5

Procrustes Analysis

5.1 Introduction

This chapter outlines various methods based on Procrustes superimposition, which are very useful tools for analysing landmark data. We have already seen Procrustes methods in two dimensional shape analysis in Chapter 3, where complex arithmetic can be used. Procrustes methods were also seen to be useful for assessing distances between shapes in Chapter 4. In this chapter we provide a more complete treatment of Procrustes methods suitable for two and higher dimensional shape analysis.

5.2 Ordinary Procrustes Analysis

5.2.1 Full ordinary Procrustes analysis

Let us first consider the case where two configuration matrices X_1 and X_2 are available (both $k \times m$ matrices of coordinates from k points in m dimensions) and we wish to match the configurations as closely as possible, up to similarity transformations. **In this chapter we assume without loss of generality that the configuration matrices X_1 and X_2 have been centred** using Equation (1.4).

Definition 5.1 *The method of* **full ordinary Procrustes analysis** *(full OPA) involves the least squares matching of two configurations using the similarity transformations. Estimation of the similarity parameters γ, Γ and β is*

carried out by minimizing the squared Euclidean distance

$$D^2_{OPA}(X_1, X_2) = \|X_2 - \beta X_1 \Gamma - 1_k \gamma^{\mathrm{T}}\|^2, \qquad (4.1)$$

where $\|X\| = \{\mathrm{trace}(X^{\mathrm{T}}X)\}^{1/2}$ *is the Euclidean norm,* Γ *is an* $(m \times m)$ *rotation matrix (*$\Gamma \in SO(m)$*),* $\beta > 0$ *is a scale parameter and* γ *is an* $(m \times 1)$ *location vector. The minimum of Equation (4.1) is written as* $OSS(X_1, X_2)$*, which stands for* **Ordinary (Procrustes) Sum of Squares**.

In Section 4.2.1, when calculating distances, we were interested in the minimum value of an expression similar to Equation (4.1) except X_1 and X_2 were of unit size.

Result 5.1 *The full ordinary Procrustes solution to the minimization of Equation (4.1) is given by* $(\hat{\gamma}, \hat{\beta}, \hat{\Gamma})$ *where*

$$\hat{\gamma} = 0 \qquad (4.2)$$

$$\hat{\Gamma} = UV^{\mathrm{T}} \qquad (4.3)$$

where

$$X_2^{\mathrm{T}} X_1 = \|X_1\| \, \|X_2\| V \Lambda U^{\mathrm{T}}, \;\; U, V \in SO(m) \qquad (4.4)$$

with Λ *a diagonal* $m \times m$ *matrix of positive elements except possibly the last element defined in Result 4.1. Furthermore,*

$$\hat{\beta} = \frac{\mathrm{trace}(X_2^{\mathrm{T}} X_1 \hat{\Gamma})}{\mathrm{trace}(X_1^{\mathrm{T}} X_1)}, \qquad (4.5)$$

and

$$OSS(X_1, X_2) = \|X_2\|^2 \sin^2 \rho(X_1, X_2), \qquad (4.6)$$

where $\rho(X_1, X_2)$ *is the Procrustes distance of Equation (1.22).*

Proof: We wish to minimize

$$D^2_{OPA} = \|X_2 - \beta X_1 \Gamma - 1_k \gamma^{\mathrm{T}}\|^2 = \mathrm{trace}(\|X_2\|^2 + \beta^2 \|X_1\|^2 - 2\beta X_2^{\mathrm{T}} X_1 \Gamma) + k\gamma^{\mathrm{T}}\gamma,$$

where X_1 and X_2 are centred. It simple to see that we must take $\hat{\gamma} = 0$. If

$$Z_i = HX_i/\|X_i\|, \quad i = 1, 2,$$

are the pre-shapes of X_i, then we need to minimize

$$\mathrm{trace}(\|X_2\|^2 + \beta^2\|X_1\|^2 - 2\beta\|X_1\|\,\|X_2\|Z_2^{\mathrm{T}}Z_1\Gamma)$$

and so from Result 4.1 we find the minimizing Γ from Equation (1.16). Differentiating with respect to β we obtain:

$$\frac{\partial D^2_{OPA}}{\partial\beta} = 2\beta\mathrm{trace}(X_1^{\mathrm{T}}X_1) - 2\mathrm{trace}(\|X_1\|\|X_2\|Z_2^{\mathrm{T}}Z_1\hat{\Gamma}).$$

