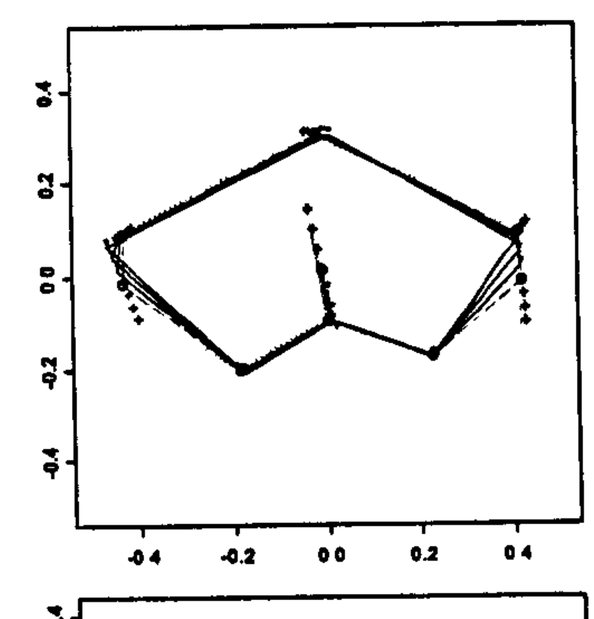
50.3%

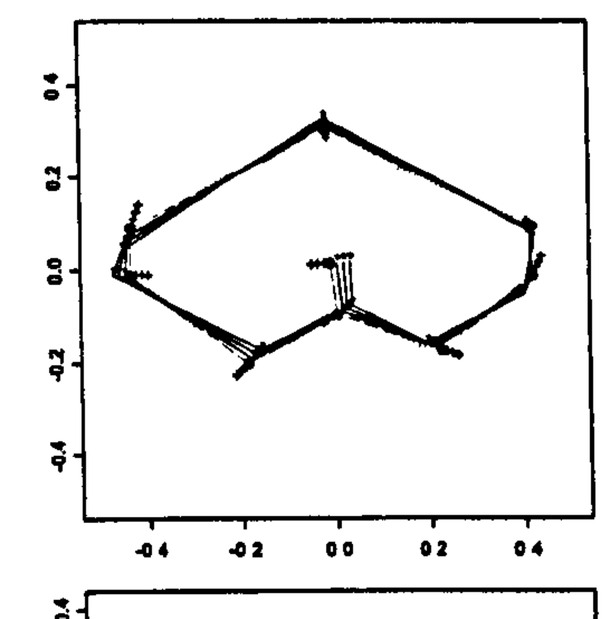




19.5%

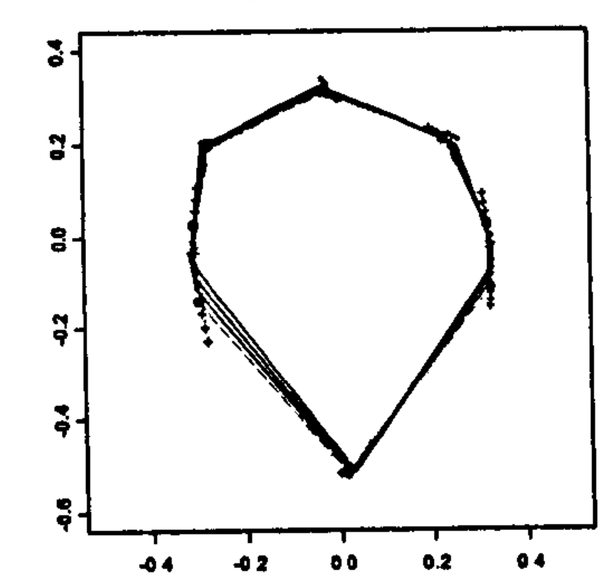
46.3%

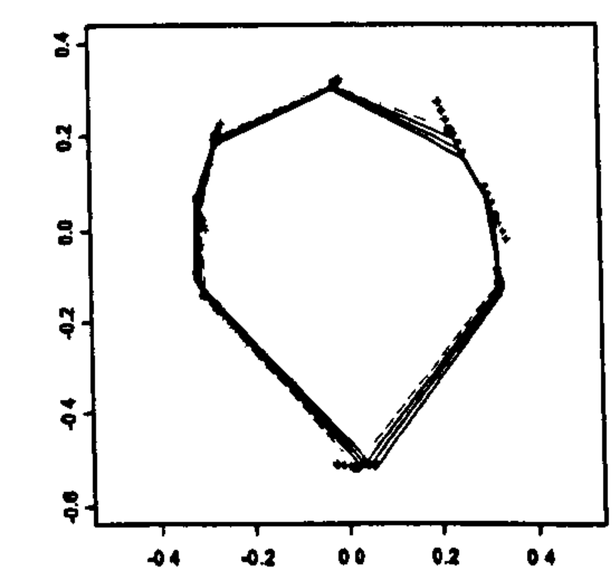




16.0%

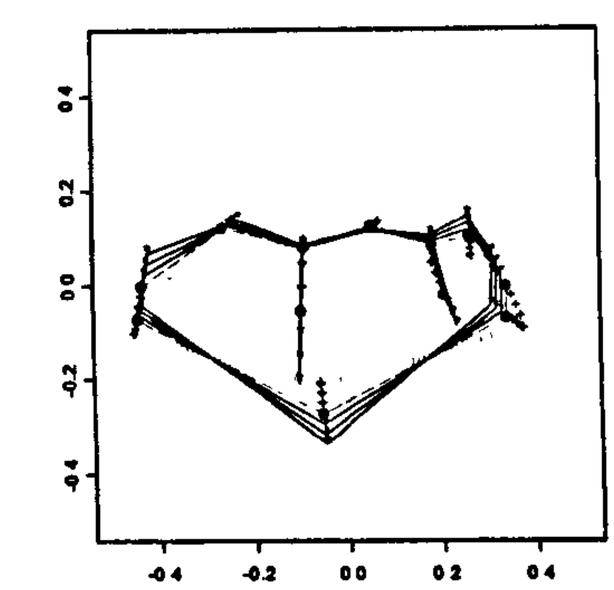
35.8%

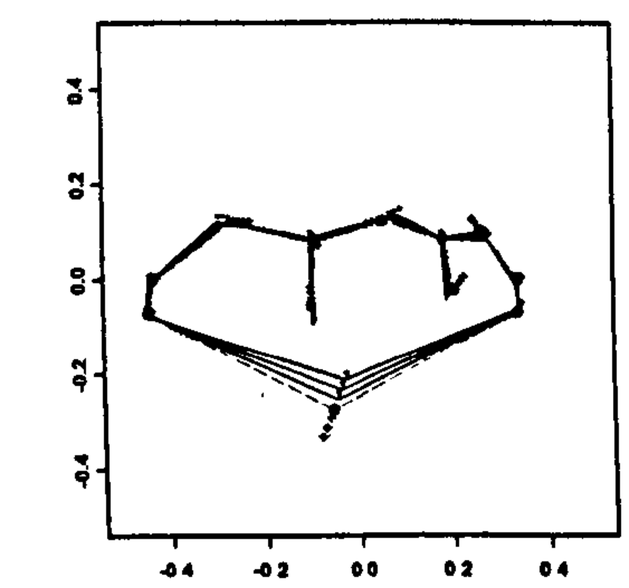




20.6%

50.8%





<10%

Fig 4.6: First (left column) and second (right column) principal components of within case variation from buccal surfaces. Lines represent configurations at -3,-2,-1 (dotted) and +1,+2,+3 (solid) standard deviations either side of the mean shape (dashed), along each PC. (Top row) Upper right central incisor. (Second row) Upper right first pre-molar. (Middle row) Lower left canine. (Fourth row) Lower left first molar. (Bottom row) Lower left second molar. (All rows) Labial landmarks at top, mesial landmarks on right. Percentage variation described by each PC indicated next to each plot.

For the upper first and lower second molar, the first PCs account for over 46% and 50% of variation within cases, respectively. The main source of within case variation appears to be in the vertical positions of the MD endpoints again, here in relation to the start of the buccal grooves. When the MD width is located more gingivally/occlusally, longer/shorter buccal grooves are recorded. These observations would be consistent with orientation inconsistencies at the imaging stage in an occlusal-gingival (OG) direction, before the positions are identified on screen. For several cases, an examination of the operators' images appeared to support this, although the same pattern of variation was also evident on many images that appeared to be consistently orientated in the OG direction. For the upper first molar, the smaller second PC (16% of within case variation) contrasts the relative widths of the mesial and distal cusps. This again suggests possible orientation inconsistencies, but this time in the mesio-distal direction and independent of the variation on the first PC. Examination of the operator's images appeared to confirm this.

Finally, note that on the basis of the first two PCs for all of the buccal surfaces, the endpoints of the LACC and cusp tips appear to be the more reliable of the landmarks considered.

4.5.3.2 Occlusal surfaces

For the upper central incisor and lower canine the largest PC in fig 4.7 indicates within case variation in the position of the incisal edge or cusp edge (and hence also the MD diameter) along the bucco-lingual (BL) width. For the upper central incisor this variation alone, accounts for 73% of variance within cases and 54% for the lower canine. For the upper first premolar, the second PC contrasts the position of the buccal cusp tip along the BL dimension, in a similar manner to the first PCs of the lower canine and upper central incisor. For the lower first molar and lower second molar each of the first PCs contrast the positions of all four/five cusp tips relative to the endpoints of the MD width, accounting for over 35% of variation within cases in both instances. For both of these surfaces, the cusp tips are all located more lingually when the mesio-distal width is placed more towards the outer buccal surface. Consequently, four of the five first PCs and the second PC for the one remaining surface (upper first pre-molar), appear to indicate consistent orientation differences in the imaging of these occlusal

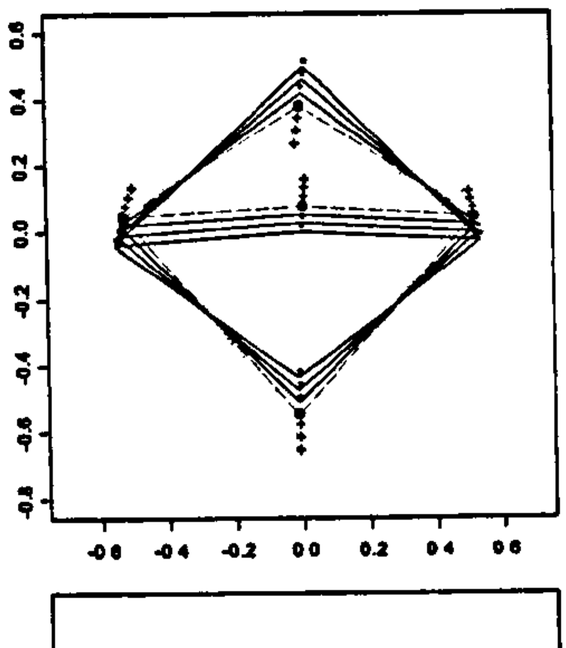
surfaces in the BL direction and/or inconsistencies in where operators place the MD width having acquired the image. Examination of the operator's images appeared to suggest the former. The first component of the upper first pre-molar describes within case variation in the position of the mesio-distal width and the position of the fissure, both also in the bucco-lingual direction. When the MD diameter is located nearer towards the buccal cusp tip, the fissure is located more towards the lingual cusp tip. Since this first component is independent of the second, it would suggest that differences in the position of the MD width occur independently of variations in the orientation of this surface at the imaging stage. One reason for this may be because the MD width is not defined across the cusp or incisal edge, as it is for the lower canine and upper first molar. Together these components account for over 55% of variation within cases for this surface.

For the lower canine, the second PC describes variation in the angle between the BL and MD dimensions, which is not evident in either of the first two PCs of any of the other occlusal surfaces. Endpoints of the MD dimension on the lower canine were reported as being particularly more difficult to locate in this view because of the more rounded shape of its outline. Additionally, the natural position of the maximum BL width is often far from being perpendicular to MD as on other teeth, although it is easier to locate. Along with the first component for this surface, 75% of within case variation is accounted for by the first two PCs.

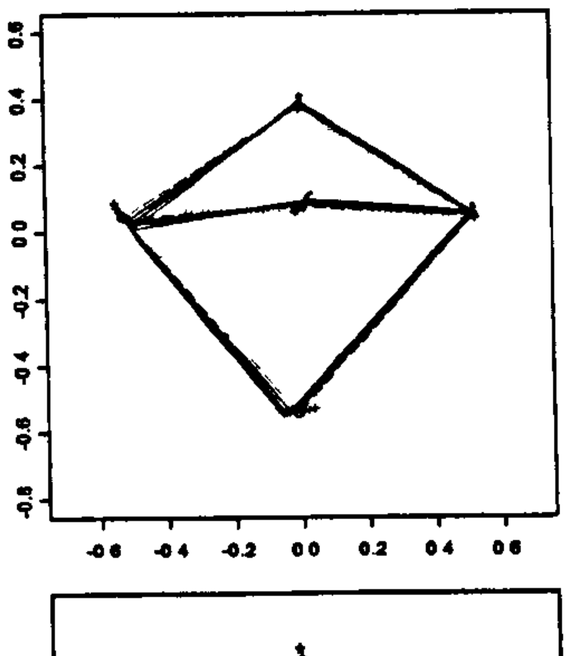
4.6 Summary & discussion of points to be addressed

The overall reliability measures in table 4.3 indicate how reliable any single operator would be in representing the shape of a tooth surface if he/she were to carry out a future Procrustes investigation of shape, using these landmark sets. For the occlusal surfaces overall reliability was found to be particularly poor (<0.3 for all surfaces considered). For the buccal surfaces, the figures were higher than the occlusal, with the canine and two incisor teeth producing the better results, $\hat{R}(C_{All}) = 0.52$ to 0.65.

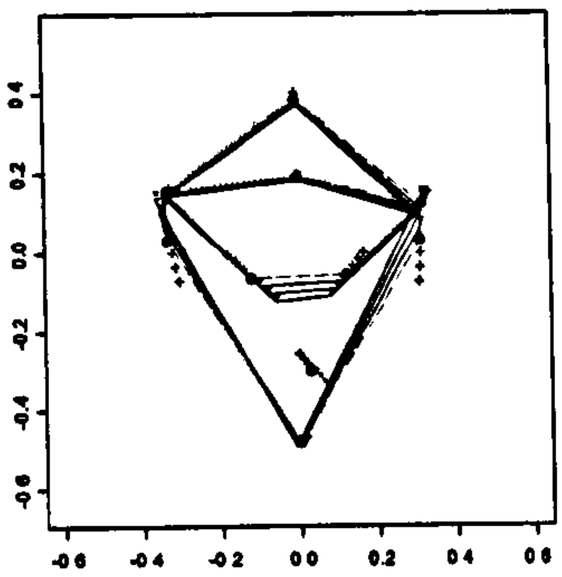
72.6%



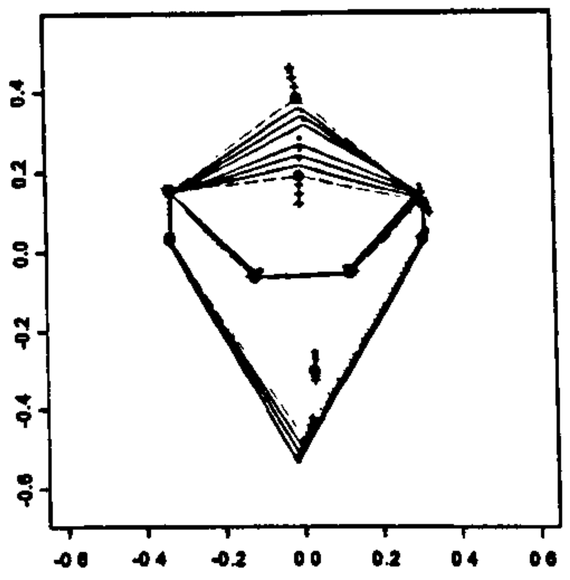
<10%



38.6%



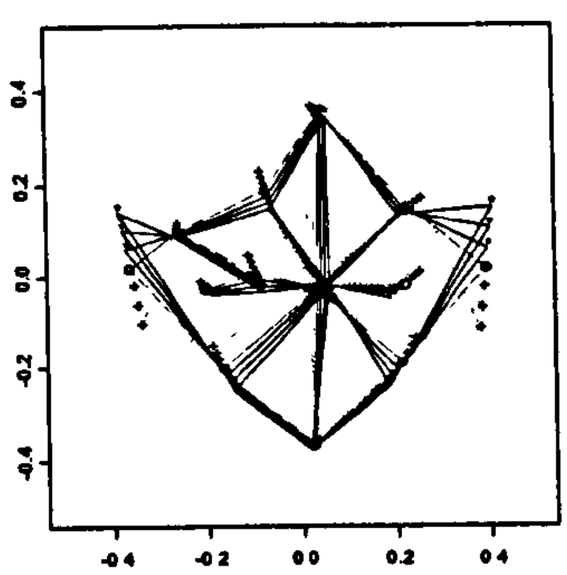
15.9%



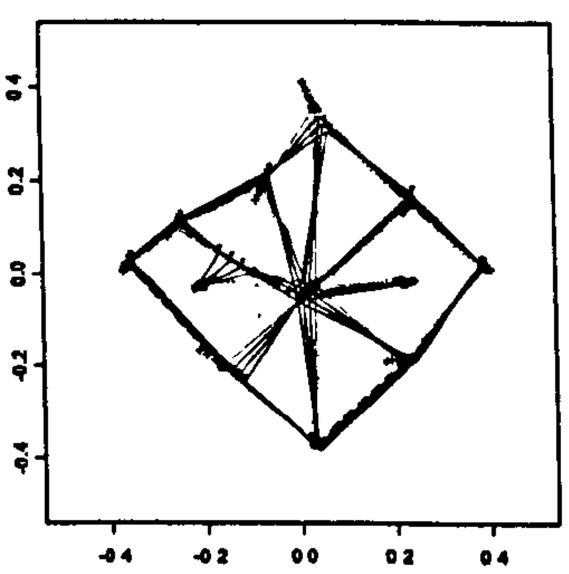
53.5%

21.2%

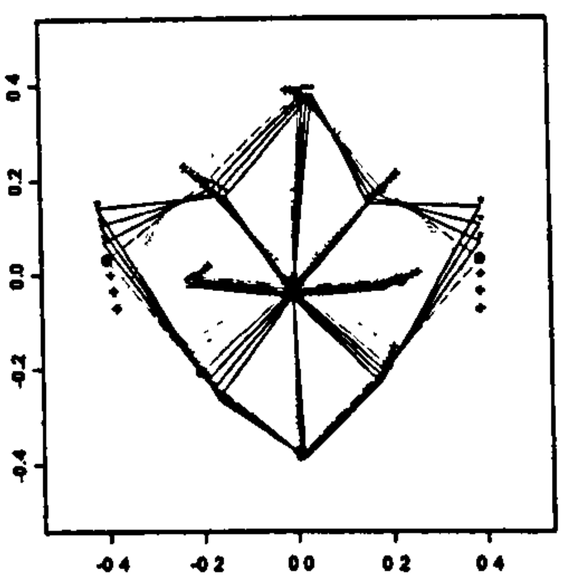
35.0%



<10%



37.5%



<10%

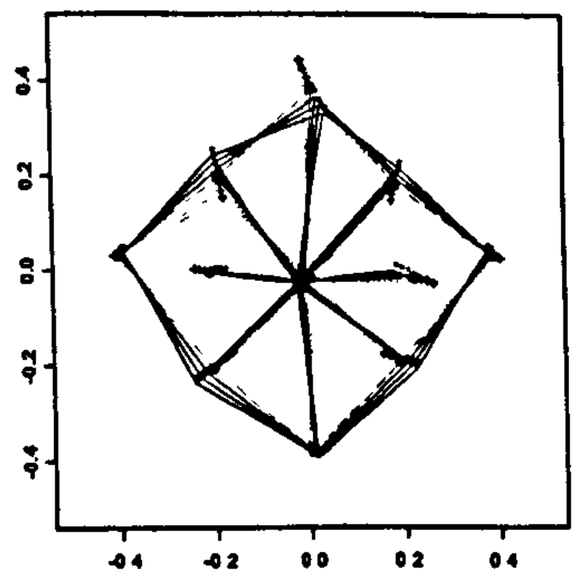


Fig 4.7: First (left column) and second (right column) principal components of within case variation from occlusal surfaces. Lines represent configs. at -3,-2,-1 (dotted) and +1,+2,+3 (solid) standard deviations either side of the grand mean shape (dashed), along each PC. (Top row) Upper right central incisor. (Second row) Upper right first molar. (Middle row) Lower left central incisor. (Fourth row) Lower left canine. (Bottom row) Lower left first molar. (Rows 1 and 2) Gingival landmarks at top, mesial on right. (Rows 3 to 5) Gingival landmarks at bottom, mesial on left. Percentage variation described by each PC indicated next to each plot.

In Chapter 3 a significant difference in mean buccal surface shape was found when comparing the upper central incisors of a group of hypodontia patients with a corresponding control group. While some statistical power clearly remains, the overall figures referred to above suggest the potential for major problems, particularly where smaller differences in shape are to be investigated.

An examination of the 'within case (patient)' covariance structures in the Procrustes fits helped to identify sources and patterns of unwanted variation. The largest principal components for the occlusal configurations suggest that consistent orientation differences prior to imaging in the bucco-lingual direction may be a particular factor in the poor reliability figures. We would expect that small changes in the orientation of occlusal surfaces would have a greater impact on the recorded landmark configurations than in the buccal view, as these surfaces are considerably more three dimensional. Looking down onto an occlusal surface, slight movements cause features such as cusp tips and incisal edges, which are nearer to the camera to move more than those around the gum and edges of the tooth, such as the endpoints of the bucco-lingual and mesio-distal widths, which are further away. An examination of the images for the occlusal surfaces did appear to confirm variation in orientation, particularly in comparison to the buccal surfaces, which would clearly influence their representation as configurations of landmarks. The relatively flatter, more two dimensional buccal surfaces would not suffer from this problem to such an extent, although orientation effects were suggested for the buccal surfaces of the molar teeth, in the occlusal-gingival direction and for the upper first molar, in the mesio-distal direction. The source of unwanted variation in the OG direction was unclear. An examination of operators' images found that same pattern of within case variation may have also resulted from inconsistent (on screen) positioning of landmarks, once similarly orientated images had been obtained.

The within cases analysis and feedback from the operators also suggested difficulty in identifying certain landmarks once a consistently-orientated image had been obtained. For example, for the upper central incisors, lower central incisors and lower canines, positions around the outline of the buccal surface such as the those at the ends of the mesio-distal width, the papilla endpoints and the corners of the incisal edge were found to be more difficult to locate in accordance with other operators than (say) the

landmarks at the ends of the LACC or cusp tips. This issue also has implications in an analysis where each landmark carries equal 'importance'.

Improved definitions for the location of landmarks may help reduce positioning inconsistencies in some of these points and further standardisation for orientation of the surfaces at the imaging stage may reduce other sources of variation. However, it may also be possible to still use information on directions of unwanted variation in other ways as we discuss below.

Note however that whilst these inconsistencies are a problem for comparing the shapes of configurations of landmarks, inter-landmark distances such as the mesio-distal, bucco-lingual or occlusal-gingival widths still produce reliable measurements. The reproducibility of inter-landmark measurements is able to be quantified on the landmark data collected here, since each image was also calibrated by each operator. Having converted the pixel distances between landmark coordinates into scaled measurements, reliability measures were obtained following the same process used for each univariate dataset of PC scores in section 4.4.1. Results for the MD, OG and BL measurements are presented below:

Buccal surfaces	MD	OG	Occlusal surfaces	MD	BL
Upper central incisor	0.74	0.87	Upper central incisor	0.74	0.80
Upper first molar	0.50	0.72	Upper first pre-molar	0.50	0.79
Lower central incisor	0.86	0.90	Lower canine	0.76	0.72
Lower canine	0.91	0.93	Lower first molar	0.82	0.80
Lower first molar	0.80	0.66	Lower second molar	0.87	0.86

Table 4.6: Reliability figures for mesio-distal, bucco-lingual and occlusal-gingival diameters.

In addition to recording inconsistencies, other 'nuisance' variation results from differences between patients in the position of the gingival margin and papilla, as noted in chapter 1 regardless of the reproducibility of points in these regions. Teeth may have the same shape but differences when represented as landmark configurations due to differences in the position of the gum around each tooth. The reliability figures here have no way of taking this into account. Any such differences in these regions will still be regarded as actual variation in shape, but in reality this would be of no interest.

Since we require methods suitable for use on unextracted teeth, the effect of gingival margin irregularity between different individuals must also be considered.

Papilla and gingival landmarks are still useful however, because they provide the best indication we have of the cemento-enamel junction and the relative dimensions of the teeth in these areas. Similarly, points which do not match exactly, but lie (say) around the outline of a tooth or along a recognisable direction of variation, still describe differences in shape perpendicular to these directions. For example, variation around the outline of tooth such as seen for the first two PCs of within case variation for the upper central incisor is of no interest, since we expect most of variation between patients to be in directions normal to the outline for this surface.

Methods of analysis need to accommodate the problems discussed here if landmark data are to be used to describe variations in tooth shape. As outlined in chapter 1, this must only be based on information from available the visible edges and surfaces of a tooth. Superimposition, registration and subsequent descriptive or inferential techniques need to regard teeth which are identical in shape but have differences in Procrustes coordinates because of unreliable positioning of landmarks or differences in the position of a patients gingival tissue as the same shape. Different surfaces each present their own particular difficulties and so a variety of solutions may be required.

In the following chapter we investigate the use of semi-landmarks (Bookstein, 1996a,d,e), an extension of the standard Procrustes methodology, which recognises that certain landmarks may be known to lie along particular lines or curves, but are difficult to locate precisely. The technique uses an additional standardisation step in the matching procedure, allowing these landmarks to move in specified directions away from their Procrustes positions, typically along 'escribed chords'. It is hoped that the information gained from the datasets in this chapter will help in our investigation and development of these ideas, providing useful information and background on possible directions in which the landmarks could be allowed to move.

Chapter 5

Semi-landmarks

5.1 Introduction

In the preceding chapter, it was found that operator inconsistencies in the location of landmarks present a particular problem if teeth are to be represented and analysed as landmark configurations. In addition, variation in the gingival margin and inter-dental papilla positions between individuals will also affect the usefulness of data collected in this region, leading to variations in shape which are of no interest.

Recent work on 'semi-landmarks' (Bookstein, 1996a,d,e) suggests one way in which lack of precise landmark correspondence may be overcome, where location ambiguities are in specific directions, typically along a curve or outline. The technique introduces an additional standardisation step in the matching procedure, combining Procrustes superimposition with the use of 'pairs of thin plate spline' (PTPS) transformations, another established tool for describing and visualising differences in shape of landmark data, already briefly described in 2.2.3.3.1. The PTPS set-up is modified to allow certain landmarks (the semi-landmarks) to move along lines to assign point-to-point correspondences. The configurations are then re-registered by GPA and the process iterated until the semi-landmarks stop moving. The final resulting set of Procrustes registered data may then be analysed in the same way as the results of any GPA, as described in section 3.4 (Bookstein, 1996d, 1998).

The interpolating spline transformations and their use in shape analysis (using a pair of thin-plate splines) are introduced in sections 5.2 and 5.3. This brings together results and descriptions from the somewhat fragmented statistical literature on splines, as well as from more difficult, often abstract, mathematical texts and aims to present these ideas in a more comprehensive, unified and self-contained fashion. Details of how the formulation of the PTPS mapping is adjusted to allow points to move in specified

directions when determining the optimal mapping between configurations then follow, introducing the idea of 'semi-landmarks' in section 5.4. In section 5.3 and 5.4, analogies of the PTPS and semi-landmark formulations with generalised least squares (GLS) are also identified and explored.

Implementation and use of the semi-landmark procedure for the analysis of tooth shape is then considered and discussed in sections 5.5 and 5.6. At the start of these investigations (early 2000), no readily available routines or software existed to perform the computations and so specific routines had to be written, using S-plus. This subsequently highlighted an undocumented issue, regarding the level of convergence that can be achieved when using the method. However, as we discuss, with most datasets this matter will typically be unimportant.

For problems associated with operator inconsistency, use is made of the datasets from the preceding chapter from multiple operators and various tooth types. Effectiveness of the semi-landmark may be judged by how much of an improvement in the reliability figures can be achieved by using the technique to filter out location ambiguities. However, for addressing the issue of unwanted variation due to the position of a patients gum, the semi-landmark routine can be seen to produce some undesirable results, the causes and implications of which are considered in section 5.7.

5.2 Interpolating splines

A spline is an interpolating function for a real valued multivariate data set, the theory of which began Schoenberg (1946). A more recent comprehensive description can be found in Wahba (1990).

5.2.1 The single spline on R^m

5.2.1.1 Definition

Consider data y_j, $\in R^1$, j=1,...,k, associated with 'sites' $t_j=(t_{j[1]},...,t_{j[m]})^T \in R^m$, where [1],...,[$m$] denote the m-dimensional coordinate system in R^m, e.g. in m=2

dimensions, [1]=x and [2]=y (say), with $t=(t_x,t_y)$. In the vast majority of texts, t_j is referred to as the 'source' or 'reference' data and y_j as the univariate 'target' data.

Given an integer-valued smoothness index r, the spline in m-dimensions is the 'smoothest possible' function $\Phi(t)=y$, $t=(t_{[1]},\ldots,t_{[m]})^T\in R^m$, $y\in R^1$, with $\Phi(t_j)=y_j$, $j=1,\ldots,k$, which minimises the roughness or 'bending' penalty below, based on the sum of integrated squared partial derivatives of order r.

$$J_r^m(\Phi)=\sum_{\alpha_1+\ldots+\alpha_m=r}\frac{r!}{\alpha_1!\ldots\alpha_m!}\int_{R^m}\left(\frac{\delta^r\Phi}{\delta t_{[1]}^{\alpha_1}\ldots\delta t_{[m]}^{\alpha_m}}\right)^2\prod_{l=1}^{m}dt_{[l]} \qquad (5.1)$$

where the summation is over all possible combinations of integers $\alpha_1,\ldots,\alpha_d\geq 0$, such that $\alpha_1+\ldots+\alpha_m=r$. One property of this penalty is that it is invariant under translations or rotations of the space of t or y (Kent & Mardia, 1994).

The null space of the penalty function $J_r^m(\Phi)$ is the $M=\binom{m+r-1}{m}$ dimensional space spanned by $\phi_1,\ldots,\phi_M$, the polynomials in m variables of total degree less than or equal to r-1. For example,

if m=1, r=2, then M=2 and the null space is spanned by $\phi_1(t)=1$, $\phi_2(t)=t_{[1]}$;

if m=2, r=2, then M=3 and the null space is spanned by $\phi_1(t)=1$, $\phi_2(t)=t_{[1]},\phi_3(t)=t_{[2]}$;

if m=2, r=3, then M=6 and the null space is spanned by $\phi_1(t)=1$, $\phi_2(t)=t_{[1]}$, $\phi_3(t)=t_{[2]},\phi_4(t)=t_{[1]}^2,\phi_5(t)=t_{[2]}^2$, $\phi_6(t)=t_{[1]}\,t_{[2]}$.

Kent & Mardia (1994) and (originally) Duchon (1976), showed that providing $r>m/2$ and subject to the constraints in (5.3) below, $J_r^m(\Phi)$ has a unique minimiser $\Phi(t)$, given by:

$$\Phi(t)=\Phi((t_{[1]},\ldots,t_{[m]})^T)=\sum_{v=1}^{M}a_v\phi_v(t)+\sum_{j=1}^{k}w_i\sigma_{r,m}(t-t_j)=\phi(t)^T a+s(t)^T w_p \qquad (5.2)$$

with

$$\sum_{j=1}^{k}\phi_v(t)w_j = 0 \text{ for } v=1,\ldots,M \tag{5.3}$$

where

$$s(t)=(\sigma_{r,m}(t-t_1),\ldots,\sigma_{r,m}(t-t_k))^T,\ \phi(t)=(\phi_1(t),\ldots,\phi_M(t))^T,$$

$$a=(a_1,\ldots,a_M)^T,\ w=(w_1,\ldots,w_k)^T,$$

$$\sigma_{r,m}(t-t_j) = \theta_{r,m}\left\|t-t_j\right\|^{2r-m}\log\left\|t-t_j\right\| \quad \text{if } 2r\text{-}m \text{ is even} \tag{5.4}$$

$$= \theta_{m,d}\left\|t-t_j\right\|^{2r-m} \quad \text{otherwise}$$

and

$$\theta_{r,m} = \frac{(-1)^{(m/2)+1+r}}{2^{2r-1}\pi^{m/2}(r-1)!(r-m/2)!} \quad \text{if } 2r\text{-}m \text{ is even}$$

$$= \frac{\Gamma((m/2)-r)}{2^{2r}\pi^{m/2}(r-1)!} \quad \text{otherwise.}$$

The function defines a mapping of the entire space of t onto the space of values y (with $\Phi(t_j)=y_j, j=1,\ldots,k$) and consists of an overall or 'global' component of linear terms in $\phi_v(t)$ and a 'local' component, comprising a weighted sum of $\sigma_{r,m}(t-t_j)$ terms.

$\sigma_{r,m}(t-t_j)$ is known as a Green's function for the r-iterated Laplacian or differential operator Δ^r. In m dimensions the Laplacian is $\Delta^r = \left(\frac{\delta^2}{\delta t_{[1]}^2}+\ldots+\frac{\delta^2}{\delta t_{[m]}^2}\right)^r$ so that $\Delta^r\,\sigma_{r,m}(t-t_j) = \delta_{t_j}$ where δ_{t_j} is the Dirac delta function. This ensures that although $J_r^m(\Phi)$ is defined over all of R^m, $\Delta^r\,\Phi(t)=0$ for $t\neq t_j$, $i=1,\ldots,k$, so that only the observed values of t contribute to the penalty function. Other properties of $\sigma_{r,m}(t-t_j)$ are:

- $\sigma_{r,m}$ is the conditionally positive definite covariance function corresponding to an intrinsic ordinary random field in R^m (see section 6.2.2.2.2).
- $\sigma_{r,m}$ are also known as 'radial basis functions' in the numerical analysis literature (see e.g. Powell, 1987).

- $\sigma_{r,m}$ is 'self similar', i.e. $\sigma_{r,m}(ct-ct_i)\equiv c^{2r-1}\sigma_{r,m}(t-t_i)$ for $c>0$ where $\equiv$ means equal up to an even polynomial in $(t-t_i)$ of degree $2r$-1. Thus $\sigma_{r,m}(t-t_i)$ and $\sigma_{r,m}(ct-ct_i)$ yield the same fitted values for $\Phi(t)$.
- $-\Delta\sigma_{r,m}(t-t_i)=\sigma_{r-1,m}(t-t_i)$.

Using the first and final properties above, it can be shown that the value of the minimised roughness penalty in equation (5.1) may be calculated as:

$$J_r^m(\Phi)=(2\pi)^m\sum_{i=1}^{k}w_i\Phi(t_i)=(2\pi)^m\sum_{j=1}^{k}\sum_{j'=1}^{k}w_jw_{j'}\sigma_{r,m}(t_j-t_{j'})+\sum_{v=1}^{M}a_v\sum_{i=1}^{k}w_i\phi_v(t_i).$$

By satisfying (5.3), $w_1,\ldots,w_k$ are said to provide define an 'M-order increment' or the 'generalised divided difference of order M'; they 'annihilate' all polynomials $\phi_1,\ldots,\phi_M$ of degree less than r. Consequently, the second term in the equation above is 0, leaving:

$$J_r^m(\Phi)=(2\pi)^m\sum_{j=1}^{k}\sum_{j'=1}^{k}w_jw_{j'}\sigma_{r,m}(t_j-t_{j'}). \qquad (5.5)$$

Note that given data $t_1,\ldots,t_k\in R^m$ (and hence $\sigma_{r,m}(t_j-t_{j'})$, $j,j'=1,..,k$), only the w_j terms in (5.2) contribute to the roughness penalty and not the a_v, $v=1,\ldots,M$. In practice the $\theta_{r,m}$ term in (5.4) and the factor of $(2\pi)^m$ above are also usually dropped.

Writing $y_{obs}=(y_1,\ldots,y_k)^T$, the parameters $a=(a_1,\ldots,a_M)^T$ and $w=(w_1,\ldots,w_k)^T$ in (5.2) are found by solving the equations:

$$y_{obs}=Qa+Sw \qquad (5.6)$$

$$Q^Tw=0 \qquad (5.7)$$

where

$$Q=\begin{pmatrix}\phi_1(t_1) & \cdots & \phi_M(t_1)\\ \vdots & \ddots & \vdots\\ \phi_1(t_{1k}) & \cdots & \phi_M(t_k)\end{pmatrix},\ S=\begin{pmatrix}\sigma_{r,m}(t_1-t_1) & \cdots & \sigma_{r,m}(t_1-t_k)\\ \vdots & \ddots & \vdots\\ \sigma_{r,m}(t_k-t_1) & \cdots & \phi_{r,m}(t_k-t_k)\end{pmatrix},$$

with solution given by:

$$a = \Gamma^{21} y_{obs},\ w = \Gamma^{11} y_{obs} \tag{5.8}$$

where $\Gamma^{21} = (Q^T S^{-1} Q)^{-1} Q^T S^{-1}, \Gamma^{11} = S^{-1} - S^{-1} Q (Q^T S^{-1} Q)^{-1} Q^T S^{-1}$. Consequently, the spline from $t \in R^m$ to $y = \Phi(t) \in R^1$ is given by:

$$\Phi(t) = \phi(t)^T B + s(t)^T W = \phi(t)^T \Gamma^{21} y_{obs} + s(t)^T \Gamma^{11} y_{obs} \tag{5.9}$$

with, from (5.5),

$$J_r^m(\Phi) = tr(w^T S w) = tr(y_{obs}^T \Gamma^{11} y_{obs}). \tag{5.10}$$

5.2.1.2 Examples

The most popular choice for *r*, the order of derivatives, is $r=2$, yielding the cubic spline in $m=1$ dimensions and the thin plate spline for $m=2$.

5.2.1.2.1 Cubic spline ($m=1$, $r=2$)

For data $t_j \in R^1$, $j=1,\ldots k$, with associated univariate values or 'heights' $y_1,\ldots, y_k$, the cubic spline is the 'smoothest possible' curve $\Phi(t)$ with constraints $\Phi(t_j) = y_j$, $j=1,\ldots,k$, which uniquely minimises the 'roughness' or 'bending' penalty given by equation (5.1) above as,

$$J_2^1(\Phi) = \sum_{\alpha_1=2} \frac{2!}{2!} \int_{R^1} \left(\frac{\delta^2 \Phi}{\delta t^2} \right)^2 dt = \int_{R^1} \left(\frac{\delta^2 \Phi}{\delta t^2} \right)^2 dt .$$

i.e. the sum of the rates of change of curvature along the curve.

From equation (5.2), with $r=2$, $m=1$, the cubic spline is given by:

$$\Phi(t) = \phi(t)^T a + s(t)^T w = (1 \quad t) \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + (\sigma_{2,2}(t-t_1) \quad \cdots \quad \sigma_{2,2}(t-t_k)) \begin{pmatrix} w_1 \\ \vdots \\ w_k \end{pmatrix}$$

$$= a_1 + a_2 t + \sum_{j=1}^{k} w_i \sigma_{2,1}(t-t_j)$$

where $\sigma_{2,1}(t-t_i)=\frac{1}{12}\|t-t_i\|^3$ (or just $\|t-t_i\|^3$), with constraints, from (5.3), $\sum_{i=1}^{k} w_i = 0$ and $\sum_{i=1}^{k} t_i w_i = 0$. The vectors a, w and hence $\Phi(t)$, are given by equations (5.6) to (5.9), with $r=2$, $m=1$.

Green & Silverman (1994) relate the NCS to a mechanical spline; a thin length of flexible wood, equipped with sliding weights which constrain the spline to pass through data pairs (t_i, y_i) on a drawing board. Elsewhere it is free to fall into any shape and so takes up a position of minimum strain energy, the leading terms of this energy being proportional to $J_2^1(\Phi)$, see fig. 5.1.

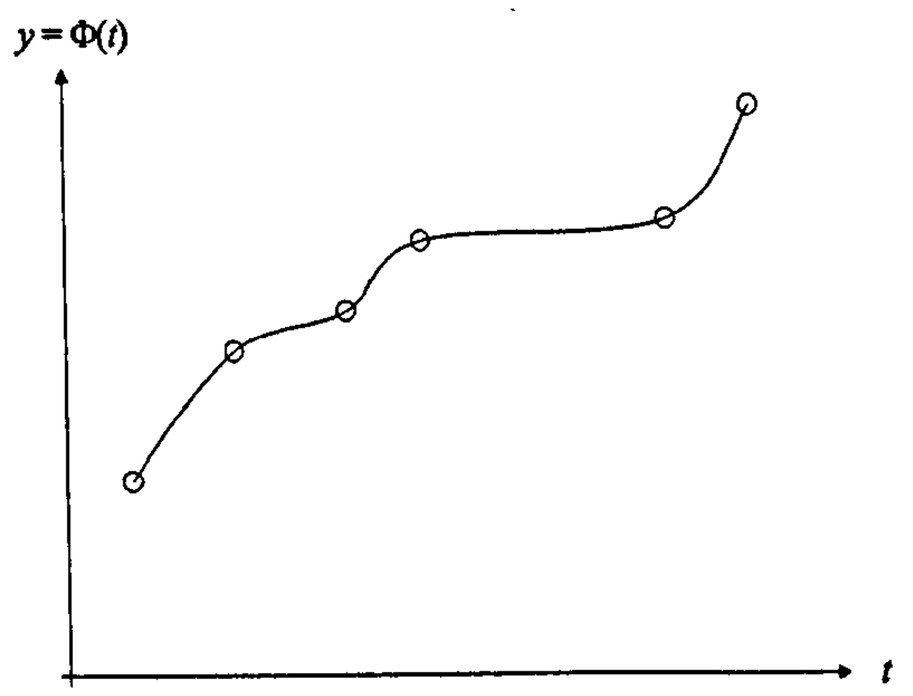


Fig. 5.1: Example of a cubic spline

Since $J_2^1(\Phi)$ depends only on the integrated second derivative, constant or linear terms will not affect its value. A shift of the data in the direction of y or t or a scalar multiplication of y (corresponding to a vertical stretch of the curve) will still give the same measure of 'roughness' (it is independent of $a=(a_1,a_2)^T$).

5.2.1.2.2 Thin Plate Spline (m=2, r=2)

Consider now data $t_1,\ldots,t_k \in R^2$, e.g. coordinates $(t_{1x},t_{1y})^T$, ..., $(t_{kx},t_{ky})^T$ in the (x,y) plane, again with associated real univariate values (heights) $y_1,\ldots, y_k$. As above, the aim is to find a smoothest possible function $\Phi(t)$, which passes through the given heights, so that $\Phi(t_i)=y_i$, $i=1,\ldots,k$. Here Φ describes a surface above the (x,y) plane,

rather than a curve, which (from (5.1) with m=2, r=2) has a minimum 'roughness' or 'bending' measured by the net quadratic variation, which is defined as:

$$J_2^2(\Phi)=\sum_{\alpha_1+\alpha_2=2}\frac{2!}{\alpha_1!\alpha_2!}\int_{R^2}\left(\frac{\delta^2\Phi}{\delta t_{x_1}^{\alpha_1}\delta t_{x_2}^{\alpha_2}}\right)^2\prod_{j=1}^{2}dt_{x_j}=\int_{R^2}\left(\frac{\delta^2\Phi}{\delta t_x^2}\right)^2+2\left(\frac{\delta^2\Phi}{\delta t_x\delta t_y}\right)^2+\left(\frac{\delta^2\Phi}{\delta t_y^2}\right)^2 dt_x dt_y. \tag{5.11}$$

Following (5.2) and (5.3) and writing $a_1=c, a_2=a_x, a_3=a_y, \Phi(t)$ may be written as:

$$\Phi(t)=\phi(t)^T a+s(t)^T w=(1 \quad t_x \quad t_y)\begin{pmatrix}c\\a_x\\a_y\end{pmatrix}+(\sigma_{2,2}(t-t_1) \quad \cdots \quad \sigma_{2,2}(t-t_k))\begin{pmatrix}w_1\\\vdots\\w_k\end{pmatrix} \tag{5.12}$$

$$=c+a_x t_x+a_y t_y+\sum_{i=1}^{n} w_i\sigma_{2,2}(t-t_i)$$

where

$$\sigma_{2,2}(t-t_i)=\frac{1}{8\pi}\|t-t_i\|^2\log\|t-t_i\| \text{ (or } \|t-t_i\|^2\log\|t-t_i\|) \tag{5.13}$$

with constraints, from (5.3):

$$\sum_{i=1}^{k} w_i=0,\ \sum_{i=1}^{k} t_{ix}w_i=0,\ \sum_{i=1}^{k} t_{iy}w_i=0. \tag{5.14}$$

The solutions for a, w and hence $\Phi(t)$, are again given by equations (5.6) to (5.9), here with r=2, m=2.

Work on thin-plate splines began with Duchon (1976) and Meinguet (1979). Instead of a thin length of flexible wood, consider a thin steel plate or sheet above the (x,y) plane, extending to infinity in all directions, which is to be deformed to satisfy the conditions above. The terms in c, a_x and a_y represent the overall tilting of the plate (and the behaviour of the function at infinity). $J_2^2(\Phi)$ measures rapid variation in the surface, departures from local linearity or flatness and so will be larger if the function exhibits high 'local' curvature. As noted in 5.2.1, only the w_i terms contribute to $J_2^2(\Phi)$. If the plate is only tilted, it need not bend (all w_i=0 and hence $J_2^2(\Phi)$=0). Rotating or translating the $t_1,\ldots,t_k$, does not affect the value of $J_2^2(\Phi)$ either. Otherwise, to bend

the plate requires energy (against elasticity). The sharper the bending, the greater the second derivatives of the surface and energy required.

Note that some authors also write $\sigma_{2,2}(t-t_j)$ as $\sigma_{2,2}(r)=r^2\log r^2$, where $r=\sqrt{(t_x-t_{jx})^2+(t_y-t_{jy})^2}=\|t-t_j\|$, the Euclidean distance in the 2D plane between t and t_j. The function $\sigma_{2,2}(r)=r^2\log r$ is plotted in fig. 5.2 below.

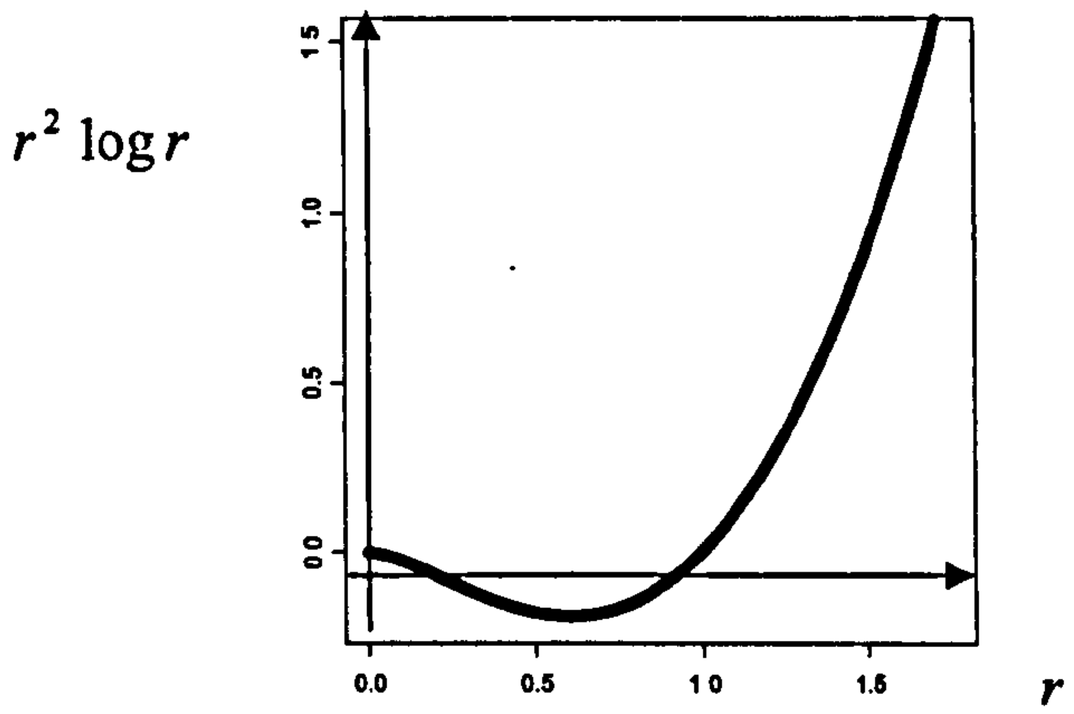


Fig. 5.2:
$r^2\log r$ vs. r

5.2.2 Multiple splines on R^m

So far we have only considered interpolation of data $t_1,...,t_k \in R^m$, with $t_j=(t_{j[1]},...,t_{j[m]})^T$, $j=1,...,k$, to target values $y_1,...,y_k \in R^1$. For target data $y_1,...,y_k \in R^d$ ($d\geq 1$), with $y_j=(y_{j[1]},...,y_{j[d]})^T$, $j=1,...,k$, associated with reference 'sites' $t_1,...,t_k \in R^m$, we can use d separate spline functions of the form of (5.2) (one for each of d univariate sets of target data $y_{j[1]},...,y_{j[d]}$, $j=1,...,k$), with common smoothness index (order of derivatives) $r>m/2$, to compute a function of the form:

$$\Phi(t)=(\Phi_1(t),...,\Phi_d(t))^T=(\Phi_1((t_{[1]},...,t_{[m]})^T),...,\Phi_d((t_{[1]},...,t_{[m]})^T)=(y_{[1]},...,y_{[d]})^T=y$$

with $\Phi(t_j)=y_j$, $j=1,...,k$. By construction, $\Phi(t)$ minimises the (sum of) roughness penalties of the form of (5.1), from each of the $\Phi_1(t),...,\Phi_d(t)$, i.e.

$$J_r^m(\Phi)=J_r^m(\Phi_1,...,\Phi_d)=J_r^m(\Phi_1)+...+J_r^m(\Phi_d). \tag{5.15}$$

Again $J_r^m(\Phi)$ is invariant under rotations and translations of the space of t or y (Kent & Mardia, 1994). This has obvious implications for its use with shape data, which we consider in section 5.3.3.

For each of the p=1,…,d splines we have, from (5.2) and (5.3),

$$\Phi_p(t)=\phi(t)^T a_p+s(t)^T w_p,$$

$$\sum_{j=1}^{k}\phi_v(t)w_{jp}=0 \text{ for } v=1,...,M, \tag{5.16}$$

where $s(t)$ and $\phi(t)$ are as in section 5.2.1.1, $a_p=(a_{1p},...,a_{Mp})^T$ and $w_p=(w_{1p},...,w_{kp})^T$. Therefore for $\Phi(t)$, we can write:

$$\Phi(t)^T=(\Phi_1(t),...,\Phi_d(t))=\phi(t)^T B+s(t)\,W \tag{5.17}$$

where

$$B=(a_1,...,a_d)=\begin{pmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & \ddots & \vdots \\ a_{M1} & \cdots & a_{Md} \end{pmatrix} \text{ and } W=(w_1,...,w_d)=\begin{pmatrix} w_{11} & \cdots & w_{1d} \\ \vdots & \ddots & \vdots \\ w_{k1} & \cdots & w_{kd} \end{pmatrix}.$$

Again the function defines a mapping of the entire space of t onto the space of y (with $\Phi(t_j)=y_j$, j=1,…,k) and consists of the sum of an overall or 'global' component of linear terms, and a 'local' component, here given by $\phi(t)^T B$ and $s(t)^T W$ respectively. The local component only exists relative to its corresponding global part as the displacement $s(t)^T W$ is defined to be added to the corresponding result of $\phi(t)^T B$.

For each of the p=1,…,d splines, $J_r^m(\Phi_p)=\sum_{j=1}^{k}\sum_{j'=1}^{k}w_{jp}w_{j'p}\sigma_{r,m}(t_j-t_{j'})$, from (5.5) and so by (5.16),

$$J_r^m(\Phi)=\sum_{p=1}^{d}J_r^m(\Phi_p)=\sum_{p=1}^{d}\sum_{j=1}^{k}\sum_{j'=1}^{k}w_{jp}w_{j'p}\sigma_{r,m}(t_j-t_{j'}). \tag{5.18}$$

Again, note that given data $t_1,...,t_k \in R^m$, (and hence $\sigma_{r,m}(t_j - t_{j'})$, j, j'=1,..,k), $J_r^m(\Phi)$ is dependent only on the $w_p = (w_{1p},...,w_{kp})^T$, p=1,..,d, terms.

Writing

$$Y = \begin{pmatrix} y_1^T \\ \vdots \\ y_k^T \end{pmatrix} = \begin{pmatrix} y_{1[1]} & \cdots & y_{1[m]} \\ \vdots & \ddots & \vdots \\ y_{k[1]} & \cdots & y_{k[m]} \end{pmatrix},$$

the solutions for B and W, p=1,…,d for any choice of r and m are found by solving the equations:

$$Y = QB + SW \tag{5.19}$$

$$Q^T W = 0 \tag{5.20}$$

with:

$$B = \Gamma^{21} Y \,,\; W = \Gamma^{11} Y \tag{5.21}$$

where

$$Q = \begin{pmatrix} \phi_1(t_1) & \cdots & \phi_M(t_1) \\ \vdots & \ddots & \vdots \\ \phi_1(t_k) & \cdots & \phi_M(t_k) \end{pmatrix} \text{ and } S = \begin{pmatrix} \sigma_{r,m}(t_1 - t_1) & \cdots & \sigma_{r,m}(t_1 - t_k) \\ \vdots & \ddots & \vdots \\ \sigma_{r,m}(t_k - t_1) & \cdots & \sigma_{r,m}(t_k - t_k) \end{pmatrix},$$

as in (5.6) and (5.7). Q and S each depend only on the $t_1,...,t_k$ and as in (5.8),

$$\Gamma^{21} = (Q^T S^{-1} Q)^{-1} Q^T S^{-1},\; \Gamma^{11} = S^{-1} - S^{-1} Q (Q^T S^{-1} Q)^{-1} Q^T S^{-1}.$$

Therefore from (5.17) the multiple spline from $t \in R^m$ to $y = \Phi(t) \in R^d$ is given by:

$$\Phi(t) = \phi(t)^T B + s(t)^T W = \phi(t)^T \Gamma^{21} Y + s(t)^T \Gamma^{11} Y \tag{5.22}$$

with, from (5.18),

$$J_r^m(\Phi) = \sum_{p=1}^{d} tr(w_p^T S w_p) = tr(W^T S W) = \sum_{p=1}^{d} tr(y_p^T \Gamma^{11} y_p) = tr(Y^T \Gamma^{11} Y). \tag{5.23}$$

The local component of $\Phi(t)$, depends on the global part, since, from (5.21),

$$W = \Gamma^{11} Y = S^{-1}Y - S^{-1}Q\,(Q^T S^{-1} Q)^{-1} Q^T S^{-1} Y = S^{-1}Y - S^{-1}Q\,\Gamma^{21}\,Y = S^{-1}(Y - QB). \quad (5.24)$$

i.e. W depends on B. Using (5.24) it can also be shown that by minimising (5.23), the transformation also minimises:

$$tr(W^T S W) = tr((Y - QB)^T (S^{-1})^T S S^{-1} (Y - QB)) = tr((Y - QB)^T S^{-1} (Y - QB)). \quad (5.25)$$

i.e. a generalised sum of squares between the result of the global component of the transformation for each t_j and the y_j. We explore the link of spline transformations with generalised least squares (GLS) in section 5.3.5. In particular we show that when r=2 and d=m, the solution for B in (5.21), optimising (5.25) corresponds to the GLS estimator in a weighted affine superimposition of T to Y, minimising a generalised sum of squares of the same form as (5.25).

Note that for d=1, equations (5.15) to (5.23) are then the same as those in 5.2.1.1 for obtaining a single interpolating spline $\Phi(t)$.

In the following section we describe how a pair of thin plate splines (d=2, m=2, r=2) may be used to display and describe differences in shape between two landmark configurations. It is a modification of this formulation that is used to determine the positions of the semi-landmarks, which we describe in section 5.4.

5.3 Pairs of thin plate splines and shape analysis

For m=2, d=2, we have reference and target data $t_1,...,t_k$ and $y_1,...,y_k \in R^2$, with $t_j = (t_{j[1]}, t_{j[2]})^T$ and $y_j = (y_{j[1]}, y_{j[2]})^T$, j=1,...,k. Writing $t_{[1]} = t_x, t_{[2]} = t_y, y_{[1]} = y_x$ and $y_{[2]} = y_y$, let t_j and corresponding y_j, denote the k corresponding landmarks of two configurations T and Y, where (in the notation of previous chapters),

$$T=\begin{pmatrix} t_{1x} & t_{1y} \\ t_{2x} & t_{2y} \\ \vdots & \vdots \\ t_{kx} & t_{ky} \end{pmatrix} \quad Y=\begin{pmatrix} y_{1x} & y_{1y} \\ y_{2x} & y_{2y} \\ \vdots & \vdots \\ y_{kx} & y_{ky} \end{pmatrix}. \tag{5.26}$$

Use of a pair of thin plate splines in plane-to-plane transformations and its application to sets of landmarks, was first introduced by Bookstein (1989). Using smoothness index (order of derivatives) r=2, in (5.17) with d=2, m=2, a function $\Phi(t)$, comprising a pair of thin plate interpolating splines (PTPS), may be used to describe a mapping of the 2D space in which T lies onto that of Y. By optimising (5.18), the mapping minimises the integral quadratic variation or 'bending energy' over the entire plane, with $\Phi(t_j)=y_j$, j=1,...,k, so that landmarks in T are transformed exactly to their corresponding positions in Y.

5.3.1 Mathematical details

From (5.17), a pair of thin plate splines transformation from $t=(t_x,t_y)^T$ to $\Phi(t)=(y_x,y_y)^T=y$ with $\Phi(t_j)=y_j, j=1,\ldots,k$, is given by:

$$\Phi(t)^T=(\Phi_1(t),\Phi_2(t))=\phi(t)^T B+s(t)^T W \tag{5.27}$$

where here

$$\phi(t)=(1,t_x,t_y)^T, s(t)=(\sigma_{2,2}(t-t_1),\ldots,\sigma_{2,2}(t-t_k))^T,$$

$$\sigma_{2,2}(t-t_j)=((t_x-t_{jx})^2+(t_y-t_{jy})^2)\log\sqrt{(t_x-t_{jx})^2+(t_y-t_{jy})^2},$$

$$B=(a_1,a_2)=\begin{pmatrix} c_1 & c_2 \\ a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, W=(w_1,w_2)=\begin{pmatrix} w_{11} & w_{12} \\ \vdots & \vdots \\ w_{k1} & w_{k2} \end{pmatrix},$$

so that

$$\Phi(t)=\begin{pmatrix} \Phi_1(t) \\ \Phi_2(t) \end{pmatrix}=\begin{pmatrix} c_1+a_{11}t_x+a_{21}t_y \\ c_2+a_{12}t_x+a_{22}t_y \end{pmatrix}+\begin{pmatrix} \sum_{j=1}^{k} w_{1j}\sigma_{2,2}(t-t_j) \\ \sum_{j=1}^{k} w_{2j}\sigma_{2,2}(t-t_j) \end{pmatrix}, \tag{5.28}$$

with constraints, from equation (5.16):

$$\sum_{j=1}^{k} w_{jp} = 0,\ \sum_{j=1}^{k} t_x w_{jp}\ 0,\ \sum_{j=1}^{k} t_y w_{jp} = 0,\ p = x, y. \tag{5.29}$$

Equation (5.28) shows how displacements in the (x,y) coordinate plane from $t=(t_x, t_y)^T$ to $y=(y_x, y_y)^T$ are described by a separate TPS for each coordinate direction; one TPS ($\Phi_1(t)=y_x$) is used for obtaining the new x-coordinates, the other ($\Phi_2(t)=y_y$), for new y-coordinates.

Recall that in section 5.2.1.2.2, the resulting value produced by a single TPS was considered to be the height of a surface normal to the (x,y) plane in which each point t lies. Instead, here there are two sets of values, $\Phi_1(t)$ and $\Phi_2(t)$ describing displacements to be applied to either t_x or t_y, when mapping to new positions y_x and y_y. Each $\Phi_p(t)$, p=1,2, however, still describes the same TPS mapping, with same minimised value for $J_2^2(\Phi_p)$, that would be required if each set of target values, y_{jx} and y_{jy}, j=1,…,k, was normal to the (x,y) plane, rather than displacements within it.

By equation (5.18), $\Phi(t)=(\Phi_1, \Phi_2)^T$ minimises the quantity:

$$J_2^2(\Phi) = J_2^2(\Phi_1) + J_2^2(\Phi_1) = \sum_{p=1}^{2} \left\{ \int_{R^2} \left(\frac{\delta^2 \Phi_p}{\delta t_x^2} \right)^2 + 2 \left(\frac{\delta^2 \Phi_p}{\delta t_x \delta t_y} \right)^2 + \left(\frac{\delta^2 \Phi_p}{\delta t_y^2} \right)^2 dt_x dt_y \right\}$$

$$= \sum_{j=1}^{k} \sum_{j'=1}^{k} w_{j1} w_{j'1} \sigma_{2,2}(t_j - t_{j'}) + \sum_{j=1}^{k} \sum_{j'=1}^{k} w_{j2} w_{j'2} \sigma_{2,2}(t_j - t_{j'}). \tag{5.30}$$

i.e. the sum of the integral quadratic variation from each thin plate spline or 'bending energy' of the PTPS transformation (see section 5.3.2 below), which given T (and hence the $\sigma_{2,2}(t_j - t_{j'})$), j,j'=1,…,k) depends only on the w_{j1} and w_{j2} terms.

By (5.19) and (5.20), the solutions for B and W, d=1,2 for r=2 and m=2 are given by solving:

$$Y = QB + SW,\ Q^T W = 0$$

where B and W are as above and here (using r=2,m=2, M=3 in (5.21)),

$$Q=\begin{pmatrix}1 & t_{1x} & t_{1y}\\ \vdots & \vdots & \vdots\\ 1 & t_{kx} & t_{ky}\end{pmatrix},\ S=\begin{pmatrix}\sigma_{2,2}(t_1-t_1) & \cdots & \sigma_{2,2}(t_1-t_k)\\ \vdots & \ddots & \vdots\\ \sigma_{2,2}(t_k-t_1) & \cdots & \sigma_{2,2}(t_k-t_k)\end{pmatrix}. \qquad (5.31)$$

From (5.21):

$$B=\Gamma^{21}Y\ ,\ W=\Gamma^{11}Y$$

with $$\Gamma^{21}=(Q^TS^{-1}Q)^{-1}Q^TS^{-1},\ \Gamma^{11}=S^{-1}-S^{-1}Q\,(Q^TS^{-1}Q)^{-1}Q^TS^{-1} \qquad (5.32)$$

and so from (5.22),

$$\Phi(t)^T=(\Phi_1(t),\Phi_2(t))=\phi(t)^T\,\Gamma^{21}Y\ +s(t)\,\Gamma^{11}Y\ , \qquad (5.33)$$

with, by (5.23),

$$J_2^2(\Phi)=\sum_{p=1}^{2}tr(w_p^TSw_p)=tr(W^TSW)=\sum_{p=1}^{2}tr(y_p^T\Gamma^{11}y_p)=tr(Y^T\Gamma^{11}Y). \qquad (5.34)$$

5.3.2 Affine and non-affine components and bending energy

As with each of the single and multiple spline transformations in 5.2, $\Phi(t)$ consists of an overall 'global' component of linear terms, here in $\phi_1(t)=1$, $\phi_2(t)=t_x$ $\phi_3(t)=t_x$, (the polynomials in t of degree <2) and a 'local' component of weighted $\sigma_{2,2}(t-t_j)$ terms (representing functions of distances between t and each of the landmarks of T, as described in section 5.2.1.2.2).

Here the global part of the transformation, given by:

$$\phi(t)^TB=\begin{pmatrix}1 & t_x & t_y\end{pmatrix}\begin{pmatrix}c_1 & c_2\\ a_{11} & a_{12}\\ a_{21} & a_{22}\end{pmatrix}=\begin{pmatrix}c_1+a_{11}t_x+a_{21}t_y\\ c_2+a_{12}t_x+a_{22}t_y\end{pmatrix}$$

is known as the 'affine' component (as it is for any multiple spline with d=m, r=2) and describes the part of the transformation corresponding to scalar multiplication or shifting of the t_x and t_y in calculating the new coordinate positions (y_x, y_y). Affine

transformations on the 2D plane transform squares into parallelograms (parallel lines in the space of T remain parallel in the space of Y) and circles into ellipses. Transformations of rotation, translation and scale are therefore also examples of affine transformations.

The remainder of the transformation is then made up of a 'local', or here, 'non-affine' series of vector displacements, given by:

$$s(t)^T W=(\sigma_{2,2}(t-t_1),\ldots,\sigma_{2,2}(t-t_k))\begin{pmatrix} w_{11} & w_{12} \\ \vdots & \vdots \\ w_{k1} & w_{k2} \end{pmatrix}=\begin{pmatrix} \sum_{j=1}^{k} w_{1j}\sigma_{2,2}(t-t_j) \\ \sum_{j=1}^{k} w_{2j}\sigma_{2,2}(t-t_j) \end{pmatrix}.$$

For each point $t=(t_x,t_y)^T$, this represents a displacement to be added to the resulting coordinates of the affine component, to give the new coordinates $y=(y_x,y_y)^T$. (For reference and target data t_j and $y_j, j=1,\ldots,k$, this then ensures that $\Phi(t_j)=y_j$).

Again, by using second derivatives in the penalty function for $\Phi(t)$, the linear terms in t in $\phi(t)^T$ vanish and so the component of the PTPS mapping accounted for by the affine transformation contributes zero to the roughness penalty $J_2^2(\Phi)$. The effect of tilting or shifting the metal thin-plate corresponding to either TPS (Φ_1 or Φ_2), described in 5.2.3, is now analogous to scalar multiplication or shifting of the values y_x or y_y, leaving $J_2^2(\Phi_1)$ or $J_2^2(\Phi_2)$ (and hence $J_2^2(\Phi)$) unchanged, as such transformations can be incorporated into the affine component.

As can be seen from equation (5.34), the 'bending energy' of the mapping is calculated from only the non-affine components of transforming points $t=t_1,\ldots,t_k$, the landmarks of configuration T, to $y_1,\ldots,y_k$, the landmarks of configuration Y. As with any spline transformation, only the observed data values contribute to the penalty function. Re-writing (5.34), $J_2^2(\Phi)$ is given by:

$$\sum_{p=1}^{2} tr(w_p^T S w_p)=\sum_{j=1}^{k}\sum_{j'=1}^{k} w_{1j}w_{1j'}\sigma_{2,2}(t_j-t_{j'})+\sum_{j=1}^{k}\sum_{j'=1}^{k} w_{2j}w_{2j'}\sigma_{2,2}(t_j-t_{j'})=tr(W^T SW), \quad (5.35)$$

a weighted sum of the squares and products of the non-affine components of $\Phi(t_j)=y_j$, j=1,…,k. Since, from (5.25), $tr(W^T SW)=tr((Y-QB)^T S^{-1}(Y-QB))$, here the PTPS transformation minimises

$$J_2^2(\Phi)=\left(\begin{pmatrix} y_{1x} \\ \vdots \\ y_{kx} \end{pmatrix}-Q\begin{pmatrix} c_1 \\ a_{11} \\ a_{21} \end{pmatrix}\right)^T S^{-1}\left(\begin{pmatrix} y_{1x} \\ \vdots \\ y_{kx} \end{pmatrix}-Q\begin{pmatrix} c_1 \\ a_{11} \\ a_{21} \end{pmatrix}\right)+\left(\begin{pmatrix} y_{1y} \\ \vdots \\ y_{ky} \end{pmatrix}-Q\begin{pmatrix} c_2 \\ a_{12} \\ a_{22} \end{pmatrix}\right)^T S^{-1}\left(\begin{pmatrix} y_{1y} \\ \vdots \\ y_{ky} \end{pmatrix}-Q\begin{pmatrix} c_2 \\ a_{12} \\ a_{22} \end{pmatrix}\right)$$

i.e. a generalised sum of squares between the result of the affine component of the transformation for each t_j and the y_j. Therefore in minimising the 'bending energy' of the mapping of T to Y, the PTPS try to represent the transformation as 'globally' as possible (at largest possible scale across the 2D plane), using the affine component. Correlated differences in the shape of Y, compared to T (i.e. differences from T applying equally to all landmarks in Y) are able to be incorporated into the affine part of the transformation, and so do not contribute to bending energy. This ensures that $s(t_j)^T W$ terms in (5.35) are as small as possible and thereby contribute less to $J_2^2(\Phi)$.

It turns out that the solution for B in (5.32), minimising $J_2^2(\Phi)$, corresponds to the GLS estimator in a weighted affine superimposition of T to Y minimising a generalised sum of squares identical to that above. We explore this link further in 5.3.5.

For shape analysis Bookstein (1991) notes that bending energy (BE) measures 'localisation', how nearby landmarks appear to have changed relative to those at a greater distance. Consider the mapping of the x-coordinates of T to those in Y by $\Phi_1(t)$. The following example illustrates how, if the difference in x-coordinates between two neighbouring landmarks j and j' is a units less in Y compared to T, then this requires more BE in the PTPS transformation from T to Y, the closer the two landmarks are in T. This is illustrated in fig. 5.3, which shows $\Phi_1(t)=\Phi_1(t_x)$ for two configurations of 4 landmarks. The configurations differ only in the position of landmark j'. (Of course since $\Phi_1(t)$ is a TPS, this should also depends on the t_y,

however for illustrative purposes, here $\Phi_1(t)=\Phi_1(t_x)$ and fig. 5.3 takes the appearance of a cubic spline transforming points t_x to y_x).

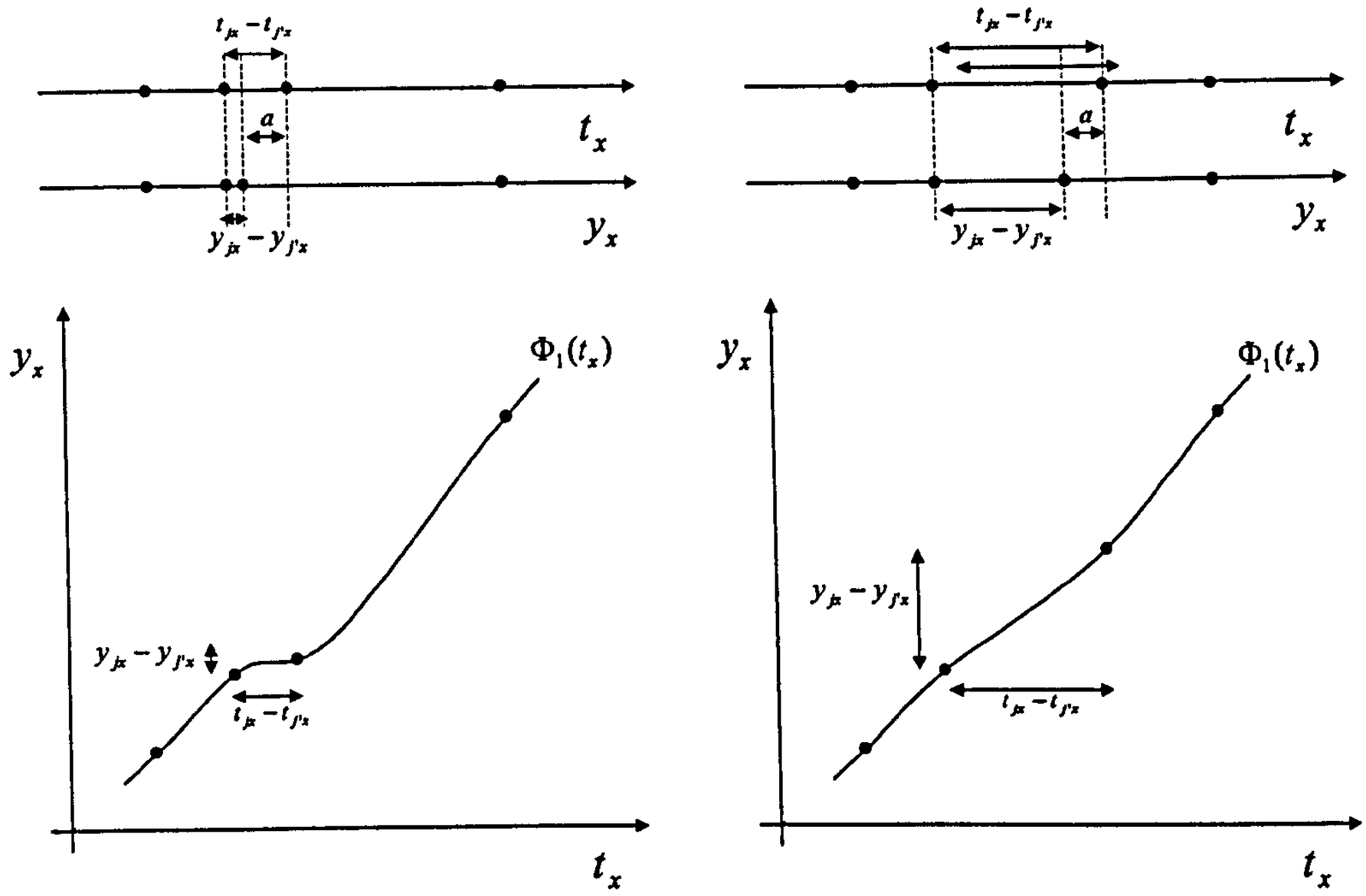


Fig 5.3: Illustration of how bending energy measures 'localisation' using a single spline mapping of the x-coordinates of two landmark configurations T and Y (differing only in the position of landmark j').

Typically landmarks that are closer together in reference shape T (more localised) will require more energy to move apart, when matching to corresponding positions in the target shape Y, unless those at a greater distance also move by the same transformation.

Bending energy may also be calculated as $tr(Y^T\Gamma^{11}Y)$, as in (5.23). The $(k\times k)$ matrix Γ^{11} (dependent only on T), is often called the 'bending energy matrix', as it is used to calculate the amount of bending required to transform T to any given configuration Y.

5.3.3 Deformation grids

The PTPS mapping can be visualised as the transformation of a square grid, placed on configuration T, to a deformed grid on configuration Y. At each junction where the lines of the square grid on T cross, the corresponding position in the deformed image is calculated using $\Phi(t)=(\Phi_1,\Phi_2)^T$ and the lines between the points re-drawn, so that corresponding landmarks are located in corresponding grid blocks. The resulting

deformed grid then tells us where and how the two configurations differ. Fig 5.4 shows the deformation from one hypothetical upper central incisor configuration to another. (The idea was briefly illustrated in section 2.2.3.3.1, using upper central incisors, one from a patient with hypodontia and the other from a control patient).

Each square is transformed to a quadrilateral. The minimum bending energy property minimises the variation of the 'affine derivative'; the shape of the individual grid cells with respect to their neighbours. When a change can be managed over a larger interval, its contribution to the integral sums of squares is lower and so the PTPS will try to represent deformations as 'globally' as possible, with the smallest variation in the shape of the individual grid cells.

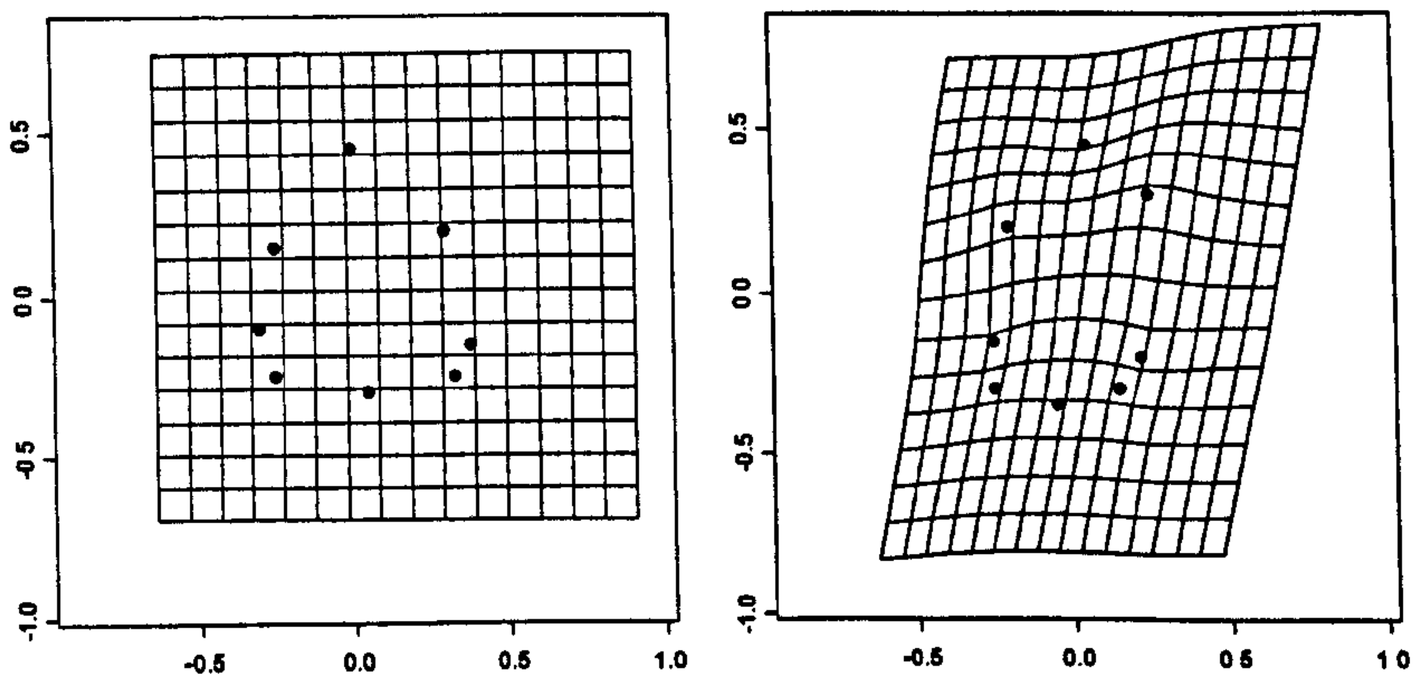


Fig. 5.4. PTPS transformation between two upper central incisor configurations. (Left) reference shape T and (right) target shape Y.

To illustrate the 'global' and 'local' components of the transformation, a separate affine and non-affine mapping can be obtained and the result for configuration T visualised by considering their effect on a square grid placed on T.

The affine transformation of T can be visualised by plotting the deformation grid corresponding to the mapping of points on a square grid (placed on T) to points given by:

$$\phi(t)^T B = \phi(t)^T \Gamma^{21} Y$$

and by displaying the 'affine configuration' (the result for T):

$$QB = Q\,\Gamma^{21}Y \tag{5.36}$$

on the grid. This is illustrated in fig. 5.5 (left) for the same 'target' and 'reference' configurations as in fig 5.4. Note how the affine mapping transforms the squares of the regular grid to parallelograms.

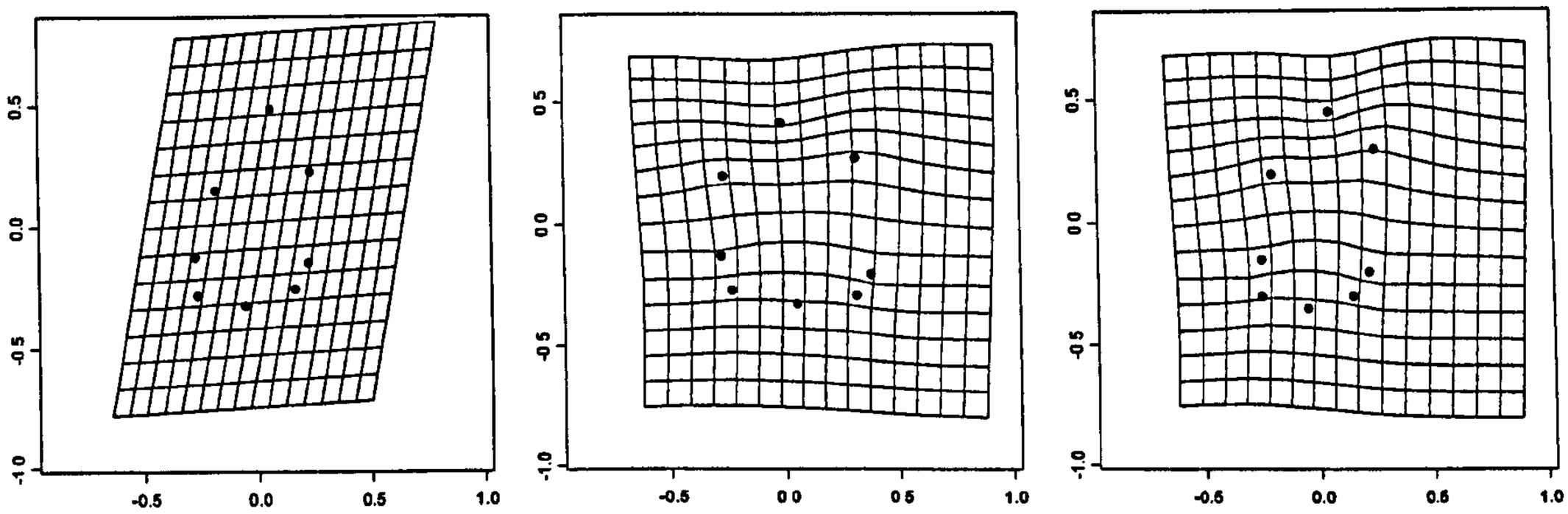


Fig. 5.5: Illustrating the affine (left) and non-affine (middle & right) components of the PTPS transformation from T to Y, shown in fig 5.4. (Left) Mapping of points t on a square grid placed on T by $\phi(t)^T\ \Gamma^{21}Y$, to locations in the space of configuration $Q\Gamma^{21}Y$ (shown). (Middle) Mapping of points t on a square grid placed on T by $t+s(t)^T\ \Gamma^{11}Y$, to locations in the space of configuration $T+Q\Gamma^{21}Y$. (Right) Mapping of points t' on a square grid placed on $Q\Gamma^{21}Y$ by $t'+s(A^{-1}(t'-c))^T\ \Gamma^{11}Y$ to locations in the space of configuration Y.

Following Dryden & Mardia (1998), a 'non-affine' mapping and corresponding configuration, illustrating the non-affine displacements of the PTPS transformation, can be produced by constructing the deformation grid corresponding to the mapping of points t on a square grid (placed on T) to locations given by:

$$t+s(t)^T\ \Gamma^{11}Y\ .$$

The configuration displayed on the grid, is the result for T, i.e.

$$T+S\,W = T+S\,\Gamma^{11}Y \tag{5.37}$$

(see fig 5.5 (middle)). However, the configuration $T+S\,\Gamma^{11}Y$ bears little relation to the non-affine components it aims to illustrate, since the displacements $s(t)^T W$ are defined to be added to points $\phi(t)^T\,\Gamma^{21}Y$ above, and not just their original locations t. (The non-affine displacements cannot be defined separately, as described in 5.3.2).

Our suggested alternative is to construct the deformation of a square grid placed on $QB=Q\,\Gamma^{21}Y$ to points in the space of the configuration $Q\Gamma^{21}Y+S\,\Gamma^{11}Y$, i.e. Y. Writing

$$\phi(t)^T B=c+At, \tag{5.38}$$

where $c=(c_1 \quad c_2)^T$, $A=\begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}$ and $t=(t_x \quad t_y)^T$, for points $t'=c+At$, in space of $Q\,\Gamma^{21}Y$ (i.e. in fig. 5.5(left)), we can calculate new positions for the t' as:

$$t'+s(A^{-1}(t'-c))^T\,\Gamma^{11}Y\,. \tag{5.39}$$

The transformation of points by $A^{-1}(t'-c)$ is necessary since $s(t')\,\Gamma^{11}Y$ does not give the correct non-affine displacements in $\Phi(t)$ (the PTPS mapping from T to Y), which are to be added to points $t'=\phi(t)^T B=c+At$ (including, for T, the affine configuration, $QB=Q\,\Gamma^{21}Y$). These are given by $s(t)^T\,\Gamma^{11}Y$, where the t denotes the point corresponding to t' in the original space of T, given by $A^{-1}(t'-c)$.

For the landmarks of configuration T, $t'_j=\phi(t_j)^T B=c+A\,t_j$, $j=1,\ldots,k$, gives the affine configuration, $QB=Q\,\Gamma^{21}Y$, as described above. The corresponding non-affine components to be added to the t'_j (or any other points t' in space of $QB=Q\,\Gamma^{21}Y$, given by the affine component), are given by $s(A^{-1}(t'_j-c))^T\,\Gamma^{11}Y=s(t_j)^T\,\Gamma^{11}Y$.

Fig 5.5(right) shows the deformation grid for transforming points t' on a square grid placed over $Q\Gamma^{21}Y$, to points given by (5.38). The configuration $Q\Gamma^{21}Y+S\,\Gamma^{11}Y$ (=Y) can then be displayed on the plot, illustrating how the shape of the affine configuration $Q\Gamma^{21}Y$ is changed when the non-affine displacements are added.

5.3.4 Effects of similarity transformations on the PTPS components and bending energy

Table 5.1 details the effects of a rotation, translation, or resizing of the reference or target configuration on the different components of the PTPS transformation and bending energy. For later reference in 6.3.2.2, details of how the affine and non-affine configurations, defined in (5.36) and (5.37) change are also given.

For two configurations, T and Y, each of k landmarks, the PTPS for transforming the landmarks T to those of Y is given by:

$$QB + SW = 1_k c^T + TA^T + SW \tag{5.40}$$

where 1_k is a vector of 1's of length k, Q and S are as in (5.31), W is as in (5.27) and c and A are as in (5.38). The original components and values in the first column of the table relating to the original PTPS mapping from $T = T_0$ to $Y = Y_0$ are subscripted with '0'. Subsequent columns then show how the these are changed by either a vector shift α, rescaling β, or rotation by $\Gamma = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$, $0 < \theta \leq 2\pi$, of either T_0 or Y_0.

If T_0 or Y_0 is shifted by α (to become $T = T_0 + \alpha$ or $Y = Y_0 + \alpha$), the affine component changes to accommodate this and provide the same 'fit' to Y_0 as before (measured by the bending energy of the PTPS mapping), since both S and W are unchanged.

If T_0 is rotated (to become $T = T_0\Gamma$), the new affine component rotates it back to the same position as for T_0 and both S, W and the rest of the mapping, including the BE, are unchanged. If Y_0 is rotated (to $Y = Y_0\Gamma$) the affine and non-affine components are simply rotated versions of the original components in the mapping from T_0 to Y_0 (by Γ) and the bending energy is again unchanged.

$T =$	T_0	$T_0 + \alpha$	T_0	$T_0\Gamma$	T_0	βT_0	T_0
$Y =$	Y_0	Y_0	$Y_0 + \alpha$	Y_0	$Y_0\Gamma$	Y_0	βY_0
c	c_0	$c_0 - B_0\alpha$	$\alpha + c_0$	-	$c_0\Gamma$	See text	$\beta\, c_0$
A	A_0	-	-	$A_0\Gamma$	$A_0^T\Gamma$	$(1/\beta)\,A_0$	$\beta\, A_0$
Affine config. $1_k c^T + TA^T$	$1_k c_0^T + T_0 A_0^T$	-	$\alpha + 1_k c_0^T + T_0 A_0^T$	-	$(1_k c_0^T + T_0 A_0^T)\Gamma$	See text	$\beta\,(1_k c_0^T + T_0 A_0^T)$
S (fn. of inter point dist's in T)	S_0	-	-	-	-	See text	-
W	W_0	-	-	-	$W_0\Gamma$	$(1/\beta^2)W_0$	$\beta\, W_0$
SW	S_0W_0	-	-	-	$S_0W_0\Gamma$	See text	$\beta\, S_0W_0$
Non-affine config. $T + SW$	$T_0 + S_0W_0$	$\alpha + T_0 + S_0W_0$	-	$T_0\Gamma + (W_0^T S_0)^T$	$T_0 + S_0W_0\Gamma$	See text	$T_0 + \beta S_0W_0$
BE matrix Γ^{11} (dep. only on T)	Γ_0^{11}	-	-	-	-	$(1/\beta^2)\Gamma_0^{11}$	-
BE $tr(W^T SW)$	BE_0	-	-	-	-	$(1/\beta^2)BE_0$	$\beta^2\, BE_0$

Table 5.1: Effect of similarity transformations on PTPS components, affine and non-affine configurations and bending energy. '-' denotes unchanged.

However, unlike the Procrustes distance, BE is not invariant to changes in the scale of either T_0 or Y_0. If Y_0 is rescaled by a factor β to $Y=\beta Y_0$, S is unchanged (since this depends only on T), but the affine component and non-affine displacements are both also rescaled by a factor of β to accommodate this. Since $W=\beta W_0$, bending energy $tr(W^T SW)$ is then also increased/decreased by a factor of β^2. For T=βT_0, the situation is slightly more complex. Unlike with changes in location or rotation, changes to the scale of T will affect the matrix S of $\sigma_{2,2}(t_j - t_{j'})$, terms depending on inter-landmark distances $r=\sqrt{(t_{jx}-t_{j'x})^2+(t_{jy}-t_{j'y})^2}$ in T. Values of $\sigma_{2,2}(\beta r)$ are produced from a different position along the curve in fig 5.2 and will weight distances between pairs of landmarks differently from before. The change in S cannot therefore, be described in terms of a simple linear transformation of S_0. However, since $\sigma(r)$ is self-similar, with $\sigma(\beta r)\equiv\beta^2\sigma(r)$, the mapping still has the same appearance as the transformation from T_0 to Y_0 (as with any of the changes to T_0 or Y_0 investigated here), with A and W being re-scaled by $1/\beta$. Since $W=(1/\beta)W_0$, BE is then also increased/decreased by $1/\beta^2$.

5.3.5 Generalised least squares, affine matching & link to PTPS

In section 5.3.2, it was noted that any multiple d-spline transformation, with $d=m$, from

$$T=(t_1,\ldots,t_k)^T=\begin{pmatrix} t_{1[1]} & \cdots & t_{1[m]} \\ \vdots & \ddots & \vdots \\ t_{k[1]} & \cdots & t_{k[m]} \end{pmatrix} \text{ to } Y=(y_1,\ldots,y_k)^T=\begin{pmatrix} y_{1[1]} & \cdots & y_{1[m]} \\ \vdots & \ddots & \vdots \\ y_{k[1]} & \cdots & y_{k[m]} \end{pmatrix}, \quad (5.41)$$

with $k\geq m+1$, minimises a generalised sum of squares, given by equation (5.25) as:

$$J_r^m(\Phi)=\sum_{p=1}^{d} J_r^m(\Phi_p)=tr(W^T SW)=tr((Y-QB)^T S^{-1}(Y-QB))$$

$$=(vec(Y)-(I_d\otimes Q)(vec(B))^T(I_d\otimes S)^{-1}(vec(Y)-(I_d\otimes Q)(vec(B)) \quad (5.42)$$

where Q, S, $B=(a_1,\ldots,a_d)$ and W are as defined by equations (5.17) and (5.21).

In this section we show how with r=2, (5.42) corresponds to the objective function, optimised in a generalised 'affine' match of T to Y and that the estimator of the parameters of this superimposition, corresponds precisely to the solution for matrix B, obtained by simultaneously solving equations (5.19) and (5.20) for a PTPS (or any multiple spline mapping, with d=m, r=2).

Dryden & Mardia (1998) similarly noted that the solutions for the affine parameters of each TPS in a PTPS mapping are generalised least squares estimators, but give no further details. Other than this brief mention, there has been very little exploration of this connection in the literature.

Here we describe and investigate this link further, in order to improve our understanding of the mechanism by which the components of a PTPS mapping are determined. In section 5.4, where we introduce the idea of semi-landmarks, the minimum bending energy property of a PTPS mapping is used to determine new positions of certain landmarks in the target configuration Y. An understanding of how bending energy is minimised, will therefore also help us to explain why particular results occur with this method, when we investigate its use for the analysis of tooth shape, in section 5.6.

5.3.5.1 Preliminary definitions

5.3.5.1.1 Generalised least squares

Before considering ordinary and then generalised least squares (OLS & GLS) affine matching, we first describe the standard GLS formulation, commonly used in univariate linear regression problems (see for example, Seber, 1977 or Weatherill, 1986).

Consider linear regression of a vector of observations y=$(y_1,\ldots,y_N)^T$ on N×P data or 'design' matrix X_D (comprising values for each observation on P-1 variables following a first column of 1's) and an N-vector of errors $\varepsilon=(\varepsilon_1,\ldots,\varepsilon_N)^T$, that is,

$$y=X_D\,\beta+\varepsilon,$$

where β is the parameter vector of P components. In GLS, the elements of ε are assumed to have covariance structure $\sigma^2 V$ where V is a known $N\times N$ matrix of weights (for ordinary least squares $V=I_N$). The GLS estimator for β, which minimises the generalized sum of squares:

$$\varepsilon^T V^{-1}\varepsilon=(y-X_D\beta)^T V^{-1}(y-X_D\beta) \tag{5.43}$$

is given by

$$\hat{\beta}=(X_D^T V^{-1}X_D)^{-1}X_D^T V^{-1}y\,. \tag{5.44}$$

5.3.5.1.2 Ordinary least squares affine matching of two configurations

For an ordinary least squares affine match of configuration T to configuration Y, with T and Y as defined in (5.41), we use:

$$Y=Q_A B_A+E \tag{5.45}$$

or equivalently,

$$vec(Y)=(I_m\otimes Q_A)\,vec(B_A)+vec(E)\,, \tag{5.46}$$

where

$$Q_A=\begin{pmatrix}1 & t_{11} & \cdots & t_{1m}\\ \vdots & \vdots & \ddots & \vdots\\ 1 & t_{k1} & \cdots & t_{km}\end{pmatrix},\ B_A=\begin{pmatrix}c_1 & \cdots & c_m\\ a_{11} & \cdots & a_{1m}\\ \vdots & \ddots & \vdots\\ a_{m1} & \cdots & a_{mm}\end{pmatrix}=(a_{A1},\ldots,a_{Ad}),\ E=\begin{pmatrix}e_{11} & \cdots & e_{1m}\\ \vdots & \ddots & \vdots\\ e_{k1} & \cdots & e_{km}\end{pmatrix}$$

and E is a zero-mean $k\times m$ random error matrix, with $vec(E)\sim$Multivariate Normal $(0, I_{km}\sigma^2)$. i.e. the elements of E are independent, with common variance σ^2.

The parameters B_A are estimated by minimising the sum of squares:

$$\begin{aligned}tr(E^T E)&=tr((Y-Q_A B_A)^T(Y-Q_A B_A))=vec(E)^T vec(E)\\ &=(vec(Y)-(I_m\otimes Q_A)vec(B_A))^T(vec(Y)-(I_m\otimes Q_A)vec(B_A))\\ &=\sum_{p=1}^{d}tr((y_{1[p]},\ldots,y_{k[p]})^T-Qa_{Ap})^T((y_{1[p]},\ldots,y_{k[p]})^T-Qa_{Ap})\,.\end{aligned} \tag{5.47}$$

i.e. an equation of the form of (5.43) where $N=km$, $y=vec(Y)$, $X_D=I_m \otimes Q_A$, $\beta=vec(B_A)$, $\varepsilon=vec(E)$ and $V=I_{km}$. The solution, by analogy with the GLS regression case in 5.3.5.1.1 (or by differentiating with respect to $vec(B_A)$, setting result=0 and solving), is given by:

$$vec(\hat{B}_A)=((I_m \otimes Q_A)^T(I_m \otimes Q_A))^{-1}(I_m \otimes Q_A)^T vec(Y)$$

and in matrix form

$$\hat{B}_A=vec_m^{-1}(\hat{B}_A)=(Q_A^TQ_A)^{-1}Q_A^TY\,. \qquad (5.48)$$

Note that throughout the above it is assumed that Q_A and $Q_A^TQ_A$ are full rank (m+1). (For further details, see Dryden & Mardia, 1998 or Rohlf & Slice, 1990). For applications involving OLS affine matching, see for example, Goodall & Green (1986), Rohlf & Slice (1990) or Horgan *et al.* (1992).

5.3.5.2 Generalised least squares affine matching of two configurations with block diagonal error covariance

In 5.3.5.2.1, it was assumed that $Cov(vec(E))=I_{km}\sigma^2$. Instead, let $Cov(vec(E))=V\sigma^2$ as in 5.3.5.1.1, where V is a $km\times km$ positive-definite matrix of weights. In matching T to Y by an affine transformation, using (5.45) or (5.46), the parameters of B_A are now estimated by minimising the generalised sum of squares:

$$vec(E)^TV^{-1}vec(E)=(vec(Y)-(I_m \otimes Q_A)vec(B_A))^TV^{-1}(vec(Y)-(I_m \otimes Q_A)vec(B_A)) \qquad (5.49)$$

with solution, by analogy with (5.43) or by differentiation with respect to $vec(B_A)$, given by:

$$vec(\hat{B}_A)=((I_m \otimes Q_A)^TV^{-1}(I_m \otimes Q_A))^{-1}(I_m \otimes Q_A)^TV^{-1}vec(Y)\,. \qquad (5.50)$$

If V is chosen as a block diagonal matrix of m blocks S, given by $(I_m \otimes S)$, with $\sigma^2=1$, then from (5.49), B_A is estimated as the set of parameters which minimise:

$$vec(E)^T (I_m \otimes S)^{-1} vec(E) = (vec(Y) - (I_m \otimes Q_A))^T (I_m \otimes S)^{-1} (vec(Y) - (I_m \otimes Q_A))$$

$$= tr(E^T S^{-1} E) = tr((Y - Q_A B_A)^T S^{-1} (Y - Q_A B_A)) \quad (5.51)$$

$$= \sum_{p=1}^{d} tr((y_{1[p]},...,y_{k[p]})^T - Q_A a_{Ap})^T S^{-1} ((y_{1[p]},...,y_{k[p]})^T - Q_A a_{Ap}).$$

Note than unless V is of the block diagonal form above, the generalised sum of squares in (5.51) cannot be separated into individual summations involving corresponding columns of Y and B_A, as shown above.

The solution, from (5.50) with $V=(I_m \otimes S)$, is given by:

$$vec(\hat{B}_A) = ((I_m \otimes Q_A)^T (I_m \otimes S)^{-1} (I_m \otimes Q_A))^{-1} (I_m \otimes Q_A)^T (I_m \otimes S)^{-1} vec(Y) \quad (5.52)$$

and in matrix form (since $V=(I_m \otimes S)$)

$$\hat{B}_A = vec_m^{-1}(\hat{B}_A) = (Q_A^T S^{-1} Q_A)^{-1} Q_A^T S^{-1} Y. \quad (5.53)$$

The generalised (or weighted) 'affine fit' of T to Y and matrix R, of residuals at each pair corresponding landmarks of the 'fitted T' and Y, are then given by:

$$Q_A \hat{B}_A = Q_A (Q_A^T S^{-1} Q_A)^{-1} Q_A^T S^{-1} Y \quad (5.54)$$

and

$$R = Y - Q_A (Q_A^T S^{-1} Q_A)^{-1} Q_A^T S^{-1} Y. \quad (5.55)$$

5.3.5.3 Link with interpolating splines

5.3.5.3.1 Multiple splines with $d=m$ and $r=2$

Clearly, if S is chosen as

$$S = \begin{pmatrix} \sigma_{2,m}(t_1 - t_1) & \cdots & \sigma_{2,m}(t_1 - t_k) \\ \vdots & \ddots & \vdots \\ \sigma_{2,m}(t_k - t_1) & \cdots & \sigma_{2,m}(t_k - t_k) \end{pmatrix}$$

with Q_A and B_A as in (5.46), then by (5.42), equation (5.51) is precisely the penalty function $J_r^m(\Phi)$ for a multiple interpolating spline $\Phi(t)$ with d=m and r=2, for transforming the landmarks of configuration $T \in R^m$ to those of $Y \in R^m$ (T and Y as defined in (5.41)). To see this, recall that the parameters of a multiple (d) spline for transforming points in the space of T to that of Y, given by:

$$\Phi(t)^T = \phi(t)^T B + s(t)W,$$

minimise

$$J_r^m(\Phi) = tr(W^T SW) = tr((Y - QB)^T S^{-1}(Y - QB))$$

$$= (vec(Y) - (I_d \otimes Q)(vec(B))^T (I_d \otimes S)^{-1}(vec(Y) - (I_d \otimes Q)(vec(B))$$

where

$$Q = (\phi(t_1),\ldots,\phi(t_k))^T = \begin{pmatrix} \phi_1(t_1) & \cdots & \phi_M(t_1) \\ \vdots & \ddots & \vdots \\ \phi_1(t_k) & \cdots & \phi_M(t_k) \end{pmatrix},$$

$$S = (s(t_1),\ldots,s(t_k))^T = \begin{pmatrix} \sigma_{r,m}(t_1 - t_1) & \cdots & \sigma_{r,m}(t_1 - t_k) \\ \vdots & \ddots & \vdots \\ \sigma_{r,m}(t_k - t_1) & \cdots & \sigma_{r,m}(t_k - t_k) \end{pmatrix},$$

and

$$B = (a_1,\ldots,a_d) = \begin{pmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & \ddots & \vdots \\ a_{M1} & \cdots & a_{Md} \end{pmatrix}, \quad W = (w_1,\ldots,w_d) = \begin{pmatrix} w_{11} & \cdots & w_{1d} \\ \vdots & \ddots & \vdots \\ w_{k1} & \cdots & w_{kd} \end{pmatrix}$$

are unknowns determined by solving $Y = QB + SW$ and $Q^T W = 0$.

For d=m and r=2, (and hence M=m+1), we have $\phi_1(t)$=1, $\phi_2(t) = t_{[1]},\ldots,\phi_{m+1}(t) = t_{[m]}$ (the polynomials in m variables of total degree less than or equal to r-1=1), Q=Q_A, B=B_A and S with elements $\sigma_{2,m}(t_j - t_{j'})$, j, j'=1,...,k. Therefore the value of the penalty function optimised by a multiple spline transformation $\Phi(t)$, from T to Y, comprising d=m splines, each with order of derivatives r=2, has objective function identical to that of a generalised affine superimposition of T to Y, with covariance matrix $(I_m \otimes S)$. We discuss the case, m=d=2, r=2 (PTPS mapping) in 5.3.5.3.2 below.

Note that by (5.51), $J_2^m(\Phi)=\sum_{p=1}^{d} J_r^m(\Phi_p)$ (the sum of penalty functions for each spline with r=2) is a summation of d generalised sums of squares, corresponding to a GLS regression of $Q=Q_A$ onto each column of Y, with weighting matrix S, optimising (5.43). The block diagonal structure in (5.51) means that the same covariance is used for each sum of squares, and that each regression is independent, so that for p=1,…,d,

$$J_2^m(\Phi_p)=tr((y_{1[p]},\ldots,y_{k[p]})^T-Q_A a_{Ap})^T S^{-1}((y_{1[p]},\ldots,y_{k[p]})^T-Q_A a_{Ap}).$$

Recall that the multiple splines are each independent and depend on the same matrix S.

In addition, the solution for parameter matrix $B=B_A$ obtained by solving the multiple spline equations above, with $Q=Q_A$, is given by equation (5.21) as $(Q_A^T S^{-1} Q_A)^{-1} Q_A^T S^{-1} Y$, corresponding precisely to the estimate of B_A, given by (5.53). Note that the GLS affine parameters B_A are 'estimated' (to minimise a generalised sum of squares) whereas $B=B_A$ in the multiple spline transformation is determined by 'solving' a set of equations.

For m=d=2, the generalised 'affine fit' of T to Y given by equation (5.54) as $Q_A (Q_A^T S^{-1} Q_A)^{-1} Q_A^T S^{-1} Y$, is the same as the 'affine configuration', given by (5.36), in section 5.3.3, representing the result of the applying the 'affine component' of a PTPS mapping (r=2, m=d=2) of the space of T to that of Y, to T.

To complete the multiple spline mapping so that the landmarks of T map exactly to those of Y, local components, SW are added to the result of applying the 'global' ('affine' when r=2) part of a multiple spline mapping to T. So, for $Q=Q_A$ (m=d, r=2),

$$Y=Q_A (Q_A^T S^{-1} Q_A)^{-1} Q_A^T S^{-1} Y+SW,$$

from which

$$SW=Y-Q_A (Q_A^T S^{-1} Q_A)^{-1} Q_A^T S^{-1} Y=R, \qquad (5.56)$$

the matrix of residuals of the generalised affine fit of T to Y defined in (5.55). Consequently, if the affine parameters in a multiple spline mapping with $m=d$, $r=2$, $B=B_A$ can be found as the estimator of the affine parameters in a generalised affine fit of T to Y, then the other unknown matrix W, required to calculate the local components of the spline, can then be found by solving (5.56), with $W=S^{-1}R$.