Hence,

$$\hat{\beta} = \frac{\|X_2\|\mathrm{trace}(Z_2^{\mathrm{T}}Z_1\hat{\Gamma})}{\|X_1\|} = \frac{\|X_2\|}{\|X_1\|}\mathrm{trace}(\Lambda) = \frac{\|X_2\|}{\|X_1\|}\cos\rho(X_1, X_2). \tag{4.7}$$

Substituting $\hat{\gamma}, \hat{\Gamma}$ and $\hat{\beta}$ into Equation (4.1) leads to

$$OSS(X_1, X_2) = \|X_2\|^2 + \hat{\beta}^2 \|X_1\|^2 - 2\hat{\beta}\|X_1\| \, \|X_2\| \cos\rho$$

and so the result of Equation (4.6) follows. □

Note that λ_m will be negative in the cases where an orthogonal transformation (reflection and rotation) would produce a smaller sum of squares than just a rotation. In practice, for fairly close shapes λ_m will usually be positive – in which case the solution is the same as minimizing over the orthogonal matrices $O(m)$ instead of $SO(m)$.

Definition 5.2 *The* **full Procrustes fit** *(or full Procrustes coordinates) of X_1 onto X_2 is*

$$X_1^P = \hat{\beta} X_1 \hat{\Gamma} + 1_k \hat{\gamma}^{\mathrm{T}},$$

where we use the superscript 'P' to denote the Procrustes

superimposition. The **residual matrix** *after Procrustes matching is defined as*

$$R = X_2 - X_1^P.$$

Sometimes examining the residual matrix can tell us directly about the difference in shape, e.g. if one residual is larger than others or if the large residuals are limited to one region of the object. In other situations it is helpful to use further diagnostics for shape difference such as the partial warps from thin-plate spline transformations, discussed later in Section 10.3.3.

In general if the rôles of X_1 and X_2 are reversed, then the ordinary Procrustes superimposition will be different. Writing the estimates for the reverse order case as $(\hat{\gamma}^R, \hat{\beta}^R, \hat{\Gamma}^R)$ we see that $\hat{\gamma}^R = -\hat{\gamma}, \quad \hat{\Gamma}^R = (\hat{\Gamma})^{\mathrm{T}}$ but $\hat{\beta}^R \neq 1/\hat{\beta}$ in general. In particular, as we noted in Chapter

3:

$$OSS(X_2, X_1) \neq OSS(X_1, X_2)$$

unless the figures are both of the same size, and so one cannot use $\sqrt{OSS}(X_1, X_2)$ as a distance. If the figures are normalized to unit size, then we see that

$$OSS(X_1/\|X_1\|, X_2/\|X_2\|) = 1 - \left\{ \sum_{i=1}^{m} \lambda_i \right\}^2 = \sin^2 \rho(X_1, X_2) = d_F^2(X_1, X_2)$$

and in this case $\sqrt{OSS}(X_1/\|X_1\|, X_2/\|X_2\|) = \sin \rho(X_1, X_2)$ is a suitable choice of shape distance, and was denoted as $d_F(X_1, X_2)$ in Equation (1.15) – the full Procrustes distance. In Example 3.1 we saw the ordinary Procrustes superimposition of a juvenile and an adult sooty mangabey in two dimensions.

Since each of the figures can be rescaled, translated and rotated (the full set of similarity transformations) we call the method **full** Procrustes analysis. There are many other

variants of Procrustes matching and these are discussed in Section 5.4.

As remarked in Chapter 3, the term **ordinary** Procrustes refers to Procrustes matching of one observation onto another. Where at least two observations are to be matched the term **generalized** Procrustes analysis is used.

While the use of many adjectives appears cumbersome it does enable us to differentiate between various similar methods.

Full Procrustes shape analysis for two dimensional data is particularly straightforward using complex arithmetic and details were given in Chapter 3.

5.3 Generalized Procrustes Analysis

5.3.1 Introduction

Consider now the general case where $n \geq 2$ configuration matrices are available $X_1, ..., X_n$. For example, the configurations could be a random sample from a population with mean μ, and we wish to estimate the shape of the population mean $[\mu]$ with an 'average' shape from the sample. In order to define what is meant by a population mean shape we have to specify a model for the population.