The minimised value of (5.51) is then given by:

$$tr(R^T S^{-1} R)=tr((Y-Q_A\hat{B}_A)^T S^{-1}(Y-Q_A\hat{B}_A))=tr((SW)^T S^{-1} SW)=tr(W^T SW)=J_2^m(\Phi).$$

5.3.5.3.2 Pair of thin plate splines ($d=m=2$, $r=2$)

Directly from the results of 5.3.5.4, we see that for $d=m=2$, with,

$$T=\begin{pmatrix} t_{1x} & t_{1y} \\ \vdots & \vdots \\ t_{kx} & t_{ky} \end{pmatrix}, Y=\begin{pmatrix} y_{1x} & y_{1y} \\ \vdots & \vdots \\ y_{kx} & y_{ky} \end{pmatrix},$$

the affine mapping from T to Y, given by $Y=Q_A B_A+E$, here has

$$Q=\begin{pmatrix} 1 & t_{1x} & t_{1y} \\ \vdots & \vdots & \vdots \\ 1 & t_{kx} & t_{ky} \end{pmatrix}, B_A=\begin{pmatrix} c_1 & c_2 \\ a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, E=\begin{pmatrix} \varepsilon_{1x} & \varepsilon_{1y} \\ \vdots & \vdots \\ \varepsilon_{kx} & \varepsilon_{ky} \end{pmatrix},$$

i.e.

$$vec(Y)=(I_2\otimes Q_A)vec(B_A)+vec(E)=\begin{pmatrix} y_{1x} \\ \vdots \\ y_{kx} \\ y_{1y} \\ \vdots \\ y_{ky} \end{pmatrix}=\begin{pmatrix} 1 & t_{1x} & t_{1y} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_{kx} & t_{ky} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & t_{1x} & t_{1y} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & t_{1k} & t_{ky} \end{pmatrix}\begin{pmatrix} c_1 \\ a_{11} \\ a_{21} \\ c_2 \\ a_{12} \\ a_{22} \end{pmatrix}+\begin{pmatrix} e_{1x} \\ \vdots \\ e_{kx} \\ e_{1y} \\ \vdots \\ e_{ky} \end{pmatrix}.$$

If we assume that:

$$Cov(vec(E))=(I_2\otimes S)=\begin{pmatrix}\sigma_{2,2}(t_1-t_1) & \cdots & \sigma_{2,2}(t_1-t_k) & 0 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ \sigma_{2,2}(t_k-t_1) & \cdots & \sigma_{2,2}(t_k-t_k) & 0 & \cdots & 0\\ 0 & \cdots & 0 & \sigma_{2,2}(t_1-t_1) & \cdots & \sigma_{2,2}(t_1-t_k)\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & 0 & \sigma_{2,2}(t_k-t_1) & \cdots & \sigma_{2,2}(t_k-t_k)\end{pmatrix},$$

then $\hat{B}_A=(Q_A^T S^{-1}Q_A)^{-1}Q_A^T S^{-1}Y$ minimises (5.47), with value:

$$tr((Y-Q_A\hat{B}_A)^T S^{-1}(Y-Q_A\hat{B}_A))=J_2^2(\Phi_1)+J_2^2(\Phi_2), \qquad (5.57)$$

the BE of the PTPS mapping of landmarks of T to those of Y. Bending energy may therefore be thought of as measuring the degree of fit of a GLS affine superimposition of T to Y, or between Y and the result of applying the affine component of the PTPS mapping from T to Y, to T.

Setting residuals $R=Y-Q_AB_A=SW$ (the non-affine components of the PTPS mapping) and solving, then gives the complete PTPS mapping of T to Y, with:

$$\begin{pmatrix}y_{1x} & y_{1y}\\ \vdots & \vdots\\ y_{kx} & y_{ky}\end{pmatrix}=\begin{pmatrix}1 & t_{1x} & t_{1y}\\ \vdots & \vdots & \vdots\\ 1 & t_{kx} & t_{ky}\end{pmatrix}\begin{pmatrix}c_1 & c_2\\ a_{1x} & a_{2x}\\ a_{1y} & a_{2y}\end{pmatrix}+\begin{pmatrix}\sum_{j=1}^{k}w_{1j}\sigma(t_1,t_j) & \sum_{j=1}^{k}w_{2i}\sigma(t_1,t_j)\\ \vdots & \vdots\\ \sum_{j=1}^{k}w_{1j}\sigma(t_k,t_j) & \sum_{j=1}^{k}w_{2i}\sigma(t_k,t_j)\end{pmatrix}$$

and from (5.57), $tr((Y-Q_A\hat{B}_A)^T S^{-1}(Y-Q_A\hat{B}_A))=tr(W^T SW)$.

For the same configurations T and Y as in section 5.3.3, fig. 5.10 shows the GLS affine fit of T to Y as well as the mapping of points on a square grid placed over T by the same transformation (equivalent to the mapping produced by the affine component of the PTPS mapping from T to Y). The filled circles are the 'fitted values' defined above and also correspond to the affine configuration defined by equation (5.36) in section 5.3.3. The crosses are configuration Y, displayed on the same plot so as one can visualise the sizes and directions of the residuals (the components of matrix R), corresponding to the

'non-affine' components of the PTPS transformations from T to Y, as explained above. Notice how the figure in the right panel of fig 5.5, in section 5.3.3, showing a mapping of points on a square grid placed on $Q_A (Q_A^T Q_A)^{-1} Q_A^T Y$ to locations in the space of Y, clearly illustrates the non-affine displacements/residuals seen here more effectively than the centre panel of fig. 5.5.

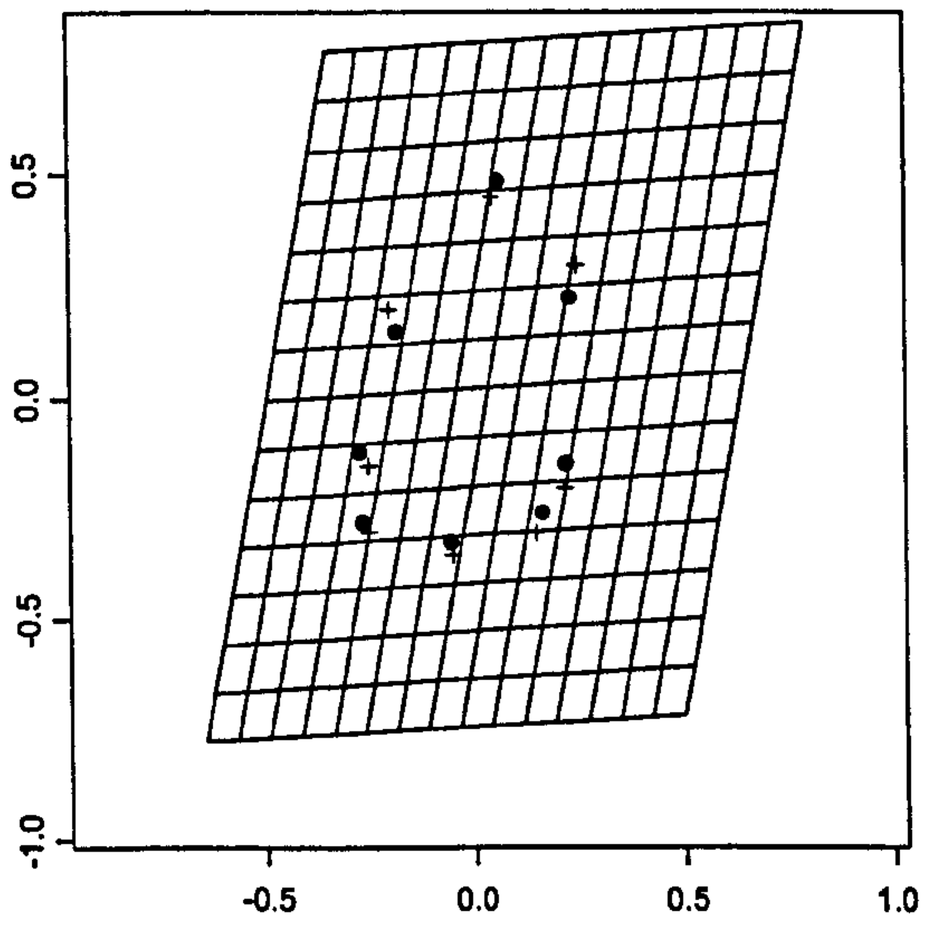


Fig 5.10: Mapping of points on a square grid, placed on T by the GLS affine transformation used to fit T to Y (or by the affine component of a the PTPS transformation from T to Y). (Filled circles) GLS affine fit of T to Y (crosses). Filled circles also correspond to the 'affine' configuration defined in 5.3.3.

5.3.6 Other uses of PTPS transformations

5.3.6.1 Image warping

One of the main uses of the PTPS technique is in 'image warping' or 'morphing'. Let D_T and D_Y denote the regions of R^2 enclosed by the boundaries of images containing configurations T and Y. Let $f_T(t)$, $t \in D_T$ and $f_Y(y)$, $y \in D_Y$ define the pixel grey levels on D_T and D_Y respectively and $\Phi(t)$ denote the PTPS mapping from points $t \in D_T$ to $y \in D_Y$, with $\Phi(T) = Y$. By using the correspondence between points $t \in D_T$ and $y \in D_Y$, provided by $\Phi(t)$, the grey levels of one image D_T may be assigned to those of D_Y or *vice versa*.

If only the landmarks of T are available, an image for D_T may be constructed from the grey levels of D_Y by calculating $f_T(t) = f_Y(\Phi(t))$, $t \in D_T$. For an illustration of this

idea, where an image 'template' of an upper canine is used to construct images of teeth using landmarks obtained from x-rays, see Enciso *et al.* (2003). Conversely, if only the landmarks of Y are available in D_Y, an image can be constructed from D_T by letting $f_Y(y)=f_T(\Phi^{-1}(t))$, where $\Phi^{-1}(t)$ denotes the reverse PTPS mapping from points $y \in D_Y$ to $t \in D_T$. Dryden & Mardia (1998) describe how this idea may be extended to compute an average image for the Procrustes mean shape. If $f_{X_i}((x,y))$, i=1,...,n denote the grey levels of images, containing landmark configurations $X_1,\ldots,X_n$, which have full Procrustes estimate of mean shape $\hat{\mu}$ and $\Phi_i(\hat{\mu})$ denotes the PTPS mapping from $\hat{\mu}$ to each X_i, then the average image corresponding to $\hat{\mu}$ has grey levels at each pixel location t given by $f_T(t)=\frac{1}{n}\sum_{i=1}^{n} f_{X_i}(\Phi_i(t))$.

5.3.6.2 Principal and partial warps

Using principal and partial warps (Bookstein, 1989, 1991), the non-affine components of a PTPS transformation may be decomposed into a series of large and small scale components, based on an eigenvalue decomposition of the BE matrix, Γ^{11} of T.

The principal warps are univariate functions of points $t \in R^2$, used to construct an orthogonal basis for re-expressing a PTPS transformation from T and are given by:

$$P_\rho(t)=\gamma_\rho^T s(t),\ \rho=1,\ldots,k-3,$$

where γ_ρ are the $k-3$ eigenvectors of Γ^{11}, corresponding to eigenvalues λ_ρ. The partial warps are then bivariate functions based on the $P_\rho(t)$, which may be calculated for a configuration Y (with homologous landmarks to those of T) and are given by:

$$R_\rho(t)=Y^T\lambda_\rho\gamma_\rho P_\rho(t)=Y^T\lambda_\rho\gamma_\rho\gamma_\rho^T s(t),\ \rho=1,\ldots,\ k-3,$$

with lth bivariate partial warp score for Y (from T) given by $Y^T\gamma_\rho$, ρ=1,..., $k-3$. The ρth partial warp corresponds to the change in landmark locations (from T to Y) which

are most weighted in the lth principal warp. Since $W^T s(t) = \sum_{\rho=1}^{k-3} R_\rho(t)$, the partial warps are a decomposition of the non-affine part of the PTPS transformation, the ρth partial warp scores indicating the contribution of the ρth principal warp to the deformation from T to Y, in each Cartesian axes. The partial warps are ordered according to the amount of bending required to move points in T to those in Y and so with increasing ρ, represent the deformation at larger to smaller scales. Dryden & Mardia (1998) suggest displaying the partial warps as deformations of points t, on a square grid placed over T, with new positions for each t given by $t + R_\rho(t)$.

5.3.6.3 Relative warps

As already mentioned in section 2.2.3.3.2, an alternative to performing PCA on the covariance matrix of the vectorised Procrustes fits, is to obtain PCs or 'Relative warps' (Bookstein, 1991) with respect to the bending energy matrix Γ^{11} of the mean shape $\hat{\mu}$.

Let l_q and f_q, q=1,...,rank(B_2) be the non-zero eigenvalues and corresponding eigenvectors of:

$$(B_2^-)^{\alpha/2} S_c (B_2^-)^{\alpha/2}$$

where $B_2 = I_2 \otimes \Gamma^{11}$ with generalised inverse B_2^- (e.g. the Moore-Penrose generalised inverse in Mardia *et al.*, 1979), $(B_2^-)^{\alpha/2} = \sum_{q=1}^{rank(B_2)} \lambda_q^{-\alpha/2} \gamma_q \gamma_q^T$, where λ_q and γ_q, r=1,...,rank(B_2) are the eigenvalues and corresponding eigenvectors of B_2, S_c is the $2k \times 2k$ sample covariance matrix of the centred vectorised tangent coordinates and α is a weighting value (see below).

The f_q, are called the relative warps. The effect of the qth relative warp can be viewed by plotting:

$$\mu \pm c B_2^{\alpha/2} f_q \sqrt{l_q}$$

for various values of c, where $B_2^{\alpha/2} = \sum_{q=1}^{rank(B_2)} \lambda_q^{\alpha/2} \gamma_q \gamma_q^T$. For a sample of i=1,...,n vectorised Procrustes fits, relative warp scores may then be obtained along these directions, and used as shape variables in a subsequent multivariate analyses, by computing:

$$f_q^T (B_2^-)^{\alpha/2} vec(X_i^P) \ , q=1,..., \text{rank}(B_2), i=1,...,n.$$

The parameter α adjusts the weights for the linear combinations and is chosen depending on the particular application. α=+1 emphasises large scale shape variability and α=-1 emphasises small scale variability. α=0 corresponds to $B_2^0 = I_{2k}^0$ and so gives the same result as ordinary PCA of the v_i with respect to the Euclidean metric.

5.4 Semi-landmarks

The PTPS transformation can be modified to allow some of the landmarks in the target configuration Y to move away from their original positions along particular directions. (Bookstein, 1996a,d,e). New locations are found for these landmarks, which minimise the bending energy of the PTPS transformation from the reference configuration T, as the landmarks in Y vary along these directions. These landmarks are known as the 'semi-landmarks'.

For a sample of configurations, Bookstein (1996a,d,e) proposed that the technique be used as an additional standardisation step for each configuration, following registration by GPA. For each configuration, selected landmarks are allowed to move along pre-specified chord directions so as to minimise the bending energy of the PTPS mapping from the full Procrustes estimate of mean shape. The selected landmarks are those which are known to lie on a particular line or curve, but along these directions, cannot be assured to correspond between cases, within reasonable ranges of variation. Once the new positions of the semi-landmarks have been determined, a new GPA is performed, a new Procrustes mean is obtained and the process repeated until the semi-landmarks stop moving. By allowing certain landmarks to move along unwanted directions of variation during the iterative registration procedure, it is hoped that the

unwanted variation in the original Procrustes coordinates along these directions can be reduced or eliminated. The final sets of Procrustes registered fits may then be analysed in the same way as for fixed landmarks in any subsequent analysis.

5.4.1 Determination of new semi-landmark positions

5.4.1.1 Modification of PTPS formulation

The PTPS mapping from reference configuration T to target configuration Y, described in detail in 5.3, may be modified as follows to allow a sublist of landmarks in Y to move along particular directions. If we consider

$$Y^0 = \begin{pmatrix} y_{1x}^0 & y_{1y}^0 \\ \vdots & \vdots \\ y_{kx}^0 & y_{ky}^0 \end{pmatrix}$$

as the nominal (original) positions of the set of landmarks Y, and allow a sublist $j_{(l)}$, $l = 1,\ldots,L$ of the $j=1,\ldots,k$ landmarks to move away from these positions in specified unit directions $u_{j_{(l)}} = (u_{jx_{(l)}}, u_{jy_{(l)}})^T$, the new target set of landmark positions in Y is given by:

$$\begin{pmatrix} y_{jx_{(l)}}^{new} \\ y_{jy_{(l)}}^{new} \end{pmatrix} = \begin{pmatrix} y_{jx_{(l)}}^0 \\ y_{jy_{(l)}}^0 \end{pmatrix} - \lambda_{(l)} \begin{pmatrix} u_{jx_{(l)}} \\ u_{jy_{(l)}} \end{pmatrix}$$

where $\lambda_{(l)}$ is some scalar. Note that a '–' is used rather than a '+' as in Bookstein (1996a,d,e), to maintain the generalised least squares analogy (section 5.4.4). Writing,

$$vec(Y^0) = (y_{1x}^0, \ldots, y_{kx}^0, y_{1y}^0, \ldots, y_{ky}^0)^T,$$

$$vec(Y^{new}) = (y_{1x}^{new}, \ldots, y_{kx}^{new}, y_{1y}^{new}, \ldots, y_{ky}^{new})^T,$$

as the vectorised old and new landmark coordinates we have:

$$vec(Y^{new}) = vec(Y^0) - U\lambda \qquad (5.58)$$

with $$Y^{new} = vec_2^{-1}(vec(Y^0) - U\lambda) = Y^0 - vec_2^{-1}(U\lambda)$$

where U is a $2L\times m$ matrix in which the $(j_{(l)}, j')$th entry is $u_{x_{(l)}}$, the $(k+j_{(l)},\ j')$th entry is $u_{y_{(l)}}$ and other entries are zero, with $\lambda=(\lambda_{(1)},\ldots,\lambda_{(L)})^T$. Recall from 3.4.1 that the subscript '2' denotes that the reverse vectorisation, of $vec(Y^{new})$ into matrix Y^{new}, is into m=2 columns. However, since the semi-landmark method is for 2D data only, we will drop this subscript from hereon.

For example, if the 2nd, 3rd and 5th landmarks of a configuration with k=5 landmarks are allowed to move along lines, we have:

$$vec(Y^{new}) = \begin{pmatrix} y^0_{1x} \\ y^0_{2x_{(1)}} - \lambda_1 u_{x_{(1)}} \\ y^0_{3x_{(2)}} - \lambda_2 u_{x_{(2)}} \\ y^0_{4x} \\ y^0_{5x_{(3)}} - \lambda_3 u_{x_{(3)}} \\ y^0_{1y} \\ y^0_{2y_{(1)}} - \lambda_1 u_{y_{(1)}} \\ y^0_{3y_{(2)}} - \lambda_2 u_{y_{(2)}} \\ y^0_{4y} \\ y^0_{5y_{(3)}} - \lambda_3 u_{y_{(3)}} \end{pmatrix} = \begin{pmatrix} y^0_{1x} \\ y^0_{2x_{(1)}} \\ y^0_{3x_{(2)}} \\ y^0_{4x} \\ y^0_{5x_{(3)}} \\ y^0_{1y} \\ y^0_{2y_{(1)}} \\ y^0_{3y_{(2)}} \\ y^0_{4y} \\ y^0_{5y_{(3)}} \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 \\ u_{x_{(1)}} & 0 & 0 \\ 0 & u_{x_{(2)}} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & u_{x_{(3)}} \\ 0 & 0 & 0 \\ u_{y_{(1)}} & 0 & 0 \\ 0 & u_{y_{(2)}} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & u_{y_{(3)}} \end{pmatrix} \begin{pmatrix} \lambda_{(1)} \\ \lambda_{(2)} \\ \lambda_{(3)} \end{pmatrix}, \quad Y^{new} = \begin{pmatrix} y^0_{1x} & y^0_{1y} \\ y^0_{2x_{(1)}} - \lambda_1 u_{x_{(1)}} & y^0_{2y_{(1)}} - \lambda_1 u_{y_{(1)}} \\ y^0_{3x_{(2)}} - \lambda_2 u_{x_{(2)}} & y^0_{3y_{(2)}} - \lambda_2 u_{y_{(2)}} \\ y^0_{4x} & y^0_{4y} \\ y^0_{5x_{(3)}} - \lambda_3 u_{x_{(3)}} & y^0_{5y_{(3)}} - \lambda_3 u_{y_{(3)}} \end{pmatrix}$$

We then seek the minimum BE transformation from T to $Y^{new}=vec^{-1}(vec(Y^0)-U\lambda)$, allowing certain landmarks to move away from their original positions along directions defined by U. By equation (5.34), this is achieved by finding vector λ minimising:

$$(y^{new}_{1x},\ldots,y^{new}_{kx})^T\,\Gamma^{11}\begin{pmatrix} y^{new}_{1x} \\ \vdots \\ y^{new}_{1x} \end{pmatrix} + (y^{new}_{1y},\ldots,y^{new}_{ky})^T\,\Gamma^{11}\begin{pmatrix} y^{new}_{1y} \\ \vdots \\ y^{new}_{1y} \end{pmatrix} = vec(Y^{new})^T \begin{pmatrix} \Gamma^{11} & 0_{k\times k} \\ 0_{k\times k} & \Gamma^{11} \end{pmatrix} vec(Y^{new})$$

$$= (vec(Y^0)-U\lambda)^T \begin{pmatrix} \Gamma^{11} & 0_{k\times k} \\ 0_{k\times k} & \Gamma^{11} \end{pmatrix} (vec(Y^0)-U\lambda) \qquad (5.59)$$

as the landmarks of the sublist vary along their associated directions and where Γ^{11} is the BE matrix of reference form T. Equation (5.59) is a generalised sum of squares in the form of (5.43) and so the solution is achieved for:

$$\lambda = \left(U^T \begin{pmatrix} \Gamma^{11} & 0_{k\times k} \\ 0_{k\times k} & \Gamma^{11} \end{pmatrix} U \right)^{-1} U^T \begin{pmatrix} \Gamma^{11} & 0_{k\times k} \\ 0_{k\times k} & \Gamma^{11} \end{pmatrix} vec(Y^0). \qquad (5.60)$$

5.4.1.2 Example with escribed chords

The choice of directions (and hence potential new positions) for each of the selected landmarks in a configuration Y depends upon the problem being considered. For example, certain landmarks may be known to be located around the outline of an object, but lack correspondence between cases along this outline. A way to model this situation (Bookstein, 1996,a,d,e) is to approximate the direction of the outline at each of these points by an 'escribed chord': a line through the point, parallel to that joining its immediate neighbours, see fig. 5.7 (left).

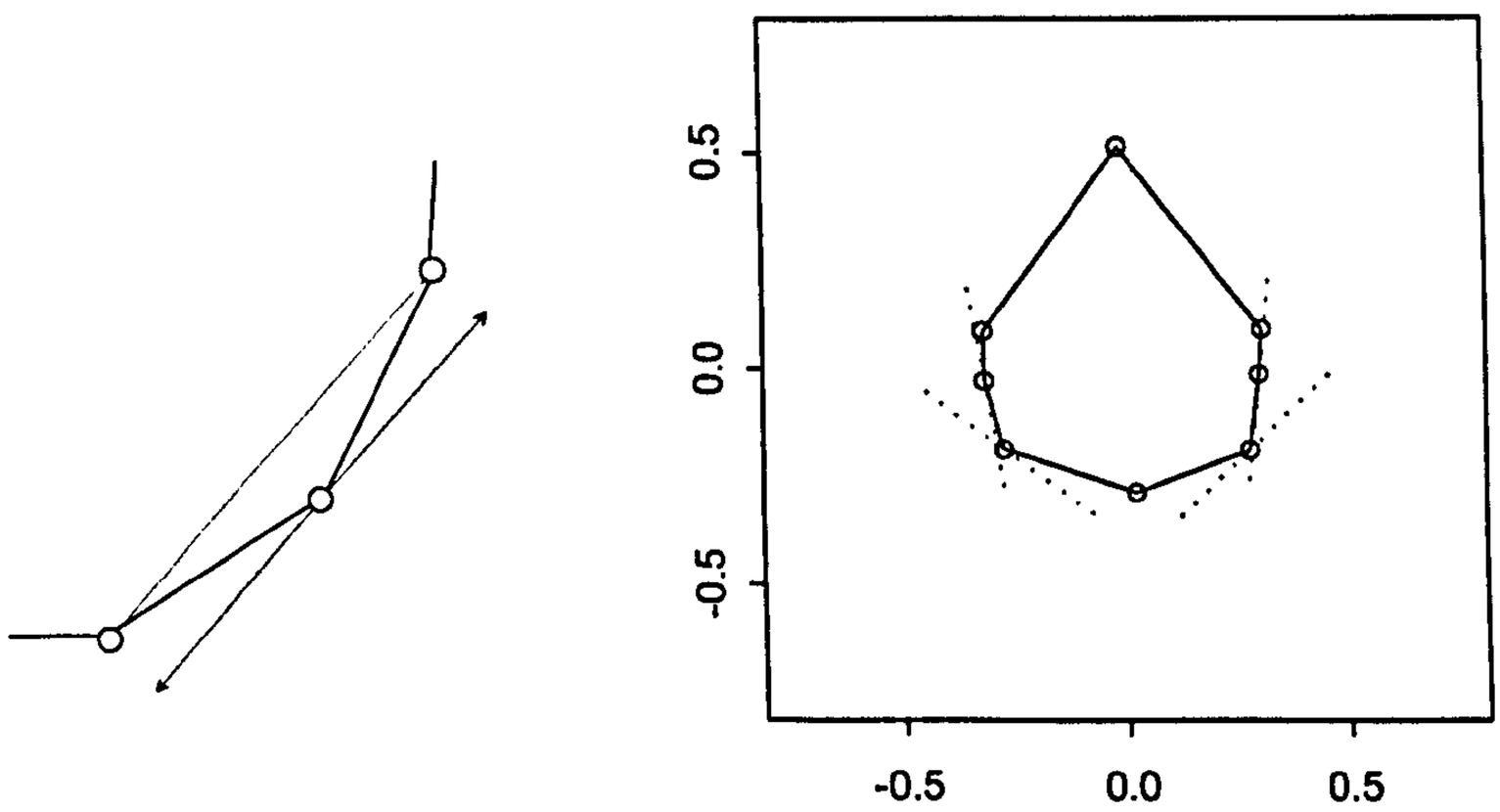


Fig 5.7: (Left) Escribed chord through a point. (Right) Chords through the MD endpoints and incisal corner landmarks of an upper central incisor configuration

Fig. 5.7(right) shows an individual upper central incisor configuration, *Y*, taken from the dataset of control cases used in 2.2.3.2.1. Four landmarks have been selected as semi-landmarks: the MD endpoints and incisal corners. Using the full Procrustes mean shape of this sample as *T*, the process of obtaining new semi-landmark positions in *Y* is illustrated in fig. 5.8 below. The left panel in fig. 5.8(a) is a square grid placed on *T*. The centre panel displays the standard (minimum BE) PTPS transformation of *T* to the original landmark positions (Y^0) of the individual case. The right panel is the minimum bending energy PTPS transformation from *T*, achieved by allowing the four

semi-landmarks to move along their escribed chords. This now requires much less bending. The original and new positions of the landmarks are displayed in fig. 5.8(b).

As the points move along their chords, the new, minimum BE positions adopted are those which produce the smallest possible variation in the affine derivative (shapes of the individual grid cells with respect to their neighbours). Positions are adopted so that any such changes are cast at the largest possible spatial scale, with the smallest possible variation in the shape of the grid cells, since then the contribution to the integral sum of squares is lower.

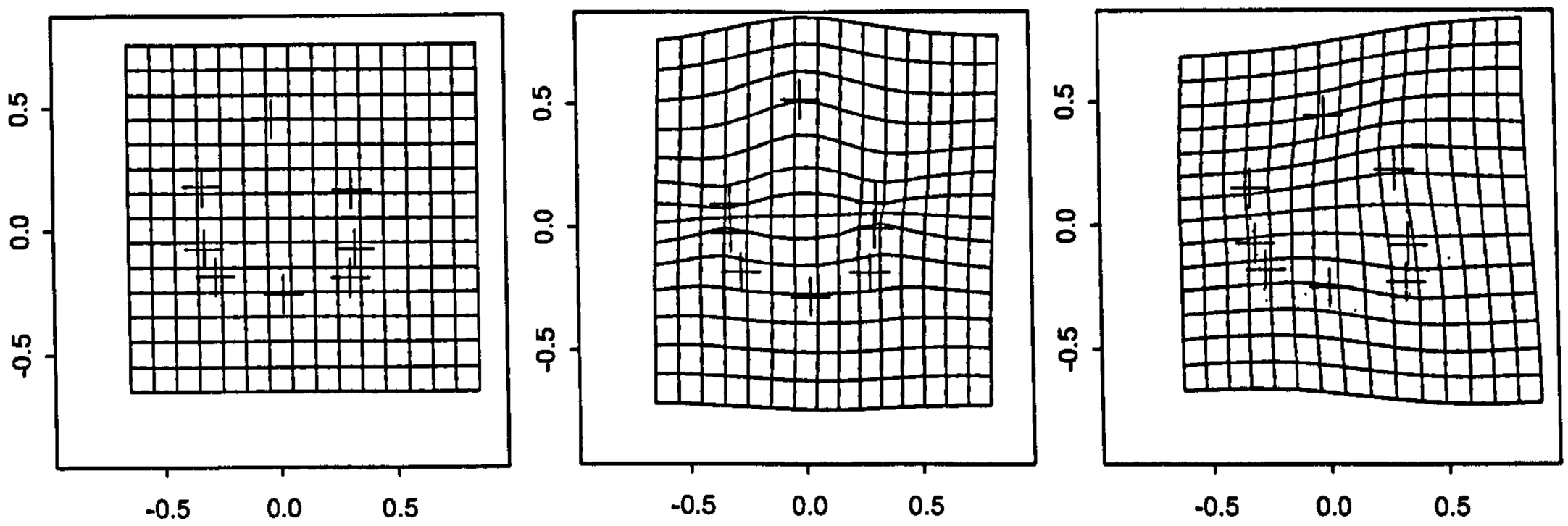


Fig. 5.8 (a): PTPS transformation from Procrustes average (T) (left) to the nominal positions Y^0 of case Y (middle) and to the new (min BE) positions of the same case having allowed four landmarks to move along their escribed chords

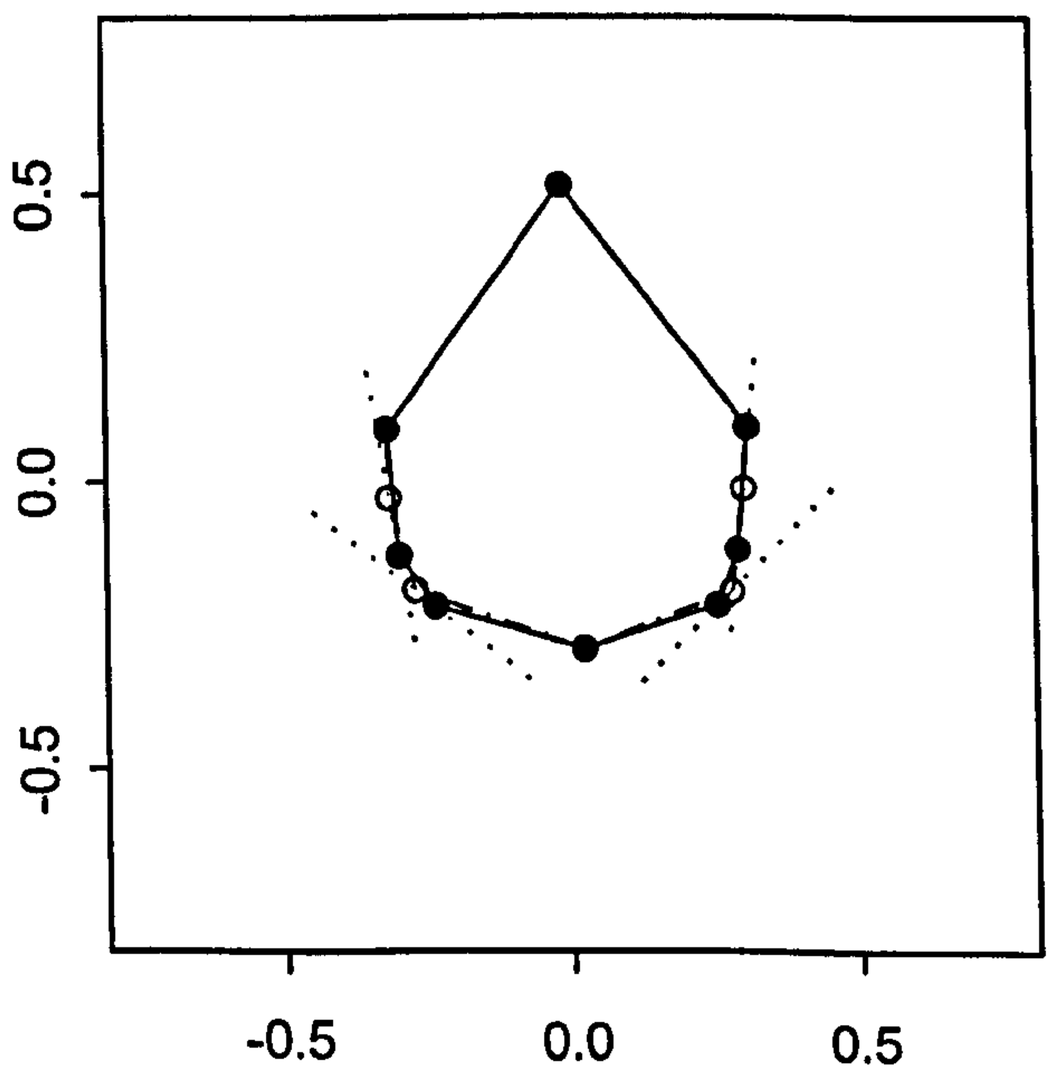


Fig 5.8(b): Nominal (o) and new (•) positions of landmarks following 'relaxation' along their escribed chords.

Where possible, regions and directions requiring most energy in the original PTPS transformation from T to Y^0 will be relaxed preferentially in relation to low energy regions (where the selected landmarks and associated chord directions allow this). In the example above, the most bending required in transforming T to Y^0 was between the gum landmarks and the endpoints of MD, these being much closer on this particular tooth, relative to those in the Procrustes average. When the MD landmarks are allowed to move in a near vertical direction, the relaxation is concentrated here, these points moving further away from those at the papilla.

5.4.1.3 Effect of similarity transformations

Note that the resulting shape of $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$ is unaffected by the registration of either the reference or target configuration, T or Y. A change in location, scale or rotation of either configuration, will still result in the landmarks moving to positions along their chords, which result in the same value for $d_F^2(Y^{new}, T)$.

As explained in 5.3.4, rotations or translations of T have no effect on the BE matrix Γ^{11} (dependent only on T). Since (5.59) is then also unchanged, the new positions of the semi-landmarks in Y^{new} are also unaffected. In terms of fig. 5.5(a) such transformations of T change only the affine component of any PTPS mapping from T to Y^{new} (as explained in 5.3.4). Therefore the minimum BE mapping from a rotated or translated T to $vec^{-1}(vec(Y^0) - U\lambda)$ is simply a rotated or translated version of the same mapping as before, resulting in the same new positions for the semi-landmarks in Y^{new}.

If T changes in size to βT, where β is some scalar, then from table 5.1 $\Gamma^{11} = (1/\beta^2)\Gamma^{11}$ and from (5.59) we would seek to optimise:

$$(vec(Y^0) - U\lambda^*)^T \, (I_2 \otimes (1/\beta^2)\Gamma^{11}) \, (vec(Y^0) - U\lambda^*)$$

$$= (1/\beta^2) \, (vec(Y^0) - U\lambda^*)^T \, (I_2 \otimes (1/\beta^2)\Gamma^{11}) \, (vec(Y^0) - U\lambda^*).$$

The solution $\lambda^* = \lambda$, as in (5.60), again resulting in the same configuration for Y^{new}.

If Y is either translated, re-sized or rotated, by matrix $\alpha = 1_k^T(c_1, c_2)$, scalar β or rotation matrix Γ, the initial nominal positions of the landmarks in Y become $Y^0+\alpha$, βY^0 or $Y^0\Gamma$, the latter causing the unit directions in U to each change by the same rotation, to give a new matrix U^*. Since T and Γ^{11} are unchanged, any PTPS mapping from T to $vec^{-1}(vec(Y^0+\alpha)-U\lambda^*)$, $vec^{-1}(vec(\beta Y^0)-U\lambda^*)$ or $vec^{-1}(vec(Y^0\Gamma)-U^*\lambda^*)$, including that which minimises the BE to $vec^{-1}(vec(Y^0)-U\lambda)$ as the semi-landmarks vary along their chords, is simply a translation α, rescalling (by β) or rotation Γ of the affine, or both affine and non-affine components of the corresponding mapping from T to $vec^{-1}(vec(Y^0)-U\lambda)$, see table 5.1. Consequently, if Y^{new*} denotes the new version of $Y+\alpha$, βY or $Y\Gamma$, then for the shift, the minimisation is of:

$$(vec(Y^0+\alpha)-U\lambda^*)^T\,(I_2\otimes\Gamma^{11})\,(vec(Y^0+\alpha)-U\lambda^*)$$

with

$$\lambda^* = (U^T(I_2\otimes\Gamma^{11})U)^{-1}U^T(I_2\otimes\Gamma^{11})vec(Y^0+\alpha) = \Lambda vec(Y^0+\alpha)$$

and $$Y^{new*} = vec^{-1}(vec(Y^0+\alpha)-U\lambda^*) = Y^{new} + vec^{-1}(vec(\alpha)-\Lambda).$$

For βY, minimisation is of:

$$(vec(\beta Y^0)-U\lambda^*)^T\,(I_2\otimes\Gamma^{11})\,(vec(\beta Y^0)-U\lambda^*)$$

with

$$\lambda^* = (U^T(I_2\otimes\Gamma^{11})U)^{-1}U^T(I_2\otimes\Gamma^{11})vec(\beta Y^0),\ Y^{new*} = \beta vec^{-1}(vec(Y^0)-U\lambda) = \beta Y^{new}.$$

Finally, for $Y\Gamma$ minimisation is of:

$$(vec(Y^0\Gamma)-U^*\lambda^*)^T\,(I_2\otimes\Gamma^{11})\,(vec(Y^0\Gamma)-U^*\lambda^*)$$

with

$$\lambda^* = (U^T(I_2\otimes\Gamma^{11})U)^{-1}U^T(I_2\otimes\Gamma^{11})vec(Y^0\Gamma)$$

and $$Y^{new*} = vec^{-1}(vec(Y^0\Gamma - U^*\lambda^*)\Gamma = vec^{-1}(vec(Y^0 - U\lambda)\Gamma = Y^{new}\Gamma.$$

Therefore, if Y is rotated, shifted or rescaled, the resulting configuration, after allowing the semi-landmarks to move, is simply a rotated, resized or translated version of the same result for Y and therefore has the same shape.

5.4.1.4 Determination of new semi-landmark positions by GLS affine matching

The formulation for determining the new positions of semi-landmarks in $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$, with reference to configuration T, can also be expressed as a problem of matching T to $vec^{-1}(vec(Y^0) - U\lambda)$ by a generalised affine transformation, as the semi-landmarks vary along their chords.

Consider the regression equation

$$Y^{new} = vec^{-1}(vec(Y^0) - U\lambda) = Q_A B_A + E,$$

or equivalently:

$$vec(Y^0) - U\lambda = (I_2 \otimes Q_A)vec(B_A) + vec(E) \qquad (5.61)$$

with

$$Cov(vec(E)) = (I_2 \otimes S),$$

where

$$Q = \begin{pmatrix} 1 & t_{1x} & t_{1y} \\ \vdots & \vdots & \vdots \\ 1 & t_{kx} & t_{ky} \end{pmatrix}, \; B_A = \begin{pmatrix} c_1 & c_2 \\ a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \; E = \begin{pmatrix} e_{1x} & e_{1y} \\ \vdots & \vdots \\ e_{kx} & e_{ky} \end{pmatrix}.$$

Rearranging (5.61), to gather all parameters in a single vector, we have

$$vec(Y^0) = (I_2 \otimes Q_A : U)\begin{pmatrix} vec(B_A) \\ \lambda \end{pmatrix} + vec(E).$$

For the example in 5.4.1.1, where the 2nd, 3rd and 5th landmarks of a configuration with k=5 landmarks are allowed to move along lines, this gives:

$$\begin{pmatrix} y^0_{1x} \\ y^0_{2x_{(1)}} \\ y^0_{3x_{(2)}} \\ y^0_{4x} \\ y^0_{5x_{(3)}} \\ y^0_{1y} \\ y^0_{2y_{(1)}} \\ y^0_{3y_{(2)}} \\ y^0_{4y} \\ y^0_{5y_{(3)}} \end{pmatrix} = \begin{pmatrix} 1 & t_{1x} & t_{1y} & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & t_{2x} & t_{2y} & 0 & 0 & 0 & u_{x_{(1)}} & 0 & 0 \\ 1 & t_{3x} & t_{3y} & 0 & 0 & 0 & 0 & u_{x_{(2)}} & 0 \\ 1 & t_{4x} & t_{4y} & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & t_{5x} & t_{5y} & 0 & 0 & 0 & 0 & 0 & u_{x_{(3)}} \\ 0 & 0 & 0 & 1 & t_{1x} & t_{1y} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & t_{2x} & t_{2y} & u_{y_{(1)}} & 0 & 0 \\ 0 & 0 & 0 & 1 & t_{3x} & t_{3y} & 0 & u_{y_{(2)}} & 0 \\ 0 & 0 & 0 & 1 & t_{4x} & t_{4y} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & t_{5x} & t_{5y} & 0 & 0 & u_{y_{(3)}} \end{pmatrix} \begin{pmatrix} c_1 \\ a_{1x} \\ a_{1y} \\ c_2 \\ a_{2x} \\ a_{2y} \\ \lambda_{(1)} \\ \lambda_{(2)} \\ \lambda_{(3)} \end{pmatrix} + \begin{pmatrix} e_{1x} \\ e_{2x} \\ e_{3x} \\ e_{4x} \\ e_{5x} \\ e_{1y} \\ e_{2y} \\ e_{3y} \\ e_{4y} \\ e_{5y} \end{pmatrix}.$$

To determine the GLS affine fit of T to $vec^{-1}(vec(Y^0)-U\lambda)$, allowing the semi-landmarks in the target configuration to move along their chords, we seek estimates of the affine parameters and vector λ simultaneously, so as to minimise:

$$vec(E)^T (I_2 \otimes S)^{-1} vec(E) \tag{5.62}$$

$$=(vec(Y^0)-(I_2 \otimes Q_A : U)\begin{pmatrix} vec(B_A) \\ \lambda \end{pmatrix})^T (I_2 \otimes S)^{-1} vec(Y^0)-(I_2 \otimes Q_A : U)\begin{pmatrix} vec(B_A) \\ \lambda \end{pmatrix})$$

which, by (5.56), will then have minimum value equal to the least bending energy PTPS mapping from T to $vec^{-1}(vec(Y^0)-U\lambda)$.

By analogy with (5.43) and (5.44), the estimator for parameter vector $(c_1, a_{11}, a_{21}, c_1, a_{12}, a_{22}, \lambda_1, ..., \lambda_{SL})^T$, is given by:

$$\begin{pmatrix} vec(\hat{B}_A) \\ \hat{\lambda} \end{pmatrix} = ((I_2 \otimes Q : U)^T (I_d \otimes S)^{-1} (I_2 \otimes Q : U))^{-1} (I_2 \otimes Q : U)^T (I_d \otimes S)^{-1} vec(Y^0). \tag{5.63}$$

Again note the slight abuse of terminology. In 5.5.1.1, the parameters are 'solved' whereas here they are 'estimated'. $\hat{B}_A$ then corresponds to the solution for the matrix B in the least bending energy PTPS mapping from T to $vec^{-1}(vec(Y^0)-U\hat{\lambda})$, given by (5.27). By (5.56), the corresponding non-affine components can then be determined by solving:

$$SW = vec^{-1}(vec(Y^0)-U\hat{\lambda}) - Q_A \hat{B}_A. \tag{5.64}$$

5.4.2 Iterative registration and subsequent analysis

For a sample of configurations, Bookstein (1996a,d,e) proposed that the technique be used iteratively to overcome lack of precise landmark correspondence between cases, when obtaining their Procrustes registered fits.

Having obtained initial fits, X_i^P, i=1,...,n and an estimate of mean shape $\hat{\mu}$, using GPA, the same set of semi-landmarks are chosen on each configuration, as free to move away from their nominal positions along homologous, pre-specified chord directions (typically defined by neighbouring landmarks in each X_i^P).

Each iteration of the registration process then involves the following steps:

1. (Semi-landmark movement) For each configuration in the sample, new semi-landmark positions are determined by following the procedure described in 5.4.1. The mean shape $\hat{\mu}$ is used as the reference shape T and each X_i^P (considered as some perturbation of $\hat{\mu}$), as the target Y, whose new semi-landmark positions are to be determined. Each set of semi-landmarks move along their chords to positions which minimise the BE of the PTPS mapping from $\hat{\mu}$, producing a new set of configurations X_i^{new}, i=1,..,n.
2. (GPA registration to updated $\hat{\mu}$) The resulting X_i^{new}, i=1,..,n are then no longer Procrustes fits to $\hat{\mu}$ and $\hat{\mu}$ is no longer the full Procrustes mean of the new configurations X_i^{new}. A new GPA is therefore performed to obtain a new, updated estimate of mean shape and corresponding fits X_i^P.

However, since $\hat{\mu}$ has changed (along with the chord directions, if these depend on the semi-landmarks positions in X_i^P), the semi-landmarks in the new X_i^P may now no longer lie at positions along their chords which optimise the BE of the PTPS mapping from $\hat{\mu}$. The process is therefore repeated from the first step above, to allow the semi-landmarks to move to new positions along their new chords with minimum bending energy from the updated $\hat{\mu}$ and then a new GPA (step 2) performed. This process is then repeated until the semi-landmarks on all configurations remain stationary.

By allowing the semi-landmarks to move along unwanted directions of variation during an iterative GPA, it is hoped that any unwanted variation in the coordinates of the initial Procrustes fits (obtained with all landmarks fixed), along these directions, can be reduced or eliminated.

At each movement step it is hoped that the configurations do not deviate too far from their original shape since there are no bounds on the magnitude of movement allowed along the chosen directions. Indeed, (Bookstein, 1996d) commented that "the 'escribed chords' method, while somewhat crude, tends to work well except where the curve on which they are placed is turning rapidly".

Following iteration, the final set of Procrustes fits may then be used for subsequent analysis, using the descriptive and inferential techniques described in section 3.4. (Bookstein, 1996d, 1998). Note however, that Goodall's F test, which assumes isotropic variance in the coordinates, does not apply to semi-landmark registered data, due to the spatial correlation in landmark positions induced by the method (Bookstein, 1998).

The final positions of semi-landmarks in the X_i^P, i=1,...,n, now lie in a lower dimension than the other fixed landmarks. The iterative GPA has allowed these landmarks to move along chords during the superimposition and this is why they are referred to as 'semi' landmarks. They are not true landmarks as they cannot be defined on a single image and exist only in the context of the final sample mean $\hat{\mu}$. (An alternative could be to use an arbitrary external configuration as the reference T which, unlike the Procrustes mean, would not keep changing). However, the semi-landmarks still 'correspond' by virtue of minimising the BE of the PTPS representing the deviation of these points in the final Procrustes mean (or arbitrary external configuration).

5.4.3 Previous applications

Up to early 2004, the only publications documenting use of the semi-landmark technique, are those that have Bookstein as the main author; an observation confirmed by Bookstein himself, in a recent posting on 'morphmet' (summer 2003). Nearly all of this material considers analysis of the shape of the midsagittal section of the corpus

callosum (found in the centre of the human brain), obtained from magnetic resonance (MR) images.

Such data was first used in the original 1996 articles, which introduced the semi-landmark technique and provided most of the mathematical details here. In Bookstein (1996a,d,e) differences in callosum shape between schizophrenic and control patients were examined. Due to the lack of any homologously identifiable features, twenty-six landmarks were collected on each image, consisting of points placed at equal distance around the midsagittal section outline. In the following iterative registration, each landmark was allowed to move along its escribed chord. The same datasets were also used in Bookstein (1999).

In more recent articles (Bookstein *et al.*, 2001, 2002a,b), the shape of the midsagittal sections of the corpus callosum was examined in patients with heavy fetal alcohol exposure and compared mean shape of a corresponding control group. The same method of data collection and iterative registration was used, although this time callosal shapes were represented by 40 equally spaced semi-landmarks around each outline.

A study using the semi-landmark method for a different application is that of Bookstein *et al.* (1999). Here the technique was used to perform comparisons of shape between the interior and exterior frontal bone profiles of archaic and modern humans, obtained from computerised tomography (CT) scans. For the inner profiles, 2 fixed landmarks and 10 equally placed semi-landmarks along the curve of the outline between them, were used to represent the bone outline. For the outer profiles 10 equally spaced semi-landmarks between 2 or 3 fixed landmarks, were used.

In each of these investigations it was found that any variation in the placement of these points around the callosum outlines or along curve of the frontal bone was considerably reduced, whereas variation normal to outline or curve remained, allowing more meaningful comparisons of shape to be made.

In section 5.6 we investigate how well this method performs in addressing some of the reliability issues identified in chapter 4 and the problem of unwanted variation resulting from differences in the position of a patient's gingival tissue.

5.5 Implementation

Although the papers by Bookstein (1996a,d,e) provide the mathematical details, at the start of our investigations (early 2000), no readily available routines or software were available to perform the computations to implement the semi-landmark method. Consequently, new specific routines had to be written, using S-plus, of which details are given below. An important undocumented issue in Bookstein's papers concerns how to assess convergence of the method, including the criteria that should be used.

5.5.1 Choice of GPA method

Recall from section 3.2.2.5 that, depending on the GPA method used, differences in the sizes of the estimated mean shape $\hat{\mu}$ and Procrustes fits X_i^P will result, although the shape of $\hat{\mu}$ and variance in shape, measured by:

$$\sum_{i=1}^{n} d_F^2(X_i^P, \hat{\mu}) = \sum_{i=1}^{n} \left\| X_i^P - \hat{\mu} \right\|^2 = RSS(\hat{\mu})$$

$$(= \sum_{i=1}^{n} \left\| X_i^P - \overline{X}^P \right\|^2 = RSS \text{, if GPA method (ii) in section 3.2.2 is used, since then } \hat{\mu} = \overline{X}^P)$$

are still identical. As was described in section 5.4.1.3, the resulting shape of $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$, after allowing movement of the semi-landmarks to new positions, is unaffected by differences in the registration of Y or T. Therefore at each movement step of the iteration, it does not matter which GPA method has been used to align the configurations beforehand. Any differences in the registration (including size) of $T = \hat{\mu}$ and $Y = X_i^P$, resulting from the GPA method used will still result in the same configurations $Y^{new} = X_i^{new}$ being produced, along with the same value of the summary measure $RSS(\hat{\mu})$ above.

Recall however, that if the iterative GPA method of Gower/Rohlf & Slice (method (ii) in section 3.2.2) is used, for obtaining the final Procrustes fits and corresponding mean shape, a partition of $\sum_{i=1}^{n} \left\| X_{i)}^P - \overline{X}_i^P \right\|^2$ $(= RSS)$ into between cases, operators and errors,

is then a partition of the value $\sum_{i=1}^{n} d_F^2(X_i^P, \hat{\mu})$ $(=RSS(\hat{\mu}))$, whereas if GPA method (i) is used, $RSS \neq RSS(\hat{\mu})$, see section 4.4.2.2. If results are to be analysed and presented in the same way as in chapter 4, then GPA method (ii) may therefore be more beneficial.

5.5.2 Convergence assessment

To assess whether or not the method has converged, i.e. whether the landmarks have essentially stopped moving, there are several possibilities. Let $X_{i(r)}^{new}$, i=1,...,n, denote the new configurations produced on iteration r after allowing the semi-landmarks of each Procrustes fit $X_{i(r-1)}^{P}$, to move to new positions with respect to Procrustes mean shape $\hat{\mu}_{(r-1)}$. Each set of semi-landmarks move distances given by vector $\lambda_{i(r)}$ along unit chord directions defined by $U_{i(r)}$. The configurations $X_{i(r)}^{P}$, i=1,...,n, and $\hat{\mu}_{(r)}$ are then the new Procrustes fits and corresponding mean shape of the $X_{i(r)}^{new}$, obtained by GPA at the end of the rth iteration.