We consider three perturbation models for the $k \times m$ configuration matrices X_i,

$$X_i = \mu + E_i, \tag{4.8}$$

$$X_i = (\mu + E_i)\Gamma_i + 1_k\gamma_i^{\mathrm{T}}, \tag{4.9}$$

$$X_i = \beta_i(\mu + E_i)\Gamma_i + 1_k\gamma_i^{\mathrm{T}}, \tag{4.10}$$

where E_i are zero mean $k \times m$ independent random error matrices, μ is the $k \times m$ matrix of the mean configuration and β_i, Γ_i and γ_i are nuisance parameters for scale, rotation and translation.

We can estimate μ directly under the measurement error model (4.8), although this model is rarely applicable in practice. We cannot estimate all of μ under models (4.9) and (4.10). We can estimate the shape $[\mu]$ or size-and-shape $[\mu]_S$ under model (4.9), and we can estimate the shape $[\mu]$ under model (4.10).

For shape analysis our objects need not be commensurate in scale, and so model (4.9) or (4.10) can be used. However, for size-and-shape analysis our objects do need to be commensurate in scale, and so model (4.9) is appropriate.

A least squares approach to finding an estimate of $[\mu]$ is that of generalized Procrustes analysis, a direct generalization of ordinary Procrustes analysis.

Definition 5.3 *The method of* **full generalized Procrustes analysis** *(full GPA) involves translating, rescaling and rotating the configurations relative to each other so as to minimize a total sum of squares, and the procedure is appropriate under model (4.9) or (4.10). We minimize a quantity proportional to the sum of squared norms of pairwise differences,*

$$G(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=i+1}^{n} \|(\beta_i X_i \Gamma_i + 1_k \gamma_i^{\mathrm{T}}) - (\beta_j X_j \Gamma_j + 1_k \gamma_j^{\mathrm{T}})\|^2 \tag{4.11}$$

subject to a constraint on the size of the average,

$$S(\bar{X}) = 1, \tag{4.12}$$

where $\Gamma_i \in SO(m), \beta_i > 0, \|X\| = \sqrt{\text{trace}(X^{\mathrm{T}}X)}$ *and* $S(X)$ *is the centroid size and the average configuration is*

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n}(\beta_i X_i \Gamma_i + 1_k \gamma_i^{\mathrm{T}}).$$

We write $G(X_1, ..., X_n)$ *for the minimum of Equation (4.11), subject to the constraint of Equation (4.12), and* $G(X_1, \ldots, X_n)$ *is called the* **generalized (Procrustes) sum of squares**.

Full generalized Procrustes matching involves the superimposition of all configurations placed 'on top of each other' in optimal positions by translating, rotating and rescaling each figure so as to minimize the sum of squared Euclidean distances. The constraint of Equation (4.12) prevents the $\hat{\beta}_i$ from all becoming close to 0. Note

that

$$\begin{aligned} G(X_1,\ldots,X_n) &= \inf_{\beta_i,\Gamma_i,\gamma_i} \frac{1}{n}\sum_{i=1}^{n}\sum_{j=i+1}^{n} \|(\beta_i X_i\Gamma_i + 1_k\gamma_i^{\mathrm{T}}) - (\beta_j X_j\Gamma_j + 1_k\gamma_j^{\mathrm{T}})\|^2 \\ &= \inf_{\beta_i,\Gamma_i,\gamma_i} \sum_{i=1}^{n} \|(\beta_i X_i\Gamma_i + 1_k\gamma_i^{\mathrm{T}}) - \frac{1}{n}\sum_{j=1}^{n}(\beta_j X_j\Gamma_j + 1_k\gamma_j^{\mathrm{T}})\|^2. \end{aligned}$$

It follows that the minimization can be alternatively viewed as another constrained estimation problem, where the mean shape $[\mu]$ is to be estimated, i.e.

$$G(X_1,\ldots,X_n) = \inf_{\mu:S(\mu)=1}\sum_{i=1}^{n} OSS(X_i,\mu) = \inf_{\mu:S(\mu)=1}\sum_{i=1}^{n}\sin^2\rho(X_i,\mu). \tag{4.13}$$

Definition 5.4 *The* **full Procrustes fit** *(or full Procrustes coordinates) of each of the X_i is given by*

$$X_i^P = \hat{\beta}_i X_i\hat{\Gamma}_i + 1_k\hat{\gamma}_i^{\mathrm{T}}, \quad i=1,\ldots,n, \tag{4.14}$$

where $\hat{\Gamma}_i \in SO(m)$ (rotation matrix), $\hat{\beta}_i > 0$ (scale parameter), $\hat{\gamma}_i^{\mathrm{T}}$ (location parameters), $i = 1, ..., n$, are the minimizing parameters.