Let ε be some pre-chosen level of tolerance. The semi-landmark process should be stopped at the end of the rth complete iteration, i.e. following GPA, to ensure that the configurations are Procrustes registered in preparation for subsequent analysis, when either:

$$\sum_{i=1}^{n} \left\| X_{i(r)}^{new} - X_{i(r-1)}^{P} \right\| < \varepsilon, \sum_{i=1}^{n} d_F^2(X_{i(l)}^{new}, X_{i(l-1)}^{P}) < \varepsilon \text{ or } \sum_{i=1}^{n} \left\| \lambda_{i(r)} \right\| < \varepsilon \qquad (5.65)$$

(the three conditions will be equivalent), or if:

$$\sum_{i=1}^{n} d_F^2(X_{i(r)}^{P}, \hat{\mu}_{(r)}) - \sum_{i=1}^{n} d_F^2(X_{i(r-1)}^{P}, \hat{\mu}_{(r-1)}) = \sum_{i=1}^{n} \left\| X_{i(r)}^{P} - \hat{\mu}_{(r)} \right\|^2 - \sum_{i=1}^{n} \left\| X_{i(r-1)}^{P} - \hat{\mu}_{(r-1)} \right\|^2$$

$$= RSS(\hat{\mu}_{(r)}) - RSS(\hat{\mu}_{(r-1)}) = \Delta RSS(\hat{\mu}_{(r)}) < \varepsilon \qquad (5.66)$$

which, if iterative GPA method (ii) is used, will be the same as considering:

$$\sum_{i=1}^{n} \left\| X_{i(r)}^{P} - \bar{X}_{i(r)}^{P} \right\|^2 - \sum_{i=1}^{n} \left\| X_{i(r-1)}^{P} - \bar{X}_{i(r-1)}^{P} \right\|^2 = RSS_{(r)} - RSS_{(r-1)} < \varepsilon \qquad (5.67)$$

since $RSS(\hat{\mu}) = RSS$. In practice, the different methods of assessment will essentially give identical results. However, the options given by (5.66) and (5.67) are preferred to those in (5.65), since then the assessment is made on the actual Procrustes fits on which subsequent analysis will be based and is in keeping with the methods of convergence assessment used in the iterative GPA methods, which are based on the same measure of sample variation in shape.

For each of the investigations in 5.6, a tolerance level of $\varepsilon = 0.001$ is used to check convergence of $RSS(\hat{\mu})$, between successive iterations. This is the same as the stricter of the two tolerance levels suggested by Gower (1975) and Rohlf & Slice (1990) for assessing convergence of the same measure between successive iterations of the GPA procedure (see section 3.2.2.2.1). Further discussion of the choice of an appropriate level of convergence, is made in section 5.5.6.

Another alternative could be to assess convergence on the total bending energy (BE) of the PTPS mappings from $\hat{\mu}$ to each X_i^P or X_i^{new}. For example, if $\Gamma^{11}_{(r-1)}$ denotes the bending energy of $\hat{\mu}_{(r-1)}$, then the process would be stopped if:

$$\sum_{i=1}^{n} tr(X^P_{i(r)}{}^T \Gamma^{11}_{(r)} X^P_{i(r)}) - \sum_{i=1}^{n} tr(X^P_{i(r-1)}{}^T \Gamma^{11}_{(r-1)} X^P_{i(r-1)}) < \varepsilon$$

or

$$\sum_{i=1}^{n} tr(X^{new\,T}_{i(r)} \Gamma^{11}_{(r-1)} X^{new}_{i(r)}) - \sum_{i=1}^{n} tr(X^P_{i(r-1)}{}^T \Gamma^{11}_{(r-1)} X^P_{i(r-1)}) < \varepsilon .$$

However, this is less desirable than the options above, since convergence of total BE does not always tell us whether or not the Procrustes registration has converged. For instance, one problem with the second option is that it is always possible for semi-landmarks to move to new positions which leave the BE from $\hat{\mu}_{(r-1)}$ essentially unchanged, but still change the shape of a configuration quite notably, for example when the difference in shape between $X^{new}_{i(r)}$ and $X^P_{i(r-1)}$ is affine. The first option should be able to detect this, but will not do so if the change in $\hat{\mu}$ is also affine.

5.5.3 S-plus routine

The S-plus routine, 'semi.it', custom written during 2000/01 and used for the investigations in section 5.6, can be found in the appendix. The arguments which may be supplied, along with their default values, are described below:

`configs` $k\times2\times n$ real array of raw landmark configuration data. The j=1,...,k landmarks should be in the same order for each of the n configurations.

`l1, l2` Mean shape and Procrustes fits always rotated so that landmarks $l1$ and $l2$ (1,...,k) in $\hat{\mu}$ lie horizontal ($l1$=left, $l2$=right). (Default: rotate $\hat{\mu}$ so major axis vertical).

`joinline` Vector sequence of landmark numbers 1,...,k in order to be joined by a continuous straight line edge whenever shapes plotted. (Default: no line).

`proc.opt` GPA method to use for obtaining Procrustes mean $\hat{\mu}$ and corresponding fits X_i^P.

1=Complex-eigenvector method (calls on 'procrustes.2d' routine in Dryden's 'shape' package), 2=Iterative procedure of Gower (1975), Rohlf & Slice (1990). (Default: 2)

`GPA.crit` Convergence criterion for iterative GPA method. (Default: 0.0001)

`DM` A $k\times2$ matrix in which the jth row is either a pair of zero's, indicating that landmark j is not a semi-landmark, or a pair of landmark numbers, indicating that it is (i.e. j is always one of the sublist $j_{(l)}$). Each pair of landmark number is used to specify the unit chord directions $u_{j_{(l)}}\ (u_{jx_{(l)}}, u_{jy_{(l)}})^T$ for each semi-landmark, by calculating the unit direction vector between each pair of specified landmarks in each $X^P_{i(r-1)}$.

(Default: matrix of 0's, i.e. no-semi-landmarks).

`conv.crit` Convergence criterion ε for $\Delta RSS(\hat{\mu}_{(r)}) = RSS(\hat{\mu}_{(r)}) - RSS(\hat{\mu}_{(r-1)}) < \varepsilon$.

(Default: 0.001 – Same as suggested by Gower (1975), Rohlf & Slice (1990) for GPA).

`max.it` Number of iterations to stop after, if convergence not achieved. (Default: 20).

`ind.pics` Plots for each configuration at each iteration. (Default: no pics). Possible values:

"on"- Plot of $X^{new}_{i(r)}$ (filled circles) and $X^P_{i(r-1)}$ (empty circles) on same figure, with chord directions shown as dotted lines through the nominal semi-landmark positions in $X^P_{i(r-1)}$. Lines extended in either direction to length specified by 'mag' (see below).

"tps" – Pair of deformation grids showing PTPS mapping from $\hat{\mu}_{(r-1)}$ to $X^P_{i(r-1)}$ and $\hat{\mu}_{(r-1)}$ to $X^{new}_{i(r)}$, as in centre and right panels of fig. 5.8(a).

"tpscomps" - Plots of affine and non-affine components of each PTPS transformation above (as described in section 5.3.3 and in fig. 5.5, left and right panels).

"affinefits" - Plots of GLS affine fit of $\hat{\mu}_{(r-1)}$ to $X^P_{i(r-1)}$ and $\hat{\mu}_{(r-1)}$ to $X^{new}_{i(r)}$, as described in 5.3.5.4 and fig. 5.11. Displayed on a deformation grid.

`mag` See above. (Default=$\|\hat{\mu}\|/4$).

`final.pics` "on" - Plots of initial and final scatters of Procrustes fits about mean shape, ignoring "joinline" argument. (Default: no plots).

"B&A" – n individual plots of each original and final X^P_i on same figure.

The output values from the routine then comprise:

`$newconfigs` $k\times2\times n$ real array of the final Procrustes fits.

`$mshape` Final Procrustes mean shape ($k\times2$ matrix).

`$totalrss` Vector of values of $RSS(\hat{\mu}_{(r)})$ obtained after initial and each GPA and at the end of each iteration.

`$changerss` Vector of values of $\Delta RSS(\hat{\mu}_{(r)}) = RSS(\hat{\mu}_{(r)}) - RSS(\hat{\mu}_{(r-1)})$.

`$iter` Number of iterations performed.

5.5.4 Recent software developments (TPSrelw v.1.29-1.32)

In spring 2003, an update was made to the TPSrelw program on the Stony Brook website, to include semi, or 'sliding' landmarks as they are referred to (see section 2.5 for web link). Here a separate file must be used to specify the semi-landmarks ('sliders'), whose associated chord directions can be defined interactively by selecting pairs of landmarks on a display of the initial estimate of mean shape.

The accompanying documentation states that the iterative method described by Rohlf & Slice (1990) is used to perform the GPA steps. However, it is unclear as to whether the semi-landmark movements steps are performed at the end of each GPA, as in our routine and as described in the TPSrelw help file, or during each iteration of the GPA registration, as is stated in the accompanying methodology description. (Out of interest a variation on the 'semi.it' routine was written to perform the calculations in this way and was found to produce identical results). In addition, part of the documentation states that the chord directions are obtained by calculating directions between pairs of landmarks in $\hat{\mu}$, rather than on each individual configuration, but then the following

section says that individual configurations are used. To use directions based on the individual configurations would clearly make more sense, given that we are trying to represent the outlines or curves on each object along which the landmark location ambiguities occur. Convergence is said to be assessed in terms of the sum of Procrustes distances to the mean shape, although no details are given of the level of tolerance used and it is the change in bending energy that is output at each iteration.

Once the semi-landmarks have stopped moving, the final estimated mean is then scaled to size 1 and the final Procrustes fits re-scaled by default to each have size $\cos\rho_i$, where ρ_i is the closest great circle distance between rotations of $X_i^P / \|X_i^P\|$ and unit size $\hat{\mu}$ on the pre-shape sphere (see section 3.3.2). Recall from 3.2.2. that for 2D data this additional step converts the results of GPA method (ii) or (iii) into those of method (i), obtained using the complex-eigenvector technique. Scaling the final X_i^P to size 1 is another option to give partial Procrustes fits to unit size $\hat{\mu}$.

Despite ambiguities over how exactly the iterations are performed, it was found that when using the same data as in our investigations (see 5.6), the program always produced a new set of configurations with practically identical shapes to those from our routine when using a tolerance of ε =0.001 to assess convergence. Furthermore, if GPA method (i) is selected in the argument for 'proc.opt', described above, or if the resulting configurations and mean shape from selecting from GPA method (ii) are subsequently scaled to sizes $\cos\rho_i$ and 1 respectively, then the resulting configurations have coordinates identical (to 3 decimal places), to those produced by TPSrelw.

5.5.5 Convergence issues

With a tolerance of ε =0.001, the procedure was always found to converge in usually 2–4 iterations, both with our application data of 5.6.1 (with semi-landmarks and chord directions as defined in fig. 5.9) and on various 'trial' datasets of <100 configurations, comprising shapes with a variety of combinations of landmarks, semi-landmark and chord directions, used while developing of the 'semi.it' routine. However, during

further investigation of the method, it was noticed that with smaller (stricter) values of ε, the routine frequently failed to converge.

For each dataset, it appears that there is always a tolerance level ε_B<0.001, below which the process will not converge (at least within the first few hundred iterations). Following the first few iterations (by which time $\Delta RSS(\hat{\mu}_{(r)})$ had fallen below 0.001), the values of $\Delta RSS(\hat{\mu}_{(r)})$ would then remain at the same magnitude ($>\varepsilon_B$), decreasing or, as we found with most of our examples, even increasing, only very slightly between iterations. This is illustrated in fig. 5.9, which shows the typical pattern of values of $\Delta RSS(\hat{\mu}_{(r)})$, with any of our datasets and choices of semi-landmarks and chord directions, over the first 30 iterations. The method converges at $\varepsilon=\varepsilon_A$ after just a few iterations, but with $\varepsilon=\varepsilon_B$, $\varepsilon_B<\varepsilon_A$, it does not.

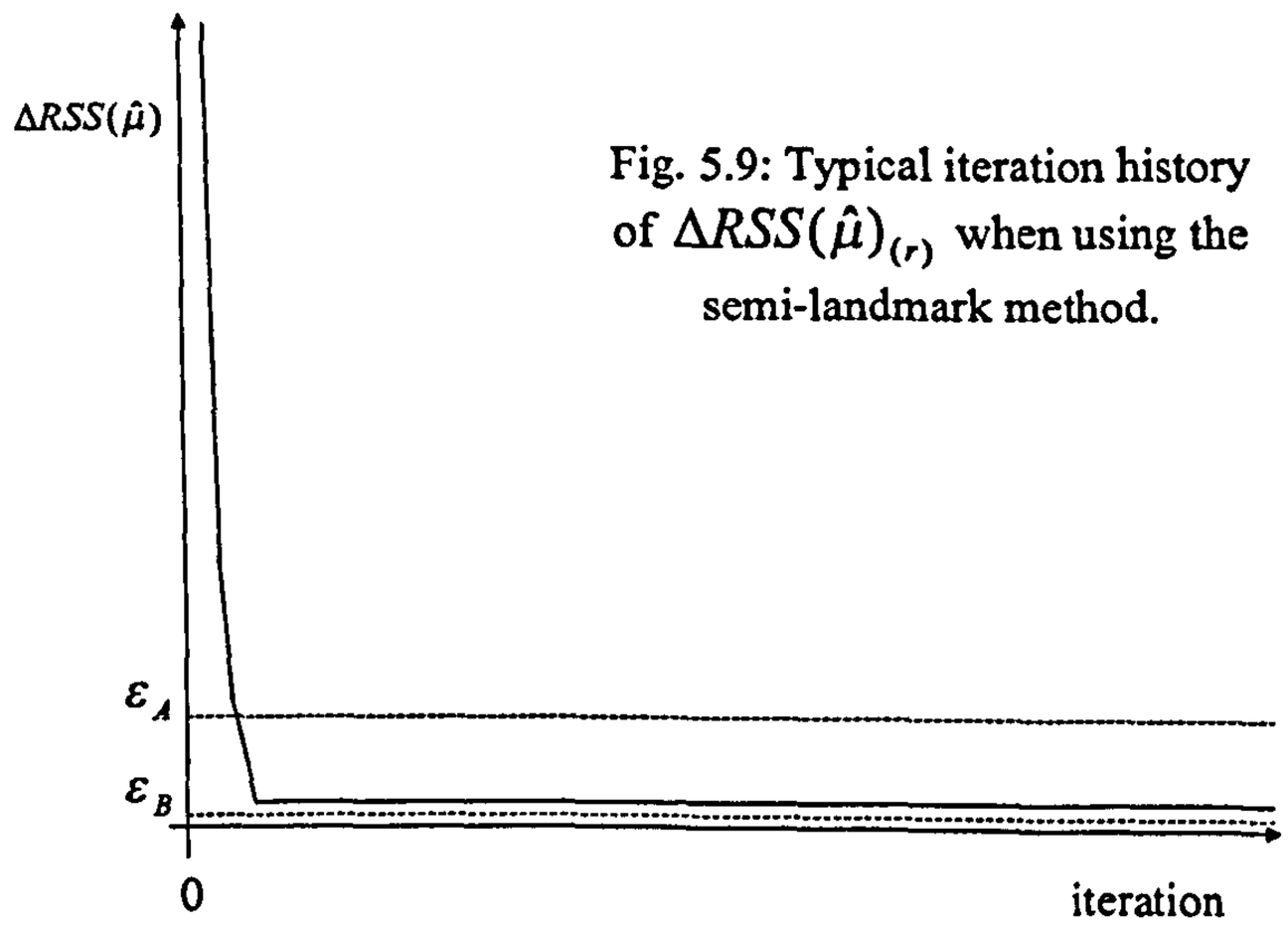


Fig. 5.9: Typical iteration history of $\Delta RSS(\hat{\mu})_{(r)}$ when using the semi-landmark method.

The problem occurs when the change in shape of the $X^P_{i_{(r-1)}}$ to $X^{new}_{i_{(r)}}$, i=1,...n, resulting from each configuration's semi-landmarks moving to new positions which minimise the BE from $\hat{\mu}_{(r-1)}$, always results in new set of configurations with new mean shape $\hat{\mu}_{(r)}$ and Procrustes fits $X^P_{i_{(r)}}$, such that,

$$\Delta RSS(\hat{\mu}_{(r)})=\sum_{i=1}^{n}\left\|X^P_{i_{(r)}}-\hat{\mu}_{(r)}\right\|-\sum_{i=1}^{n}\left\|X^P_{i_{(r-1)}}-\hat{\mu}_{(r-1)}\right\|>\varepsilon_B .$$

This is because the semi-landmarks in each $X^{P}_{i_{(r-1)}}$ do not necessarily move to positions which reduce the sample variation in shape (only bending energy). In addition, any change in their locations may result in a worse fit at the fixed and/or semi-landmarks when registered to $\hat{\mu}_{(r)}$ at the end of the iteration (compared to the previous). This is because the OLS superimposition distributes any change in the coordinates of $X^{P}_{i_{(r-1)}}$ about the residuals between all landmarks of $X^{P}_{i_{(r)}}$ and $\hat{\mu}_{(r)}$. As $\Delta RSS(\hat{\mu}_{(r)}) > \varepsilon_B$ another iteration is performed. Since the mean has changed, the $X^{P}_{i_{(r)}}$ now need to move to new positions $X^{new}_{i_{(r+1)}}$ which minimise the BE from $\hat{\mu}_{(r)}$, but once again, this results in a new set of configurations with new mean shape $\hat{\mu}_{(r+1)}$ and Procrustes fits $X^{P}_{i_{(r+1)}}$, such that $\Delta RSS(\hat{\mu}_{(r+1)}) > \varepsilon_B$.

What appeared to happen with our example datasets, was that once the process has reached the flatter part of the graph in fig. 5.9, the semi-landmarks begin to move in the same direction on each iteration, by very small, but consistent amounts. This then changes the mean, obtained at the end of the iteration, in some same systematic way, which then requires that semi-landmarks move in the same way again on the following iteration and so on, resulting in a small, but consistent increase or decrease in $RSS(\hat{\mu})$ at the end of each iteration. Any examination of the configurations after 100-200 iterations of failing to converge also revealed that if the semi-landmarks are consistently moving in the same directions like this, for so many iterations, the shapes of the configurations can become notably distorted. Note also that it makes no difference which GPA method is used, as the same shaped configurations and values of $RSS(\hat{\mu})$, on which convergence is assessed, are always produced.

The inability of the method to converge whenever ε is set small enough was discussed with both Bookstein and Rohlf during early 2003, following the appearance of the TPSrelw update to include semi-landmarks. In TPSrelw, the level of convergence is set within the program to a high enough level to ensure that this does not occur. However, Rohlf stated that he had encountered similar problems when adding the feature to his software with small, continual increases in $\Delta RSS(\hat{\mu}_{(r)})$ or very slow convergence, being a problem with most of the datasets he had tried, at very small tolerance levels. The

most recent version of the accompanying instructions to TPSrelw now states that "the amount of progress at each iteration can often be slow and the measure of convergence may sometime increase, rather than decease". For this reason, an option to specify the maximum number of iterations to be performed has also since been included (as in our 'semi.it' routine), although the level of tolerance used, still remains unchangeable, as well as undocumented.

Bookstein also confirmed that he had encountered this issue, referring to the phenomenon as "discovering the interaction between the semi-landmark mean locations and the bending energy formalism". He went on to suggest that this is less likely to happen, the higher the number of landmarks each configuration has (both semi and fixed), although no further explanation was given. Like ourselves, the issue was un-noticed at first, including when the original 1996 semi-landmark papers were published. His suggested solution was to simply stop the algorithm prior to its entry into the 'uninformative regime' or 'slow manifold', the flat region of the plot in fig 5.9, where $\Delta RSS(\hat{\mu}_{(r)})$ begins to remain unchanged. A key question is then whether or not the level of convergence achieved prior to this, is then satisfactory, i.e: when are these unchanging values of $\Delta RSS(\hat{\mu}_{(r)})$ actually small enough that they can be ignored? Bookstein stated "while he had no unifying theory on when the remaining slow manifold is negligible", he had always found the level of convergence achieved "satisfactory" for the data in question. Given the precision of the recording technique, number of landmarks and sample size, all of which affect the value of $RSS(\hat{\mu})$ and $\Delta RSS(\hat{\mu})$, we consider what might be an appropriate choice of tolerance to use for ε in section 5.5.6 below. The results lend support to the recommendation of ε=0.001 or ε=0.01, suggested by Gower (1975) and Rohlf & Slice (1990) for assessing convergence of the iterative GPA method, based on the same measure.

5.5.6 Convergence level, pixel resolution and sample size

What level of convergence actually matters in both semi-landmark and the iterative GPA method, can be related to the pixel resolution of the recorded landmark positions and number of configurations. For the Sheffield system, recall that images are typically 1000×1000 pixels, with (say) the LACC of the buccal surface of an upper central

incisor, typically being 700 pixel units. When each configuration in a sample is registered to an estimate of mean shape by GPA, a 1-unit change in the pixel coordinates of any number of landmarks will result in a change in Procrustes fit at all landmarks, measured by its contribution to $RSS(\hat{\mu})$ and this is what we consider here in establishing how small a change in $RSS(\hat{\mu})$ is actually important.

We first consider the effect of a 1 pixel unit change in the original location of a randomly selected landmark, on a randomly selected configuration from a sample of upper central incisors, when performing a GPA. Note that this may result in an increase or decrease in RSS, as the new shape may be able to be matched more or less closely to the mean of the sample, than before. Samples of two different sizes were considered, one comprising the 19 cases of one operator from the reliability data of chapter 4, the other all 76 configurations from all 4 operators. For each sample, the procedure was performed 100 times. To randomly change a landmark location by one pixel, a randomly selected change of +/-1 was made to either the x or y coordinate (also randomly selected). We then repeated the process, considering the effect on $RSS(\hat{\mu})$ of a 1 pixel unit change in the original locations of all eight landmarks, on a randomly selected configuration and then of a 1 pixel unit change on all landmarks on all 19 or 76 configurations in each sample. Again 100 sets of simulations were produced for each set of changes being investigated. The ranges and mean values of the change in *RSS* produced after performing GPA on each simulated dataset are presented in table 5.2.

	Sample size=19			Sample size=76		
Change	Range		Mean	Range		Mean
1 pixel, 1 config..	0.0000009	**0.0002**	0.00004	0.0000009	**0.0002**	0.00004
1 pixel, all configs	0.00009	**0.0003**	0.0001	0.00003	**0.001**	0.0004
8 pixels, all configs.	0.00002	**0.0008**	0.0003	0.00007	**0.005**	0.001

Table 5.2: Change in $RSS(\hat{\mu})$ by disturbing 1 or all 8 landmarks of one or all upper central incisor configurations in each sample, by 1 pixel unit. Results based on 100 simulations of each set of changes.

The results in bold indicate that a change in $RSS(\hat{\mu})$ as large as 0.005, using a sample size of 76 may result from just a 1 pixel change in each of the raw landmarks of each configuration. (Similarly a change in $RSS(\hat{\mu})$ of 0.001 can be attributed to just a 1

pixel change in the location of any one landmark on each configuration). This ties in with recommendations of Gower (1975) and Rohlf & Slice (1990) that with samples of size <100, a change in $RSS(\hat{\mu})$ of 0.001 or 0.01 is adequate for the convergence of the iterative GPA method. (Both papers actually use samples of size <50). The results above are also based on landmark coordinates being accurate to a single pixel, with no additional operator or other errors, which in practice there will be. Consequently, on sample of size <100, failure of the semi-landmark method to converge to levels of <0.001 in practice does not matter.

Note that the value and change in *RSS* depends on the number of configurations in a sample with $RSS(\hat{\mu})$ being naturally greater, the larger the sample size. When working with the change in $RSS(\hat{\mu})$ for assessing convergence, it would make sense to consider the effects of sample size when pre-specifying a tolerance level. For (say) around 50 configurations, a convergence criteria of $\Delta RSS(\hat{\mu})$<0.001 corresponds to a change in mean residual sum of squares of <0.00002. For 1000 configurations, the same mean square criteria would therefore correspond to a change in $RSS(\hat{\mu})$ of 0.02 and so on. The value of $RSS(\hat{\mu})$ will also depend on the number of landmarks in each configuration, although for the applications to be considered here, this is not expected to influence the results that much, since there are typically similar numbers of landmarks on each of the different tooth surfaces.

5.6 Application to tooth shape problems

In chapter 3, certain landmarks were identified as being difficult to locate precisely or lacking meaningful correspondence between cases along particular directions of variation. For example, notable difficulties were identified in the placement of certain points around the lower (incisal) outline of buccal tooth surfaces. These recording inconsistencies will contribute to increased variation in shape; the 'within-patient' variation inflating observed variance in these regions. In addition, recall that 'nuisance' variation results from differences between patients in the positions of their gingival margin and inter-dental papilla. While highly reproducible within cases, landmarks in

these regions lack meaningful point-to-point correspondence between cases and again produce variation which is of no interest.

Despite such problems however, these landmarks are still useful. Each is known to lie along a particular line or curve, i.e. around the lower outline of the buccal surface or along the sides of the tooth and so (in a lower dimension) still provide important information on shape in these regions, namely the relative dimensions of the teeth normal to these directions. In practice, differences in shape along these directions are of no interest but actual biological variation between patients, which is, will be perpendicular to these directions.

In this section we investigate the use of 'semi-landmarks' to address such difficulties, in the hope that we may filter out unwanted variation along these directions at the same time as we filter out differences due to size, rotation and location. For problems associated with operator inconsistency, use is made of the dataset of buccal surface configurations of upper central incisors from the preceding chapter, obtained by different operators. After choosing an appropriate set of chord directions to represent the unwanted directions of variation at the certain landmarks, the technique is applied to each of the problems described above. The resulting configurations from each application are then examined to see whether or not moving the semi-landmarks along their chords to minimise the bending energy of each configuration from the mean shape, is successful in removing the unwanted patterns of variation in these landmarks.

For the reproducibility investigation, effectiveness of the method may be judged by how much of an improvement in the reliability figures can be achieved. So that the results are comparable with those of chapter 4, here we use iterative GPA method of Gower/Rohlf & Slice (method (ii) in section 3.2.2) for obtaining the Procrustes fits and corresponding mean shape, with $RSS(\hat{\mu}) = RSS$

Following the discussion in 5.5.6, for each of the investigations, the semi-landmark process is regarded as converged, when:

$$\Delta RSS(\hat{\mu}_{(r)}) = RSS(\hat{\mu}_{(r)}) - RSS(\hat{\mu}_{(r-1)}) < \varepsilon = 0.001.$$

5.6.1 Landmarks around the lower outline

In section 4.5.3.1, a pictorial examination of the within-case variation in Procrustes fits (and operator feedback) suggested that most unwanted variation in shape resulted from inconsistencies in the positions of the mesio-distal endpoints and landmarks at the corners of the incisal edge, in directions around the lower outline of the tooth. On each configuration these directions of nuisance variation, may be represented using escribed chords directions as shown in fig 5.10.

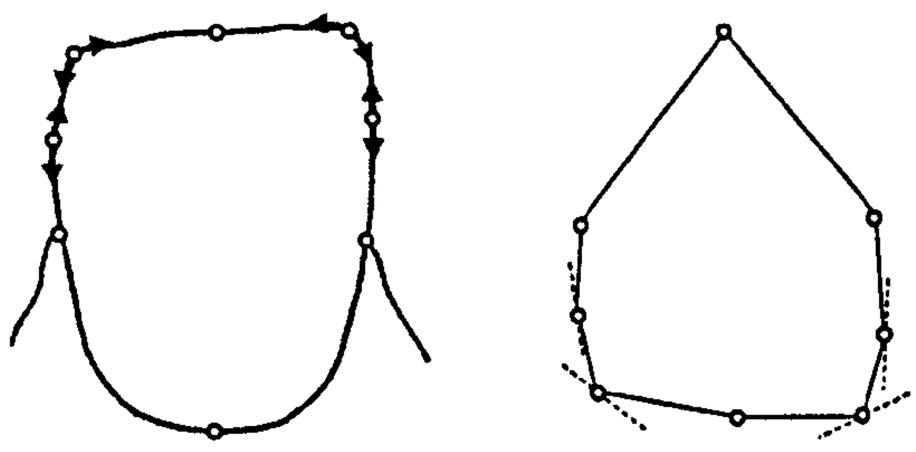


Fig 5.10: (Grey arrowed lines) Directions of landmark location inconsistencies around/along buccal outline. (Dashed lines) Unwanted variation represented as' escribed chords'; lines through each chosen semi-landmark, parallel to that joining its immediate neighbours.

At each movement step of the semi-landmark procedure, each of the four landmarks of each of the 4 operators representations of each of 19 upper central incisors from chapter 4, was allowed to move along these chords and adopt new positions which produce the PTPS mapping with least BE from the current estimate of mean shape. It is then hoped that any differences in shape due to the positions of these points along their chords is reduced or eliminated.

Since we are trying to eliminate unwanted patterns of variation due to operator inconsistency, a reduction in the 'within-case' variation in the Procrustes fits (between operators and in errors) will clearly be desirable. However, we must also be aware that by allowing landmarks to move relative to an overall mean from many different cases, this also allows variation to be removed between cases (as well as within) and doing so may be at the expense of removing important biological variation between cases in these directions. In this application this should not be a problem. Clinically, variation around the outline is of no interest but differences normal to the outline are,

representing (say) how tapered the buccal surface of the upper central incisor becomes towards the incisal edge, as is observed in hypdontia populations (see section 1.3.1.2). In general the success of using the semi-landmark technique on reliability data, will depend upon the extent to which any reduction of within-case variation is achieved at the expense of a reduction in between case or total variation and so a recalculation of the overall reliability figure for this tooth surface should enable us to establish whether or not this has been the case.

Fig 5.11 shows the original/initial and final sets of Procrustes registered configurations after using the semi-landmark procedure. As desired, the method has been successful in reducing the variation around the outline at the MD endpoints and at the corners of the incisal edge, yet variation normal to the outline appears to be retained.

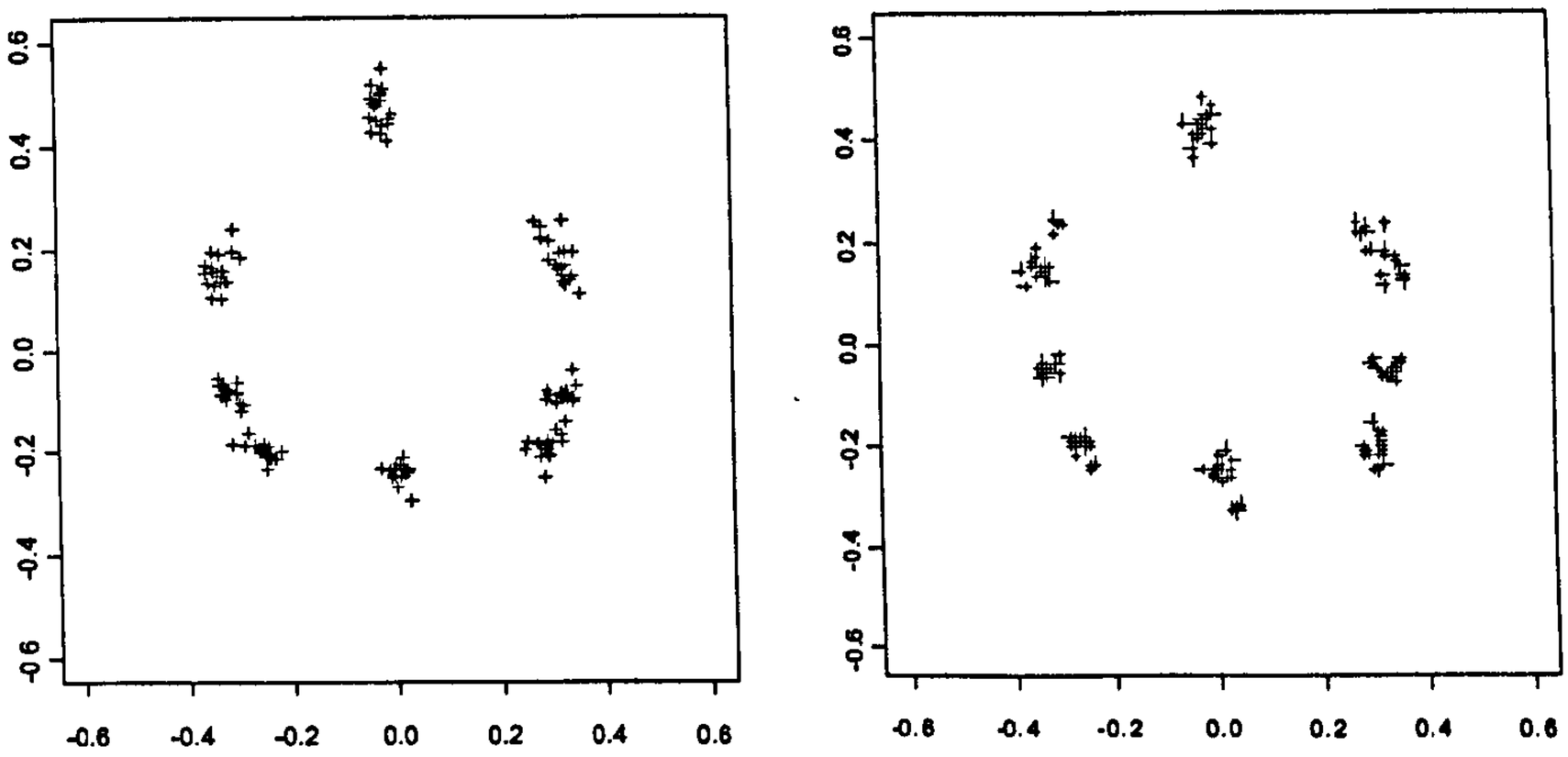


Fig. 5.11: (Left) Original Procrustes fits. (Right) Procrustes fits following semi-landmark procedure

The effectiveness of the technique may also be judged on how much of an improvement to the reliability figures of chapter 4 can be achieved. In table 4.3, the overall reliability figure $\hat{R}(C_{All})$, expressing the proportion of variation in shape, attributable to actual differences between cases, rather than operator inconsistencies was calculated using the 2-way ANOVA partition of $RSS = \sum_{i=1}^{n} \| X_i^P - \hat{\mu} \|^2$ described in 4.4.2.2. The value was found to be 0.52 (with RSS =0.92). A recalculation of the figure, based on the final Procrustes fits from the semi-landmark procedure now gives a value $\hat{R}(C_{All})$=0.70 (with RSS =0.65). Although the total variation in shape has been reduced by around a

third, the $\hat{R}(C_{All})$ values indicate that a larger proportion of variation removed was 'within cases' (between operators and errors) rather than between cases. There was also found to be no notable change in the average shape produced by the semi-landmark procedure, suggesting the absence of any biasing effects (Procrustes distance between original and final mean shapes was 0.004).

In view of these findings, the use of the semi-landmark method around the lower outline of the buccal surface appears encouraging.

5.6.2 Gingival landmarks

In addition to recording inconsistencies, other 'nuisance' variation in landmark locations results from differences between patients in the position of the gingival margin and inter-dental papilla, as described in sections 1.5.2 and 4.6. Teeth may have the same shape, but differences when represented as landmark configurations due to the extent to which the visible part of the tooth crown is obscured by the position of the gingival tissue. Such variation is again of no interest but would affect the recorded positions of the gingival margin and papilla landmarks as shown in fig 5.12 (left) below. However, as with the MD and incisal corner landmarks, the gingival landmarks are still useful, as they provide the best indication we have of the relative dimensions of the teeth in these areas and any variation normal to the sides of the teeth will still be of interest.

As in 5.6.1, the semi-landmarks and chords are chosen to represent the unwanted variation in the locations of these landmarks. However, as fig 5.10 (middle) shows, the 'escribed chord' approach is inappropriate here, since lines through any of the three points parallel to lines joining their immediate neighbours do not approximate the 'nuisance' directions of variation. Instead the directions may be represented as chords based on other landmarks in the configuration (see fig. 5.12 (right)). For the central gingival margin landmark, at the end of the long axis of the clinical crown (LACC), the direction of the LACC itself may be used to represent the direction of vertical variation in its position. For the papillae landmarks, chords from the mesio and distal endpoints of the MD width through these locations can be used to approximate directions of

variation up and down the sides of the teeth. (We assume that the sides of the teeth continue to follow the direction of the chords beneath the papillae).

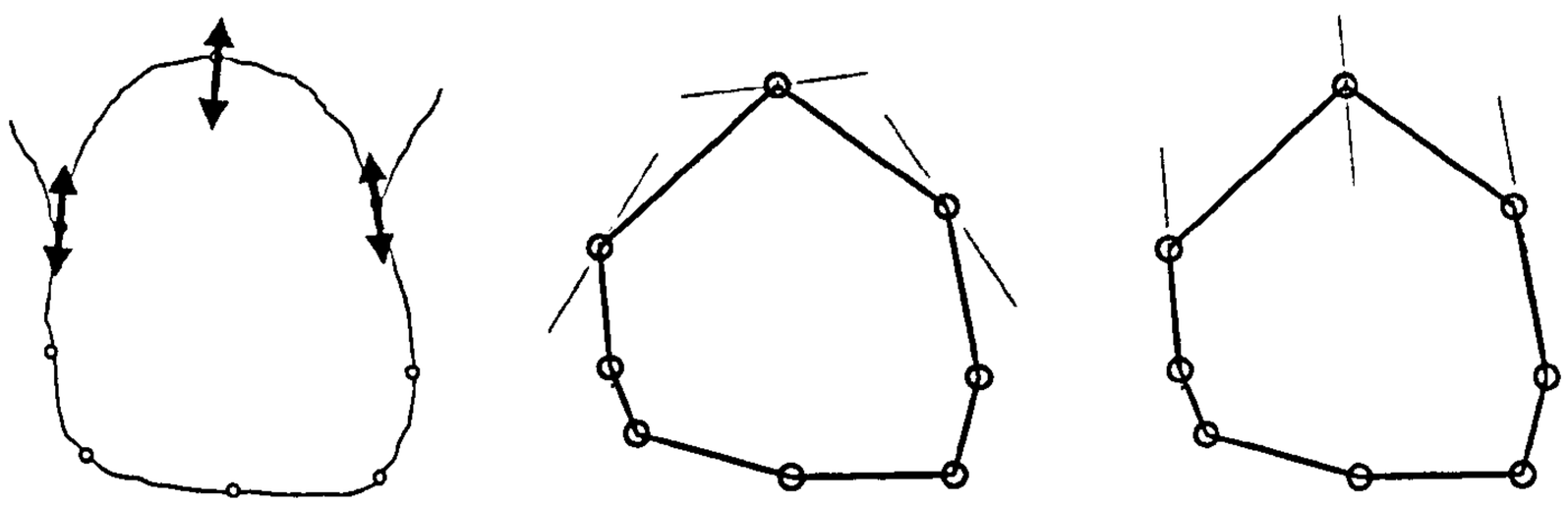


Fig 5.12 (Left) Nuisance directions of variation resulting from differences between patients in the position of their gingival tissue. (Middle) Escribed chords through gum landmarks. (Right) Representation of variation in left panel as chords through gingival landmarks, based on shape of sides of tooth above MD and on the LACC.

In addition to registration differences we then also consider variation in the gingival landmark positions, along these directions, as an additional equivalence class to be eliminated by the semi-landmark procedure. Note that any such variation is only present 'between cases', since the data were obtained from stone casts of patients teeth. A re-assessment of the reliability figures is not regarded as appropriate here, since the multiple representations by different operators carry the same information on gum position, since they were obtained from the same patients. Any variation in the gingival landmarks 'within cases' would be due only to inconsistencies in their location by different operators and this was found to be negligible. Operators reported no difficulties in identifying gingival landmark positions and the analysis of 4.5.3.1 revealed nothing to suggest that these landmarks were a particular problem to identify consistently. For this reason, we consider only one operator's representations of each of the 19 cases to investigate the semi-landmark method.

Following the iterative procedure, we would then hope that any variation between the configurations in the positions of the gingival landmarks, along these directions will have been be removed. However, after the first iteration, the variation in shape had more than doubled from RSS =0.157 to 0.396 and so the process was stopped. By allowing the gum landmarks to move in directions corresponding to gingival variation, unrealistic shapes were produced for many configurations, as can be seen in fig. 5.13.

This shows the original and new semi-landmark positions for two upper central incisors, with chord directions as defined in fig. 5.12(right), following the first movement step.

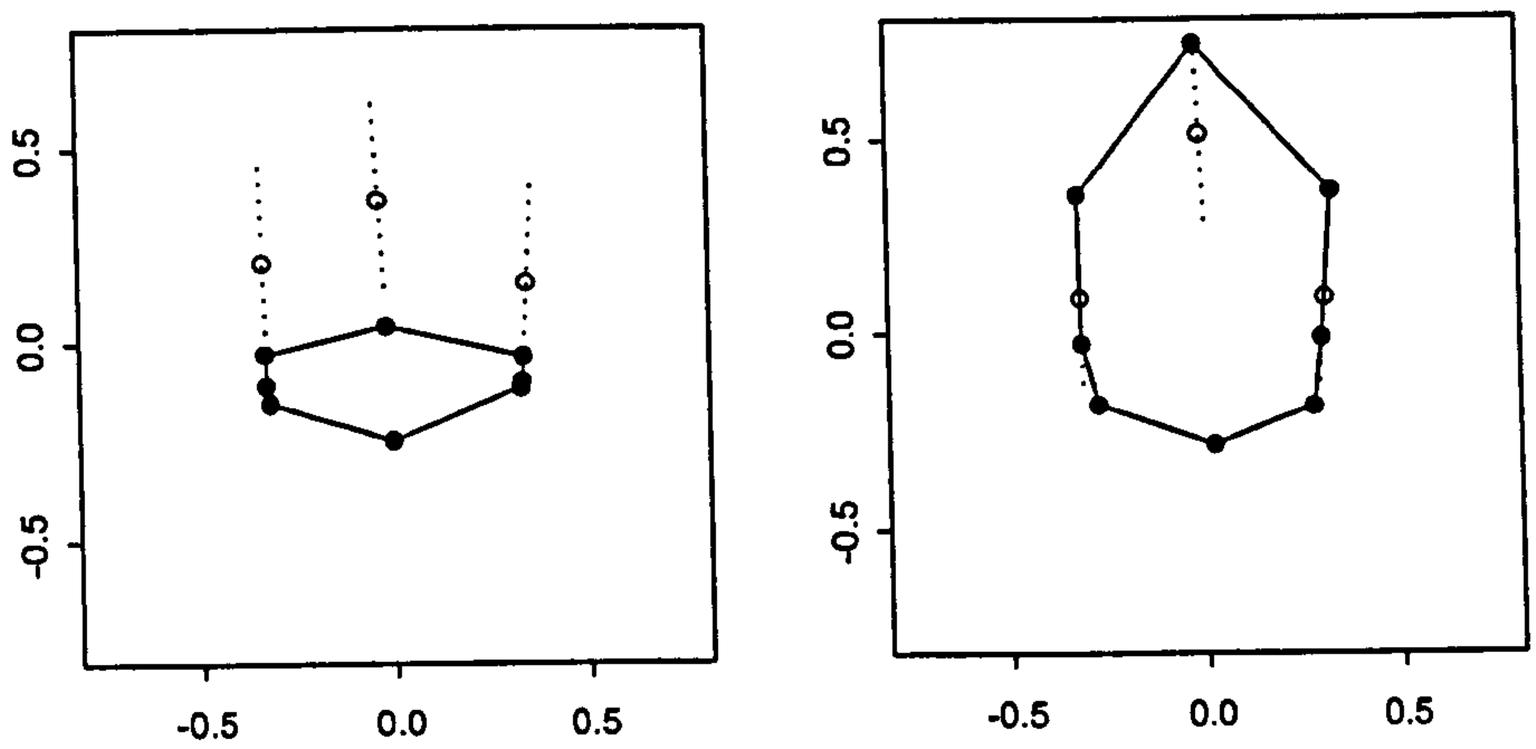


Fig 5.13: Original (o) (dashed line) and new (•) (solid line) configurations for two individual central incisors following first iteration.

The increased variation in shape between configurations, along these directions, can be see in the plots of the original and new Procrustes fits in fig. 5.14 below.

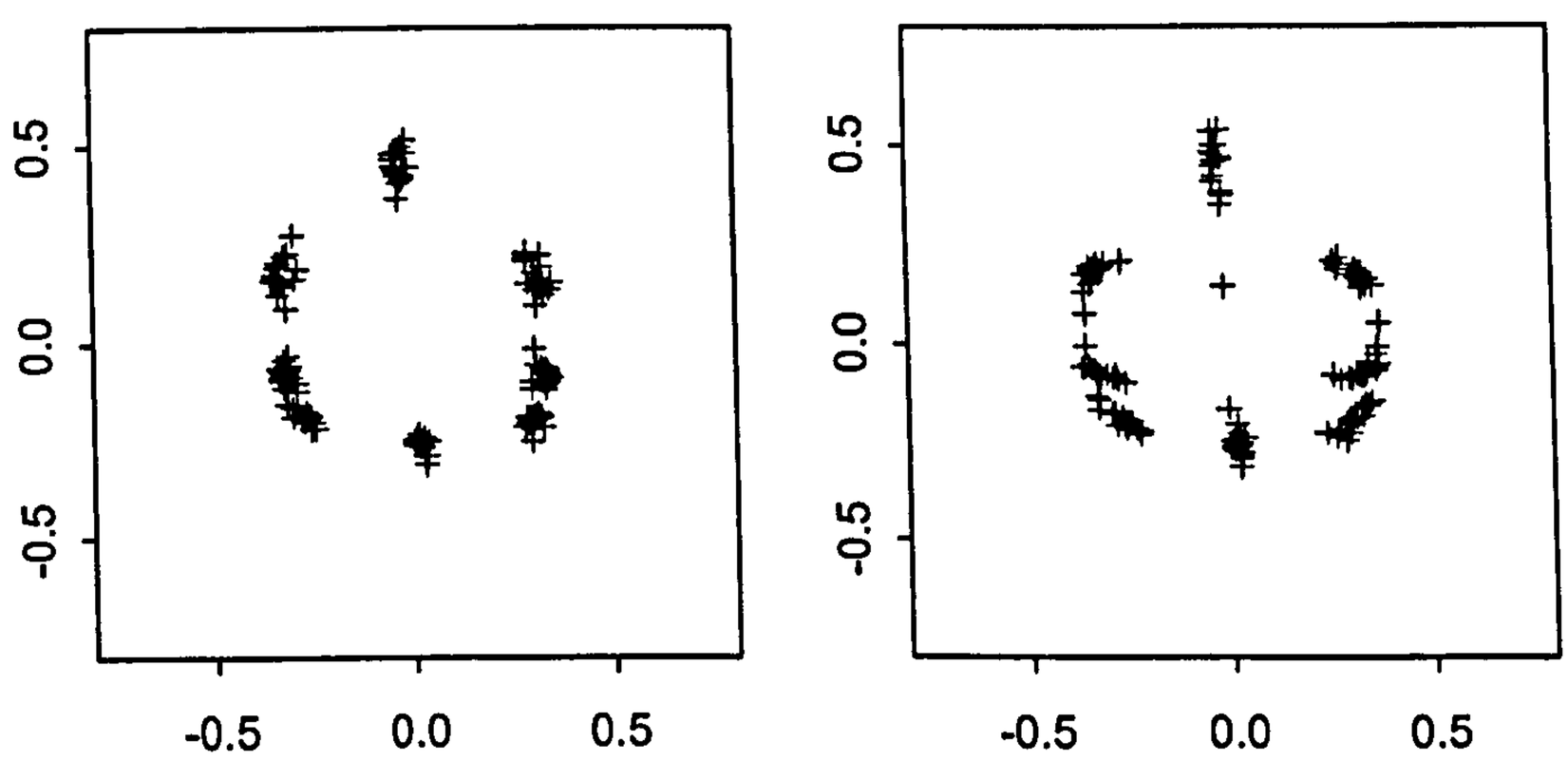


Fig. 5.14: (Left) Original Procrustes fits. (Right) Fits after first iteration of semi-landmark procedure.

5.6.3 Further investigation: Why large shape changes occur

The use of semi-landmarks is clearly not a suitable approach to the gingival variation problem. In this section we investigate which particular aspects of the minimum bending energy criterion gives rise to the unusual shapes in fig. 5.13 and identify other potential situations where the method breaks down. In 5.7 we then discuss possible

ways in which the method can be modified or developed to address some of these difficulties, before investigating these ideas in greater detail in chapter 6.

Consider one of the cases where the method produced an unusual configuration at the end of the first iteration (the configuration in the left of fig 5.13 above). Let $Y^0 = X^P$ be the original (Procrustes registered) version of this configuration and $T = \hat{\mu}$ the initial Procrustes mean from the same sample. The PTPS mapping for transforming $T = \hat{\mu}$ to $Y^0 = X^P$ is defined by equation (5.33) as $Q\Gamma^{21}X^P + S\Gamma^{11}X^P$, where Q, S, Γ^{21} and Γ^{11} (each dependent only on T) are as defined in equations (5.31) and (5.32).

Fig 5.15(a)(left and middle) displays the PTPS transformation from $T = \hat{\mu}$ to $Y^0 = X^P$, as a deformation grid, indicating how a square grid placed over $\hat{\mu}$ is deformed so that the landmarks are mapped exactly to those of X^P. Recall from 5.3.3 that in order to minimise the bending energy of the mapping from T to Y^0, the PTPS try to represent the transformation as 'globally' as possible using the affine component, so that there is the smallest possible variation in the shape of the grid cells. Recall also that minimisation of BE can also be thought of as finding the GLS affine fit of $T = \hat{\mu}$ to $Y^0 = X^P$, with error covariance matrix given by $(I_2 \otimes S)$. Here the 'affine fit', given by $Q\Gamma^{21}X^P$, is displayed in fig 5.15(a)(right) with crosses indicating fitted affine points to the fixed target landmarks in X^P. Where the affine fits do not match their targets, residual non-affine components, given by $S\Gamma^{11}X^P$, are required in addition to the affine part, so that $\hat{\mu}$ is mapped to X^P exactly. These produce the variations in the shape of the grid cells in fig 5.15(a)(middle). The sum of these squared 'local' components then make up the bending energy of the transformation.

Now consider the same two configurations when some of the landmarks in X^P (the gingival landmarks) are free to move along directions as defined in fig. 5.12, but the remaining landmarks are still required to be matched exactly. Fig 5.15(b)(left and middle) shows the new PTPS transformation from $T = \hat{\mu}$ to $Y^{new} = vec^{-1}(vec(X^P - U\lambda)) = X^{new}$, which has minimum BE of all PTPS

transformations, as the semi-landmarks vary by distances λ along chord directions defined by U. Fig 5.15(b) (right) shows the affine component of this mapping as a deformation grid and the 'affine fit' of $Q\Gamma^{21}\ X^{new}$ to X^{new}.

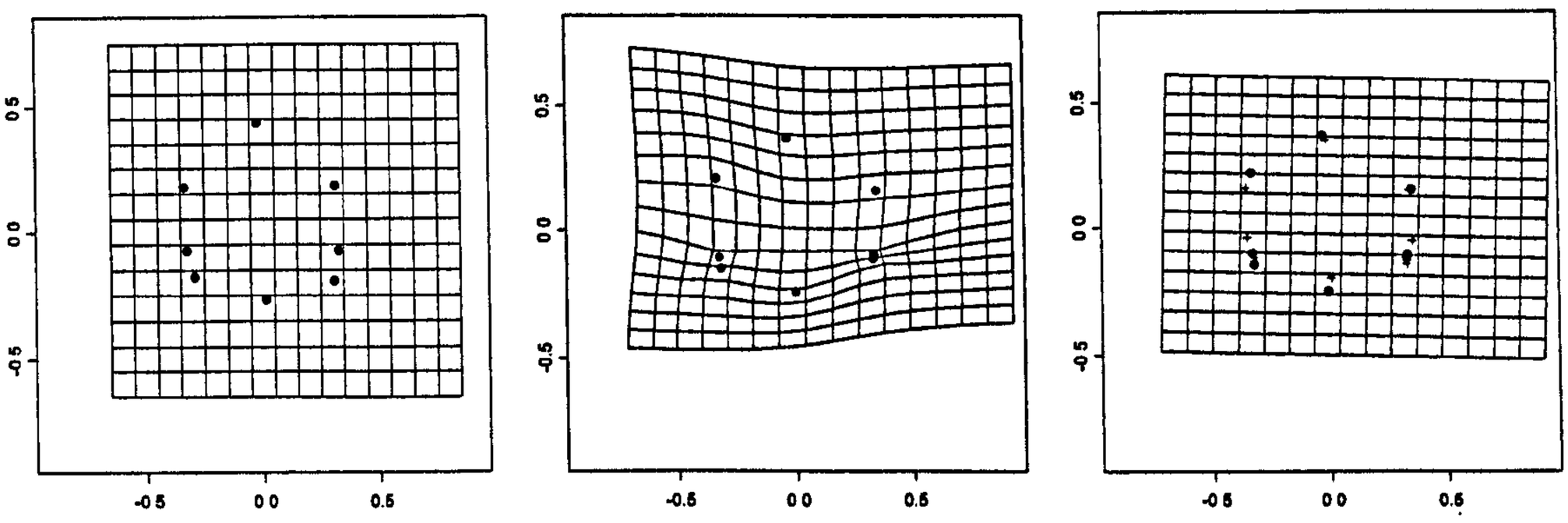


Fig 5.15 (a): (Left) Square grid placed over initial Procrustes average $\hat{\mu}$. (Middle) PTPS transformation from $\hat{\mu}$ to Procrustes registered configuration X^P (Right) Affine component of the transformation. (Crosses) indicates the fitted affine points, given by $Q\Gamma^{21}X^P$, to landmarks (filled circles) of X^P.

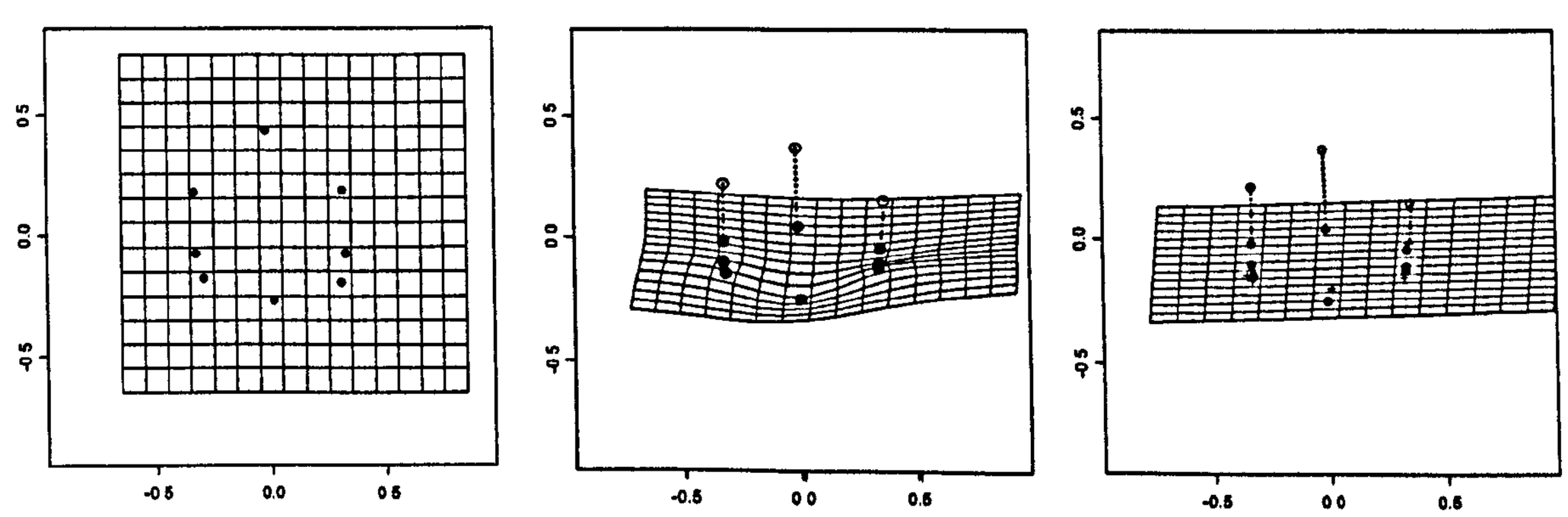


Fig 5.15 (b): (Left) Square grid placed over $\hat{\mu}$ as in (a) above. (Middle) Minimum BE PTPS transformation from $\hat{\mu}$ to $X^{new} = vec^{-1}(vec(X^P - U\lambda))$. (Right) Affine component. (Crosses) indicates the fitted affine points, given by $Q\Gamma^{21}X^{new}$, to landmarks (filled circles) of X^{new}. (Empty circles) nominal positions of semi-landmarks, as in corresponding panels of (a) above.

In order to produce the minimum BE PTPS transformation from $\hat{\mu}$ to X^{new}, the semi-landmarks move to positions which result in the mapping which gives the closest possible 'affine fit' (in terms of the sum of squared generalised residuals or non-affine components) of $\hat{\mu}$ to the fixed landmarks and positions along the semi-landmark's

chords in X^{new} (see fig. 5.15(b)(right)). In other words, they move so that there is the smallest possible variation in the shape of the grid cells in fig 5.15(b)(middle) when the corresponding non-affine or local components, $S\Gamma^{11}X^{new}$ are added to $Q\Gamma^{21}X^{new}$, to ensure that points in $\hat{\mu}$ are mapped exactly to the fixed landmarks and chord positions in X^{new}. By allowing some of the landmarks to move, the shape of the grid cells in the target plane of $X^{new}=vec^{-1}(vec(X^P-U\lambda))$ need not vary as much as they did in the original PTPS mapping from $\hat{\mu}$ to X^P. A new, 'better fitting' affine transformation of $\hat{\mu}$, in that it has a smaller sum of squared local residual components (bending energy), is always able to be found.

In this example, the affine transformation simultaneously attempts to fit to both the fixed landmarks around the lower (incisal) outline and to any positions on the semi-landmarks chords, representing possible positions of the three gingival landmarks. A 'better fitting' affine component for mapping the lower fixed landmarks of $\hat{\mu}$ to those of X^{new} is able to be cast across the entire plane in which X^{new} lies since the semi-landmarks can adopt positions on their chords in accordance with this affine mapping. Specifically, the vertical distances between the MD endpoints and landmarks along the incisal edge are relatively larger in $\hat{\mu}$, when considered alongside the other vertical distances between landmarks, than in X^P. Consequently, since the semi-landmarks are able to move in a predominantly vertical direction, the vertical affine component shrinks $\hat{\mu}$ vertically to minimise the GLS residuals (sum of squared non-affine components) between $Q\Gamma^{21}X^{new}$ and the fixed landmarks and chord positions in X^{new}, resulting in a downward (incisal) movement of the gingival landmarks from their nominal positions.

However, is it because the semi-landmarks end up moving such large distances that the configurations produced in fig 5.12 end up as unrealistic representations of shape. In fig. 5.12 (right) positions are adopted that are higher (cervical) than where the top of the clinical crown, hidden beneath the gum, would be. (This was observed for several other cases in the sample). However, these unlikely shapes are entirely possible within the definition of how new semi-landmark positions are determined, since there are no constraints placed on the extent of movement of the semi-landmarks in either direction.

In any application, the only constraints as such are determined by the locations of the fixed landmarks and/or the directions in which the semi-landmarks are able to move.

In general, when all semi-landmarks are free to move in roughly the same direction (or in accordance with a similarity transformation), the determining factor for where the landmarks move to is the PTPS mapping between the fixed landmarks (in these directions). In such situations, large movements of the semi-landmarks will occur when a notable change in the affine component of the PTPS mapping from the mean shape, (determined by the mapping between the fixed landmarks), is able to occur in the directions of the semi-landmarks chords, and there are no nearby fixed landmarks or other semi-landmarks whose movement is restricted to more perpendicular directions, to stop this occurring. The actual distance moved will then depend on how much of a change in the affine component (cast across the entire plane of Y^{new}) occurs. Conversely, when there are nearby fixed landmarks or differences in chord directions result in certain semi-landmarks being effectively fixed in the direction of others semi-landmarks' chords, the affine component is less able to change, since the PTPS mapping still has to map to positions along these chord directions with as small as set of local non-affine components as possible. This restricts the ability of the method to produce such large movements. As was seen in fig 5.15(b), in section 5.6.2, the gum landmarks are able to move in the same roughly parallel vertical directions and in accordance with the affine mapping for the fixed landmarks in these directions, as there are no other fixed or semi-landmarks with more perpendicular chord directions nearby in this region, to influence the affine component.

As another illustration, fig. 5.16 considers the mapping of just the *y*-coordinates of $T = \hat{\mu}$ to those of $Y^{new} = vec^{-1}(vec(X^P - U\lambda)) = X^{new}$ for the example in fig. 5.15 and 5.13(left). The display shows the (linear) part of the affine component of the PTPS mapping for transforming the *y*-coordinates, ignoring any dependency on the *x*-coordinates for now. The plot takes the appearance of the linear component of a cubic spline, as in the previous illustration of a PTPS mapping in fig. 5.3. Fig. 5.16(left) shows the original mapping, with all landmarks fixed. When the three gingival landmarks are allowed to move in the y direction, an improved linear fit can be achieved to all *y*-coordinates of the landmarks, determined by the fit now achievable at the lower

(incisally located) fixed landmarks. See fig 5.16(middle). This produces the large movement in the *y*-direction of the three semi-landmarks. The right panel of fig. 5.15 shows what would happen if the movement of one of the three semi-landmarks (say the one on the gingival margin), is either fixed or restricted to only move in the *x*-direction. The linear part of the affine component for transforming the y-coordinates is now not so different from in the original mapping on the left and so the semi-landmarks do not move by as great a distance. Note that because of the GLS criteria, the semi-landmarks would not move exactly onto the line.

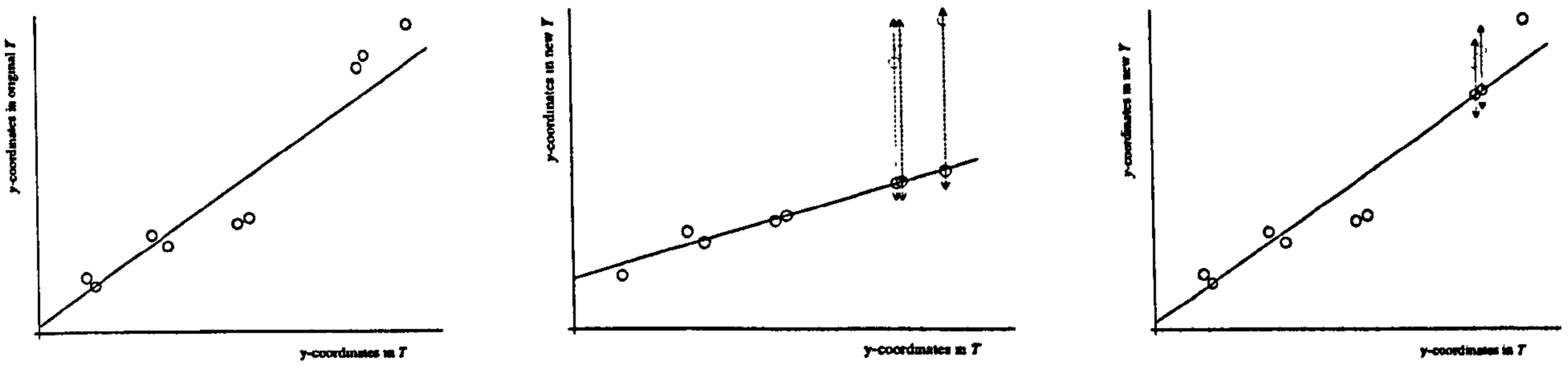


Fig 5.15: Linear part of the affine component for transforming *y* coordinates from configuration *T* to *Y*. (Left) When all landmarks in target configuration *Y* are fixed. (Middle) When gingival landmarks allowed to move along chords in *y*-direction. (Right) If one of the semi-landmarks fixed or constrained to only move in the x-direction.

In the application of 5.6.1 there is a fixed landmark in the vicinity of the semi-landmarks around the lower outline of the tooth surface (at the bottom of the LACC) and the different chord directions of the MD endpoints and incisal edge corners are always at an angle of between 90 and 135 degrees, which must be matched onto exactly by the PTPS transformation from $\hat{\mu}$ to $X^{new} = vec^{-1}(vec(X^P - U\lambda))$. This means that the shape of the configuration cannot change too much, whereas when the gingival landmarks are semi-landmarks, there are no fixed landmarks or differences in the chord directions to constrain the large change in the affine component of the PTPS mapping and stop the shapes 'collapsing' or 'stretching' as seen in fig 5.13.

Similarly, in Bookstein's (1996a,d,e) original application, large changes in shape were not a problem since the choice of directions attached to each landmark restricted this possibility. Each landmark was allowed to move along its escribed chord, representing

the outline of the corpus callosum. When escribed chords are used for landmarks around the outline of an object, as in Bookstein's example and 5.6.1, large changes in shape cannot occur because the points are effectively constrained to stay around the outline and the affine component (which produces most of the change in shape) is not really able to change, except perhaps by a rotation.

5.7 Discussion

In the application to gum landmarks the minimum bending energy criterion is clearly not the best option for removing unwanted variation in shape, because large unrealistic movements of the semi-landmarks to unrealistic positions can result from large changes in the affine component of the PTPS transformation from $\hat{\mu}$. There may be many applications where non-affine similarities are to be regarded as part of the same equivalence class as rotation, location and scale and so removing any shape differences of this form, while allowing for affine shape changes during the registration, may have its uses. However, if we were to ignore affine differences in shape in the study of tooth shape, this would exclude many interesting and important dissimilarities which may occur.

Although the technique has been found to be useful in removing unwanted variation in shape when landmarks are free to move around the outline of the object, clearly the semi-landmark method is in its current form is not the best way of addressing problems where the variation is to be removed is perpendicular to the outline of an object, with nothing to constrain the extent of movement which may occur on any iteration.

In the following chapter we consider modifications to Bookstein's technique and look to develop a methodology suitable for removing patterns of variation in directions other than around the outline, our main focus being on unwanted landmark variation resulting from differences between patients in the position of their gingival tissue. Since there is nothing in the technique itself or in the necessary choice of chords' directions for these semi-landmarks to constrain the affine part of the (minimum BE) PTPS transformation or movements resulting from it, it seems natural to focus our investigations on ways in which the technique can be modified to do this.

In sections 5.3.5.3 and 5.4.1.4 links of the PTPS mapping with GLS were investigated and in chapter 6 we again look to links with other topics and ideas related to the spline methodology, which may provide a basis for further exploration or development.

Large unrealistic movements of semi-landmarks, producing unnatural shapes, were seen to occur on configurations where the minimum bending energy PTPS mapping from $T=\hat{\mu}$ to $Y^{new}= vec^{-1}(vec(X^P)-U\lambda)=X^{new}$ consists of an affine transformation, which produces a large affine change in the shape of $\hat{\mu}$. We therefore could consider ways of penalising movements of the semi-landmarks which require such a change in the shape of $\hat{\mu}$, for example, by constraining the parameters of the affine mapping from $T=\hat{\mu}$ to $Y=Y^{new}$ in some way, or developing a measure of the extent of stretching/collapsing produced by the mapping, which may then be penalised.

In addition, there may be other possible transformations and optimisation criteria, other than using a PTPS mapping and bending energy, that we could consider for determining the new positions of semi-landmarks, when matching $T=\hat{\mu}$ to $Y = vec^{-1}(vec(Y^0)-U\lambda)$ (or *vice versa*). Here we have seen how the minimum BE criteria often produces movements of the semi-landmarks in directions away from what would be expected if variation were to be reduced about the mean shape and so other options may prove more to be useful in this regard.

We also consider the suitability of the other existing variants of Procrustes and spline superimposition (other than Bookstein's semi-landmark method) that have been proposed for use when there is lack of precise correspondence of landmarks along particular directions or where one would wish to account for the effects of certain unwanted differences in shape in some way.

Chapter 6

Proposal and evaluation of new semi-landmark methods

6.1 Introduction

In the previous chapter, it was seen that extension of the Procrustes method to an iterative process, where certain landmarks on each configuration were allowed to move along pre-specified chord directions, failed to provide a method of registration which accounted for the unwanted patterns of variation due to differences in the locations of patients' gingival landmarks. Allowing each set of these landmarks to move to new positions determined by the minimum bending energy pairs of thin plate splines mapping from the current estimate of mean shape, it was found that large movements often occurred in certain configurations, producing very unrealistic shapes.

In this chapter we consider novel ways in which the semi-landmark and/or registration procedure can be modified to stop the 'collapsing/stretching' problem described in 5.6.3. In particular we examine other possible mappings and objective functions between the current mean and each individual configuration that may be optimised when determining the new positions of the semi-landmarks.

We start in section 6.2 by considering possible modifications to the PTPS transformation used in Bookstein's original (1996a,d,e) semi-landmark method, investigating the use of different possible mappings and penalty functions other than bending energy. We consider how pairs of smoothing splines may be used to prohibit large movements of the semi-landmarks away from their original positions, and discuss the extension of the smoothing spline by Rohr and co-workers (Rohr *et al.*, 1997, Rohr, 1998) to include anisotropic errors, which can be shown to offer a generalisation of the semi-landmark method. We also consider the use of higher order splines and their

associated roughness penalties and mappings which may be produced by kriging, a method of spatial prediction, of which splines can be shown to be a special case. When modified to include semi-landmarks, this allows use of a more general class of polynomial functions in the 'global' part of the mapping and other possible generalised least squares penalty functions. Finally in this section, we consider different ways in which we can place constraints on or penalise the influence of the affine component, when determining the optimal PTPS mapping to new semi-landmark positions, so that we may restrict the possibility of unrealistic changes in shape being produced. Recall that large movements of the semi-landmarks were seen to occur in instances when the change in optimal PTPS mapping from the mean to the new configuration, consisted of a large change in the affine component of the transformation.

Section 6.3 then considers other transformations and optimisation criteria which are not spline mappings, such as the minimisation of the full Procrustes distance between the configurations as the semi-landmarks move along their chords, or simply moving the semi-landmarks to the nearest point along their chords to the corresponding landmarks in the mean shape (following suitable prior alignment).

Before summarising and assessing which of the new methods are to be considered further in section 6.5, section 6.4 reviews and discusses some of the other novel variants of Procrustes and spline superimposition techniques that have previously been proposed to allow for lack of precise correspondence of landmarks along particular directions or where one would wish account for certain unwanted differences in shape in some standardised way. The ideas of weighted (generalised) Procrustes superimposition are presented, here using pre-specified covariance functions to reflect the relative accuracy of the different landmarks, along with Green's (1996) extension of the semi-landmark technique to include curves and outlines.

6.2 Modifications of the PTPS mapping

In this first section we consider how the pair of thin-plate splines (PTPS) formulation and/or its associated penalty function, the bending energy (BE), may be modified or

extended to allow alternative choices of optimisation criteria and mappings to be used for the determination of new semi-landmark positions along their chords.

We start by investigating the use of existing variants of interpolating splines, such as smoothing splines, higher order splines and kriging predictors, all of which may also be used for plane-to-plane mappings, in the same way as a PTPS transformation. As part of the section on (pairs of) smoothing splines, we describe Rohr's (1998) extension to include anisotropic covariance matrices at each of the landmark locations and show how the original semi-landmark method may be considered as a special case of this approach. After describing the details of each of the techniques, we illustrate how the formulation may be modified to include semi-landmarks in the 'target' configuration and show how optimisation of each method's associated objective function, leads to a set of equations and solution of the same form as the original semi-landmark method.

We then propose and evaluate different possible penalties and constraints, aimed at controlling the influence of the affine component when determining the optimal PTPS mapping from configuration T to the target configuration $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$. (In practice $Y^0 = X^P$, the full Procrustes fit to mean shape $T = \hat{\mu}$). As was seen in 5.6.3, it is the fact that affine component can produce large 'stretching/collapsing' effects when minimising the BE of the transformation of T to Y^{new} that often leads to the semi-landmarks adopting unrealistic positions. By developing suitable measures, these may then be included as part of the objective function, so that movements resulting from mappings comprising large affine changes of T are penalised.

6.2.1 Penalty functions based on smoothing splines

In addition to data interpolation, the other main use of spline functions is in smoothing. As well as a pair of interpolating splines, Dryden & Mardia (1998) considered how a pair of ordinary smoothing splines may be used to produce a mapping from one configuration to another in situations where it is not essential that the landmarks be matched exactly. Here a function comprising the residual sum of squares between the mapping of T and Y as well as a roughness penalty (typically the BE of the

transformation) is optimised, allowing the amount of bending to be penalised relative to the degree of fit achieved.

However, if T is to be mapped by a pair of smoothing splines to a configuration of the form $vec^{-1}(vec(Y^0)-U\lambda)$, then as λ varies, there is no constraint that points in T will be mapped to corresponding positions of the fixed landmarks or along the semi-landmarks chords exactly, unless we place restrictions on the form of the resulting configurations. (That the configuration of the more reliable, fixed landmarks is retained and that the semi-landmarks may only move along directions of unwanted variation corresponding to those we specify, is one of the aspects of the semi-landmark method we would wish to retain). This then allows pre-specification of a parameter, which penalises the total sum of squared movements of the semi-landmarks along their chords relative to the bending energy of the mapping. The idea of penalising the amount of movement of landmarks is appealing since we may then be able to restrict the large movements of semi-landmarks found to occur when minimising the BE alone. We then extend the method to allow different penalty weights to be attached to the movement of different semi-landmarks.

Alternatively, we may make use of the modification to the (pair of) smoothing splines mapping developed by Rohr (1998), Rohr *et al.* (1997, 1999). Instead of assuming that the error covariance at all of the landmarks is isotropic (as in the standard formulation), Rohr and co-workers extended the technique to allow specification of different covariance structures at each of the landmarks. This offers a generalisation of Bookstein's semi-landmark method to allow movement within ellipses of different shape and size around each landmark, rather than along straight lines (although chords may also be represented as ellipses stretching to infinity in one direction and zero in the other). Consequently, with suitable choices of covariance matrices it is also possible to use this technique to produce a mapping where the positions of the fixed landmarks are retained and the new positions of the semi-landmarks are still along their chords, but now with the ability to penalise the different amount of movement of each of the semi-landmarks. However, as we show in 6.2.1.3, the form of solution based on this method is much less simple to work with than that which we derive directly.

6.2.1.1 Ordinary smoothing spline

The ordinary (isotropic) 'smoothing' or 'approximation' spline was considered by Duchon (1976) and Wahba (1990), and its uses in various applications such as regression analysis have been considered by, for example, Green & Silverman (1991). In keeping with section 5.2, we first describe the general smoothing spline for configurations T and Y, comprising data points t_j and y_j, j=1,...,k, in m-dimensions.

If $\Phi(t)$ is some (multivariate) function defined in m-dimensions, with smoothness index (order of derivatives) r and $\alpha>0$ some 'smoothing' parameter, then instead of equation (5.15) the objective function to be minimised is:

$$\frac{1}{k}\sum_{j=1}^{k}(y_i-\Phi(t_j))^T(y_j-\Phi(t_j))+\alpha J_r^m(\Phi) \qquad (6.1)$$

combining an ordinary (unweighted) sum of squares term measuring lack of fit between the fitted values $\Phi(t_j)$ and data points y_{ji} and a roughness penalty $J_r^m(\Phi)$, measuring the 'roughness' of the transformation as defined in (5.1).

If we write $\Phi(t_j)=z_j$, j=1,...,k say, then in general $z_j \neq y_j$. Since Φ was free to be any function, with square integrable rth order derivatives, then by the same argument as in section 5.2.2, the optimal choice of function would be a set of m spline mappings (one for each dimension) given by $\Phi(t)=(\Phi_1(t),..,\Phi_m(t))^T$, mapping the t_j to z_j. From equation (5.19), the multiple spline mapping of a $k\times m$ matrix of points $T=(t_1^T,...,t_k^T)$ to a $k\times m$ matrix of points $Z=(z_1^T,...,z_k^T)$ is given by:

$$(\Phi(t_1)^T,...,\Phi(t_k)^T)^T=Z=QB+SW \qquad (6.2)$$

where Q and S are as defined in (5.19) with columns $(\phi_1(t_j),...,\phi_M(t_j))^T$ and $s(t_j)$ $=(\sigma_{r,m}(t_j,t_1),...,\sigma_{r,m}(t_j,t_k))^T$ respectively, and where

$$B = \Gamma_S^{21} Z \text{, with } \Gamma_S^{21} = (Q^T S^{-1} Q)^{-1} Q^T S^{-1} \tag{6.3}$$

$$W = \Gamma_S^{11} Z \text{, with } \Gamma_S^{11} = S^{-1} - S^{-1} Q (Q^T S^{-1} Q)^{-1} Q^T S^{-1}, \tag{6.4}$$

therefore:

$$Z = Q\Gamma_S^{21} Z + \Gamma_S^{11} ZS . \tag{6.5}$$

Recall that Γ_S^{21} and Γ_S^{11} depend only on Q and S (which depend only on T), whereas B and W depend on the points Z being mapped to. Here we have denoted Γ^{21} and Γ^{11} as Γ_S^{21} and Γ_S^{11} to indicate that they were determined using matrix S (see below).

To determine Z, equation (6.1) can be re-phrased (for any m) as:

$$\min_Z \text{ trace}((Y-Z)^T (Y-Z)) + \alpha \text{ trace}(Z^T \Gamma_S^{11} Z)$$

which is achieved for $Z = (I + \alpha\Gamma_S^{11})^{-1} Y$. Substituting this value back into (6.3) to (6.5) then gives the parameters of the interpolating spline mapping to Z, i.e.

$$B = \Gamma_S^{21} (I + \alpha\Gamma_S^{11})^{-1} Y \tag{6.6}$$

$$W = \Gamma_S^{11} (I + \alpha\Gamma_S^{11})^{-1} Y \tag{6.7}$$

with

$$\Phi(t) = (\phi_1(t), \ldots, \phi_M(t))^T B + s(t) W$$

and

$$Z = Q\Gamma_S^{21} (I + \alpha\Gamma_S^{11})^{-1} Y + \Gamma_S^{11} (I + \alpha\Gamma_S^{11})^{-1} YS . \tag{6.8}$$

However, by expanding (6.8), using (6.3) and (6.4), it can be shown that this is of the form of an interpolating spline from T to Y with matrix $(S + \alpha I)$ instead of S and so by analogy with equations (5.19) and (5.20), the parameters B and W of the smoothing spline may also be computed by solving:

$$Y = QB + (S + \alpha I)W \tag{6.9}$$

$$Q^T W = 0. \tag{6.10}$$

This gives alternative expressions for parameters B and W in smoothing spline $\Phi(t)$, as:

$$B=\Gamma^{21}_{S+\alpha I}Y,\ \Gamma^{21}_{S+\alpha I}=(Q^T(S+\alpha I)^{-1}Q)^{-1}Q^T(S+\alpha I)^{-1} \tag{6.11}$$

$$W=\Gamma^{11}_{S+\alpha I}Y,\ \Gamma^{11}_{S+\alpha I}=(S+\alpha I)^{-1}-(S+\alpha I)^{-1}Q\,(Q^T(S+\alpha I)^{-1}Q)^{-1}Q^T(S+\alpha I)^{-1} \tag{6.12}$$

with
$$Z=(I+\alpha\Gamma^{11}_S)^{-1}Y=QB+SW=Q\Gamma^{21}_{S+\alpha I}Y+S\Gamma^{11}_{S+\alpha I}Y\,. \tag{6.13}$$

The parameters $\Gamma^{21}_{S+\alpha I}$ and $\Gamma^{11}_{S+\alpha I}$ are of the same form as Γ^{21} and Γ^{11} in (5.21) but here with $(S+\alpha I)$ instead of S. (In fact, $\Gamma^{21}_{S+\alpha I}=\Gamma^{21}_S(I+\alpha\Gamma^{11}_S)^{-1}$ and $\Gamma^{11}_{S+\alpha I}=\Gamma^{11}_S(I+\alpha\Gamma^{11}_S)^{-1}$.) Although $\Phi(t)$ is the interpolating spline from the space of T to that of Z, it is referred to as the smoothing spline from T to Y (with fitted values Z).

Following the arguments in 5.3.5, the global parameters B of the smoothing spline may also be determined by the GLS equation $\min_B tr(Y-QB)^T(S+\alpha I)^{-1}(Y-QB)$ and the local parameters W, then obtained by solving $(S+\alpha I)W=Y-QB$.

Returning to (6.1), it can be seen that scalar α controls how 'smooth' the smoothing spline from T to Y (with fitted values Z) is and is usually pre-specified. When α is large, the fit resembles an affine transformation since optimisation of (6.1) is then essentially minimisation of the roughness of $\Phi(t)$, as much less weight is given to the actual 'closeness' of the fitted z_j to the y_j. Conversely, when α is small, the size of the non-affine components is penalised to a lesser degree and the mapping will resemble a set of interpolating splines, fitting the y_j exactly when α=0, since $\Phi(t)$ can then make the first term in (6.1) zero as well.

6.2.1.2 Smoothing spline with anisotropic landmark uncertainties

As with the Procrustes technique, the ordinary smoothing spline described above assumes isotropic errors at each of the landmark positions when superimposing T to Y. (Interpolating splines assume no variance at the landmark positions.) In practice this assumption is often unlikely and so Rohr (1998) and Rohr *et al.* (1997, 1999) extended the idea of the smoothing spline to produce a method which includes local estimates of 'anisotropic' landmark uncertainties, in the form of covariance matrices, which may be

user defined. To do this, the interpolation spline equations of section 5.2 first require some re-formulation. Instead of writing the parameters of multiple d interpolating spline $\Phi(t)$ as matrices $B=(a_1,...,a_d)$ and $W=(w_1,...,w_d)$, they can instead be written in vector form, with $B^*=(a_{11},...,a_{d1},...,a_{1M},...,a_{dM})^T=vec(B^T)$ and $W^*=vec(W^T)=(w_{11},...,w_{d1},...,w_{1M},...,w_{dM})^T$. Equations (5.19) and (5.20) then become:

$$S^*W^*+Q^*B^*=Y^* \tag{6.14}$$

$$Q^{*T}W^*=0 \tag{6.15}$$

where $S^*=S\otimes I_d, Q^*=Q\otimes I_d$, $Y^*=vec(Y^T)$ and equation (5.21) for B and W has Q and S replaced by Q^* and S^* respectively (and so $\Gamma_S^{21*}=\Gamma_S^{21}\otimes I_d, \Gamma_S^{11*}=\Gamma_S^{11}\otimes I_d$).

If landmark errors are represented by covariance matrices Σ_j, for example, in d=m=2 dimensions:

$$\Sigma_j=\begin{pmatrix}\varepsilon_{xx,j} & \varepsilon_{xy,j}\\ \varepsilon_{yx,j} & \varepsilon_{yy,j}\end{pmatrix} \text{ with } \varepsilon_{xy,j}=\varepsilon_{yx,j}, \tag{6.16}$$

then the objective function for an anisotropic smoothing spline is:

$$\sum_{j=1}^{k}(y_j-\Phi(t_j))^T\Sigma_i^{-1}(y_j-\Phi(t_j))+\alpha J_r^2(\Phi). \tag{6.17}$$

Optimisation of (6.17) then allows landmarks and directions for which covariance is smaller to be matched preferentially to those where covariance is larger.

Again it can be shown that there is a unique solution for $\Phi(t)$ in the form of an interpolating spline as in 6.2.1.1. Denoting the vector of values $(\Phi(t_1),...,\Phi(t_k))^T$ by Z^* and re-writing (6.17) as:

$$\min_{Z^*}\ (Y^*-Z^*)^T R^{-1}(Y^*-Z^*)+\alpha(Z^{*T}\Gamma_S^{11*}Z^*) \tag{6.18}$$

where $R^{-1}=diag(\Sigma_1,...,\Sigma_k)^{-1}=diag(\Sigma_1^{-1},...,\Sigma_k^{-1})$,

optimal positions Z^* (and therefore parameters B^* and W^*) can be shown to be obtainable by matrix differentiation (of (6.18)) or by solving a system of linear equations as in (6.15) and (6.16), here with S^* replaced by $S^* + \alpha R$, i.e.

$$Y^* = Q^* B^* + (S^* + \alpha R) W^* \tag{6.19}$$

$$Q^{*T} W^* = 0 \tag{6.20}$$

to give

$$\Phi(t) = ((\phi_M(t), \ldots, \phi_M(t)) \otimes I_2)\, B^* + (s(t)^T \otimes I_2)\, W^*$$

and

$$Z^* = (I_{2k} + \alpha R \Gamma_S^{11*})^{-1} Y^* = Q^* B^* + S^* W^*, \tag{6.21}$$

with B^* and W^* given by equations (6.3) and (6.4) with Q, S and Z replaced by Q^*, $S^* + \alpha R$ and Z^*. So for a pair (d=1,2) of thin-plate smoothing splines in m=2 dimensions, with roughness index r=2, we would have $B^* = (c_1, c_2, a_{11}, a_{21}, a_{12}, a_{22})^T$, $W^* = (w_{11}, w_{21}, \ldots, w_{1k}, w_{2k})^T$ and $Y^* = (y_{1x}, y_{1y}, \ldots, y_{kx}, y_{ky})^T$ with:

$$Q^* = \begin{pmatrix} 1 & 0 & t_{1x} & 0 & t_{1y} & 0 \\ 0 & 1 & 0 & t_{1x} & 0 & t_{1y} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & t_{kx} & 0 & t_{ky} & 0 \\ 0 & 1 & 0 & t_{kx} & 0 & t_{ky} \end{pmatrix} \quad R = \begin{pmatrix} \varepsilon_{xx,1} & \varepsilon_{xy,1} & 0 & 0 & \cdots & \cdots & 0 & 0 \\ \varepsilon_{yx,1} & \varepsilon_{yy,1} & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & \varepsilon_{xx,2} & \varepsilon_{xy2} & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \varepsilon_{yx,2} & \varepsilon_{yy,2} & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & 0 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 0 & \varepsilon_{xx,k} & \varepsilon_{xy,k} \\ 0 & 0 & \cdots & \cdots & 0 & 0 & \varepsilon_{yx,k} & \varepsilon_{yy,k} \end{pmatrix}$$

$$S^* = \begin{pmatrix} \sigma(t_1 - t_1) & 0 & \sigma(t_1 - t_2) & 0 & \cdots & \cdots & \sigma(t_1 - t_k) & 0 \\ 0 & \sigma(t_1 - t_1) & 0 & \sigma(t_1 - t_2) & \cdots & \cdots & 0 & \sigma(t_1 - t_k) \\ \sigma(t_2 - t_1) & 0 & \sigma(t_2 - t_2) & 0 & \cdots & \cdots & \vdots & \vdots \\ 0 & \sigma(t_2 - t_1) & 0 & \sigma(t_2 - t_2) & \cdots & \cdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \sigma(t_k - t_1) & 0 & \cdots & \cdots & \cdots & \cdots & \sigma(t_k - t_k) & 0 \\ 0 & \sigma(t_k - t_1) & \cdots & \cdots & \cdots & \cdots & 0 & \sigma(t_k - t_k) \end{pmatrix}.$$

Note however, that $\Phi(t)$ can now no longer be separated into separate spline functions of the form of (5.2), with $\Phi(t)=(\Phi_1(t),\ldots,\Phi_d(t))$ and $J_r^m(\Phi)=\sum_{p=1}^{d} J_r^m(\Phi_p)$. Note also that the Σ_j actually represent the localisation errors of corresponding landmark pairs in both Y and T and so result from combining the covariance matrices of corresponding landmarks. Rohr *et al.* (1999) state that if that the two covariance matrices can be assumed to depend only slightly on the non-affine part of the transformation, the matrices can be combined by applying a rotation, scaling and/or re-sizing to one of the matrices before adding them. Alternatively, if the objects have approximately the same registration then the two covariance matrices can simply be added together.

The method extends the work in Rohr *et al.* (1996) where localisation errors were approximated by scalar weights σ_j^2 representing the circular variance at each landmark (and so $\Sigma_j = diag(\sigma_j^2,\sigma_j^2)$ in (6.17)). Here the objective function to be minimised was:

$$\sum_{j=1}^{k} \frac{(y_j-\Phi(t_j))^T(y_j-\Phi(t_j))}{\sigma_j^2} + \alpha J_r^m(\Phi). \qquad (6.22)$$

It is also easy to see that if $\Sigma_j = diag(1,1)$, equations (6.18) and (6.19) reduce to the ordinary smoothing spline equations (6.9) and (6.10) (but in the new notation) and that the (pair of) interpolating splines may also be written in this form by specifying $\Sigma_j=0$ (no variance at each landmark).

For an application involving shape, Rohr *et al.* (1999) used the approach of (6.18) for the registration of two or three dimensional tomographic MR images of the human brain. Green & Silverman (1994) also describe weighted smoothing for d=1, using a function of the form of (6.18) above with weights set inversely proportional to the variance of the observations. They also considered use of non-diagonal weighting matrices, for models which include additional explanatory variables and where the errors in the observations are correlated, having been taken sequentially over time.

6.2.1.3 Semi-landmark method as anisotropic smoothing spline

Rohr *et al.* (1999) briefly noted how the anisotropic smoothing spline for m=2, r=2 generalised Bookstein's semi-landmark method, allowing movement within ellipses around each selected landmark during superimposition, rather than along straight lines. Bookstein's technique can be thought of as a special case of this method, since for straight lines the variance in one direction is zero whereas in the direction of the chord it is infinity. However no mathematical details were given and so we explore this link further here.

To produce the same results as the semi-landmark method, matrices Σ_j^{-1}, must be pre-specified so that the mapping of corresponding points in $k\times2$ matrix T to positions along the unit chord directions (u_x, u_y) of the semi-landmarks are un-penalised, whereas any positions $\Phi(t)=(z_x, z_y)$ for which $(z_x, z_y)\neq\lambda(u_x y_x, u_y y_y)$ are penalised heavily. In addition, points in T corresponding to the fixed landmarks in configuration Y must map exactly to their positions in Y, and so Σ_j^{-1} for these landmarks must also be specified appropriately. The contribution from each landmark to the left hand side of (6.18) is:

$$\varepsilon_{11}^{-1}(y_x - z_x)^2 + 2\varepsilon_{12}^{-1}(y_x - z_x)(y_y - z_y) + \varepsilon_{22}^{-1}(y_y - z_y)^2. \qquad (6.23)$$

For the semi-landmarks we require that:

$$\frac{y_y - z_y}{y_x - z_x} = \frac{u_y}{u_x}, \text{ i.e. } \left((y_y - z_y) - \frac{u_y}{u_x}(y_x - z_x)\right)^2 = 0,$$

$$\text{i.e. } (y_y - z_y)^2 - 2\frac{u_y}{u_x}(y_x - z_x)(y_y - z_y) + \left(\frac{u_y}{u_x}\right)^2 (y_x - z_x)^2 = 0 \qquad (6.24)$$

Comparing coefficients in (6.23) and (6.24), for the semi-landmarks we must have:

$$\Sigma^{-1} \propto \begin{pmatrix} (u_y/u_x)^2 & -u_y/u_x \\ -u_y/u_x & 1 \end{pmatrix}, \text{ i.e. } \Sigma^{-1} = \kappa \begin{pmatrix} (u_y/u_x)^2 & -u_y/u_x \\ -u_y/u_x & 1 \end{pmatrix} \qquad (6.25)$$

with κ extremely large so as to penalise heavily instances when (6.23) is not true. We could have also arrived at this by considering the following. If a series of points lie along a straight line with slope $\Delta=u_y/u_x$, their covariance is given by:

$$\Sigma=\tau\begin{pmatrix}1 & u_y/u_x\\ u_y/u_x & (u_y/u_x)^2\end{pmatrix} \tag{6.26}$$

where τ corresponds to the variance in the *x*-direction. However, (6.26) is singular and so cannot be inverted so instead consider what happens to a series of points for which $\Sigma=diag(v,\delta)$, where *v* and δ are the variance in the *x* and *y*-directions. As δ→0 the data will resemble a series of points lying along the *x*-axis with variance *v*. If the data are transformed to lie along a chord with slope u_y/u_x, by right multiplication by a rotation matrix with $\cos(\theta)=u_x, \sin(\theta)=u_y$ (rotating the x-axis anti-clockwise by angle θ), then using $\text{cov}(YA)=A^T\,\text{cov}(Y)A$, the new covariance matrix is:

$$\Sigma=\begin{pmatrix}u_x & -u_y\\ u_y & u_x\end{pmatrix}\begin{pmatrix}v & 0\\ 0 & \delta\end{pmatrix}\begin{pmatrix}u_x & u_y\\ -u_y & u_x\end{pmatrix}=\begin{pmatrix}vu_x^2+u_y^2\delta & u_xu_y(v-\delta)\\ u_xu_y(v-\delta) & vu_y^2+u_x^2\delta\end{pmatrix}. \tag{6.27}$$

As δ→0:
$$\Sigma\to v\begin{pmatrix}u_x^2 & u_xu_y\\ u_xu_y & u_y^2\end{pmatrix}=vu_x^2\begin{pmatrix}1 & u_y/u_x\\ u_y/u_x & u_y^2/u_x^2\end{pmatrix} \tag{6.28}$$

which is of the form of (6.26), with $\tau=vu_x^2$. The inverse of (6.26) is given by:

$$\Sigma^{-1}=\frac{1}{(vu_x^2+u_y^2\delta)(vu_y^2+u_x^2\delta)-u_x^2u_y^2(v-\delta)^2}\begin{pmatrix}vu_y^2+u_x^2\delta & -u_xu_y(v-\delta)\\ -u_xu_y(v-\delta) & vu_x^2+u_y^2\delta\end{pmatrix}. \tag{6.29}$$

As δ→0:
$$\Sigma^{-1}\to\kappa\begin{pmatrix}vu_y^2 & -vu_xu_y\\ -vu_xu_y & vu_x^2\end{pmatrix}=\kappa vu_x\begin{pmatrix}u_y^2/u_x^2 & -u_y/u_x\\ -u_y/u_x & 1\end{pmatrix}$$

which is of the form of (6.25) with κ→∞ as δ→0. So one has two choices, either specifying Σ^{-1} directly as in (6.24), with very large κ, or specifying Σ in the form of (6.27) with δ very small and *v* very large, so that Σ may be inverted.

For the fixed landmarks, $\Sigma=0$, but then Σ^{-1} is undefined. However, if $\Sigma=diag(\delta,\delta)$, for very small δ, then:

$$\Sigma^{-1}=\frac{1}{\delta^2}\begin{pmatrix}\delta & 0\\ 0 & \delta\end{pmatrix}=\begin{pmatrix}1/\delta & 0\\ 0 & 1/\delta\end{pmatrix} \tag{6.30}$$

which as $\delta\to 0$ heavily penalises any values of $(z_x,z_y)\neq(y_x,y_y)$.

Note that with appropriate choices of covariance matrices, it is then possible to penalise different extents of movement along chords, by considering the variance along each chord to be finite rather than infinite, for example, by varying the choice of v for each semi-landmark in (6.28). However, to do this, a neater form of solution and easier set of equations to work with than the above, can be derived directly.

6.2.1.4 Semi-landmarks with smoothing spline penalty function

In this section were propose a new method which penalises the amount of movement of semi-landmarks along chords, while minimising the bending energy of the mapping from $k\times 2$ configuration T to $Y^{new}=vec^{-1}(vec(Y^0)-U\lambda)$. Recall that Y^0 is the $k\times 2$ configuration of nominal landmark positions, of which a sublist $j_{(l)}, j=1,\ldots,k, l=1,\ldots,L$ are semi-landmarks, λ is a vector of scalars of length L (to be determined) and $U=(U_x:U_y)^T$ defines the chord directions of the semi-landmarks as described in 5.4.1. The method is derived by constraining the form which $\Phi(T)$ may take, in the objective function of an ordinary or anisotropic smoothing spline and optimising over λ. If the configuration Z, obtained by minimising the smoothing spline penalty function (6.1) or (6.5) is constrained to be of the form $Z=vec^{-1}(vec(Y^0)-U\lambda)$ then:

$$\min_Z \ \text{trace}((Y^0-Z)^T(Y^0-Z))+\alpha\,\text{trace}(Z^T\Gamma_S^{11}Z)$$

$$=\min_Z\,(vec(Y^0)-vec(Z))^T(vec(Y^0)-vec(Z))+\alpha\,(vec(Z)^T(I_2\otimes\Gamma_S^{11})vec(Z)^T)$$

becomes:

$$\min_\lambda\ (U\lambda)^T(U\lambda)+\alpha\,(vec(Y^0)-U\lambda)^T(I_2\otimes\Gamma_S^{11})(vec(Y^0)-U\lambda)^T \tag{6.31}$$

where the right hand side is the usual penalty function for the semi-landmark method. Differentiating (6.31) with respect to λ and setting the derivative equal to zero gives:

$$\lambda=(U^T(\alpha(I_2\otimes\Gamma_S^{11})+I_{2k})U)^{-1}U^T\alpha(I_2\otimes\Gamma_S^{11})vec(Y^0) \qquad (6.32)$$

With α very large, the optimal mapping and new positions of the semi-landmarks will resemble the result of original (minimum BE only) criterion, whereas for very small α, the semi-landmarks will not move and so α must be pre-specified somewhere in between to produce a satisfactory compromise. We explore different possible values of α in 6.2.1.5 below. This idea may also be extended to allow different penalties on the extent of movement of different semi-landmarks along their chords, inversely proportional to quantities $\sigma_1^2,...,\sigma_L^2$, by considering penalty function:

$$\min_{\lambda}(U\lambda)^T diag(\sigma_1^2,...,\sigma_L^2)^{-1}(U\lambda)+\alpha(vec(Y^0)-U\lambda)^T(I_2\otimes\Gamma_S^{11})(vec(Y^0)-U\lambda)^T \qquad (6.33)$$

which is optimised (by matrix differentiation) for:

$$\lambda=(U^T(\alpha(I_2\otimes\Gamma_S^{11})+diag(\sigma_1^2,...,\sigma_{SL}^2)^{-1}U)^{-1}U^T\alpha(I_2\otimes\Gamma_S^{11})vec(Y^0). \qquad (6.34)$$

For example we could penalise the movement in the gingival semi-landmarks away from their original positions in the hope of reducing the chance of large movements and therefore unrealistic shapes being produced, while leaving other semi-landmarks (e.g. around the lower outline) un-penalised.

In the notation introduced in 6.2.1.2, equations (6.33) and (6.34) can be written as:

$$\min_{\lambda}\ (U^*\lambda)^T R^{-1}(U^*\lambda)+\alpha(Y^*-U^*\lambda)^T\Gamma_S^{11*}(Y^*-U^*\lambda) \qquad (6.35)$$

$$\lambda=(U^{*T}(\alpha\Gamma_S^{11*}+R^{-1})U^*)^{-1}U^{*T}\alpha\Gamma_S^{11*}Y^* \qquad (6.36)$$

where

$$R^{-1}=diag(\Sigma_{j_{(1)}}^{-1},...,\Sigma_{j_{(L)}}^{-1}),\Sigma_{j_{(l)}}=diag(\sigma_{j_{(l)}}^2,\sigma_{j_{(l)}}^2),U^*=vec_{2k}^{-1}(vec((vec(U_x):vec(U_y))^T)).$$

However use of covariance matrices of the form of (6.16) makes no sense here since we only need to specify one variance value for each semi-landmark (rather than a covariance matrix) as the landmarks are constrained to move along straight lines.

6.2.1.5 Examples

Following an initial GPA to obtain the full Procrustes mean shape $\hat{\mu}$ and corresponding fits, X^P, fig 6.1 shows the results of the first semi-landmark step for four configurations, when using different possible values of α in (6.31), with the gingival landmarks as semi-landmarks, allowed to move along chords as described in fig. 5.12.

Clearly the method has some success in stopping large movements of the semi-landmarks (compare the results for α=1, 0.3, 0.05 with the original BE-only criterion in the top row). Choosing the most appropriate value of α however depends on several factors. If too small a value of α is pre-specified, the semi-landmarks will not actually move and so we will be unable to remove the unwanted variation between cases. Too large a value and the landmarks move too far again (to the minimum BE positions), leading to unrealistic shapes and increasing sample variation. Another difficulty is that the 'best' value of α for one configuration may not be the best for another, in terms of reducing variation in the semi-landmark positions about the mean shape. Depending on the actual amount by which the bending energy from $T=\hat{\mu}$ to $Y^{new}=vec^{-1}(vec(X^P)-U\lambda)=X^{new}$ can be reduced, the semi-landmarks on one configuration may not move by the same distance along their chords as on another when the same value of α is used. This in turn depends on the shape of the configuration and the directions of the semi-landmarks chords. For example as fig 6.1 shows, with α=0.05, the semi-landmarks on cases 2 and 3 hardly move, whereas on case 4, there is still considerable movement. Even though the actual movement of the landmarks is penalised by a relatively greater amount the reduction in bending energy (×0.05) achieved by moving the semi-landmarks such as distance along each of their chords is still smaller than the increase in sum of squared distances moved.

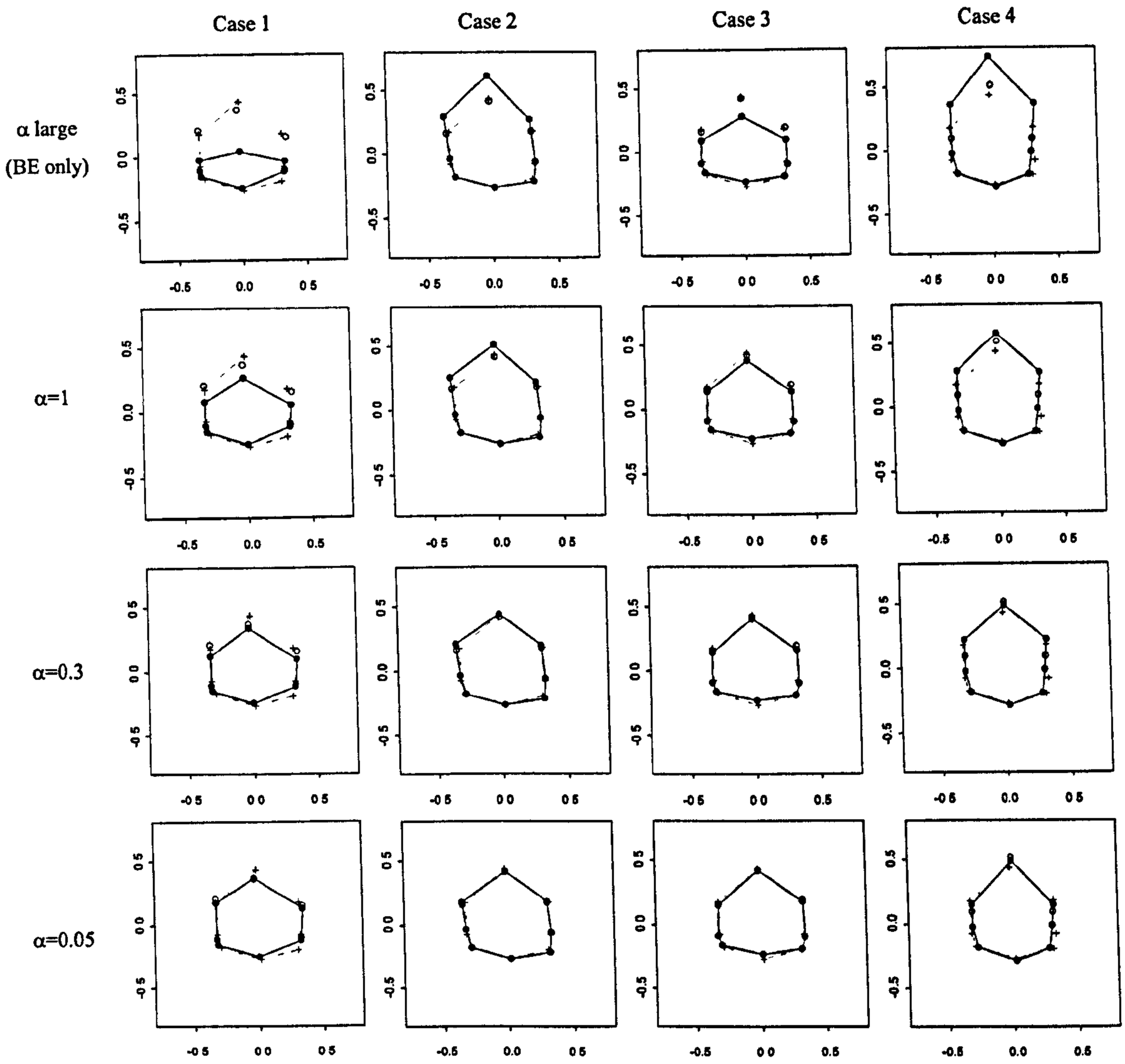


Fig 6.1: Determination of new semi-landmark positions for 4 cases (using mean from a larger sample) using the constrained smoothing spline with various values of α. (Empty circles) nominal positions of landmarks when registered to mean shape (crosses). (Filled circles) new configuration with new positions of semi-landmarks along chords.