An algorithm to estimate the transformation parameters $(\gamma_i, \beta_i, \Gamma_i)$ is described below in Section 5.3.2. The parameters $(\Gamma_i, \beta_i, \gamma_i)$ have been termed 'nuisance parameters', because they are not the parameters of primary interest in shape analysis.

Definition 5.5 *The* **full Procrustes estimate of mean shape (full Procrustes mean)** *is given by* $[\hat{\mu}]$*, where*

$$\hat{\mu} = \arg \inf_{\mu: S(\mu)=1} \sum_{i=1}^{n} \sin^2 \rho(X_i, \mu) = \arg \inf_{\mu: S(\mu)=1} \sum_{i=1}^{n} d_F^2(X_i, \mu). \tag{4.15}$$

Note that

$$G(X_1, \ldots, X_n) = \inf_{\mu: S(\mu)=1} \sum_{i=1}^{n} \sin^2 \rho(X_i, \mu).$$

Result 5.2 *The point in shape space corresponding to the arithmetic mean of the Procrustes fits,*

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i^P,$$

has the same shape as the full Procrustes mean.

Proof: The result follows because we are minimizing sums of Euclidean square distances in generalized Procrustes analysis. The minimum of

$$\sum \|X_i^P - \mu\|^2$$

over μ is given by $\hat{\mu} = n^{-1} \sum_i X_i^P$. $\square$

Hence, once a collection of objects has been matched into optimal full Procrustes position with respect to each other, calculation of the full Procrustes mean shape is simple – it is computed by taking the arithmetic means of each coordinate. We see that full generalized Procrustes analysis is analogous to minimizing sums of squared distances in the shape space d_F^2 defined in Equation (1.15).

The constraint on the sizes can be chosen in a variety of ways. For example, an alternative to Equation (4.12) is

$$\sum_{i=1}^{n} S^2(\beta_i X_i \Gamma_i + 1_k \gamma_i^{\mathrm{T}}) = \sum_{i=1}^{n} S^2(X_i). \qquad (4.16)$$

The full Procrustes mean shape has to be found iteratively for $m = 3$ and higher dimensional data, but an explicit eigenvector solution is available for two dimensional data, which was seen in Result 3.2.

5.3.2 Algorithm for higher dimensions

Algorithm: GPA

1. **Translations.** Centre the configurations to remove location. Initially let

$$X_i^P = X_i \ , \ \ i = 1, \ldots, n.$$

2. **Rotations.** For the ith configuration let

$$\overline{X}_{(i)} = \frac{1}{n-1} \sum_{j \neq i} X_j^P,$$

then the new X_i^P is taken to be the ordinary Procrustes superimposition, involving only rotation, of the old X_i^P on $\overline{X}_{(i)}$. The n figures are rotated in turn. This process is repeated until the Procrustes sum of squares of Equation (4.11) cannot be reduced further.

3. **Scaling.** Let Φ be the $n \times n$ correlation matrix of the $\mathrm{vec}(X_i^P)$ (with the usual rôles of variable and observation labels reversed) with eigenvector $\phi = (\phi_1, \ldots, \phi_n)^{\mathrm{T}}$ corresponding to the largest eigenvalue. Then from Ten Berge (1977) take

$$\hat{\beta}_i = \left(\frac{\sum_{k=1}^n \|X_k^P\|^2}{\|X_i^P\|^2} \right)^{1/2} \phi_i,$$

which is repeated for all i.

4. Repeat steps 2 and 3 until the Procrustes sum of squares of Equation (4.11) cannot be reduced further.

The algorithm usually converges quickly.

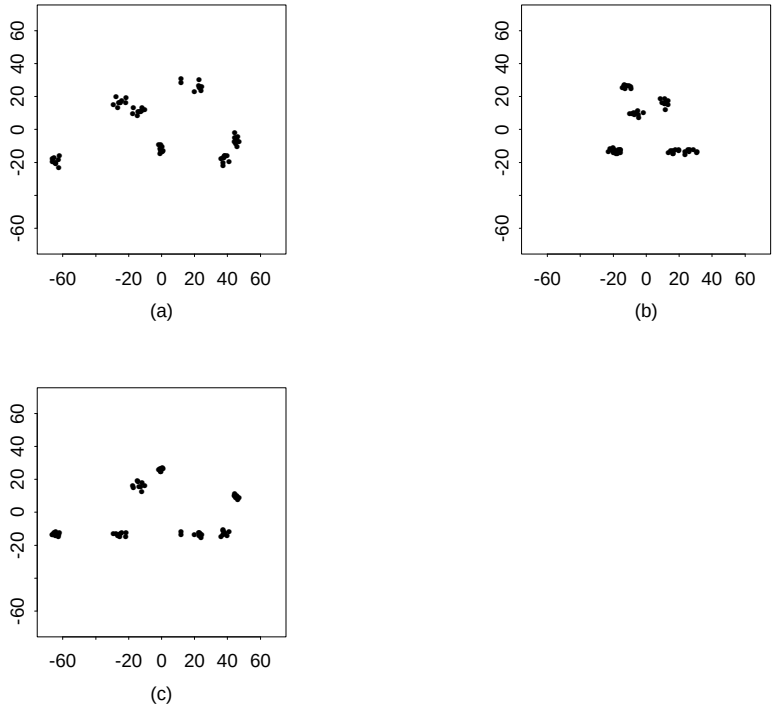

Figure 50 The male macaque skulls registered by full GPA. The three views are (a) the side view x–y, (b) the front/back view x–z and (c) the bottom/top view y–z projections.

Example 5.1 In Figures 50 and 51 we see the superim-

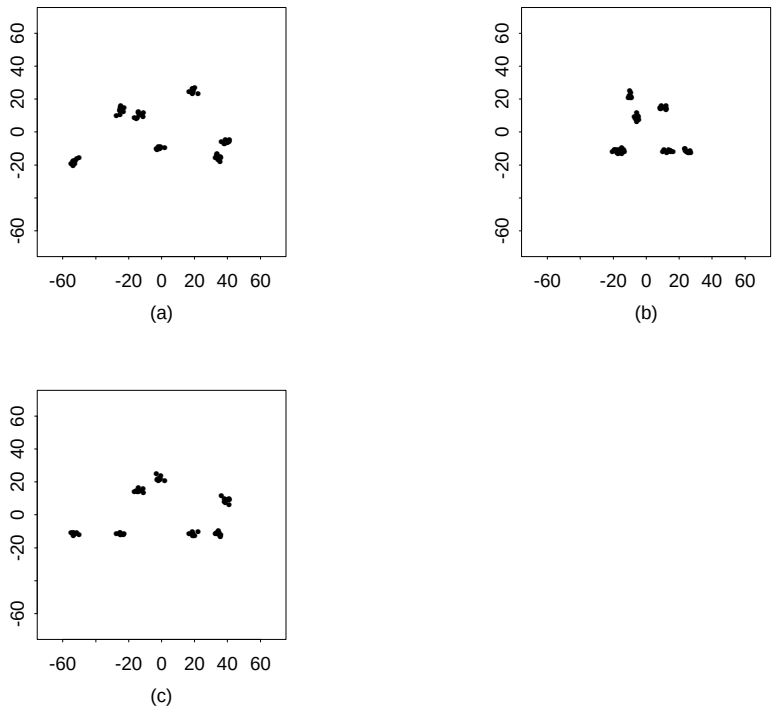

Figure 51 The female macaque skulls registered by full GPA. The three views are (a) the x–y, (b) the x–z and (c) the y–z projections.