One remaining problem however, is that even though we are able to restrict the amount of movement of the semi-landmarks along their chords, there are still cases where the movement of the semi-landmarks appears to be in directions opposite to that which we would expect if we are to achieve a reduction in variance in the positions of these landmarks about the mean shape. See for example, case 1 for α=0.3 in fig 6.1.

6.2.2 Penalty functions based on higher order splines and kriging

In this section we consider other deformations in m=2 dimensions using different choices of functions for $\sigma_{r,m}(t-t_i)$ and $\phi_1,\ldots,\phi_M$, described in section 5.2. Recall that

$\phi_1,...,\phi_M$ are the $M = \binom{m+r-1}{m}$ polynomials in m variables of total degree $\leq r$-1 which form a basis for the penalty function $J_r^m(\Phi)$ and $\sigma_{r,m}(t-t_i)$ is a Green's function for $\Delta^r = \left(\frac{\delta^2}{\delta t_{[1]}^2}+...+\frac{\delta^2}{\delta t_{[m]}^2}\right)^r$ with $\Delta^r\,\Phi(t)=0$ for $t \neq t_j, j=1,\ldots,k$. $\sigma_{r,m}(t-t_j)$ also defines a conditionally positive definite covariance function corresponding to an intrinsic random field, as we shall see in 6.2.2.2 below.

6.2.2.1 Higher order spline transformations

For data $y_1,...,y_k \in R^m$ associated with $t_1,...,t_k \in R^m$, section 5.2 showed that providing $r>m/2$, any combination of m independent spline functions of the form of (5.2), each with index r, may be used to interpolate the t_j to the y_j, with:

$$\Phi(t)=(\Phi_1(t),...,\Phi_m(t))^T=B^T\phi(t)+W^T s(t) \tag{6.37}$$

where

$$s(t)=(\sigma_{r,m}(t-t_1),...,\sigma_{r,m}(t-t_k))^T, \phi(t)=(\phi_1(t),...,\phi_M(t))^T,$$

$$\sigma_{r,m}(t-t_j)=\theta_{r,m}\left\|t-t_j\right\|^{2r-m}\log\left\|t-t_j\right\|,\ \theta_{r,m}=\frac{(-1)^{(m/2)+1+r}}{2^{2r-1}\pi^{m/2}(r-1)!(r-m/2)!} \text{ if } 2r\text{-}m \text{ even}$$

$$\sigma_{r,m}(t-t_j)=\theta_{m,d}\left\|t-t_j\right\|^{2r-m},\ \theta_{r,m}=\frac{\Gamma((m/2)-r)}{2^{2r}\pi^{m/2}(r-1)!} \text{ otherwise} \tag{6.38}$$

$$B=\begin{pmatrix} a_{11} & \cdots & a_{1d} \\ \vdots & \ddots & \vdots \\ a_{M1} & \cdots & a_{Md} \end{pmatrix}, W=\begin{pmatrix} w_{11} & \cdots & w_{1d} \\ \vdots & \ddots & \vdots \\ w_{k1} & \cdots & w_{kd} \end{pmatrix},$$

with solution for B and W for any choice of r and m, found by solving the equations:

$$Y=QB+SW,\ Q^TW=0 \tag{6.39}$$

where

$$Q=\begin{pmatrix} \phi_1(t_1) & \cdots & \phi_M(t_1) \\ \vdots & \ddots & \vdots \\ \phi_1(t_{1k}) & \cdots & \phi_M(t_k) \end{pmatrix}, Y=\begin{pmatrix} y_{11} & \cdots & y_{1m} \\ \vdots & \ddots & \vdots \\ y_{k11} & \cdots & y_{km} \end{pmatrix}, S=\begin{pmatrix} \sigma_{r,m}(t_1-t_1) & \cdots & \sigma_{r,m}(t_1-t_k) \\ \vdots & \ddots & \vdots \\ \sigma_{r,m}(t_k-t_1) & \cdots & \phi_M(t_k-t_k) \end{pmatrix} \tag{6.40}$$

with $$B=\Gamma^{21}Y\,,\ W=\Gamma^{11}Y\,, \tag{6.41}$$

$$\Gamma^{21}=(Q^TS^{-1}Q)^{-1}Q^TS^{-1} \text{ and } \Gamma^{11}= S^{-1}-S^{-1}Q\,(Q^TS^{-1}Q)^{-1}Q^TS^{-1} \tag{6.42}$$

and

$$J_r^m(\Phi) = \sum_{p=1}^{m} J_r^m(\Phi_p)=tr(W^TSW)=tr(Y^T\Gamma^{11}Y)\,. \tag{6.43}$$

In 5.3 it was seen that for a pair of interpolating splines for data $\{(t_{jx},t_{jy})\}$ and $\{(y_{jx},y_{jy})\}$, j=1,...,k in m=2 dimensions, with order of derivatives r=2, $\phi_1(t)=1$, $\phi_2(t)=t_x$, $\phi_2(t)=t_y$ and $\sigma_{2,2}(t-t_i)=\|t-t_i\|^2\log\|t-t_i\|$, with roughness penalty:

$$J_2^2(\Phi)= J_2^2(\Phi_1)+J_2^2(\Phi_2)=\sum_{p=1}^{2}\left\{\int_{R^2}\left(\frac{\delta^2\Phi_p}{\delta t_x^2}\right)^2+2\left(\frac{\delta^2\Phi_p}{\delta t_x\delta t_y}\right)^2+\left(\frac{\delta^2\Phi_p}{\delta t_y^2}\right)^2 dt_x dt_y\right\} \tag{6.44}$$

known as the 'bending energy' of the mapping. This is the quantity optimised by Bookstein's semi-landmark method, when certain landmarks in Y are allowed to move along chords. Writing $Y^{new}=vec(Y^0)-U\lambda$, with U as defined in 5.4.1, Y^0 the nominal positions of all landmarks and λ a vector of scalars, λ was shown to be found by minimising, from (6.43):

$$tr(vec(Y^0)-U\lambda)^T\Gamma^{11}(vec(Y^0)-U\lambda)) \tag{6.45}$$

with:

$$\lambda=\left(U^T\begin{pmatrix}\Gamma^{11} & 0\\ 0 & \Gamma^{11}\end{pmatrix}U\right)^{-1}U^T\begin{pmatrix}\Gamma^{11} & 0\\ 0 & \Gamma^{11}\end{pmatrix}vec(Y^0)\,. \tag{6.46}$$

However, for data $\{t_j\}$ and $\{y_j\}\in R^2$ we could also consider other roughness penalties and spline mappings with order of derivatives r>2 and obtain a solution using (6.46) since the equations above hold for any choice of r (S, Q and hence Γ^{11} are recalculated accordingly from (6.40) to (6.42)). For example, for r=3, the null space of the penalty function is the M=6 dimensional space spanned by polynomials $\phi_1(t)=1$, $\phi_2(t)=t_x$, $\phi_3(t)=t_y$, $\phi_4(t)=t_x^2$, $\phi_5(t)=t_y^2$, $\phi_6(t)=t_xt_y$, with $\sigma_{3,2}(t-t_i)=-\|t-t_j\|^4\log\|t-t_j\|$ with:

$$J_3^2(\Phi) = tr(Y^T \Gamma^{11} Y) = \sum_{p=1}^{2} \left\{ \int_{R^2} \left(\left(\frac{\delta^3 \Phi_p}{\delta t_y^3} \right)^2 + 3 \left(\frac{\delta^3 \Phi_p}{\delta t_x^1 \delta t_y^2} \right)^2 + 3 \left(\frac{\delta^3 \Phi_p}{\delta t_x^2 \delta t_y^1} \right)^2 + \left(\frac{\delta^3 \Phi_p}{\delta t_x^3} \right)^2 \right) dt_x dt_y \right\} \tag{6.47}$$

which may be used to produce alternative optimal positions for the semi-landmarks along their chords. An example follows in section 6.2.2.5.

Note however that for $r \geq 4$ and with only $k=8$ landmarks, as we have for our upper central incisors, Q is then singular and $(Q^T S^{-1} Q)^{-1}$ cannot be obtained. That is, for $r=4$ the null space of the penalty function is then the $M=10$ dimensional space spanned by the additional polynomials $\phi_7(t)=t_x^3$, $\phi_8(t)=t_y^2$, $\phi_9(t)=t_x^2 t_y$, $\phi_{10}(t)=t_x t_y^2$, leading to an 8×10 matrix for Q and so here the interpolation problem is over parameterised.

6.2.2.2 Kriging

It can be shown that interpolating splines (and ordinary smoothing splines) are particular cases of kriging, a commonly used method of prediction used in spatial statistics, named after the mining engineer D. Krige (see for example, Cressie, 1993). The link between kriging and interpolating splines has been noted by many authors, in particular, Kent & Mardia (1994) and Matheron (1981).

Given univariate data y_j at a collection of 'sites' $t_j \in R^m$, the problem is again to fit a smooth function f to interpolate the data so that $f(t_j) = y_j$. For splines, the aim is find a smooth function of the site location t which interpolates the data with minimum value of a roughness penalty, based on the sum of integrated squared partial derivatives of a given order r. With kriging the $y_1,...,y_k$, are regarded as an observed realisation of a univariate random process $Y(t)$ at sites $t_1,...,t_k \in R^m$ and the aim is to find the best unbiased linear predictor of $Y(t)$ on R^m, using the observations $y_1,...,y_k$.

We show how alternative formulation of the kriging problem leads to a set of equations of the same form as (6.39) to (6.42), for both single and multiple sets of independent

kriging predictors, with each predictor comprising a linear combination of functions of t and covariance terms. As with the interpolating splines, a mapping between two planes can also be produced by considering a pair of kriging predictors and we show how with certain choices of polynomial and covariance functions, this yields the same mapping as a pair of thin-plate splines. With kriging however, a wider range of functions may be chosen for the global part of the mapping (other than polynomials $\phi_1,...,\phi_M$) and for the $\sigma_{r,m}(t-t_j)$ terms, allowing us to explore other deformations and penalty functions, which, after modifying the kriging formulation in the same way as the PTPS mapping, may be used to determine the location of semi-landmarks.

6.2.2.2.1 Kriging predictors

Given k univariate observations $y_{obs}=(y_1,...,y_k)^T$ taken at sites $t_1,...,t_k \in R^m$, the kriging predictor of a univariate random process $Y(t)$ at an m-dimensional site is determined by finding the best unbiased linear predictor of the form $\hat{Y}(t)=\sum_{j=1}^{k}\gamma_j y_j$.

Consider the general linear model:

$$Y(t)=\sum_{v=1}^{P}\beta_i g_v(t)+\varepsilon(t) = \beta^T g(t)+\varepsilon(t) \qquad (6.48)$$

with $g(t)=(g_1(t),...,g_P(t))^T$, where $g_i(t)$ are known functions and $\varepsilon(t)$ is a random field. Dryden and Mardia (1998) show that if $((Y(t_1),...,Y(t_k))^T$ is the random vector of the process $Y(t)$ at the k sites and y_{obs} is the realisation of $(Y(t_1),...,Y(t_k))^T$, then the regression predictor $\hat{Y}(t)$ is the conditional mean of $Y(t)$ given $((Y(t_1),...,Y(t_k))^T$, i.e. $E[Y(t)\,|\,(Y(t_1),...,Y(t_k))^T]$ and so the regression predictor at site t is given by:

$$\hat{Y}(t)=(g(t)^T a+s(t)^T w)=a^T g(t)+w^T s(t) \qquad (6.49)$$

where

$$a=(D^T S^{-1} D)^{-1} D^T S^{-1},\ w=S^{-1}(I_k - D(D^T S^{-1} D)^{-1} D^T S^{-1}) \qquad (6.50)$$

$$s(t)=(\sigma(t-t_1),...,\sigma(t-t_k))^T,$$

with

$$S=\begin{pmatrix}\sigma(t_1-t_1) & \cdots & \sigma(t_1-t_k)\\ \vdots & \ddots & \vdots\\ \sigma(t_k-t_1) & \cdots & \sigma(t_k-t_k)\end{pmatrix}, D=\begin{pmatrix}g_1(t_1) & \cdots & g_p(t_1)\\ \vdots & \ddots & \vdots\\ g_1(t_k) & \cdots & g_p(t_k)\end{pmatrix}.$$

The values of w and a are found by solving the dual kriging equations:

$$aD + Sw = y_{obs} \tag{6.51}$$

$$D^T a = 0 \tag{6.52}$$

where a takes the form of a GLS estimator minimising $(y_{obs} - Da)^T S^{-1}(y_{obs} - Da)$.

6.2.2.2.2 Intrinsic random fields, universal and generalised predictors

The prediction is called 'universal' kriging when $\varepsilon(t)$ is a zero mean stationary random field with positive definite covariance function:

$$\text{cov}(\varepsilon(t),\varepsilon(t')) = \text{cov}(Y(t),Y(t')) = \sigma(t-t').$$

However, $\varepsilon(t)$ may also be an intrinsic random field. Define $(\delta_1,t_1,...,\delta_k,t_k)$ as an 'increment' with respect to a space G, of known functions, if:

$$\sum_{j=1}^{k}\delta_j h(t_j) = 0 \text{ for all } h \in G. \tag{6.53}$$

Process $Y(t)$ is an 'intrinsic random field', if, at all increments, $\sum_{j=1}^{k}\delta_j Y(t_j)$ is distributed with zero mean and variance $\sum_{j=1}^{k}\sum_{j'=1}^{k}\delta_j\delta_{j'}\sigma^*(t_j-t_{j'}) \geq 0$. $\sigma^*(t_j-t_{j'})$ is called a 'conditional positive definite' (c.p.d.) covariance function with respect to G, since at all increments it is non-negative (taking increments annihilates functions of G, see (6.53)). An intrinsic random field therefore corresponds to an equivalence class of processes $Y(t)$ which all have different $g(t)$ terms expressed as functions in G. The

'generalised' kriging predictor is then given by (6.48) but with $\sigma(t-t_j)$ a c.p.d. covariance function and $g(t)=(g(t_1),...,g(t_k))^T$ forming a basis for G.

One important class of intrinsic random processes has c.p.d. covariance function $\sigma*(t-(t+h))=\sigma_\alpha(h), h\in R^m$, indexed by scalar $\alpha>0$, with

$$\begin{aligned}\sigma_\alpha(h)&=(-1)^{[\alpha]+1}\|h\|^{2\alpha} && \alpha \text{ not an integer}\\ &=(-1)^{\alpha+1}\|h\|^{2\alpha}\log\|h\| && \alpha \text{ an integer}\end{aligned} \qquad (6.54)$$

If G_{r-1} is the space of polynomials in h of degree$\leq r$-1, then $\sigma_\alpha(h)$ defines a c.p.d. covariance with respect to G_{r-1}, provided $r-1\geq[\alpha]$, where [.] denotes the 'integer part of'. Such c.p.d. covariance functions are also 'self similar', i.e. $\sigma_\alpha(ch)\equiv c^{2\alpha}\sigma_\alpha(h)$ for $c>0$, where $\equiv$ means equivalent up to an even polynomial in h of degree $2[\alpha]$ and so $\sigma_\alpha(ch)$ and $\sigma_\alpha(h)$ yield the same predictions. The full theory and general framework for intrinsic random fields is presented in Mardia *et al.* (1996b).

6.2.2.2.3 Multiple kriging predictors and the link with splines

Now consider the prediction of an m-dimensional multivariate process $Y(t)=(Y(t)_1,...,Y(t)_m)^T$ using a collection of m independent kriging predictors each with the same covariance function (and hence $s(t)$) and same vector of functions $g(t)$. Here the k observations in R^m at sites $t_1,...,t_k\in R^m$ are written as $k\times m$ matrix:

$$Y_{obs}=\begin{pmatrix} y_{obs_{1[1]}} & \cdots & y_{obs_{1[m]}} \\ \vdots & \ddots & \vdots \\ y_{obs_{k[1]}} & \cdots & y_{obs_{k[m]}} \end{pmatrix}$$

In the same way as for multiple splines, the multiple kriging predictor can be written as:

$$(\hat{Y}(t)_1,...,\hat{Y}(t)_m)^T=\hat{Y}(t)=C^T g(t)+W^T s(t) \qquad (6.55)$$

where each predictor $\hat{Y}(t)_m$, is of the form of (6.49). If D and S are as defined in (6.50) the solutions for C and W for any choice of $g(t)$ defining D and covariance function $\sigma(t-t_j)$ defining S are found by solving the equations:

$$DC + SW = Y_{obs} \tag{6.56}$$

$$D^T W = 0 \tag{6.57}$$

with

$$C=(D^T S^{-1} D)^{-1} D^T S^{-1} Y_{obs}, \; W=S^{-1}(I_k - D(D^T S^{-1} D)^{-1} D^T S^{-1}) Y_{obs}. \tag{6.58}$$

So for m=2 the kriging predictors can be used for plane-to-plane mappings (and displayed as deformation grids). Note also that C takes the form of a generalised least squares estimator minimising $(vec(Y_{obs}) - Dvec(C))^T (I_M \otimes S)^{-1} (vec(Y_{obs}) - Dvec(C))$.

It is clear from (6.55) to (6.57) that if $g_1(t),...,g_p(t)$ form a basis for G_{r-1}, the space of polynomials in m variables of degree$\leq r$-1, we would then have a set of functions identical to the $\phi_1,...,\phi_M$ spanning the $M=\binom{m+r-1}{m}$ null space of the penalty function for an interpolating spline in m dimensions with index r. For example, for m=2, r=2, we would have $g_1(t)=1$, $g_2(t)=t_x$, $g_3(t)=t_y$ and hence D the same as matrix Q for a thin-plate spline. If the c.p.d. covariance function in (6.54) has index $\alpha=r$-1, then it is defined exactly the same as $\sigma(t-t_j)$ in (6.38) for an interpolating spline. For example, for m=2, α=1, $\sigma_\alpha(t_j - t_{j'}) = \|t_j - t_{j'}\|^{2(r-1)} \log\|t_j - t_{j'}\|$.

Therefore, a multiple spline is exactly the same as an intrinsic (generalised) multiple kriging predictor using the self-similar random field with $\alpha=r$-1, where r is the smoothness index of the corresponding spline function and c.p.d. covariance function $\sigma_{r-1}(h)$ taken with respect to G_{r-1}, the polynomials in m variables of total degree$\leq r$-1. As noted in the examples above, the thin-plate spline (m=2), with smoothness index r=2 is the same as the intrinsic kriging predictor using α=1, $\sigma_1(0)=0$ and c.p.d. covariance function taken with respect to G_1 (the linear functions). Regardless of the choice of

$g_1(t),...,g_P(t)$ and $\sigma(t-t_j)$ the solution for C and W in (6.58) for the multiple kriging predictor is exactly the same form as that for B and W for the multiple interpolating spline in (6.41), but with Q replaced by D. Where the interpolating splines optimise $tr(Y^T\Gamma^{11}Y)$, with $\Gamma^{11} = S^{-1} - S^{-1}Q\,(Q^TS^{-1}Q)^{-1}Q^TS^{-1}$ for any choice of r and m (and hence $\phi_1,...,\phi_M$ and $\sigma(t-t_j)$), the kriging predictors minimise $tr(Y_{obs}^T K Y_{obs})$ where $K = S^{-1}(I_k - D(D^TS^{-1}D)^{-1}D^TS^{-1})$ for any choice of $g_1(t),...,g_P(t)$ and $\sigma(t-t_j)$.

6.2.2.3 Semi-landmarks with kriging/higher order splines penalty functions

Based on the final statement above it is easy to propose a new generalisation of Bookstein's method and consider kriging with semi-landmarks when m=2. For two configurations, T and Y, each of k landmarks and any choice of functions $g_1(t),...,g_P(t)$ and $\sigma(t-t_j)$, the parameter estimates of a multiple (m=2) kriging predictor optimise $tr(Y^TKY)$, where here K is calculated as above, for T. Writing $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$, with U as defined in 5.4.1, Y^0 the nominal positions of all landmarks and λ a vector of scalars, we then seek the optimal value of:

$$tr(Y^{new^T} K Y^{new}) = \min_{\lambda}(vec(Y^0) - U\lambda)^T \begin{pmatrix} K & 0 \\ 0 & K \end{pmatrix} (vec(Y^0) - U\lambda) \qquad (6.59)$$

as the semi-landmarks move along their chords. By analogy with (6.46), λ is given by:

$$\lambda = \left(U^T \begin{pmatrix} K & 0 \\ 0 & K \end{pmatrix} U \right)^{-1} U^T \begin{pmatrix} K & 0 \\ 0 & K \end{pmatrix} Y_{col}^0 \qquad (6.60)$$

allowing new semi-landmark positions to be determined with respect to any suitable choices of $g_1(t),...,g_p(t)$ and $\sigma(t-t_j)$. Notice that we can also consider the suggestions made for higher order multiple interpolating splines in 6.2.2.1 within the kriging framework.

6.2.2.5 Examples

If G_{r-1} is the space of polynomials of degree r-1 in m=2 dimensions, then for r=2, G_1 comprises $g_1(t)=1$, $g_2(t)=t_x$, $g_3(t)=t_y$ and for r=3, G_2 comprises the same three functions and $g_4(t)=t_x^2$, $g_5(t)=t_y^2$, $g_6(t)=t_x t_y$. However, recall that for $r\geq 4$ and only k=8 landmarks, D would then be singular and the mapping over parameterised. Ensuring that $r-1\geq[\alpha]$ and $\alpha>0$ some of the options we could consider are then:

- G_1 with α=0.5, giving $\sigma_{1/2}(t_j-t_{j'})=\|t_j-t_{j'}\|$
- G_2 with α=1, giving $\sigma_1(t_j-t_{j'})=\|t_j-t_{j'}\|^2\log\|t_j-t_{j'}\|$
- G_2 with α=2, giving $\sigma_2(t_j-t_{j'})=-\|t_j-t_{j'}\|^4\log\|t_j-t_{j'}\|$, corresponding to the pair of splines mapping with 3rd derivatives in the roughness penalty, proposed in 6.2.2.1.
- G_1 with α=1, giving $\sigma_1(t_j-t_{j'})=\|t_j-t_{j'}\|^2\log\|t_j-t_{j'}\|$, corresponding to the PTPS mapping used in the original semi-landmark method, where minimisation of (6.59) is of the bending energy of the mapping (included here for comparison).

Fig 6.2 (top row) shows the results of using the four possible kriging/spline mappings above to determine the new positions of the semi-landmarks using the same case and semi-landmarks/chord directions as used to illustrate the 'collapsing' problem in 5.6.3. In the second row is the mapping (shown as a deformation grid) from the sample mean $T=\hat{\mu}$ to the nominal landmark positions $Y^0=X^P$ (the Procrustes fit to $\hat{\mu}$) and in the fourth row the mapping from $\hat{\mu}$ to $Y^{new}=vec^{-1}(vec(Y^0)-U\lambda)=X^{new}$.

Clearly, use of polynomials spanning G_2 is unwise here as this fails to produce mappings that are bijective (one to one). As with the spline mappings, there is no restriction in the formulation to prevent this and as is evident from the final two columns of fig 6.2, this also results completely un-meaningful semi-landmark positions.

Use of an alternative covariance function with polynomials spanning G_1 (as in the original PTPS transformation) appears to restrict the extent of 'collapsing', but not

substantially. When polynomials spanning G_1 are used with any covariance function, the mapping will still always try to minimise the sum of squared residuals of the GLS affine fit between $Q\Gamma^{21}Y^{new}$ and the chord positions and fixed landmarks of Y^{new}. Regardless of the defined covariance, collapsing/stretching still results if an improved fit to all landmarks and chords can be achieved by the semi-landmarks moving large distances. Perhaps one option could be to consider a generalised affine mapping with residuals weighted so that the influence of the fixed landmarks is reduced.

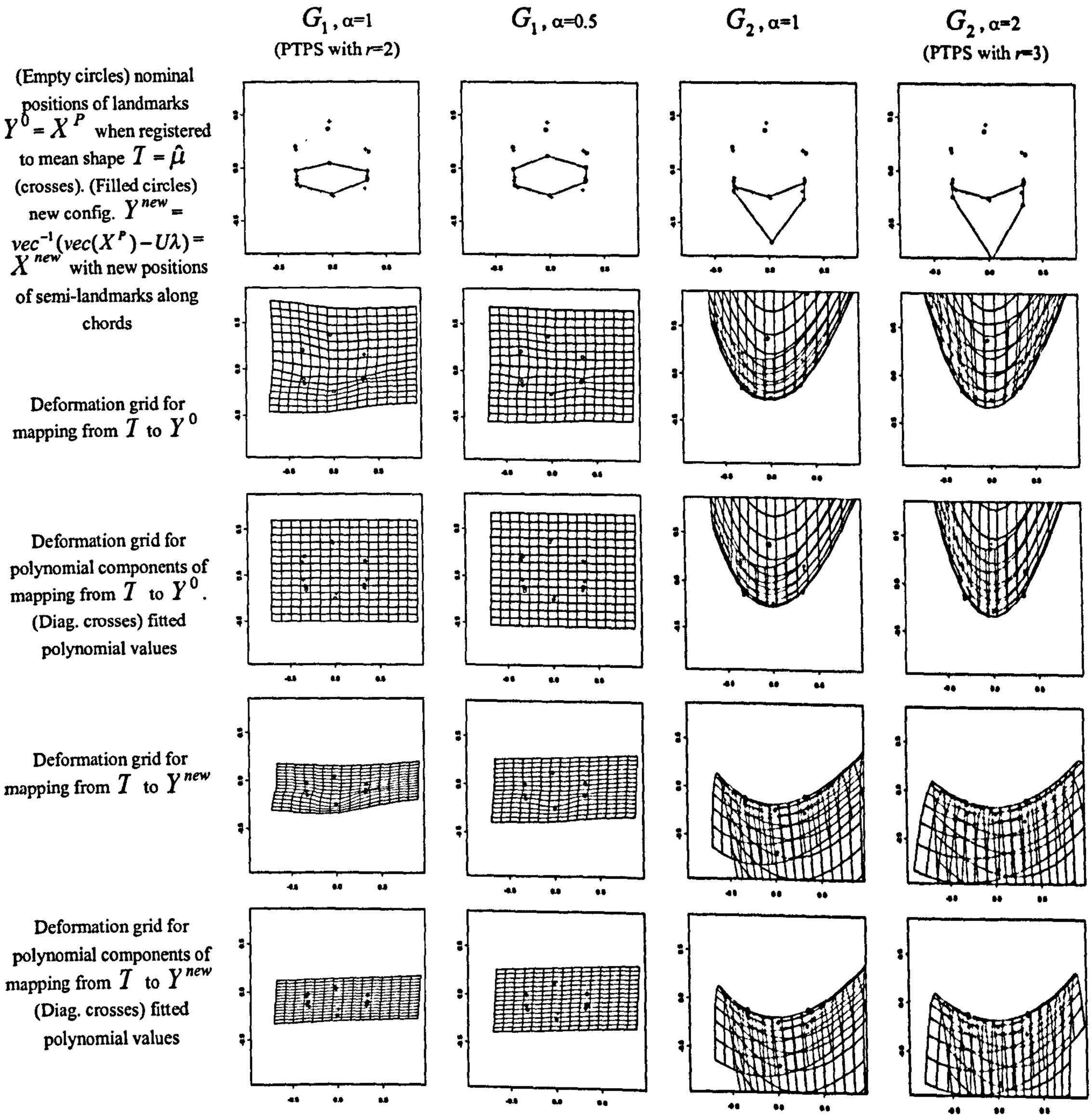


Fig. 6.2: Determination of new semi-landmark positions using kriging (and higher order spline) mappings for same configuration (Procrustes fit) and mean shape used in fig. 5.13(left).

6.2.3 Constraints and penalties on the affine component of a PTPS transformation

Using the bending energy method, large movements in the semi-landmarks, producing unrealistic shapes, were seen to be caused by large stretches or compressions produced by the affine component of the PTPS mapping from T to $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$. (Where in practice $T = \hat{\mu}$, the full Procrustes mean shape, $Y^0 = X^P$, the Procrustes fit of a configuration to $\hat{\mu}$, and $Y^{new} = X^{new}$). We could therefore propose and consider ways of penalising or constraining the extent to which this situation occurs by either constraining the parameters of the affine mapping from T in some way, or developing a measure of the extent of stretching/collapsing produced by the affine component, which may then be penalised.

Once again, we denote the PTPS mapping from T to any configuration Y as:

$$Y = QB + SW = Q\Gamma^{21}Y + S\Gamma^{11}Y$$

where $Q\Gamma^{21}Y$ is the GLS affine mapping of T to Y (with weighting matrix S), $S\Gamma^{11}Y$ are the corresponding non-affine components, completing the exact mapping to Y (and may be thought of as the residuals between 'affine fit' $Q\Gamma^{21}Y$ and Y) and S, Q, Γ^{11} and Γ^{21} are as defined in (5.21) or (6.40) and (6.42) and depend only on T.

6.2.3.1 Constraints on the affine component

For these particular configurations and choice of semi-landmarks and chord directions, the collapsing/stretching problem occurs because of movement produced by the affine component of the thin-plate spline in the y-direction. As described in 5.6.3, with the semi-landmarks free to move in roughly parallel vertical directions, the optimal (vertical) affine mapping is essentially determined by the (vertical) fit achievable to the fixed landmarks only. The semi-landmarks are free to adopt positions consistent with the mapping in this direction so as to optimise the fit between $Q\Gamma^{21}vec^{-1}(vec(Y^0) - U\lambda)$ and $vec^{-1}(vec(Y^0) - U\lambda)$. A simple modification to the PTPS formulation could

therefore be to constrain the coefficients of the affine terms of both the horizontal and vertical spline transformations (mapping to the x and y coordinates of the target configuration respectively) to have equal parameters. The mapping is then no longer a PTPS transformation, but is a constrained GLS affine match, as described in 5.3.5.4.2, with weights given by S. In equation (5.57) we can set $B=(b:b)$ with $b=(c_1,a_1,a_2)^T$ and rearrange so that we have the affine GLS superimposition problem of $(k\times 2)$ T to $(k\times 2)$ Y, given by:

$$\min_{a}(vec(Y)-\begin{pmatrix}Q\\Q\end{pmatrix}b)^T(I_m\otimes S)^{-1}(vec(Y)-\begin{pmatrix}Q\\Q\end{pmatrix}b). \qquad (6.61)$$

This is then easily modified to include semi-landmarks in Y, in the same way as equations (5.61). The vector λ, determining the new positions of the semi-landmarks in $Y^{new}=vec(Y^0)-U\lambda$, may be obtained by simultaneously solving the following for b and λ.

$$\min_{b,\lambda}\left(vec(Y^0)-\left(\begin{pmatrix}Q\\Q\end{pmatrix}:U\right)\begin{pmatrix}b\\\lambda\end{pmatrix}\right)^T\begin{pmatrix}S&0\\0&S\end{pmatrix}^{-1}\left(vec(Y^0)-\left(\begin{pmatrix}Q\\Q\end{pmatrix}:U\right)\begin{pmatrix}b\\\lambda\end{pmatrix}\right) \qquad (6.62)$$

with

$$\begin{pmatrix}b\\\lambda\end{pmatrix}=\left(\left(\begin{pmatrix}Q\\Q\end{pmatrix}:U\right)^T\begin{pmatrix}S&0\\0&S\end{pmatrix}^{-1}\left(\begin{pmatrix}Q\\Q\end{pmatrix}:U\right)\right)^{-1}\left(\begin{pmatrix}Q\\Q\end{pmatrix}:U\right)^T\begin{pmatrix}S&0\\0&S\end{pmatrix}^{-1}vec(Y^0). \qquad (6.63)$$

The quantity being minimised in (6.62) is no longer the bending energy of the mapping from T to $vec^{-1}(vec(Y^0)-U\lambda)$, but is the generalised residual sums of squares of the constrained affine superimposition of T to $vec^{-1}(vec(Y^0)-U\lambda)$, providing yet another new objective function one could use.

Another option could be to constrain the two sets of affine components in the pair of splines to always produce a similarity transformation of T when mapping to Y^{new}. This would then lead to a 'weighted' or GLS Procrustes superimposition problem, with semi-landmarks moving so as to minimise the weighted residual sums of squares between $SG(T)$ and $vec^{-1}(vec(Y^0)-U\lambda)$. We discuss this idea and the method of 'weighted Procrustes superimposition' in more detail in section 6.4.

6.2.3.2 Affine penalties

Large unrealistic movements of semi-landmarks were seen to occur on configurations where the minimum bending energy PTPS mapping from T to $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$, consists of a GLS affine transformation, $Q\Gamma^{21}Y^{new}$, which produces a large affine change in the shape of T (in practice, the full Procrustes mean). We therefore could consider ways of penalising movements of the semi-landmarks which require a large affine change in the shape of T, and explore three novel possibilities below.

Note however, that not all affine transformations of T should be penalised since we would still wish to allow for those which are rotations, translations or scalings since these do not represent any change the shape of T. It therefore make sense to use Procrustes distances to measure and penalise affine differences in shape, since this then allows for similarity transformations of T, but penalises those mappings which are not.

6.2.3.2.1 Shape change produced by the affine component (option 1)

Given two unit size configurations T and Y, Dryden & Mardia (1998) decompose the squared Procrustes distance between the two shapes into a contribution from the affine and non-affine parts as below.

$$d_F^2(Y,T) = \textit{affine contribution} + d_F^2(T, T + S\Gamma^{11}Y),$$

where $S\Gamma^{11}Y = SW$ are the 'non-affine' displacements in the PTPS mapping from T to Y, added to the landmark locations of T, rather than the corresponding affine component ($Q\Gamma^{21}Y$). One possible measure of the difference in shape between T and Y attributable to the affine contribution would then be:

$$d_F^2(Y,T) - d_F^2(T, T + S\Gamma^{11}Y) \tag{6.64}$$

or as a proportion:

$$1 - \frac{d_F^2(T, T + S\Gamma^{11}Y)}{d_F^2(Y,T)}.$$

Note the requirement that T and Y are unit size. Although the appearance of the PTPS mapping from T to Y is unaffected by their relative sizes, the PTPS mapping from T to βY, for some $\beta>0$ is given by $\beta(Q\Gamma^{21}Y+S\Gamma^{11}Y)=Q\Gamma^{21}\beta Y+S\Gamma^{11}\beta Y$ and so adding the non-affine displacements to T will produce different shapes and hence different values of $d_F^2(T,T+S\Gamma^{11}Y)$, depending on the relative sizes of T and Y. Consequently some standardisation of the size of the two configurations is required in order to produce a suitable comparable measure for different shapes. If configurations Y and T are allowed to be any size we should therefore re-write (6.64) as:

$$d_F^2(Y,T)-d_F^2(T,T/\|T\|+S\Gamma^{11}Y/\|Y\|).$$

To use such a measure for determining the new positions of semi-landmarks, we would seek:

$$\min_{\lambda} d_F^2(Y^{new},T)-d_F^2(T,T+S\Gamma^{11}Y^{new}/\|Y^{new}\|) \qquad (6.65)$$

$$=\min_{\lambda} d_F^2(vec^{-1}(vec(Y^0)-U\lambda),T)-d_F^2(T,T+S\Gamma^{11}\frac{vec^{-1}(vec(Y^0)-U\lambda)}{\|vec^{-1}(vec(Y^0)-U\lambda)\|}).$$

As the semi-landmarks move along their chords, $Y^{new}=vec^{-1}(vec(Y^0)-U\lambda)$ needs to be re-scaled to ensure it is always of unit size for the reasons stated above. If there is no constraint on the size of the target configuration the movement of the semi-landmarks will affect the size of Y^{new} (and *vice versa*) and hence the optimal positions in $T+S\Gamma^{11}Y^{new}$. (Although bending energy is dependent on the size of the two configurations, recall from section 5.4.1.3 that for the original semi-landmark method, the resulting shape of Y^{new} was unaffected). However, an analytical solution to (6.65) is not straightforward and so we must make use of an optimisation algorithm, such as that described in 7.5.2 in order to determine λ.

Unfortunately, there are certain drawbacks to using (6.65), which we illustrate using examples below. Fig 6.3 shows the results of using this criterion to determine the new semi-landmark positions of two upper central incisors, following Procrustes registration to the initial estimate of mean shape $\hat{\mu}$ from the entire sample from which they were taken. The choice of semi-landmarks and chord directions were again as in fig. 5.13.

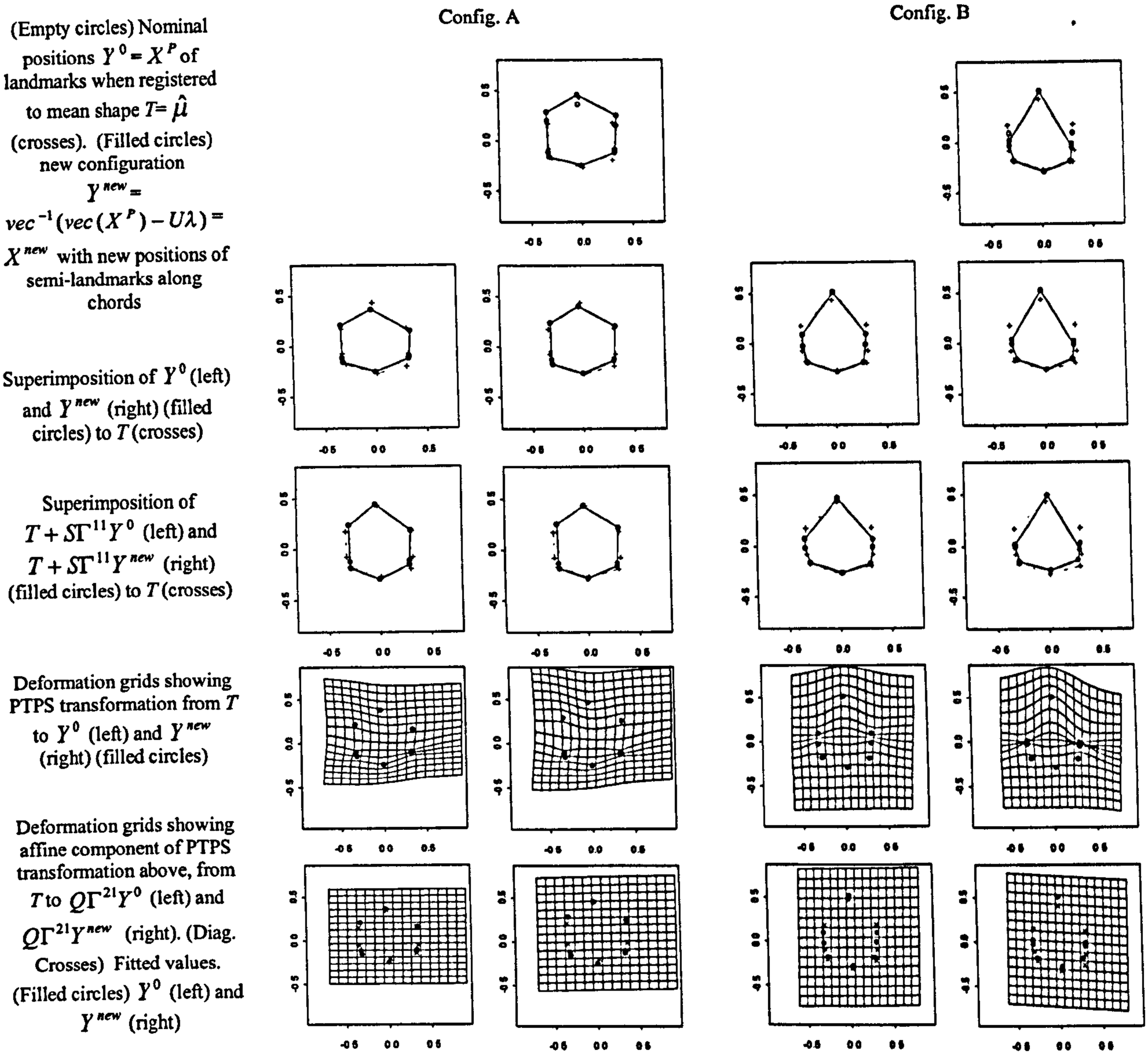


Fig 6.3: Determination of new semi-landmark positions for two cases, following initial superimposition to the mean of a larger sample, using the affine contribution penalty (option 1).

	Config. A		Config. B	
	$Y = Y^0$	$Y = Y^{new}$	$Y = Y^0$	$Y = Y^{new}$
$d_F^2(Y,T)$	0.175	0.161	0.030	0.085
$d_F^2(\hat{\mu}, \hat{\mu} + S\Gamma^{11}Y/\|Y\|)$	0.014	0.020	0.034	0.094
$d_F^2(Y,\hat{\mu}) - d_F^2(\hat{\mu}, \hat{\mu} + S\Gamma^{11}Y/\|Y\|)$	0.161	0.141	-0.004	-0.009
$tr((Y/\|Y\|)^T \Gamma^{11} Y/\|Y\|) = BE$	0.175	0.280	0.337	0.816

Table 6.1: Squared Procrustes distances and bending energy of original (to $Y^0 = X^P$) and optimal, to $Y^{new} = vec^{-1}(vec(X^P) - U\lambda) = X^{new}$ minimising $d_F^2(Y^{new},T) - d_F^2(T, T + S\Gamma^{11}Y^{new}/\|Y^{new}\|)$, PTPS mappings from $T = \hat{\mu}$.

Using $T = \hat{\mu}$, $Y^0 = X^P$ and setting $Y^{new} = vec^{-1}(vec(X^P) - U\lambda) = X^{new}$, the first row shows the movement of the semi-landmarks along their chords to new positions. The second row shows the superimposition of Y^0 and Y^{new} to T and the third row, the

superimposition of $T+S\Gamma^{11}Y^0/\|Y^0\|$ and $T+S\Gamma^{11}Y^{new}/\|Y^{new}\|$ to T, illustrating the visual appearance of the terms $d_F^2(Y,T)$ and $d_F^2(T,T+S\Gamma^{11}Y/\|Y\|)$ in the penalty function above for $T=\hat{\mu}$, $Y=X^P$ and $Y=X^{new}$. The fourth row shows the PTPS transformation from T to Y^0 and Y^{new} as deformation grids and the fifth row, the affine component of these mappings (from T to $Q\Gamma^{21}Y^0$ and $Q\Gamma^{21}Y^{new}$). Table 6.1 displays the various squared full Procrustes distances between the configurations shown in fig. 6.3 and also the bending energy of the PTPS mappings from T to $Y^0/\|Y^0\|$ and $Y^{new}/\|Y^{new}\|$. Here this is calculated by standardising Y^0 and Y^{new} to unit size so that the values are comparable (recall that bending energy depends on size) and so that the calculation uses the same non-affine displacements as in $d_F^2(T,T+S\Gamma^{11}Y/\|Y\|)$, i.e. we use:

$$tr(Y/\|Y\|-Q\Gamma^{21}Y/\|Y\|)^T S^{-1}(Y/\|Y\|-Q\Gamma^{21}Y/\|Y\|)$$

$$=tr(S\Gamma^{11}Y/\|Y\|)^T S^{-1}(S\Gamma^{11}Y/\|Y\|)=tr((Y/\|Y\|)^T\Gamma^{11}Y/\|Y\|).$$

The main finding from this sample, illustrated by the two example configurations in table 6.1, was that minimisation of (6.65) is most often achieved by maximising $d_F^2(T,T+S\Gamma^{11}Y^{new}/\|Y^{new}\|)$, relative to $d_F^2(Y^{new},T)$ and this nearly always results in a negative value for the penalty function. For both these configurations (and the rest of this sample), this results in an increase in $d_F^2(T,T+S\Gamma^{11}Y/\|Y\|)$ and typically a smaller increase in $d_F^2(Y,T)$ as well, so that Y^{new} is such that:

$$d_F^2(Y^{new},T)-d_F^2(Y^0,T)\le d_F^2(T,T+S\Gamma^{11}Y^{new}/\|Y^{new}\|)-d_F^2(T,T+S\Gamma^{11}Y^0/\|Y^0\|). \quad (6.66)$$

To do this, the semi-landmarks move to positions so that the PTPS mapping from T to $Y^{new}/\|Y^{new}\|$ comprises non-affine terms $S\Gamma^{11}Y^{new}/\|Y^{new}\|$ which make $T+S\Gamma^{11}Y^{new}/\|Y^{new}\|$ as different from T as possible, but only in a way such that inequality (6.66) results. See fig 6.3, 2nd and 3rd rows. In contrast to the situation when minimising BE, there is no requirement that the fit of the affine component to Y^{new} is optimal in any sense (minimising the size of the $S\Gamma^{11}Y^{new}/\|Y^{new}\|$ terms), and so

here the non-affine components are free to take on any values (with corresponding affine component).

Clearly however, increases in $d_F^2(Y,T)=d_F^2(X,\hat{\mu})$ $(d_F^2(Y^{new},T)-d_F^2(Y^0,T)\leq 0)$ are undesirable as we wish to reduce the variation in shape about the mean, not increase it. Note also that $d_F^2(Y,T)-d_F^2(T,T+S\Gamma^{11}Y/\|Y\|)$ can actually be negative to start with (i.e. for $Y=Y^0$) as for case *B* in table 6.1, which casts some doubt on the use (6.65) as a measure of affine contribution. In addition, with many configurations, folding can occur as for (see again case *B* in fig 6.2). Here this results in a new configuration where one of the papilla semi-landmarks has moved below the positions of the MD endpoints (see fig 6.2, fourth row), which cannot be true by the definition of these landmarks. In light of these issues, other possible measures of the shape change produced by the affine component of the PTPS transformation were considered.

6.2.3.2.2 Shape change produced by the affine component (option 2)

In this section we explore novel alternatives to the previous approach for measuring the shape change produced by the affine component of a PTPS mapping, and explore how these may be used to determine the new positions of semi-landmarks. For T and Y of unit size and PTPS transformation given by $Q\Gamma^{21}Y+S\Gamma^{11}Y$, recall from 5.3.3 that we can also construct the configuration $Q\Gamma^{21}Y$, corresponding to the affine fit of T to Y. (The $S\Gamma^{11}Y$ are a set of displacements, not a configuration). Fig 6.4 shows the different possible Procrustes distances between $Q\Gamma^{21}Y$ and T and Y we could then consider and also those based on the configuration $T+S\Gamma^{11}Y$ used in 6.2.3.2.1.

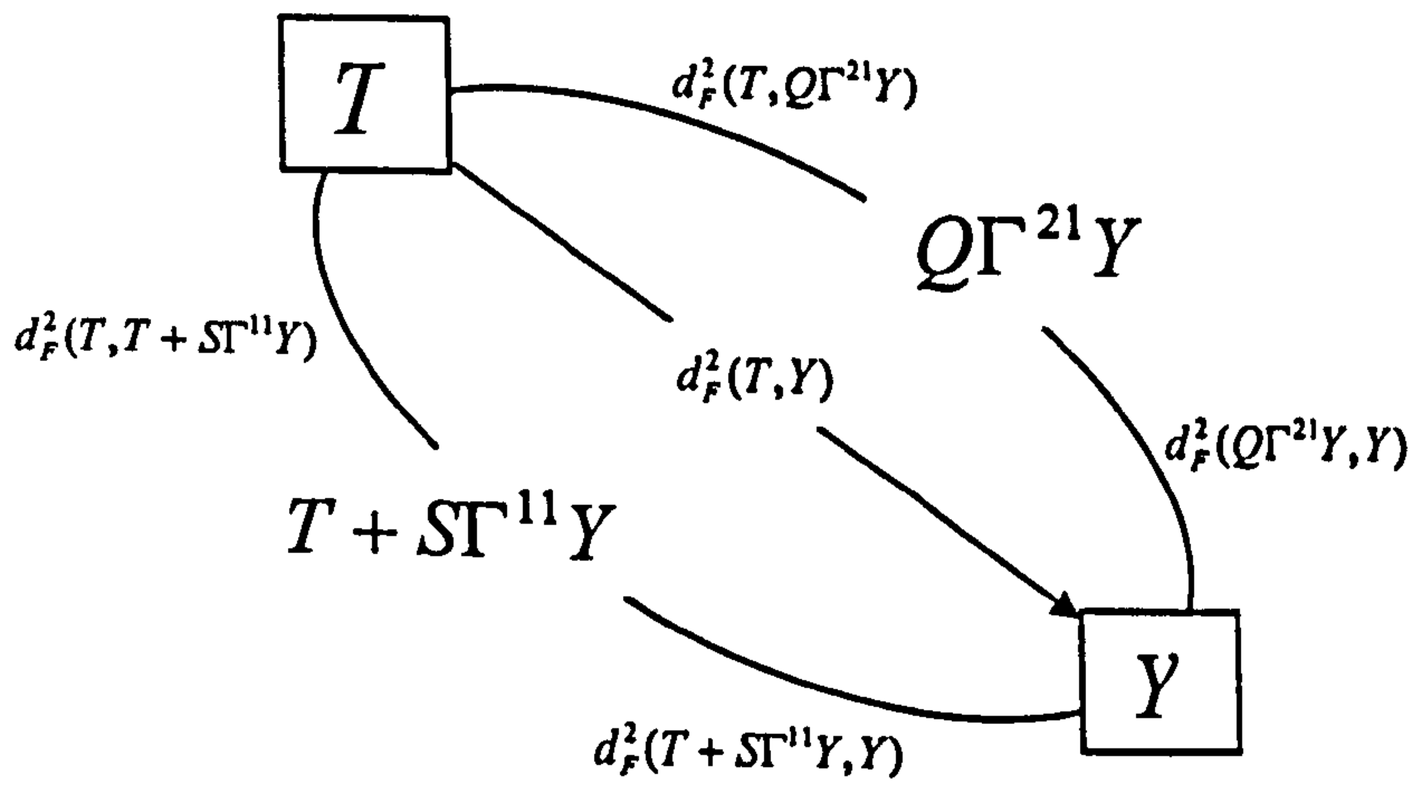


Fig 6.4: Possible (squared) Procrustes distances between configurations based on the affine and non-affine components of the PTPS mapping from T to Y.

Note that for both $Q\Gamma^{21}Y$ and $T+S\Gamma^{11}Y$ we have the inequalities:

$$d_F^2(T,Y) \leq d_F^2(T,Q\Gamma^{21}Y) + d_F^2(Q\Gamma^{21}Y,Y),$$

$$d_F^2(T,Y) \leq d_F^2(T,T+S\Gamma^{11}Y) + d_F^2(T+S\Gamma^{11}Y,Y)$$

and so the previous measure of affine contribution $d_F^2(Y,T) - d_F^2(T,T+S\Gamma^{11}Y)$ may be viewed as providing a lower bound for the size of $d_F^2(T+S\Gamma^{11}Y,Y)$. As a measure of affine shape change itself, $d_F^2(T+S\Gamma^{11}Y,Y)$ would be somewhat misleading since the difference in shape between $T+S\Gamma^{11}Y$ and Y is not equal to the affine difference in shape between T and Y, i.e. an affine mapping of $T+S\Gamma^{11}Y$ to Y would not equal Y and is not the same transformation as the affine fit of T to Y (given by $Q\Gamma^{21}Y$). Use of the configuration $T+S\Gamma^{11}Y$ as a way of representing (and displaying) the non-affine difference in shape between T and Y was also criticised in section 5.3.3 because the $S\Gamma^{11}Y$ are defined by (and to be added to) their corresponding affine component $Q\Gamma^{21}Y$, and not T.

Alternative measures of the affine difference in shape between T and Y could instead be based on $Q\Gamma^{21}Y$. As shown in fig 6.4, we could calculate:

$$d_F^2(T,Q\Gamma^{21}Y) \tag{6.67}$$

or

$$d_F^2(T,Y) - d_F^2(Q\Gamma^{21}Y,Y). \tag{6.68}$$

In contrast to (6.66), with both these options there would be no requirement that the sizes of T and Y be standardised in any way.