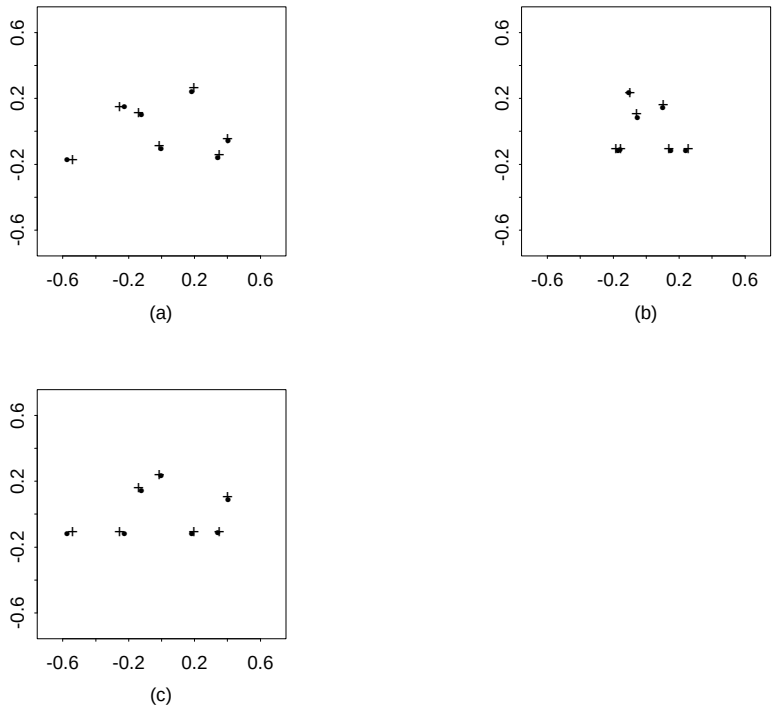

Figure 52 The male ($\cdot$) and female ($+$) Procrustes mean shapes scaled to unit size with the male registered to the female by OPA. The three views are (a) the x–y, (b) the x–z and (c) the y–z projections.

posed male and female macaques from the dataset described in Section 1.2.8, using full Procrustes superimposition. There are $k = 7$ landmarks in $m = 3$ dimensions.

In two of the males the highest landmark in the 'y' direction ($bregma$) is somewhat further away (in the 'x' direction) than in the rest of the specimens. The full Procrustes means (normalized to unit size) are displayed in Figure 52 and the male mean has been superimposed onto the female mean by OPA. The full Procrustes estimated mean shapes for the males $[\hat{\mu}_1]$ and females $[\hat{\mu}_2]$ are full Procrustes distance $d_F = 0.05$ apart, and the root mean square of d_F to the estimated mean shape within each group is 0.08 for the males and 0.06 for the females. The males are more variable in shape than the females. Formal tests for mean shape difference are considered in Chapter

7, and in Section 7.1.3 we see that this is not a significant difference in mean shape.

□

When $n = 2$ objects are available we can consider OPA or GPA to match them. The advantage of using GPA is that the matching procedure is symmetrical in the ordering of the objects, i.e. GPA of X_1 and X_2 is the same as GPA of X_2 and X_1. As we have seen in Section 5.2.1, OPA is not symmetrical in general, unless the objects have the same size.

5.4 Variants of Procrustes Analysis

There are many variants to Procrustes analysis. Partial Procrustes involves just superimposition by translation and rotation (not scaling) as opposed to full Procrustes which

involves the full set of similarity transformations. Another alternative could be matching using orthogonal matrices instead of rotation matrices (reflection Procrustes).

In the

tt shapes package in R the command `procGPA()` carries out generalized Procrustes analysis. Note `procGPA( x , scale=TRUE )` carries out full Procrustes analysis (the default), and `procGPA( x , scale=FALSE )` carries out partial Procrustes analysis.

The option `reflection=TRUE` uses reflection invariance whereas the default value does not have reflection invariance.

5.5 Shape Variability: Principal Components Analysis

The full Procrustes mean shape provides a suitable average shape. It is also of great interest to describe the variability in shape. Principal components analysis (PCA) of the sample covariance matrix in Procrustes tangent space coordinates provides a very effective means of analysing the main modes of variation in shape. In addition PCA is useful in shape analysis in order to reduce the dimensionality of a problem, as it is in multivariate analysis.

We introduced PCA of tangent space coordinates for two dimensional data in Section 3.4. In this section we consider the topic in more detail and give some more examples.

5.5.1 Tangent space PCA

Cootes et al. (1992) and Kent (1994) developed PCA in the tangent space. In particular Kent (1994) proposed PCA of the partial Procrustes tangent coordinates defined in Equation (1.28). Consider n pre-shapes $Z_1, \ldots, Z_n$ with tangent space shape coordinates given by $v_1, ..., v_n$, with a pre-shape $\hat{\mu}$ corresponding to the full Procrustes mean shape as the pole, so

$$v_i = (I_{km-m} - \text{vec}(\hat{\mu})\text{vec}(\hat{\mu})^{\text{T}})\text{vec}(Z_i\hat{\Gamma}_i), \qquad (4.17)$$

where each v_i is a real vector of length $(k-1)m$, obtained from Equation (1.35). Alternatively we could use the full Procrustes residuals r_i of Equation (3.13), or the full Procrustes tangent coordinates v_{Fi} of Equation (1.32)

which become

$$v_{Fi} = (I_{km-m} - \mathrm{vec}(\hat{\mu})\mathrm{vec}(\hat{\mu})^{\mathrm{T}})\mathrm{vec}(\hat{\beta}_i Z_i \hat{\Gamma}_i). \quad (4.18)$$

Note that $\sum_{i=1}^{n} v_{Fi} = 0 = \sum_{i=1}^{n} r_i$ and $\sum_{i=1}^{n} v_i \approx 0$.

As defined earlier in Definition 3.5, the principal components (PCs) γ_j are the orthonormal eigenvectors of the sample covariance of the tangent coordinates S_v, corresponding to eigenvalues $\lambda_j, j = 1, \ldots, p = \min(n - 1, M)$ (M is the dimension of the shape space).

The effect of the jth PC can be seen by plotting icons for for various values of the standardized PC score. In particular we examine

$$v(c, j) = \bar{v} + c\lambda_j^{1/2}\gamma_j, \quad j = 1, \ldots, p, \quad (4.19)$$

for a range of values of the standardized PC score c and then project back into configuration space using Equation

(1.36) and Equation (1.7). The linear transformation to an icon in the configuration space

$$\mathrm{vec}(X_I) = \text{block diag}(H^{\mathrm{T}}, \ldots, H^{\mathrm{T}})\{v(c,j)+\gamma\} \quad (4.20)$$

is a good approximation to the inverse projection from the tangent space to an icon, near the pole.

So, to evaluate the effect of the jth PC for a range of values of c, calculate v from Equation (4.19), project back using the inverse transformation to the pre-shape sphere and then evaluate an icon using say Equation (1.7) or (4.20) to give the centred pre-shape.

There are several ways to visualize the effect of each PC:

1. Evaluate and plot an icon for a few values of $c \in [-3, 3]$, where $c = 0$ corresponds to the full Procrustes mean shape. The plots could either be separate or

superimposed.

2. We could draw vectors from the mean shape to the shape at $c = +3$ and/or $c = -3$ say to understand the structure of shape variability. The plots should clearly label which directions correspond to positive and negative c if both values are used.
3. One could superimpose a square grid on the mean shape and deform the grid to icons in either direction along each principal component. The methods of Chapter 10 will be useful for drawing the grids, and for example the thin-plate spline deformation could be used.
4. One could animate a sequence of icons backwards and forwards along the range $c \in [-3, 3]$. This dynamic method is perhaps the most effective for displaying

each PC.

In datasets where the shape variability is small it is often beneficial to magnify the range of c in order to easily visualize the effect of each PC.

In some datasets only a few PCs may be required to explain a high percentage of shape variability. Some PCs may correspond to interpretable aspects of variability (e.g. thickness, bending, shear) although there are often many combined effects in each PC.

By carrying out PCA in the tangent space we are decomposing variability (the total sum of Procrustes distances) into orthogonal components, with each PC successively explaining the highest variability in the data, subject to being orthogonal to the higher PCs. If the structure of shape variability is that the points

are approximately independent and isotropic with equal variances, then the eigenvalues λ_j of the covariance matrix in tangent space will be approximately equal (this property is proved in Section 6.6.6). If there are strong dependencies between landmarks, then only a few PCs may capture a large percentage of the variability.

An alternative decomposition which weights points close together differently from those far apart is Bookstein's (1991) relative warps, described in Section 10.3.5.

Example 5.2 A random sample of 23 T2 mouse vertebral outlines was taken from the Small group of mice introduced in Section 1.2.1. Six mathematical landmarks are located on each outline. In Example 3.3 we saw the PCs for these landmark data. Consider taking further pseudo-landmarks around the outline. In between each pair

of landmarks 9 equally spaced pseudo-landmarks were placed (similarly to Figure 4), giving a total of $k = 60$ landmarks in $m = 2$ dimensions. In Figure 53 we see the Procrustes superimposed outlines, after initially standardizing to unit size.