For the second option, the squared Procrustes distance given by $d_F^2(Q\Gamma^{21}Y,Y)$ would represent the remaining variation in shape after fitting an affine transformation of T to Y. This is not quite the bending energy of the mapping from T to Y, since $d_F^2(Q\Gamma^{21}Y,Y) = \min_{SG} tr(SG(Q\Gamma^{21}Y) - Y)^T (SG(Q\Gamma^{21}Y) - Y)$, where $SG(.)$ denotes the 'similarity group' of transformations, whereas bending energy is given by

$tr(Q\Gamma^{21}Y-Y)^T S^{-1}(Q\Gamma^{21}Y-Y)$. (Furthermore, $\min_{SG}\left\|SG(Q\Gamma^{21}Y)-Y\right\|^2 \neq \left\|Q\Gamma^{21}Y-Y\right\|^2$ and $\left\|Q\Gamma^{21}Y-Y\right\|^2=\left\|T-(T+S\Gamma^{11}Y)\right\|^2 \neq$BE, since $Q\Gamma^{21}Y$ is determined by a GLS match to Y). Using equation (6.68), with $Y=vec^{-1}(vec(Y^0)-U\lambda)$, we would seek:

$$\min_{\lambda} d_F^2(T, vec^{-1}(vec(Y^0)-U\lambda)) - d_F^2(Q\Gamma^{21}vec^{-1}(vec(Y^0)-U\lambda), vec^{-1}(vec(Y^0)-U\lambda)). \quad (6.69)$$

However, minimisation of (6.69) is easily achieved by maximising the second term relative to the first, again typically resulting in a negative value for the penalty function, with:

$$d_F^2(Y^{new},T)-d_F^2(Y^0,T) \leq d_F^2(\hat{\mu},Q\Gamma^{21}Y^{new})-d_F^2(\hat{\mu},Q\Gamma^{21}Y^0).$$

Here the semi-landmarks move to positions along their chords so that the affine mapping of T to $vec^{-1}(vec(Y^0)-U\lambda)$ changes the shape of T as much as possible, while ensuring that the equation above holds true. Clearly this is not a sensible option, especially as this again most often results in increases in $d_F^2(X,\hat{\mu})$, when $T=\hat{\mu}$ and $Y^{new}=X^{new}$ and instances of extreme folding.

The more logical option would be to use (6.67), $d_F^2(T,Q\Gamma^{21}Y)$, measuring the affine shape change in T produced by the PTPS transformation to Y, a quantity that will always be positive. For $Y=vec^{-1}(vec(Y^0)-U\lambda)$, we would then seek:

$$\min_{\lambda} d_F^2(T, Q\Gamma^{21}vec^{-1}(vec(Y^0)-U\lambda)). \quad (6.70)$$

Note that values of $d_F^2(T,Q\Gamma^{21}Y)>d_F^2(T,Y)$ are possible, but not unreasonable, corresponding to pairs of configurations where the affine component of the PTPS transformation from T to Y produces a bigger change in the shape of T than the entire transformation. However, if using (6.70) for the determination of new semi-landmark positions, a value of $d_F^2(T,Q\Gamma^{21}vec^{-1}(vec(Y^0)-U\lambda))=0$ can always be reached by the affine mapping taking on the form of any similarity transformation of T. Since there is

no restriction on the size of the non-affine components (affine residuals) completing the mapping to Y^{new}, then as the semi-landmarks move along their chords, there are many possible solutions for which (6.70) is equal to zero. There is no criterion to fit Y^{new} as well as possible in any way, unless we introduce an additional penalty on the $S\Gamma^{11}Y^{new}$ terms, such as the BE of the mapping. For a pre-specified value α, we could then seek:

$$\min_{\lambda} d_F^2(T, Q\Gamma^{21}Y^{new}) + \alpha tr(Y^{new}\Gamma^{11}Y^{new}) \tag{6.71}$$

$$= \min_{\lambda} d_F^2(T, Q\Gamma^{21}vec^{-1}(vec(Y^0) - U\lambda)) + \alpha tr(vec^{-1}(vec(Y^0) - U\lambda)\Gamma^{11}vec^{-1}(vec(Y^0) - U\lambda))$$

This would then give a penalty function where the bending energy of the mapping from T to Y^{new} is still used, but the possibility of large affine changes in the shape of T being produced is restricted.

However, as was found for smoothing splines, choice of α is difficult. After trial and error, using a test sample of upper central incisors, with mean shape $T = \hat{\mu}$ and semi-landmarks and chord directions on each Procrustes fit $Y^0 = X^P$, as in fig. 5.12 again, a value of α can be found which reduces the overall variation in shape (*RSS*) after the first iteration, but this is at expense of worsening the fit of some configurations to the mean, while improving that of others. The success of the method for each configuration depends on the values of $d_F^2(T, Q\Gamma^{21}Y^{new}) = d_F^2(\hat{\mu}, Q\Gamma^{21}vec^{-1}(vec(X^P) - U\lambda))$ and $tr(Y^{new}\Gamma^{11}Y^{new}) = tr(vec^{-1}(vec(X^P) - U\lambda)\Gamma^{11}vec^{-1}(vec(X^P) - U\lambda))$ that can actually be attained. For example, for one configuration the smallest possible value of $tr(Y^{new}\Gamma^{11}Y^{new})$ could be reached with a PTPS mapping resembling a similarity transformation, whereas for a different configuration, to achieve the same value may require a large affine change in the shape of $\hat{\mu}$. For the first configuration, the movement of the semi-landmarks would not be penalised, whereas for the second, any movement increasing the value of $d_F^2(T, Q\Gamma^{21}Y^{new})$ would. Similarly, a small value of $d_F^2(T, Q\Gamma^{21}Y^{new})$ may require $tr(Y^{new}\Gamma^{11}Y^{new})$ to be high or low, depending on the configuration in question. Since the achievable values of $d_F^2(T, Q\Gamma^{21}Y^{new})$ and $tr(Y^{new}\Gamma^{11}Y^{new})$ will be different for each configuration in the sample, it is difficult to choose a value of α which works equally well on all configurations.

Table 6.2 and fig 6.5 show the results of using (6.71) to determine the new semi-landmark positions for three of the configurations our test sample, for different choices of α. Once more an analytical solution to the objective function was not found to be straightforward and so the optimisation algorithm described in 7.5.2 was used. Fig 6.5 (left) shows the nominal and new semi-landmark positions for each configuration for different choices of α. For comparison with the original semi-landmark method, the final row shows the results of using only the BE criterion for each of these configurations. Fig 6.5 (right) displays the affine component of each of the PTPS transformations from $T=\hat{\mu}$ to $Y^{new}=vec^{-1}(vec(X^P)-U\lambda)=X^{new}$, as a deformation grid. Also shown (top row, right) are the affine components of the original PTPS mapping of $T=\hat{\mu}$ to the nominal positions ($Y^0=X^P$) of each of the three configurations and the affine component for when only the BE of the mapping is optimised (bottom row). Table 6.2 details the values of $d_F^2(T,Q\Gamma^{21}Y^{new})$ and $tr(Y^{new}\Gamma^{11}Y^{new})$ for each configuration for each choice of α and when minimising only the BE of the PTPS mapping from $T=\hat{\mu}$ to $Y^{new}=vec^{-1}(vec(Y^0)-U\lambda)=X^{new}$. Also quoted are the values $d_F^2(T,Q\Gamma^{21}Y^0)$ and $tr(Y^0\Gamma^{11}Y^0)$ between $T=\hat{\mu}$ and the original versions of each of the configurations. The values highlighted in bold correspond to the choice of α for which $Y^{new}=vec^{-1}(vec(Y^0)-U\lambda)=X^{new}$ minimises $\left\|Y^{new}-T\right\|^2=\left\|X^{new}-\hat{\mu}\right\|^2$, the Euclidean sum of squared distances between the semi-landmarks and their corresponding positions in $\hat{\mu}$.

After running the process for a single iteration, table 6.2 (final column) shows that there are several choices of α (from those considered) which may be used successfully to reduce the variation in shape $RSS(\hat{\mu})$ due to differences in the positions of the semi-landmark along their chords. However, as described above, the best choice of α to reduce $\left\|Y^{new}-T\right\|^2$ is different for each configuration, as can be seen in fig 6.5 (left) and table 6.2. For α=0.01, configurations *B* and *C* in fig 6.5(left) shows how penalising too heavily, the affine difference in shape from $\hat{\mu}$ can actually cause the semi-landmarks to move in a way that increases, rather than decreases, $\left\|Y^{new}-T\right\|^2=\left\|X^{new}-\hat{\mu}\right\|^2$.

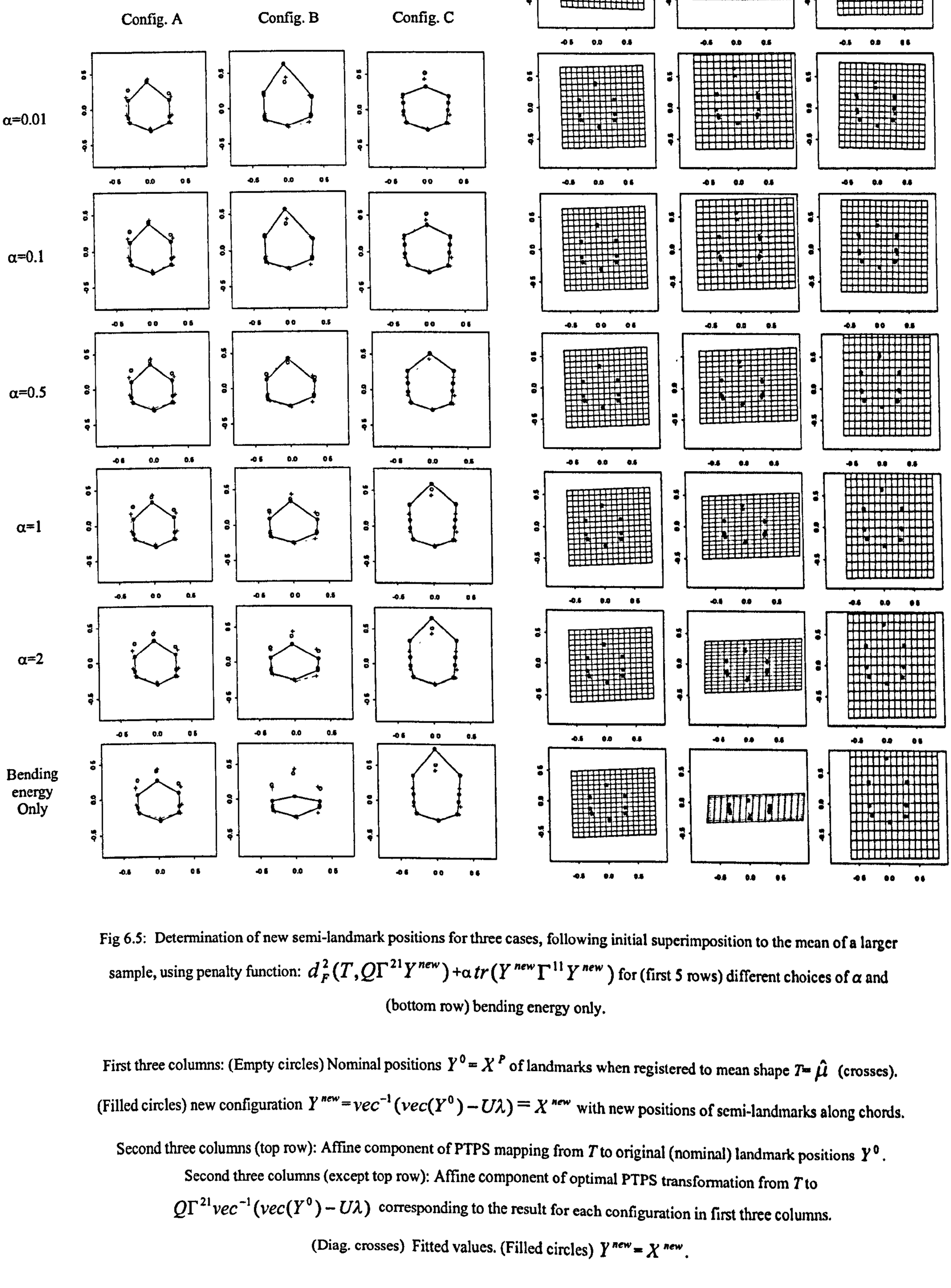


Fig 6.5: Determination of new semi-landmark positions for three cases, following initial superimposition to the mean of a larger sample, using penalty function: $d_F^2(T, Q\Gamma^{21}Y^{new}) + \alpha\, tr(Y^{new}\Gamma^{11}Y^{new})$ for (first 5 rows) different choices of α and (bottom row) bending energy only.

First three columns: (Empty circles) Nominal positions $Y^0 = X^P$ of landmarks when registered to mean shape $T = \hat{\mu}$ (crosses). (Filled circles) new configuration $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda) = X^{new}$ with new positions of semi-landmarks along chords.

Second three columns (top row): Affine component of PTPS mapping from T to original (nominal) landmark positions Y^0.
Second three columns (except top row): Affine component of optimal PTPS transformation from T to $Q\Gamma^{21}vec^{-1}(vec(Y^0) - U\lambda)$ corresponding to the result for each configuration in first three columns.
(Diag. crosses) Fitted values. (Filled circles) $Y^{new} = X^{new}$.

	Config. A		Config. B		Config. C		All configs.
	$d_F^2(T, Q\Gamma^{21}Y^{new})$	$tr(Y^{new}\Gamma^{11}Y^{new})$	$d_F^2(T, Q\Gamma^{21}Y^{new})$	$tr(Y^{new}\Gamma^{11}Y^{new})$	$d_F^2(T, Q\Gamma^{21}Y^{new})$	$tr(Y^{new}\Gamma^{11}Y^{new})$	RSS (1st it)
$Y^{new} = Y^0$	0.002	0.159	0.023	0.175	0.007	0.337	0.084
α=0.01	**0.000**	**0.028**	0.000	0.170	0.000	0.109	0.067
α=0.1	0.000	0.026	0.002	0.153	**0.001**	**0.085**	0.051
α=0.5	0.001	0.023	**0.018**	**0.118**	0.012	0.034	0.048 *
α=1	0.002	0.022	0.038	0.098	0.023	0.017	0.075
α=2	0.003	0.021	0.067	0.081	0.033	0.011	0.126
$tr(Y^{new}\Gamma^{11}Y^{new})$ only	0.009	0.021	0.187	0.065	0.046	0.007	0.284

Table 6.2: Values in affine/bending energy penalty for three configurations following initial GPA. (Bold) Best choice of α to minimise $\|Y^{new} - T\|^2 = \|X^{new} - \hat{\mu}\|^2$. * Best choice of α to minimise *RSS* for entire sample, following subsequent GPA.

One thing that is evident however, is that for some configurations, relatively high values of $d_F^2(T, Q\Gamma^{21}Y^{new})$ do not necessarily result in unrealistic positions of the semi-landmarks if $d_F^2(T, Q\Gamma^{21}Y^0)$ was large to start with. See for example, the results for configuration *B* in fig 6.5/table 6.2. Some tooth shapes may naturally have a large affine difference in shape compared with T and so why should this be penalised? On the other hand, large movements of semi-landmarks may still result when $Q\Gamma^{21}Y^{new}$ resembles a similarity transformation of T (with $d_F^2(T, Q\Gamma^{21}Y^{new}) = 0$), if this represents a large change from the shape of $Q\Gamma^{21}Y^0$. For each configuration A, B and C in fig 6.5, the best choice of α, in terms of minimising $\|Y^{new} - \hat{\mu}\|^2$ appears to be that which produces a PTPS transformation with affine component least changed from that in the original PTPS mapping from $\hat{\mu}$ to Y^0. In the following section, we therefore consider new objective functions which measure and penalise such changes.

6.2.3.2.3 Change in affine component (option 3)

Using (6.71), certain possible shapes of $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$ were penalised by measuring the change in T produced by $Q\Gamma^{21}Y^{new}$, the affine component of the PTPS mapping from T to Y^{new}. However, this does not necessarily restrict the large movement of landmarks or allow the changes in Y we would wish to. As described at the end of the previous section, it appears that it is large changes in the affine component of the PTPS the mapping from T which result in unrealistic shapes being produced, rather than just large affine differences in shape between T and Y^{new}. As

the semi-landmarks move along their chords, a better option could therefore be to penalise movements which result in a large change in the shape of $Q\Gamma^{21}Y^{new}$ compared to $Q\Gamma^{21}Y^{0}$ and so instead of (6.71), we could consider the penalty function:

$$\min_{\lambda} d_F^2(Q\Gamma^{21}Y^0, Q\Gamma^{21}Y^{new}) + \alpha\, tr(Y^{new}\Gamma^{11}Y^{new}) \tag{6.72}$$

$$= \min_{\lambda} d_F^2(Q\Gamma^{21}vec^{-1}(vec(Y^0) - U\lambda), Q\Gamma^{21}vec^{-1}(vec(Y^0) - U\lambda))$$

$$+ \alpha\, tr(vec^{-1}(vec(Y^0) - U\lambda)\Gamma^{11}vec^{-1}(vec(Y^0) - U\lambda)).$$

Obviously, if (6.72) comprised only a $d_F^2(Q\Gamma^{21}Y^0, Q\Gamma^{21}Y^{new})$ term, this would always be minimised by there being no change in the PTPS transformation from T to Y^0 and so (6.72) is a compromise between minimising the bending energy of the mapping, $(Y^{new} - Q\Gamma^{21}Y^{new})^T S^{-1}(Y^{new} - Q\Gamma^{21}Y^{new})$, and the extent to which this results in a large change in the shape of the affine component of the PTPS transformation from T. Again we use the Procrustes distance between $Q\Gamma^{21}Y^0$ and $Q\Gamma^{21}Y^{new}$ so as not to penalise changes which are similarity transformations.

Using only the bending energy criterion, the semi-landmarks move away from their nominal positions, if new positions $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$ can be found so that a reduction in bending energy is achieved, i.e. if:

$$tr(Y^{new}\Gamma^{11}Y^{new}) < tr(Y^0\Gamma^{11}Y^0) \text{ or } tr(Y^0\Gamma^{11}Y^0) - tr(Y^{new}\Gamma^{11}Y^{new}) > 0.$$

Here the semi-landmarks move to new positions, only if a reduction in BE greater than some fraction of the size of the corresponding change in the shape of the affine component of the PTPS mapping can be achieved, i.e. if:

$$\frac{1}{\alpha} d_F^2(Q\Gamma^{21}Y^0, Q\Gamma^{21}Y^{new}) + tr(Y^{new}\Gamma^{11}Y^{new}) < \frac{1}{\alpha} d_F^2(Q\Gamma^{21}Y^0, Q\Gamma^{21}Y^0)\, tr(Y^0\Gamma^{11}Y^0)$$

$$\text{or} \qquad tr(Y^0\Gamma^{11}Y^0) - tr(Y^{new}\Gamma^{11}Y^{new}) > \frac{1}{\alpha} d_F^2(Q\Gamma^{21}Y^0, Q\Gamma^{21}Y^{new}). \tag{6.73}$$

Consequently, whether or not the semi-landmarks move will clearly depend on the individual configuration in question, as we illustrate in fig. 6.6 and table 6.3.

Fig. 6.6 shows the results of using (6.72), with various values of α, to determine the new semi-landmark positions for three upper central incisor configurations from the larger test sample already described, with mean $\hat{\mu}$, Procrustes fits denoted as X^P and semi-landmarks and chord directions as in fig 5.12. Once again an optimisation algorithm (see 7.5.2) had to be used to solve (6.72) for each configuration. Fig 6.6 (left) shows the nominal ($Y^0 = X^P$) and new ($Y^{new} = vec^{-1}(vec(Y^0) - U\lambda) = X^{new}$) semi-landmark positions for each configuration and fig 6.6 (right) displays the affine component of each of the PTPS transformations from $T = \hat{\mu}$ to $Y^{new} = X^{new}$, as a deformation grid. Also shown (top row, right) is the affine component of the original PTPS mapping of $T = \hat{\mu}$ to the nominal landmark positions ($Y^0 = X^P$) of each of the three configurations. Table 6.3 details the values of $d_F^2(Q\Gamma^{21}Y^0, Q\Gamma^{21}Y^{new})$ and $tr(Y^{new}\Gamma^{11}Y^{new})$ for each configuration for each choice of α and the values $d_F^2(T, Q\Gamma^{21}Y^0)$ and $tr(Y^0\Gamma^{11}Y^0)$ between $T = \hat{\mu}$ and the original versions, $Y^0 = X^P$ of each of the configurations. Again, the values highlighted in bold correspond to the choice of α which minimises $\left\|Y^{new} - T\right\|^2 = \left\|X^{new} - \hat{\mu}\right\|^2$.

	Config. A		Config. B		Config. C	
	$d_F^2(Q\Gamma^{21}Y^0, Q\Gamma^{21}Y^{new})$	$tr(Y^{new}\Gamma^{11}Y^{new})$	$d_F^2(Q\Gamma^{21}Y^0, Q\Gamma^{21}Y^{new})$	$tr(Y^{new}\Gamma^{11}Y^{new})$	$d_F^2(Q\Gamma^{21}Y^0, Q\Gamma^{21}Y^{new})$	$tr(Y^{new}\Gamma^{11}Y^{new})$
$Y^{new} = Y^0$		0.175		0.048		0.337
α=0.01	**0.000**	**0.112**	0.000	0.037	0.000	0.337
α=0.1 *	0.000	0.104	**0.000**	**0.030**	0.000	0.337
α=0.5	0.006	0.088	0.003	0.018	0.000	0.337
α=1	0.014	0.079	0.006	0.014	0.000	0.337
α=2	0.027	0.071	0.009	0.012	0.000	0.337
$tr(Y^{new}\Gamma^{11}Y^{new})$ only	0.084	0.065	0.014	0.010	0.018	0.007

Table 6.3: Values in affine change/bending energy penalty for three configurations following initial GPA. (Bold) Best choice of α to minimise $\left\|Y^{new} - T\right\|^2 = \left\|X^{new} - \hat{\mu}\right\|^2$. * Best choice of α to minimise $RSS(\hat{\mu})$ or *RSS* for entire sample, following subsequent GPA.

For configuration *C*, table 6.3 and fig. 6.6 show that with each value of α considered there are no new possible positions for the semi-landmarks to move for which (6.73) would be true. For this particular configuration there will always be a difference in the shape between Y and T due to the position of the three gingival landmarks, which we would have wished to remove. For those configurations whose semi-landmarks are able to move (roughly 75% of this sample), an additional problem is again the choice of α.

For configurations A and B, the optimal values of α from those considered for reducing $\left\|Y^{new} - T\right\|^2 = \left\|X^{new} - \hat{\mu}\right\|^2$ were different and, as table 6.3 shows, were also different from the α minimising the variation in shape ($RSS(\hat{\mu})$) for the entire sample.

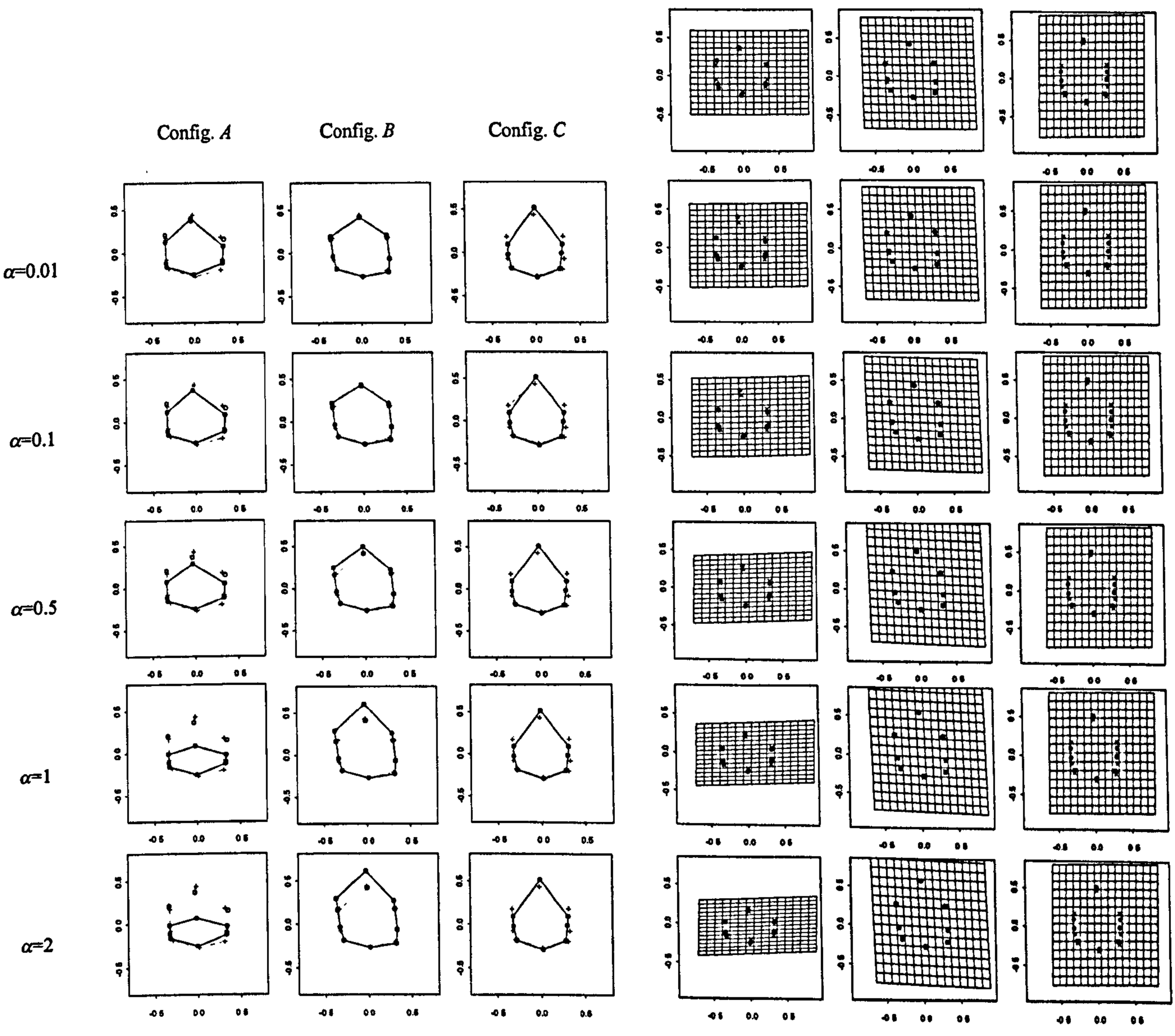


Fig 6.6: Determination of new semi-landmark positions for three cases, following initial superimposition to the mean of a larger sample, using penalty function: $d_F^2(Q\Gamma^{21}Y^0, Q\Gamma^{21}Y^{new}) + \alpha\, tr(Y^{new}\Gamma^{11}Y^{new})$ for different choices of α..

First three columns: (Empty circles) Nominal positions $Y^0 = X^P$ of landmarks when registered to mean shape T (crosses). (Filled circles) new configuration $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda) = X^{new}$ with new positions of semi-landmarks along chords. Second three columns (top row): Affine component of PTPS mapping from T to original (nominal) landmark positions Y^0. Second three columns (except top row): Affine component of optimal PTPS transformation from T to $Q\Gamma^{21}Y^{new}$ corresponding to result for each configuration in first three columns. (Diag. Crosses) Fitted values. (Filled circles) Y^{new}.

6.2.4 *Ad hoc* modifications of the semi-landmark method

One simple device to stop excessive movements of semi-landmarks that was considered, was to set bounds on the extent of allowed movement in each direction. If the distance moved from the nominal positions is greater than that pre-specified then the proposed movement would either be rejected (in which case the semi-landmark does not move) or the new position set at the bound of movement exceeded. However, with sensible bounds set according to values from the dental literature and experience of periodontologists, use of the original (minimum bending energy) semi-landmark method just produced a new sample of configurations of which a large proportion of gingival landmarks either did not move or immediately moved to the pre-specified bounds. This does nothing to remove the unwanted variation in shape between configurations due to differences in the positions of the gum.

6.3 Modifications not using PTPS mappings

Rather than considering new penalty functions based on the PTPS mapping used in Bookstein's (1996a,d,e) original semi-landmark method, we could consider other superimposition methods and optimisation criteria to determine the new positions of semi-landmarks along their chords.

In most of our previous assessments we have used quantities such as the Euclidean sum of squared distances between semi-landmarks and their corresponding positions in the mean shape and the (sum of) squared full Procrustes distances of new configurations about the mean, to judge the effectiveness of each proposed method. Here we consider use of such criteria directly for determining the new positions of semi-landmarks. Following prior registration by GPA, the new position of each selected semi-landmark, along its chord, could be calculated as simply the point at nearest Euclidean distance to the corresponding landmark in the current estimate of shape. Alternatively we could find new positions which optimise the full Procrustes distance or ordinary sums of squares over rotations, scales and translations between the configurations as the semi-landmarks vary along their chords.

The final part of this section then suggests other possible transformations and optimisation criteria which may be worth future investigation for not only the current problem, but also other applications.

6.3.1 Points at least Euclidean distance

As in 5.4.1.1, let Y^0 denote the nominal positions of the j=1,...,k landmarks in configuration Y of which a sublist $j_{(l)}$, $l = 1,...,L$ are free to move along pre-specified (unit) directions $u_{j_{(l)}} = (u_{jx_{(l)}}\ u_{jy_{(l)}})^T$. The equation of the chord through $(y^0_{jx_{(l)}}, y^0_{jy_{(l)}})$ may be written as:

$$y_{jy} = m_{dir}\ y_{jx} + k_{dir}$$

$$\text{where } m_{dir} = \frac{u_{jy_{(l)}}}{u_{jx_{(l)}}} \text{ and } k_{dir} = y^0_{jy_{(l)}} - m_{dir}\ y^0_{jx_{(l)}}.$$

The line perpendicular to this direction through the corresponding landmark (t_{jx}, t_{jy}) in the reference configuration T is given by:

$$y_y = m_{perp}\ y_x + k_{perp},$$

$$\text{with } m_{perp} = \frac{-1}{m_{dir}} \text{ and } k_{perp} = t_{jy} - m_{perp}\ t_{jx}.$$

The two lines will intersect at the point on the semi-landmark's chord at least Euclidean distance from landmark (t_{jx}, t_{jy}) with the new coordinates $(y^{new}_{jx_{(l)}}, y^{new}_{jy_{(l)}})$ given by:

$$\left(\frac{k_{dir} - k_{perp}}{m_{perp} - m_{dir}},\ m_{dir}\left(\frac{k_{dir} - k_{perp}}{m_{perp} - m_{dir}}\right) + k_{dir}\right).$$

For example, in fig. 6.7, semi-landmark $j_{(l)}$ in configuration Y is allowed to move away from its original position, $(y^0_{lx_{(l)}}, y^0_{jy_{(l)}})$, along its 'escribed chord', parallel to a line joining the nominal positions (y^0_{ax}, y^0_{ay}) and (y^0_{bx}, y^0_{by}) of neighbouring landmarks a and b, with $m_{dir} = y^0_{by} - y^0_{ay} / y^0_{bx} - y^0_{by}$.

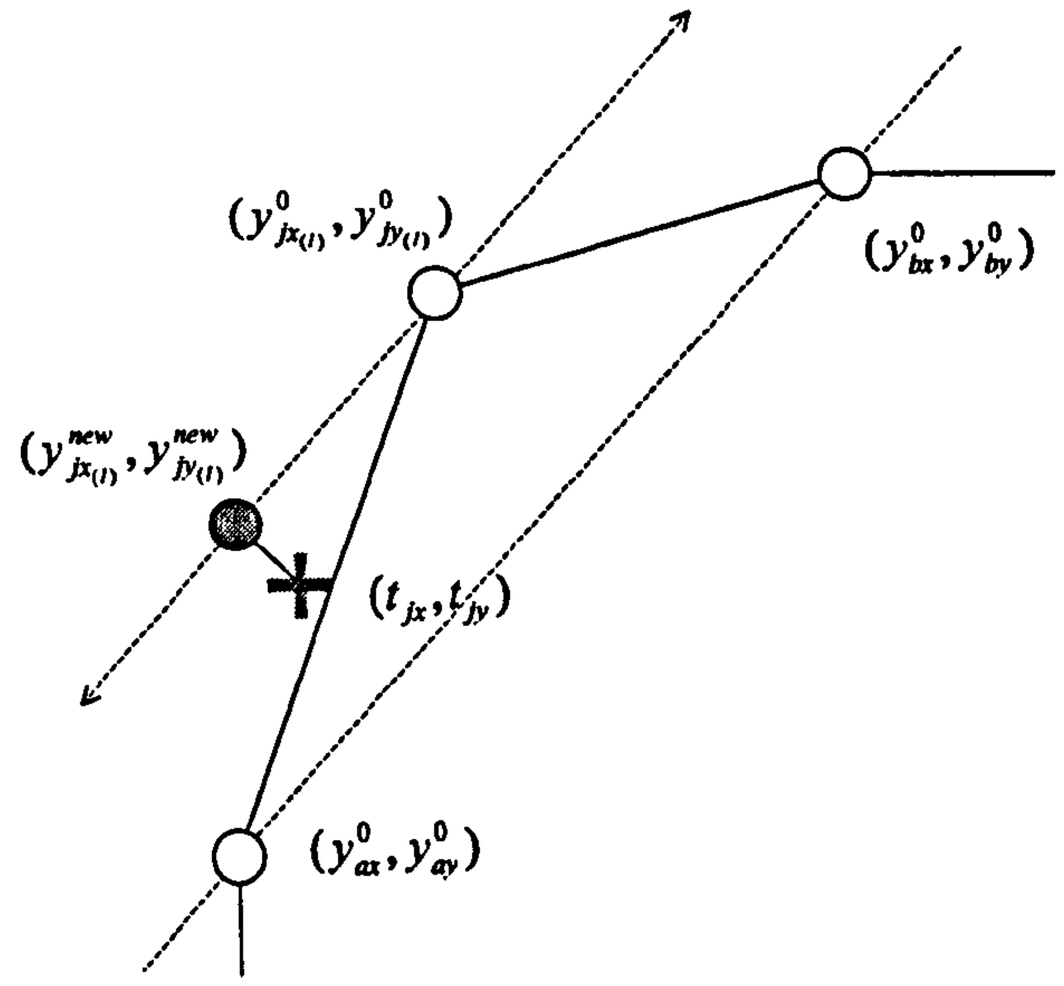


Fig 6.7 (Empty circles) Nominal positions of semi-landmark $j_{(l)}$ and neighbouring landmarks *a* and *b*, defining direction of escribed chord though $(y^0_{jx_{(l)}}, y^0_{jy_{(l)}})$ (dashed line, arrows). (Filled circle) Point on chord, at nearest Euclidean distance to corresponding landmark (t_{jx}, t_{jy}) in reference shape (cross).

For an entire configuration with target form $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$, where U is the matrix of pre-specified chord directions as defined in section 5.4.1.1 and λ a vector of unknown scalars of length L (to be determined), the new positions of all semi-landmarks with reference to a configuration T, may again be written in a least squares formulation. For each of the $j_{(l)}$, l=1,...,L, semi-landmarks and associated (unit) chord directions $u_{j_{(l)}} = (u_{jx_{(l)}}\ u_{jy_{(l)}})^T$, the new 'nearest point' position is given by:

$$\begin{pmatrix} y^{new}_{jx_{(l)}} \\ y^{new}_{jy_{(l)}} \end{pmatrix} = \begin{pmatrix} y^0_{jx_{(l)}} \\ y^0_{jy_{(l)}} \end{pmatrix} - \lambda_l \begin{pmatrix} u_{jx_{(l)}} \\ u_{jy_{(l)}} \end{pmatrix}$$

and may be determined by finding each λ_l, l=1,...,L, satisfying:

$$\begin{aligned} \min_{\lambda_l} \sqrt{(y^{new}_{jx_{(l)}} - t_{jx})^2 + (y^{new}_{jy_{(l)}} - t_{jy})^2} &= \min_{\lambda_l} \sqrt{(y^0_{jx_{(l)}} - u_{jx_{(l)}}\lambda_l - t_{lx})^2 + (y^0_{iy_{(r)}} - u_{ji_{(l)}}\lambda_l - t_{jy})^2} \\ &= \min_{\lambda_r} (y^0_{jx_{(l)}} - u_{jx_{(l)}}\lambda_l - t_{jx})^2 + (y^0_{jy_{(l)}} - x_{jy_{(r)}}\lambda_l - t_{jy})^2 . \end{aligned} \tag{6.74}$$

For the entire set of semi-landmarks, the sum of each of these terms may be written in ordinary least squares (OLS) form as:

$$\begin{aligned} \min_{\lambda} tr(Y^{new} - T)^T (Y^{new} - T) &= \min_{\lambda} (vec(Y^0) - U\lambda - vec(T))^T (vec(Y^0) - U\lambda - vec(T)) \\ &= \min_{\lambda} (vec(Y^0) - vec(T) - U\lambda)^T (vec(Y^0) - vec(T) - U\lambda) \end{aligned} \tag{6.75}$$

with solution:

$$\lambda = (U^T U)^{-1} U^T (vec(Y^0) - vec(T)). \tag{6.76}$$

Fig 6.8 shows the results of the first movement stage for an upper central incisor, with $Y^0 = X^P$, its Procrustes fit to the initial estimate of mean shape, $T = \hat{\mu}$, obtained from an entire sample of such configurations. As usual the semi-landmarks are the three gingival landmarks, with chord directions as defined in fig. 5.12, representing variation due to the position of a patient's gum. Note that while the new positions of the semi-landmarks in $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda) = X^{new}$ minimise (6.75), they do not optimise the OLS superimposition of Y^{new} to T. (We consider use of this criterion in the following section). If we were to re-superimpose Y^{new} to T, the semi-landmarks would not be at the nearest points along their chords, minimising the Euclidean distances to their corresponding landmarks in T.

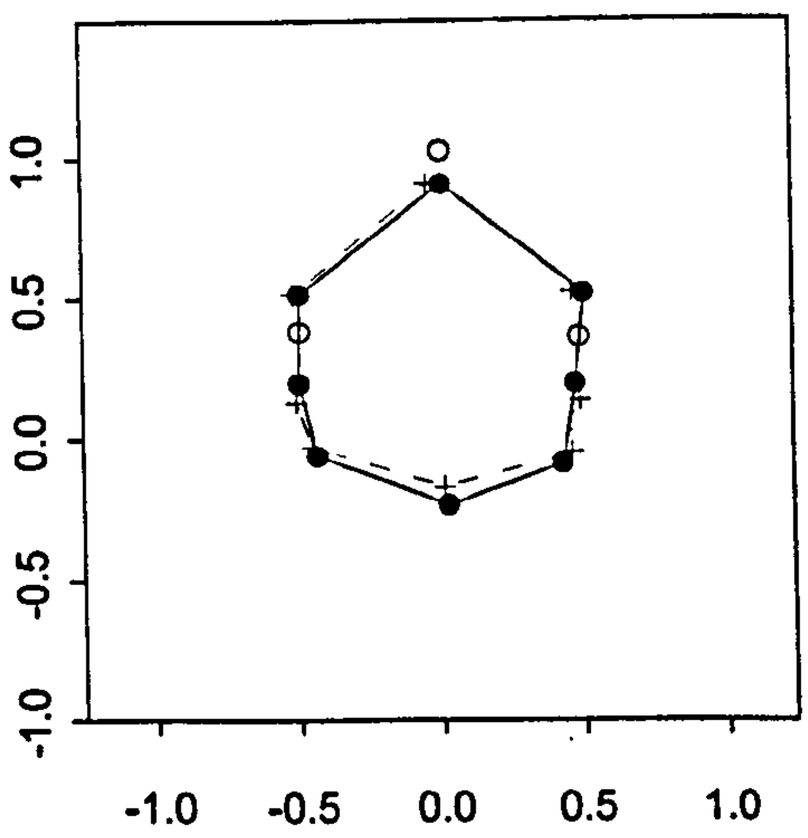


Fig 6.8: (Empty circles) Nominal positions ($Y^0 = X^P$) of landmarks when configuration registered to estimated mean shape $T = \hat{\mu}$ (crosses).
(Filled circles) New version of configuration ($Y^{new} = vec^{-1}(vec(Y^0) - U\lambda) = X^{new}$) with semi-landmarks at the nearest points along each chord (dotted lines), to the corresponding landmarks in T.

After calculating new semi-landmark positions for each configuration in a sample, a new mean and Procrustes fits are obtained and the process repeated until the landmarks stop moving. Iteration is important here since even if the mean is unchanged, two configurations differing only in the positions of their semi-landmarks along the same chord directions will not end up being the same shape in just one iteration. Where the semi-landmarks move to is dependent on the prior alignment of Y^0 to T. Any differences in the original configurations will result in a different Procrustes registration of each configuration to $\hat{\mu}$ with semi-landmarks then moving to different positions on

their chords. Over the course of the iterations however, we would expect any differences along the chord directions to be eventually filtered out.

Fig 6.9 shows the result of running this new semi-landmark procedure iteratively using the nearest point criterion, on the test sample of configurations referred to above, until convergence at $RSS(\hat{\mu})<0.001$. GPA steps were performed using the method of Gower (1975) modified by Rohlf & Slice (1990) (and so $RSS(\hat{\mu})=RSS$). As can be seen, all variation of the semi-landmarks is removed along the directions of the their chords (as defined in fig. 5.12), making this the most promising modification of the semi-landmark method so far, to take forward for further investigation.

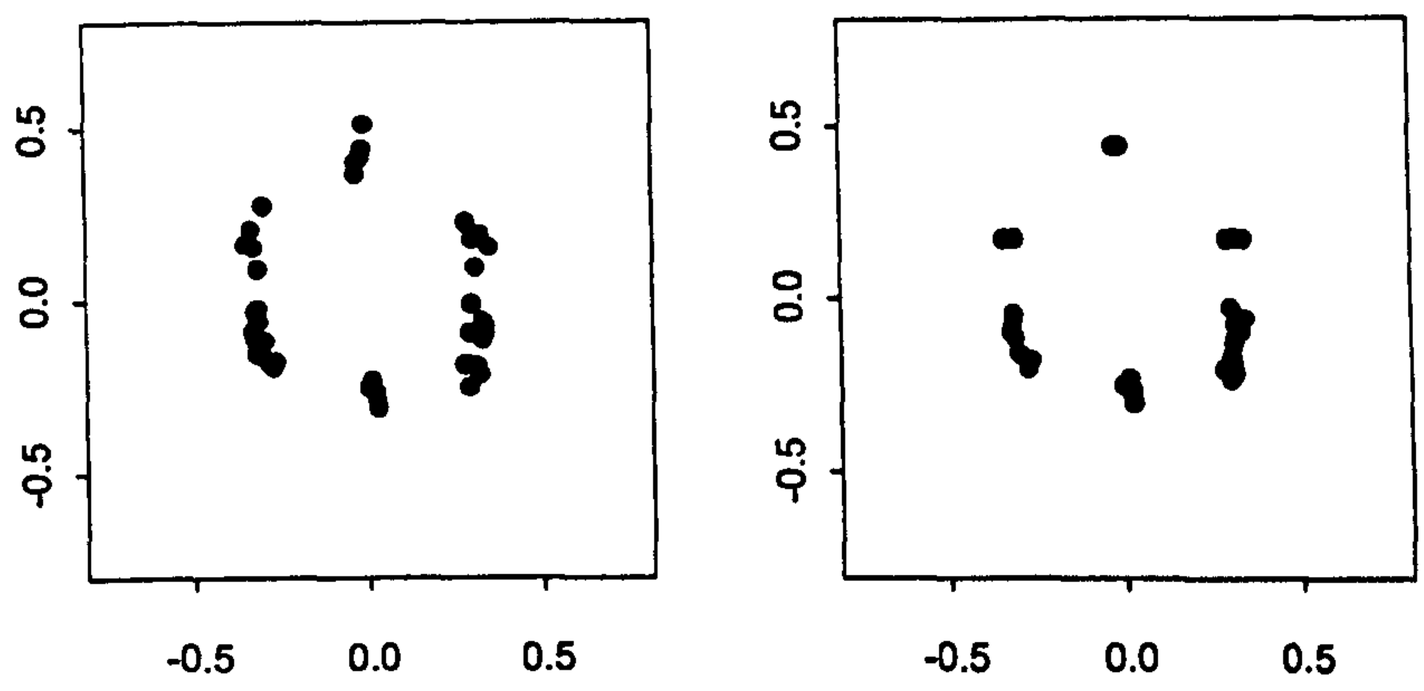


Fig 6.9: (Left) Original Procrustes fits (all landmarks fixed). (Right) Procrustes fits following iterative semi-landmark procedure, using nearest point criteria

In contrast to Bookstein's original method and any of the new proposals considered so far, the new positions of the semi-landmarks are determined by a 'local' rather than 'global' criterion at each movement step, since they depend only the position of the corresponding of the landmark in T, rather than on some function of all landmarks in Y^0 and T. However, with each of preceding methods based on PTPS transformations and the new proposal in 6.3.2, there was no dependency on where the semi-landmarks move to on the prior alignment of the two configurations, whereas here the new positions will be influenced by the registration of Y^0 and T, as noted above. One could therefore argue that the nearest point criterion is still dependent on all landmarks in T and Y^0 since it depends on the 'global' method by which they were previously aligned. We consider how differences in the sizes of $\hat{\mu}$ and X^P, produced by the different GPA methods, affects the performance of the nearest point criterion in 7.3.2.

This new procedure has some similarities with the iterative closest point method of Besl & McKay (1992). After repeated superimposition to the same unchanging mean or reference shape, the new positions of all landmarks are determined at the closest points on chords between many equally spaced landmarks around the outline of an object, to those in the reference shape. Following our exploration of the nearest point method in here, a similar technique, making use of a 'least Euclidean distance' criterion, has also been used by Andresen *et al.* (2002), where perpendicular projection onto a template configuration was used to assign homologous points on the outlines of human mandibles from computerised tomography (CT) scans. The procedure presented here however, allows one to make use of the known homologies in the fixed landmarks during prior alignment, recognising that only a subset of the landmarks are located without certainty. It also allows the user to define chords in any directions as the potential new positions of landmarks, rather than only between the equally spaced points approximating the outline of an object.

In section 6.4.2, we describe other methods where the potential new positions of landmarks fall along a string of outline co-ordinates, rather than along chords, where a nearest point criterion can be used to assign point-to-point homologies. Use of a string of coordinates to represent the possible new positions of semi-landmarks would clearly be more accurate than using chords to approximate an outline. However, for the gingival variation problem, complete outline information is unavailable as some of the possible new positions of the landmarks along the sides of the tooth are obscured by gingival tissue. Consequently we must approximate these positions by chord directions, using information based on only the visible parts of the outline in these regions.

6.3.2 Full Procrustes criterion

Another new option for where the semi-landmarks move to could be to minimise the OLS between $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$ and $T = \hat{\mu}$, under the action of the similarity group of transformations (SG) and as the semi-landmarks vary along their chords, i.e. we could seek:

$$\min_{SG,\lambda} \left\| SG(Y^{new}) - T \right\|^2 = \min_{SG,\lambda} \left\| SG(vec^{-1}(vec(Y^0) - U\lambda) - T \right\|^2 \qquad (6.77)$$

or, in keeping with the original semi-landmark method, where the PTPS transformation is of T to Y, seek:

$$\min_{SG,\lambda} \left\| SG(T) - Y^{new} \right\|^2 = \min_{SG,\lambda} \left\| SG(T) - (vec^{-1}(vec(Y^0) - U\lambda) \right\|^2. \qquad (6.78)$$

Unlike the nearest point method, where (squared) Euclidean distances to the corresponding landmarks in T are minimised as each landmark moves along its chord, here the configuration $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$ may also be rotated, scaled and translated, with all new landmark positions determined simultaneously.

Although they may produce different values, both functions will result in the semi-landmarks moving to positions which produce the same shape for Y^{new}, so in practice it does not matter which we use. Since we usually talk about the individual configuration Y being superimposed to the reference shape T (here the estimated mean $\hat{\mu}$), we will consider the first option (6.77).

In contrast to the nearest point method, here the prior registration of Y and T is irrelevant since these differences are filtered out by the penalty function. Although differences in the scale of T and Y would produce different values of (6.77) this would have no effect on the resulting shape of Y^{new}. Therefore it also does not matter which method of GPA is used to obtain $T = \hat{\mu}$ and the corresponding Procrustes fits.

In trying to obtain the solution to (6.77), we have the problem of finding:

$$\min_{\beta,\gamma,\theta,\lambda} \left\| \beta vec^{-1}(vec(Y^0) - U\lambda)\Gamma(\theta) + 1_k \gamma^T - T \right\|^2 \qquad (6.79)$$

where β is a scalar, γ a translation vector and $\Gamma(\theta)$ a 2×2 rotation matrix, as described in 3.2.1. In 3.2.1, the OLS superimposition of one configuration to another was found by first centring both configurations, then rotating one configuration so as to minimise the sum of squared Euclidean distances between corresponding landmarks and finally re-sizing one configuration to minimise this measure further. The difficulty here is that as the semi-landmarks move (by λ), the centre of one of the configurations changes,

along with the optimal rotation and scale required to minimise the sum of squared distances between corresponding landmarks. (Note also that the optimal rotation $\Gamma(\hat{\theta})$ in now no longer independent of the optimal scale $\hat{\beta}$ since both depend on λ). Consequently an explicit solution to (6.79) is difficult to find and so instead we again make use of the optimisation algorithm for non-linear functions described in 7.5.2.

Note that the resulting configuration $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$ is not the actual OLS superimposed version of Y^{new} to $T = \hat{\mu}$ minimising (6.78), as illustrated in fig 6.10. This would be given by:

$$\hat{\beta} vec^{-1}(vec(Y^0) - U\lambda)\Gamma(\hat{\theta}) + 1_k \hat{\gamma}^T \tag{6.80}$$

where $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\theta}$ are the values of β, γ and θ optimising (6.77). However, both Y^{new} and $\hat{\beta} Y^{new} \Gamma(\hat{\theta}) + 1_k \hat{\gamma}^T$ have exactly the same shape and on the following step of the iterative procedure, the same new Procrustes fit to the updated mean will result.

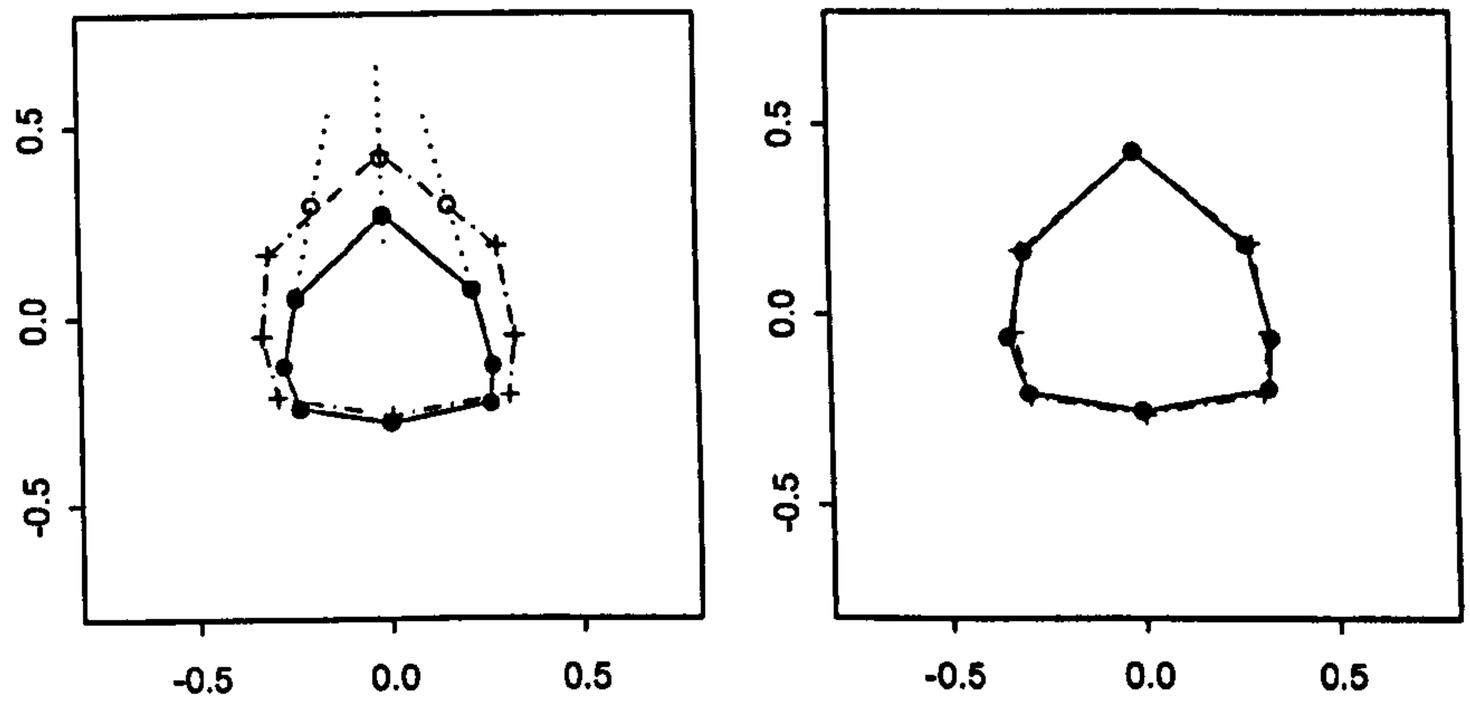


Fig 6.10: (Left) (Empty circles) Nominal positions ($Y^0 = X^P$) of landmarks when configuration registered to mean shape $T = \hat{\mu}$ (crosses). (Filled circles) New version of configuration ($Y^{new} = (vec^{-1}(vec(Y^0) - U\lambda) = X^{new}$) with semi-landmark positions moving along chords (dotted lines) to minimise $\left\| SG(Y^{new}) - \hat{\mu} \right\|^2$, as shown in panel on right. (Right) (Filled circles) OLS superimposition of Y^{new} to T (crosses).

Note also that the optimised value of penalty function (6.77) is not the full Procrustes distance between $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$ and $T = \hat{\mu}$ unless T is first scaled to unit size. If the full Procrustes mean shape $\hat{\mu}$ has been obtained by the GPA method (i),

described in 3.2.2.1, using the constraint $\|\mu\|=1$, then this will already be true. Otherwise to optimise the squared full Procrustes distance between $Y^{new} = vec^{-1}(vec(Y^0)-U\lambda)$ and T, as the semi-landmarks move along their chords, we would instead seek:

$$\min_{\lambda} d_F^2(vec^{-1}(vec(Y^0)-U\lambda),T) =$$

$$\min_{SG,\lambda} \left\| SG(vec^{-1}(vec(Y^0)-U\lambda)-\frac{T}{\|T\|} \right\|^2 = \min_{\beta,\gamma,\theta,\lambda} \left\| \beta vec^{-1}(vec(Y^0)-U\lambda)\Gamma(\theta)+1_k\gamma^T-\frac{T}{\|T\|} \right\|^2 . \quad (6.81)$$

Again, note that despite having the same shape, $Y^{new} = vec^{-1}(vec(Y^0)-U\lambda)$, is not the actual full Procrustes fit of Y^{new} to $T/\|T\|$ minimising (6.81), which would be given by (6.80), with $\hat{\beta}$, $\hat{\gamma}$ and $\hat{\theta}$ optimising (6.81) above. In addition, the resulting vector solution λ and therefore shape of Y^{new} is the same regardless of whether we seek to optimise (6.77) or (6.81) and again the initial registration of T and Y does affect the resulting shape of Y^{new}. With T scaled to unit size in (6.81), the only difference is that a different value of the penalty function results.

Fig 6.11 shows the result of running this new semi-landmark procedure with full Procrustes criterion on the same sample of configurations (with same semi-landmarks and chord directions) used in fig 6.10, until convergence at $\Delta RSS(\hat{\mu})<0.001$, using the iterative GPA method. The end results are almost identical to those produced using the new nearest point method, with all variation at the positions of the semi-landmarks removed in directions along their chords. At this level of convergence the two final sets of Procrustes fits are identical to 2 decimal places. However, it is difficult to check whether or not the two sets of data would be identical for smaller values of $\Delta RSS(\hat{\mu})$, because the slow convergence issue, found to occur with the bending energy method in 5.5.5, also occurs with these new criteria. (We investigate this further in 7.4). One notable difference between the two methods however, was that convergence at $\Delta RSS(\hat{\mu})<0.001$ took only 2 iterations using the full Procrustes criterion, whereas the nearest point method took 7 iterations.

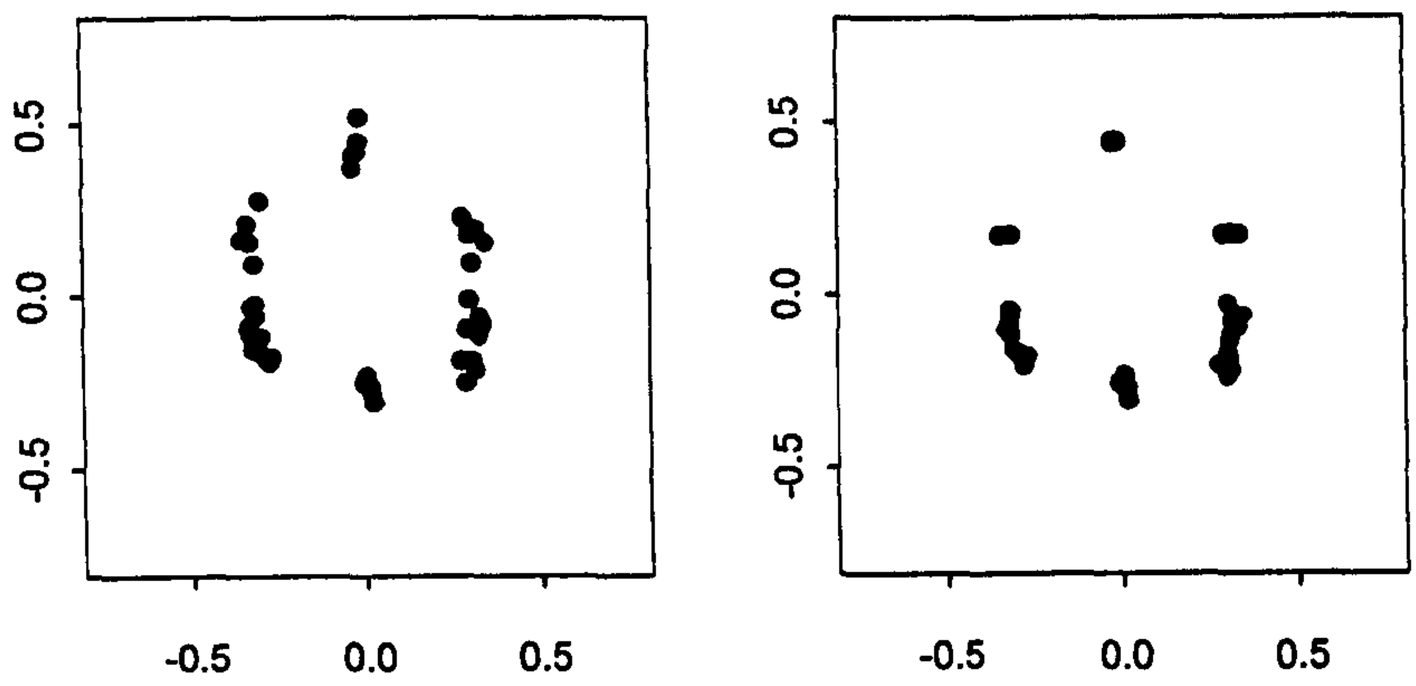


Fig 6.11: (Left) Original Procrustes fits (all landmarks fixed). (Right) Procrustes fits following iterative semi-landmark procedure, using full Procrustes/OLS criterion

So, we have another potentially useful new criterion for the determination of new semi-landmark positions to investigate, along with the nearest point method. Although the results of using each criterion are practically identical for this particular test sample and choice of semi-landmarks and chord directions, there may be other particular shapes, combinations of landmarks, semi-landmarks and chord directions for which the two methods do not give such similar results.

Finally note that if we were to run the nearest point method iteratively within each semi-landmark step, i.e. for each configuration:

(i) Find new semi-landmark positions optimising (6.8)

(ii) Re-superimpose the resulting Y^{new} to unchanged $\hat{\mu}$ by OLS

(iii) Set $Y^0 = Y^{new}$ and repeat from (i) until the semi-landmarks stop moving,

this would generally give the same new semi-landmark positions and configuration Y^{new} as using the full Procrustes criterion (6.81) or OLS criterion (6.77) at each step, providing the level of convergence used in (iii) is consistent with that set within the optimisation algorithm used to solve (6.77) or (6.81).

6.3.3 Other semi-landmark methods

Obviously there are many other possible transformations and novel optimisation criteria we could consider for the determination of new semi-landmark positions when matching $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$ to $T = \hat{\mu}$ (or vice-versa). In this section we

have considered two Euclidean metrics, the nearest point and full Procrustes criteria, as alternatives to using the original BE method or the various 'roughness' penalties proposed in section 6.2. Here configurations were matched to the mean shape over similarity transformations (or following such a match), whereas in section 6.2, the new objective criteria were optimised over pairs of splines or kriging functions from the reference (mean) shape.

Another simple criterion which may be relevant in applications other than those here, could be to move the landmarks to positions which match some feature such as a ratio of distances, or angle between a set of landmarks in the reference (mean) shape. Another option could be to determine new semi-landmark positions through values of other covariates, for which there is an established link with shape.

Analogous to the smoothing spline in 6.2.1, we could also consider smoothed Procrustes matching (see for example Dryden & Mardia, 1998). For a smoothed superimposition of (say) Y to T, the objective function to be optimised is of the form:

$$\min_{SG} \|SG(Y) - T\|^2 + \alpha J(SG(Y) - T) \tag{6.82}$$

or

$$\min_{SG} \|SG(T) - Y\|^2 + \alpha J(SG(T) - Y) \tag{6.83}$$

where J is some roughness penalty and α some pre-specified scalar. This has a link with weighted Procrustes methods (see section 6.4.1), since for $J(A) = trace(A^T \Omega^{-1} A)$, (6.82) becomes:

$$\min_{SG} (SG(Y) - T)^T (I + \alpha\Omega^{-1})(SG(Y) - T).$$

If we allow some of the landmarks in Y to move away from their original positions to locations $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$ then (6.83) for example, becomes:

$$\min_{SG,\lambda} \|SG(T) - vec^{-1}(vec(Y^0) - U\lambda)\|^2 + \alpha J(SG(T) - vec^{-1}(vec(Y^0) - U\lambda)).$$

6.4 Other novel methods allowing for lack of landmark correspondence

In this penultimate section, we discuss some of the other existing variants of Procrustes and spline superimposition (other than Bookstein's semi-landmark method) that have been proposed for use when there is lack of precise correspondence of landmarks along particular directions or where one would wish to account for the effects of certain unwanted differences in shape in some way.

The first section describes weighted Procrustes analysis, where a more generalised covariance structure is allowed when matching configurations over the similarity group, rather than an isotropic one, as implicitly assumed by the use of an ordinary sums of squares criterion. For a known or pre-specified covariance metric, the methods require only a simple adaptation of the OPA and GPA formulae and procedures described in chapter 3 and for the superimposition of one configuration to another, we outline how the technique may be modified to also include semi-landmarks. However, in practice the covariance at and between landmarks needs to be estimated and this is still very much an ongoing area of development.

The second section then considers Green's (1996) extension of the semi-landmark method to include any combination of curves and landmarks that can be identified as homologous between cases. Rather than using landmarks which may move along chords, here the parts of an outline or curving feature along which a (semi-) landmark is thought to lie are instead represented in their entirety by strings of coordinates and utilised in a manner consistent with the treatment of landmarks. Also described are the ideas of Sampson *et al.* (1996) (which may be incorporated into one of the stages of Green's method), who used the idea of normal projection from landmarks to points on closed outline as a way of determining points of correspondence.

6.4.1 Weighted Procrustes analysis

Rather than using an iterative semi-landmark procedure to remove directions and patterns of unwanted variation in shape, we could instead incorporate information on

the different covariance structures at each of the landmarks (and between the landmarks) directly into the registration procedure and subsequent analysis. Here the selected landmarks do not move from their nominal positions (although we propose a new method of determining semi-landmark positions in this way in section 6.4.1.1). Instead the directions of poor correspondence and uncertainly at these positions are taken into account when obtaining a Procrustes registration of the data, so that their effect on the estimates of covariance at the other, more reliable landmarks is reduced.

6.4.1.1 Weighted Procrustes superimposition of two configurations

For two configurations T and Y we saw in chapter 3 how the ordinary Procrustes superimposition of (say) Y to T is obtained by seeking:

$$\min_{SG} tr(SG(Y)-T)^T(SG(Y)-T) = \min_{SG} vec(SG(Y)-T)^T vec(SG(Y)-T)$$

$$= \min_{\beta,\Gamma,\gamma\in SG} vec(\beta Y\Gamma(\theta)+1_k\gamma^T-T)^T vec(\beta Y\Gamma(\theta)+1_k\gamma^T-T) \qquad (6.84)$$

where β, $\Gamma(\theta)$ and $1_k\gamma^T$ are a scalar, rotation matrix and translation vector respectively. Implicitly, the superimposition assumes isotropic errors (the same variance in all directions) at each of the landmarks, and so any dissimilarities in shape are distributed to the residuals of all pairs of corresponding landmarks. Often this can make interpretation misleading, masking the nature of the actual difference in shape. For example, for the upper central incisor configurations, any differences in gingival landmark locations due to the position of a patient's gum (which are of no interest) will be distributed over the residuals of all landmarks in the configuration.

For weighted Procrustes matching, a $2k\times 2k$ pre-specified superimposition weighting matrix Σ_S is introduced into (6.84) and the problem becomes one of matching configurations by GLS. Estimates of β, $\Gamma(\theta)$ and γ are then obtained by seeking:

$$\min_{SG} vec(SG(Y)-T)^T\Sigma_S^{-1}vec(SG(Y)-T) =$$

$$\min_{\beta,\Gamma,\gamma\in SG} vec(\beta Y\Gamma(\theta)+1_k\gamma^T-T)^T\Sigma_S^{-1}vec(\beta Y\Gamma(\theta)+1_k\gamma^T-T). \qquad (6.85)$$

When Σ_S is pre-specified, this simply involves replacing $\|A\| = \{vec(A)^T vec(A)\}^{1/2}$ with $\|A\|_{\Sigma_S} = \{vec(A)^T \Sigma_S^{-1} vec(A)\}^{1/2}$ in formulae (3.1) to (3.3) for the OLS Procrustes superimposition of two configurations (Dryden & Mardia, 1998). See also, for example, Goodall (1991) who describes the problem in m=2 dimensions, as a weighted least squares multiple regression problem with $2k$ observations, one response variable, $vec(T)$, and solution for a vector of four parameters, corresponding to each coordinate of γ, $\beta\cos\theta$ and $\beta\sin\theta$. Note that even though (6.85) represents a GLS problem, we use the term 'weighted' to avoid confusion with a 'generalised' Procrustes matching of multiple objects.

In a regression setting the matrix Σ_S corresponds to the variance of the residuals of $vec(T) - vec(\beta Y\Gamma(\theta) + 1_k\gamma^T)$. If $T=\hat{\mu}$, the full Procrustes estimate of mean shape, then as we shall see in the following section, Σ_S may be chosen to correspond to the population or sample covariance structure (defined at $\hat{\mu}$), assumed to incorporate actual variation in shape as well as that due to errors. Alternatively, Σ_S may be chosen to represent the result of combining the error covariance matrices of Y and T (at T), as described in section 6.2.1.2, for the anisotropic smoothing spline (see below).

Incorporating a more generalised covariance structure, into the matching procedure, allows the relative accuracies and confidence of correspondence between different landmarks to be represented. Landmarks and directions along which unwanted or uninteresting differences in location occur can be given less weight during the superimposition, so that those regarded as more reliable are matched preferentially. In addition Σ_S can also be specified to include the covariance between landmarks as well as at each location. For example, in m=2 dimensions, the errors or uncertainties at each landmark j could be represented by elliptical covariance matrices in the same way as in the anisotropic smoothing spline in 6.2.1.2, with

$$\Sigma_j = \begin{pmatrix} \varepsilon_{xx,j} & \varepsilon_{xy,j} \\ \varepsilon_{yx,j} & \varepsilon_{yy,j} \end{pmatrix} \text{ with } \varepsilon_{xy,j} = \varepsilon_{yx,j}$$

and Σ_S specified as a block diagonal matrix:

$$\Sigma_S = \begin{pmatrix} \varepsilon_{xx,1} & 0 & \cdots & 0 & \varepsilon_{xy,1} & 0 & \cdots & 0 \\ 0 & \varepsilon_{xx,2} & 0 & \vdots & 0 & \varepsilon_{xy,2} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 & \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \varepsilon_{xx,k} & 0 & \cdots & 0 & \varepsilon_{xy,k} \\ \varepsilon_{yx,1} & 0 & \cdots & 0 & \varepsilon_{yy,1} & 0 & \cdots & 0 \\ 0 & \varepsilon_{yx,2} & 0 & \vdots & 0 & \varepsilon_{yy,2} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 & \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \varepsilon_{yx,k} & 0 & \cdots & 0 & \varepsilon_{yy,k} \end{pmatrix}.$$

If we allow Y to be of the form $vec^{-1}(vec(Y^0)-U\lambda)$, then the objective function in (6.85) becomes:

$$\min_{SG} vec(SG(vec^{-1}(vec(Y^0)-U\lambda))-T)^T \Sigma_S^{-1} vec(SG(vec^{-1}(vec(Y^0)-U\lambda))-T))$$

allowing a weighted Procrustes superimposition to be used for the determination of new semi-landmark positions for any pre-specified Σ_S. Alternatively, (6.85) may be re-specified as a superimposition of T to Y:

$$\min_{SG} vec(SG(T)-Y)^T \Sigma_S^{-1} vec(SG(T)-Y). \quad (6.86)$$

For $\Sigma_S=(I_2 \otimes S)$, as defined in 5.3.5.2, (6.86) is then the objective function for a GLS affine superimposition of T to Y, with covariance $(I_2 \otimes S)$ (corresponding to the affine component/fit of a PTPS transformation), but with the affine parameters constrained to be a similarity transformation, as suggested in 6.2.3.1. If Y is of the form $vec^{-1}(vec(Y^0)-U\lambda)$, then the new positions of the semi-landmarks could be found by solving (for β, $\Gamma(\theta)$, γ and λ simultaneously):

$$\min_{\beta,\Gamma,\gamma \in SG,\lambda} vec(SG(T)-vec^{-1}(vec(Y^0)-U\lambda))^T S^{-1} vec(SG(T)-vec^{-1}(vec(Y^0)-U\lambda)).$$

Note that as with the full Procrustes criterion in 6.3.2, it does not matter whether we consider a weighted superimposition of T to $Y^{new}=vec^{-1}(vec(Y^0)-U\lambda)$ or *vice versa*. Both will result in $vec^{-1}(vec(Y^0)-U\lambda)$ being the same shape.

6.4.1.2 Weighted registration of a sample of configurations

Recall from chapter 3 that in an unweighted GPA, the configurations X_i, i=1,...,n, (each $k\times m$) are assumed to be perturbations from a multivariate normal model, with mean μ and covariance matrix $\sigma^2 I_{km}$, which have then been translated, rotated and re-scaled (by vector $1_k\gamma_i'^T$, matrix Γ_i' and scalar β_i') and observed as each X_i, i.e.

$$X_i = \beta_i' (\mu + E_i) \Gamma_i' + 1_k\gamma_i'^T \tag{6.87}$$

where $vec(E_i) \sim N(0, \Sigma = \sigma^2 I_{km})$. If β_i, Γ_i and $1_k\gamma_i^T$ are the reverse similarity transformations of $1_k\gamma_i'^T$, Γ_i' and β_i' then equivalently:

$$vec(\beta_i X_i \Gamma_i + 1_k\gamma_i^T) \sim N(vec(\mu), \Sigma = \sigma^2 I_{km}). \tag{6.88}$$

Estimates of μ and β_i, Γ_i and $\gamma_i \in SG$ for each configuration are found by seeking:

$$\min_{SG_i, \hat{\mu}} \sum_{i=1}^{n} vec(SG_i(X_i) - \hat{\mu})^T vec(SG_i(X_i) - \hat{\mu})$$

$$= \min_{\hat{\mu}, \beta_i, \Gamma_i, \gamma_i} \sum_{i=1}^{n} vec(\beta_i X_i \Gamma_i + 1_k\gamma_i^T - \hat{\mu})^T vec(\beta_i X_i \Gamma_i + 1_k\gamma_i^T - \hat{\mu}). \tag{6.89}$$

Fig 6.11 shows how for a sample of upper central incisors, variation in the positions of the gingival landmarks is distributed around the residuals of all landmarks of the Procrustes fits, when using an unweighted (ordinary) GPA. The left panel shows five upper central incisor configurations, constructed from the same crown outline of an extracted tooth in Wheeler (1962), using the usual k=8 landmarks as defined in fig.1.3. With the aid of an experienced dentist, the five non-gingival landmarks were located in identical positions for each configuration whereas the three gingival landmarks were recorded in five different sets of locations, corresponding to possible positions of the patient's gingival tissue. The plot on the right shows the unweighted GPA registration of the configurations, which appears to suggest considerable variation in the shape of the lower outline of the buccal surface, when in fact there is none.

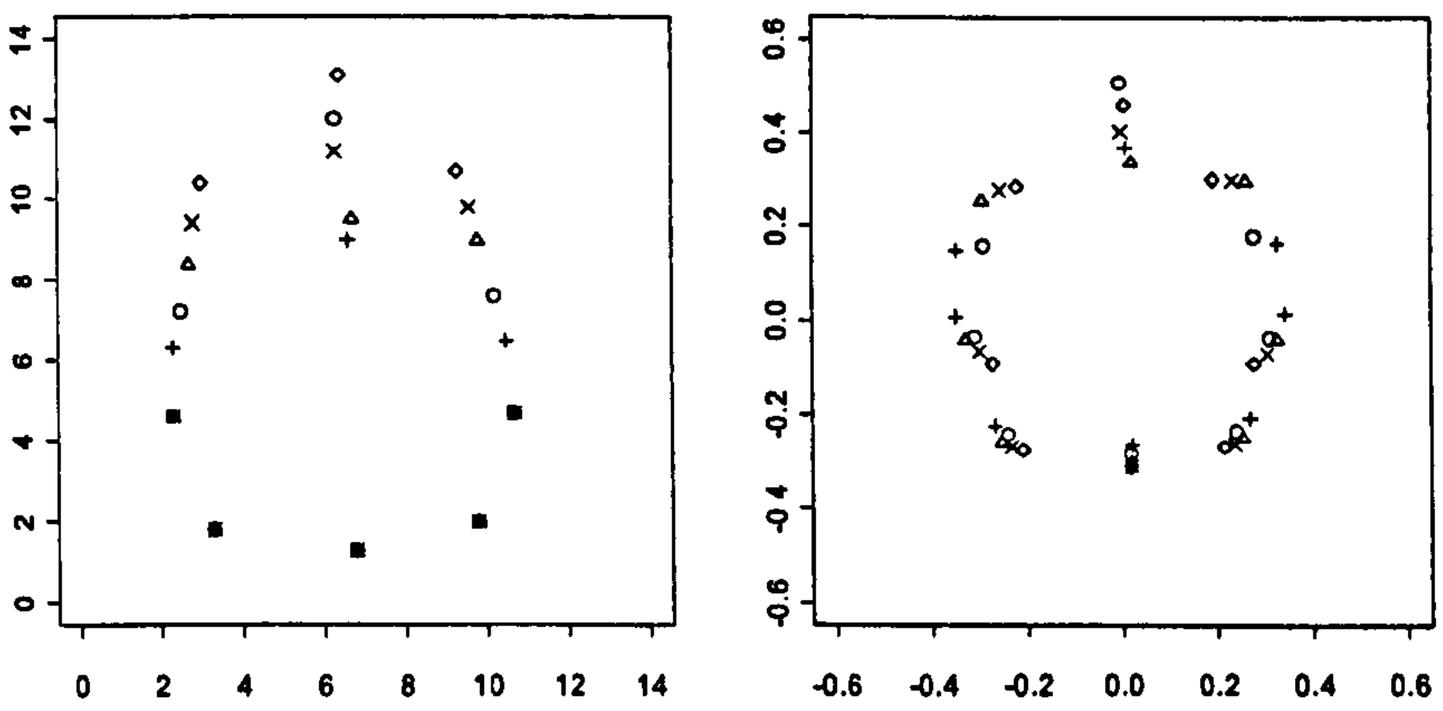


Fig. 6.11: (Left) Five configurations representing the same tooth with different possible gum positions. (Right) OLS Procrustes registration. Each configuration represented by different symbols.

The idea of the alternating GPA step in the semi-landmark routine is that as the variation along the chord directions (representing the variation in fig. 6.11 (left)), is reduced or removed, so too is its effect on the resulting variation at the coordinates of the Procrustes fits of the other fixed landmarks. Alternatively, a weighted GPA can incorporate known error information into the registration procedure directly. Here the affected landmarks do not move, but the directions of poor correspondence and uncertainly are taken into account when obtaining the Procrustes registration of the data.

In the model of (6.87) or (6.88), each displacement E_i is defined relative to μ along with the covariance Σ of the E_i. The covariance (or model metric) Σ accommodates the actual variation in shape of the configurations as well as the errors at the recorded landmarks (Goodall, 1991). (Note that, as described by Goodall, 1991, a component of each E_i will always be confounded with the β_i, Γ_i and γ_i). For the dataset of upper central incisors, the assumption of an isotropic covariance form for the E_i $(\Sigma = \sigma^2 I_{km})$, is clearly unrealistic, since we have correlated errors in the gingival landmark locations due to unwanted variation in the position of patient's gingival tissue. (In practice, error variation also occurs around the lower incisal outline, due to operator inconsistencies in the placement of the MD landmarks and those at the corners of the incisal edge).

Instead, a weighted Procrustes analysis allows a more general covariance structure for the E_i to be considered, for the registration and subsequent analysis of a sample of

configurations, with $vec(E_i) \sim N(0, \Sigma)$. The mean shape $\hat{\mu}$ and generalised least squares estimators of each β_i, Γ_i and $\gamma_i^T \in SG_i$, determining the (weighted) Procrustes fits $\hat{\beta}_i X_i \hat{\Gamma}_i + 1_k \hat{\gamma}_i^T$ to $\hat{\mu}$, are obtained by seeking:

$$\min_{SG_i, \hat{\mu}} \sum_{i=1}^{n} vec(SG_i(X_i) - \hat{\mu})^T \Sigma^{-1} vec(SG_i(X_i) - \hat{\mu})$$

$$= \min_{\hat{\mu}, \beta_i, \Gamma_i, \gamma_i} \sum_{i=1}^{n} vec(\beta_i X_i \Gamma_i + 1_k \gamma_i^T - \hat{\mu})^T \Sigma^{-1} vec(\beta_i X_i \Gamma_i + 1_k \gamma_i^T - \hat{\mu}). \qquad (6.90)$$

Again, working with a known covariance structure requires only a slight modification of the formulations for the unweighted case, for example, the algorithm of Gower (1975) modified by Rohlf & Slice (1990) in section 3.2.2.2.1.

If the actual biological variation and errors in the positions of the gingival landmarks can be incorporated into the matching procedure and subsequent analysis by specification of Σ, then this would allow the variation in shape at all landmarks (including the non-gingival ones) to be represented more accurately and realistically by the Procrustes fits, providing a better indication of the locations of the differences in shape that are of interest. For the data in fig. 6.11, the influence of the gingival landmarks on the alignment of the lower landmarks would be reduced with the fits resembling the plot of the left of fig 6.11 more than the right. Any differences in shape between configurations either in the non-gingival landmarks, or in directions where landmarks do not have as high variance, would then be able to be detected more easily. Although the resulting Procrustes fits will no longer correspond to points in Kendall's shape space (which by definition, requires that there are no remaining differences due to OLS similarity transformations), standard multivariate techniques may be used on the coordinates of the residuals or fits as described in 3.4.2.

The main difficulty with the use of weighted Procrustes registration however (and why we have not used it here) is that in practice Σ is not known and must be estimated from the data. The key problem is dealing with invariances due to similarity transformations since, as many authors point out (e.g. Fitzpatrick *et al.* (1998)), the E_i are defined relative to the population mean rather than that at the raw landmark locations and so the

covariance matrix needs to be estimated from the residuals of the Procrustes fits. Walker (2000) and Lele (1993) document some of the problems in estimating covariances after superimposition. In the former, the author investigates the ability of the ordinary Procrustes method to recover a known covariance matrix after superimposition and shows how this can be very different from the model on which it was based. In the latter, Lele recommends that estimation of Σ should be based on inter-landmark distances instead of the Procrustes fits.

Goodall (1991) and (1995) considered weighted Procrustes matching for factored covariance structures of the form $\Sigma_k \otimes \Sigma_d$, where Σ_d is the covariance between the d dimensions at each landmark and Σ_k is the covariance between landmarks (assumed identical in each dimension). For d=2, Σ_d will be of the same form as the landmark covariance matrices in the anisotropic smoothing spline with $\Sigma_1=\ldots=\Sigma_n$. However, the use of factored covariances has been criticised by many authors as being unrealistic and difficult to estimate (e.g. Lele, 1993). Subsequently, Goodall (1995) proposed refinements based on restricted maximum likelihood estimation. Another alternative is the offset normal maximum likelihood approach of Dryden & Mardia (1991, 1998).

One possibility described by both Goodall (1991) and Dryden & Mardia (1998) is to use a pre-specified 'superimposition' metric $\Sigma=\Sigma_S$ for fitting (obtaining the Procrustes fits) and an estimated (model) covariance $\hat{\Sigma}$ for testing. Both papers describe how $\hat{\Sigma}$ could obtained by an iterative technique, using for example, $\Sigma_S=I$ to start with:

(a) Obtain Procrustes fits using $\Sigma=\Sigma_S$

(b) Compute $\hat{\Sigma}$ by a suitable estimator, e.g. the maximum likelihood estimator (Goodall, 1991) in equation (3.24) with $v_i = X_i^P$.

(c) Set $\Sigma_S=\hat{\Sigma}$ and re-compute Procrustes fits

(d) Repeat from (b) until $\hat{\Sigma}$ converges.

As Adams *et al.* (2004) note, the proper estimation of the relative variability at different landmarks and patterns of covariance within and between landmarks is an important problem that needs to be resolved. Until then however, the advice they recommend is

based on that of Rohlf and Slice (1990), who suggested producing simulations using the estimated mean as a population mean to make sure that the patterns of variation in the Procrustes fits are not just artefacts of the OLS registration procedure.

6.4.2 Methods combining landmark and outline information

The final part of this chapter describes two alternative methods to the semi-landmark technique that have been proposed for the analysis of shape when one would wish to include features such as part of the outline or a curve on an object in an analysis, but where suitable candidates for inclusion as landmarks are sparse or where landmarks alone fail to capture the characteristic(s) of interest. The main method described is Green's (1996) extension of the semi-landmark technique to include any combination of curves and landmarks. Instead of representing parts of an outline or curve as a series of landmarks and associated chord directions, the curve itself is parameterised as series of vertices and treated in the same way as landmarks when superimposed onto a homologous object. The technique may also utilise the ideas presented in Sampson *et al.* (1996), who considered the iterative closest point algorithm of Besl & McKay (1992) and normal projection from landmarks to points on a closed outline, as a way of determining positions of correspondence.

Note however that none of these ideas can be used to address the gingival variation problem here, since part of the required outline information along which landmarks have poor correspondence, is obscured by gingival tissue. However, for investigations involving the lower outline of buccal tooth surfaces or the occlusal outlines of the molar or canine teeth, where there is also lack of precise landmark correspondence, these methods may be worthy of consideration, especially when the approximation of the chords to the curvature of an outline is poor and so are included here for completion.

6.4.2.1 Green's method for curves and landmarks

At any step of the semi-landmark procedure, movement of the semi-landmarks along escribed chords can lead to new positions some distance away from the actual outline of an object. Bookstein (1996d) notes that this can be a particular problem when sampling

is too sparse to suit the actual curvature of an outline. One simple improvement to the technique could be to 'project' the new position back onto the original outline by projecting normally to the chord direction. However, in some instances this may lead to positions a long way from where the landmarks initially started or the projection may miss the outline altogether (see 6.4.2.2). Alternatively, Bookstein (1996d) suggests the method of Green (1996) for ensuring that the points are constrained to the original outline during the registration procedure. Rather than using landmarks which may move along chords, here the curves along which points of correspondence are thought to lie are instead represented in their entirety during a superimposition and treated in a manner consistent with that of the landmarks.

In Green's terminology, representations of shape comprising curves and landmarks are referred to as 'figures', rather than 'configurations', which contain only landmarks. Curves are assumed to be smooth, non-intersecting and terminate at landmarks and are recorded as a list of vertices (with linear curves assumed between). Each curve can have many different parameterisations so figures are considered equivalent (in terms of shape) if one can be derived from the other by a re-parameterisation of the curves, as well as a similarity transformation.

Note however, that because of the requirement that the curves terminate at landmarks, representing the bounds or limits between which a point of correspondence is defined to lie, we are unable to use this method for the gingival variation problem. With the use of semi-landmarks and chords, the potential positions of points along each direction are infinite, whereas here the method relies on a finite sampling from each set of possible positions. For each of the papilla landmarks, this would require locating points corresponding to the ends of the cemento-enamel junction (CEJ) and for the gingival margin landmark, identification of the position on the CEJ at the top of the long axis of the crown (LAC). The lower bounds would be located at the MD endpoints and bottom of the LAC respectively. However, each of the CEJ locations will be obscured by the gingival tissue, as shown in fig 6.12 and so the curves representing the possible positions of correspondence along the sides of the tooth would be incomplete. For other landmark reliability issues, such as around the lower outline of the buccal surface or around the perimeter of the canine and molar occlusal surfaces, this problem would not arise, since curves of possible landmark positions would be able to be represented in

their entirety, e.g. between the two landmarks used to define the semi-landmarks escribed chord.

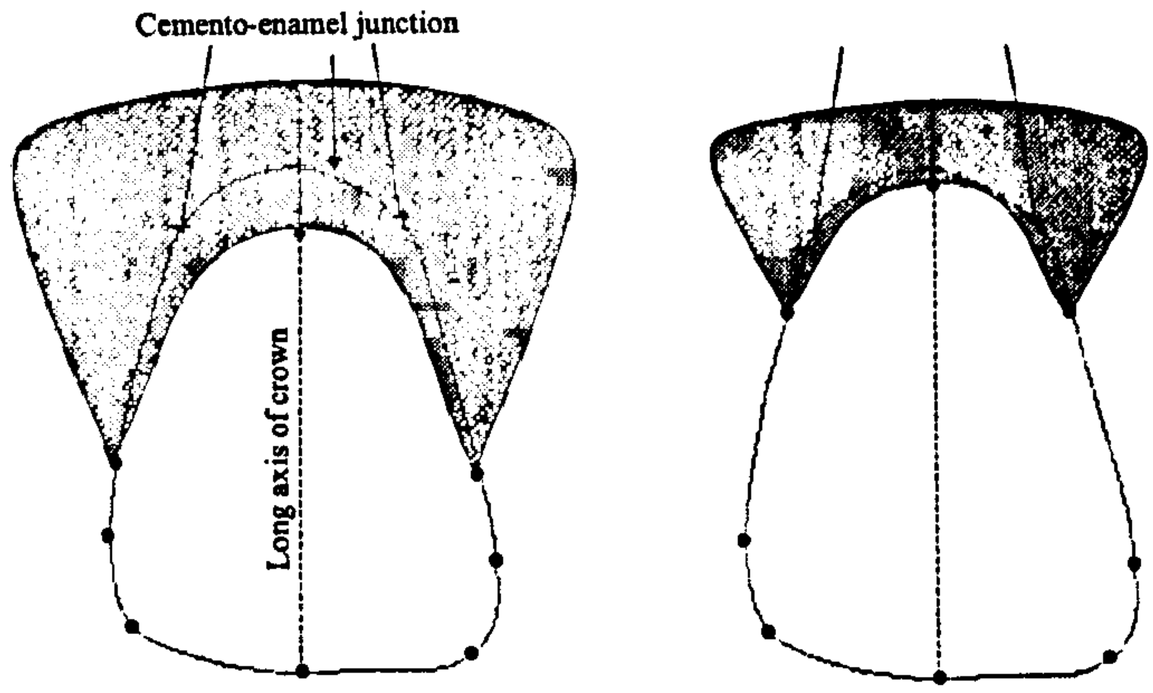


Fig 6.12: Cementoenamel junction positions (crosses) in relation to gingival landmarks (circles)

If T_F and Y_F are figures containing r curves and k landmarks, the registration of Y_F to T_F (typically the reference or mean figure $\hat{\mu}_F$) is performed in two stages:

(a) Firstly, the parameterisation of the curves in Y_F is standardised with respect to those in T_F. (The different parameterisation of the curves represents an equivalence class which must be standardised/filtered out). The curves in figure T_F are arbitrarily sampled at a finite number of points and are required to be mapped by a PTPS transformation (which includes matching of the fixed landmarks, including those at the ends of each curve) to points along the homologous curves in Y_F. A search over possible target points along each curve is performed and the positions corresponding to the mapping with least bending energy are assigned as the homologues of those sampled along the curves in T_F. This then provides a mapping so that every point along each curve in T_F has a corresponding point in the homologous curve in Y_F. Green (1996) asserts that the end result will have only weak dependence on how the points in T_F are sampled.

(b) For the superimposition of Y_F to T_F, the OLS Procrustes matching procedure is generalised to figures containing r curves and k landmarks as follows. Let y_j and t_j, j=1,...,k denote the landmark positions in Y_F and T_F and $c_p^Y(l)$ and $c_p^T(l)$ the (x,y) positions at homologous positions l along each curve p=1,...,r. Each curve p in Y_F has length L_p. The superimposition of Y_F to T_F over similarity transformations is

obtained as that which minimises the sum of squared distances between $SG(Y_F)$ and T_F, given by:

$$\alpha \frac{1}{k}\sum_{j=1}^{k}(SG(y_i)-t_i)^2+\beta \frac{1}{\sum_{p=1}^{r}L_p}\sum_{p=1}^{r}\int_{L_p}(SG(c_p^Y(l))-c_p^T(l))^2\,dl \qquad (6.91)$$

where α and β, satisfying $\alpha+\beta=1$, provide the relative weightings of landmarks and curves. The integrals are approximated by choosing a sampling again and calculating the appropriate finite sum. If the 'figures' are firstly centred and standardised to unit size, minimisation of (6.91) is only over rotations of Y_F. To do this, the centroid and centroid size of a figure (say Y_F) are given by:

$$\bar{y}_F=\alpha \frac{1}{k}\sum_{j=1}^{k}y_i+\beta \frac{1}{\sum_{p=1}^{r}L_p}\sum_{p=1}^{r}\int_{L_p}C_j^Y(l)\frac{ds}{dl}dl,$$

$$\|Y_F\|=\alpha \frac{1}{k}\sum_{j=1}^{k}(y_i-\bar{y})^2+\beta \frac{1}{\sum_{p=1}^{r}L_p}\sum_{p=1}^{r}\int_{L_p}(C_j^Y(l)-\bar{y})^2\frac{ds}{dl}dl.$$

For a sample of figures, Y_{Fi}, $i=1,\ldots,n$, an iterative process is used to obtain $\hat{\mu}_F$ and each registered version of Y_{Fi} to $\hat{\mu}_F$, starting with some arbitrary initial estimate for $\hat{\mu}_F$, e.g. one of the Y_{Fi}. At each step the Y_{Fi} are registered to $T_F=\hat{\mu}_{C_r}$ using (a) and (b) above. The average displacement of each landmark and curve point corresponding to the sampling of curves in $\hat{\mu}_F$ is calculated and these displacements then applied to $\hat{\mu}_F$ to update, until convergence is reached. For subsequent multivariate analysis, a final sampling of the mean's curves must be chosen and corresponding landmarks on the curves of each case obtained, using (a) above. This may be a different sampling from that used in the preceding stages of the registration, but must be carefully considered. A balance needs to be achieved between coarse sampling where information might be lost and over-fine sampling resulting in redundant variables (and loss of power) in the analysis which follows. (This comment obviously also applies to the semi-landmark method if arbitrary points are to be chosen to represent curves). These points are then analysed alongside the real landmarks following a final OLS registration of the data.

6.4.2.2 Sampson's method using iterative closest points and normal projection

Two alternative approaches to the parameterisation stage in (a) are the nearest point method and normal projection, both used in Sampson *et al.* (1996), who introduced a new variant of eigenshape analysis (see section 2.3.4) for the analysis of closed outline data without any landmarks. In their method, each object's outline is represented as a large number of equally spaced points Y_o, which are connected by edges. Consequently, this method would only be suitable for occlusal surfaces, where the entire outline is visible (i.e. molars and canines, but not incisors).

For outlines Y_o and T_o, the registration of (say) Y_o to T_o relies on the 'Iterative Closest Point' algorithm of Besl & McKay (1992), comprising the following steps:

i) Obtain Procrustes superimposition of Y_o to T_o using the two sets of points.

ii) Compute the closest positions on the edges of Y_o (in terms of Euclidean distance) corresponding to the points in T_o.

iii) Repeat from step i) using the nearest points as the new Y_{out} until convergence.

For a sample of outlines Y_{io}, i=1,...,n, the estimated mean outline $\hat{\mu}_o$ *and* corresponding registered versions of the Y_{io} are obtained as follows. Starting with an initial estimate for $\hat{\mu}_o$ (which could be one of the Y_{io}):

1) Match each Y_{io} to $\hat{\mu}_o$ using the ICP algorithm above with $T_o = \hat{\mu}_o$.

2) Compute new corresponding points on each Y_{io} by 'normal projection' from the equally spaced points in $\hat{\mu}_o$ (see below).

3) Compute a new estimate of $\hat{\mu}_o$ as the pointwise average of the new points in each new Y_{io} from 2).

4) Repeat from step 1 until $\hat{\mu}_o$ has converged.

Vectors of scalar deviations along normal projections from the points in the final estimate of mean outline then represent individual outlines and variation in shape is analysed by an eigen decomposition of these distances.

In the parameterisation stage (a) of Green's method, he also considered the nearest point and normal projection techniques as alternatives to assigning homologies along curves rather than by PTPS mappings. For each sampled point on a curve in the reference figure T_F, one could choose either the nearest point in the homologous curve of figure Y_F or the position normal to the curve at the sampled point in T_F as the point of correspondence, as illustrated in fig 6.13 below. (The idea of normal projection, rather than the Euclidean distance from a chord, could also be considered as another criterion for determining new position of semi-landmark along chords). However, as Green (1995, 1996) notes, 'normal projection' can often generate anomalies where nearby normals cross or where a normal misses the curve in Y_F. He also notes that both 'normal projection' and the 'nearest point' approaches depend on the registration of the two figures (as also noted in 6.3.1) and produce a homology that is no longer necessarily continuous, whereas use of a PTPS mapping produces a parameterisation that is independent of the registration of the figures and provides a continuous mapping so that all the information on the curve is retained.

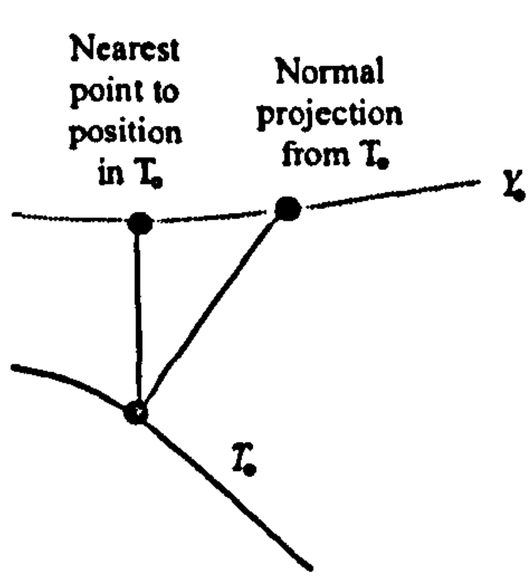


Fig. 6.13: Nearest point and normal projection to assign correspondences

6.5 Summary and discussion

In order to address the problem of large, unrealistic movements occurring when using the original semi-landmark method (with the minimum bending energy criterion) to filter out variation due to differences in the position of patient's gingival tissue, we have considered other superimposition techniques and alternative optimisation criteria that may be used to determine the new positions of semi-landmarks along their chords. Methods investigated involved optimising objective functions over superimpositions of T (in practice the full Procrustes mean $\hat{\mu}$) to $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$, where

$Y^0 = X^P$ is the Procrustes fit of a configuration to $T = \hat{\mu}$, by pairs of interpolating splines, smoothing splines and kriging predictors, or over superimpositions of $vec^{-1}(vec(Y^0) - U\lambda)$ to T by ordinary least squares.

In 6.2.1 we showed how the original semi-landmark method can be considered as a special case of the anisotropic smooth spline and how the penalty function of a weighted smoothing spline may be used to penalise large movements along chords at different semi-landmarks. However, for the gingival variation problem it was found difficult to pre-specify the relative weights of the bending energy (BE) and movement penalties, to reach a balance between the landmarks moving too far or not at all.

In 6.2.2 we described how PTPS transformations and higher order splines are all special cases of the same general kriging model where the mapping from a configuration T to configuration Y is always of the form:

$$D\Gamma^A Y + S\Gamma^B Y \tag{6.92}$$

optimising the quantity:

$$tr(Y^T \Gamma^B Y) \tag{6.93}$$

where D is the matrix corresponding to the choice of polynomials of T, S is the matrix corresponding to the chosen covariance function and Γ^A and Γ^B are always given by:

$$\Gamma^A = (D^T S^{-1} D)^{-1} D^T S^{-1}, \ \Gamma^B = S^{-1}(I_k - D(D^T S^{-1} D)^{-1} D^T S^{-1}).$$

However, while the use of a different covariance function in the PTPS mapping seemed to restrict the large movement of the gingival landmarks slightly, the use of polynomials higher than first order was found to produce several unwanted effects, including folding. The general use of affine components (polynomials of first order) with any covariance function was also criticised since the choice of chord directions for the gingival landmark problem will always result in the landmarks moving large distances if an improved affine fit (smaller value of (6.93)) to all fixed landmarks and chords can be achieved. Ways of restricting the degree of shape change produced by the affine component of mapping of the form (6.92) were therefore considered.

In 6.2.3.2, two methods of quantifying and penalising the change in shape of T produced by the affine component of the PTPS mapping to $Y^{new} = vec^{-1}(vec(Y^0) - U\lambda)$ were considered, along with a penalty on the change in the shape of the 'affine fit' from that in the mapping to the original configuration, Y^0. The first of these options explored the measure of affine contribution presented in Dryden & Mardia (1998). However, it was found that its use as a criterion for the determination of new semi-landmark positions often led to an increase in the Procrustes distance of a configuration from the mean shape $\hat{\mu}$ and folding. In addition, the penalty itself often produces negative values, even before the landmarks move. Alternative methods requiring that the BE of the mapping also be penalised in proportion to a new measure of the affine shape change in T, compared to T or from the affine fit in the original mapping to Y^0, were therefore considered. As with the smoothing spline though, it was found difficult to specify the relative weights of the two functions to work equally well on all configurations, in restricting large movements of the gingival landmarks.

Methods were then also considered that did not use a spline or kriging mappings in the objective function. In 6.3.1 the idea of simply moving the semi-landmarks in the Procrustes fit of $Y^0 = X^P$ to $T = \hat{\mu}$ to the nearest point along their chords to the corresponding landmark in the mean shape was presented. Alternatively, the new positions of semi-landmarks could be determined as those which optimise the full Procrustes distance between $Y^{new} = vec^{-1}(vec(X^P) - U\lambda)$ and $T = \hat{\mu}$. Both methods appeared to offer promising results for the problem of gingival landmark variation on buccal tooth surfaces and in chapter 8 we carry out a detailed evaluation and comparison of these new methods, compared to the original (minimum bending energy) technique, in terms of addressing this issue.

Table 6.4 presents a summary of the different options considered for determination of new semi-landmark positions, including the idea of a weighted Procrustes criterion, corresponding to a PTPS mapping with constrained affine component, suggested in 6.2.3.1 and 6.4.1.1. Although excluded from further consideration for addressing the current problem of unwanted gum variation, some of the other options presented here may well be worthy of investigation in other applications.

In a recent posting on the 'Morphmet' forum (July 2003), Bookstein himself expressed how he had always hoped that someone would explore other possible ways to generate new semi-landmark positions or "other functionals according to which to slide" and commented on the lack of literature and examples that don't have Bookstein as an author. Here we have proposed several alternative methods for the determination of new semi-landmark positions and have identified two particular options which appear to offer a potential solution to a particular problem associated with the analysis of tooth shape.

However, it may be that these methods also offer a more useful criterion than the bending energy method, for the use of the semi-landmark technique in more general situations. In chapters 9 and 10 we therefore consider use of the nearest point and full Procrustes criterion for other applications, in particular for other reliability problems in the study of tooth shape and (for illustration of applications outside dentistry) on distorted configurations generated from simple geometric forms.

In the following chapter, we first consider some of the implementation issues arising with the use of the new semi-landmark methods, as we did for the original method in section 5.5. In particular, we consider the importance of the GPA method used with the nearest point criterion, convergence issues and describe the S-plus routines used for the applications in the remaining chapters of this thesis, for future reference by other researchers.

Objective function ($\min_{\lambda}$)	Options/parameters	Solution λ

Smoothing splines, general splines and kriging

Weighted smoothing spline	$(U\lambda)^T diag(\sigma_1^2,...,\sigma_{SL}^2)^{-1}(U\lambda)$ $+\alpha((vec(Y^0)-U\lambda)^T(I_2\otimes\Gamma^{11})(vec(Y^0)-U\lambda)^T)$	$\sigma_1^2,...,\sigma_{SL}^2, \alpha$	$(U^T(\alpha(I_2\otimes\Gamma^B)+R^{-1}U)^{-1}U^T\alpha(I_2\otimes\Gamma^B)vec(Y^0)$ R=$diag(\sigma_1^2,...,\sigma_{SL}^2)^{-1}$
General splines/Kriging	$(vec(Y^0)-U\lambda)^T(I_2\otimes B)^{-1}(vec(Y^0)-U\lambda))$	*S (covariance) and D (order of polys.) in* Γ^B	$(U^T(I_2\otimes B)U)^{-1}U^T(I_2\otimes B)Y^0_{col}$

PTPS mappings with affine penalties

Affine contribution 1	$d_F^2(vec^{-1}(vec(Y^0)-U\lambda),T)-d_F^2(T,\frac{T}{\|T\|}+S\Gamma^{11}\frac{vec^{-1}(vec(Y^0)-U\lambda)}{\|vec^{-1}(vec(Y^0)-U\lambda)\|})$		Use optimisation algorithm
Affine contribution 2	$d_F^2(T,Q\Gamma\ vec^{-1}(vec(Y^0)-U\lambda))+\alpha(vec(Y^0)-U\lambda)^T(I_2\otimes\Gamma^{11})^{-1}(vec(Y^0)-U\lambda))$	α	
Change in shape of affine component	$d_F^2(Q\Gamma^{21}vec^{-1}(vec(Y^0)-U\lambda),Q\Gamma^{21}vec^{-1}(vec(Y^{new})-U\lambda))$ $+\alpha(vec(Y^0)-U\lambda)^T(I_2\otimes\Gamma^{11})^{-1}(vec(Y^0)-U\lambda))$	α	

Procrustes and nearest point superimpositions

Nearest point	$(vec(Y^0)-vec(T)-U\lambda)^T(vec(Y^0)-vec(T)-U\lambda)$ (T and Y^0 already aligned)		$(U^TU)^{-1}U^T(vec(Y^0)-vec(T))$
Full Procrustes / OLS superimposition	$\|SG(vec^{-1}(vec(Y^0)-U\lambda)-T\|^2$ or $\|SG(T)-(vec^{-1}(vec(Y^0)-U\lambda)\|^2$ or $d_F^2(vec^{-1}(vec(Y^0)-U\lambda),T)$		Use optimisation algorithm
Weighted Procrustes	$vec(SG(T)-vec^{-1}(vec(Y^0)-U\lambda))^T\Sigma_S^{-1}vec(SG(T)-vec^{-1}(vec(Y^0)-U\lambda))$	Σ_S	See 6.4.1.1 or use optimisation algorithm
	Use $\Sigma_S=(I_2\otimes S)$ for affine mapping constrained to be similarity transformation, with covariance S (as in PTPS mapping)		

Table 6.4: Summary of new criteria proposed for the determination of new semi-landmark positions on a configuration. New version of configuration Y^{new} = $vec^{-1}(vec(Y^0)-U\lambda)$, where Y^0 is the original version, U is a matrix of pre-specified unit chord directions and λ is a vector of scalars to be determined. T =reference configuration (current mean shape). For full definitions of $\sigma_1^2,...,\sigma_{SL}^2$, α, Γ^B, S, D, Σ_S see relevant sections.