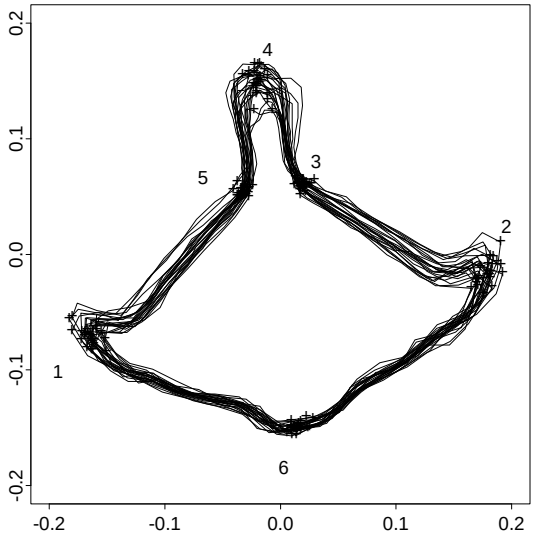

Figure 53 Procrustes rotated outlines of T2 Small mouse vertebrae. The six mathematical landmarks are marked (+) on each outline.

The sample covariance matrix in the tangent space

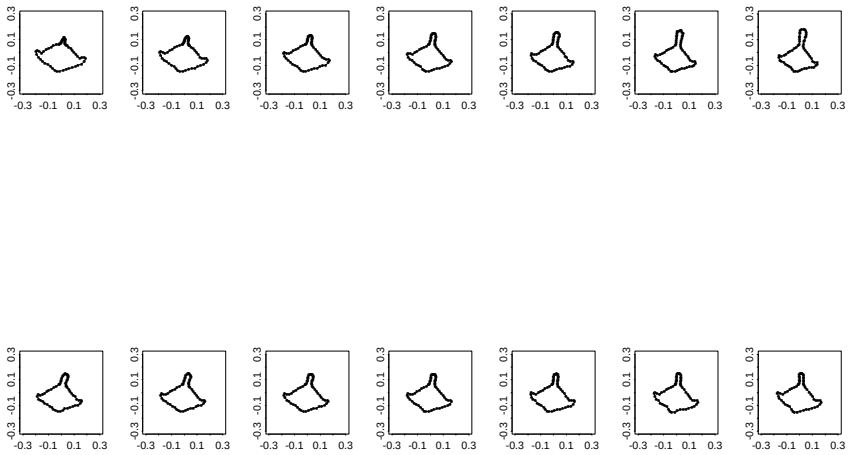

Figure 54 Two rows of series of T2 vertebral shapes evaluated along the first two PCs – the ith row shows the shapes at $c \in \{-6, -4, -2, 0, 2, 4, 6\}$ standard deviations along the ith PC. Note that in each row the middle plot ($c = 0$) is the full Procrustes mean shape. By magnifying the usual range of c by 2 the effect of each PC is more clearly illustrated.

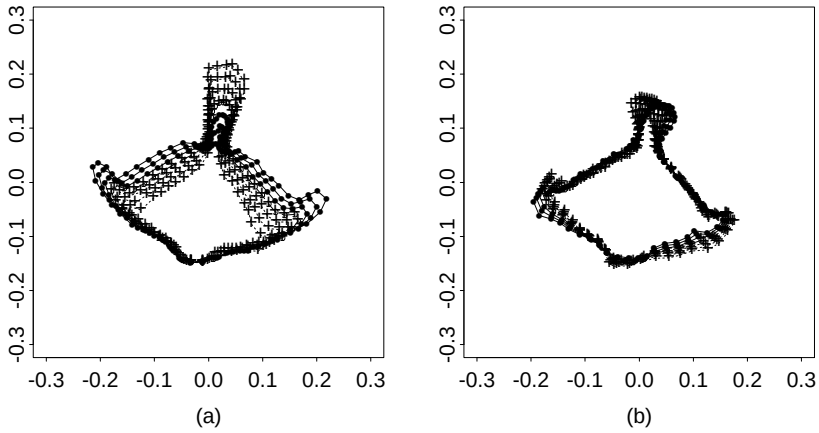

Figure 55 The first (a) and second (b) PCs for the T2 Small vertebra data. The plot shows the icons in each row from the previous figure overlaid on the same picture, i.e. each plot shows the shapes at $c \in \{-6, -4, -2\}$ (·—·—·—·), the mean shape at $c = 0$ (+ - - - - + - - - - +) and the shapes at $c \in \{+3, +2, +1\}$ (+.....+.....+.....+.....+) standard deviations along each PC.

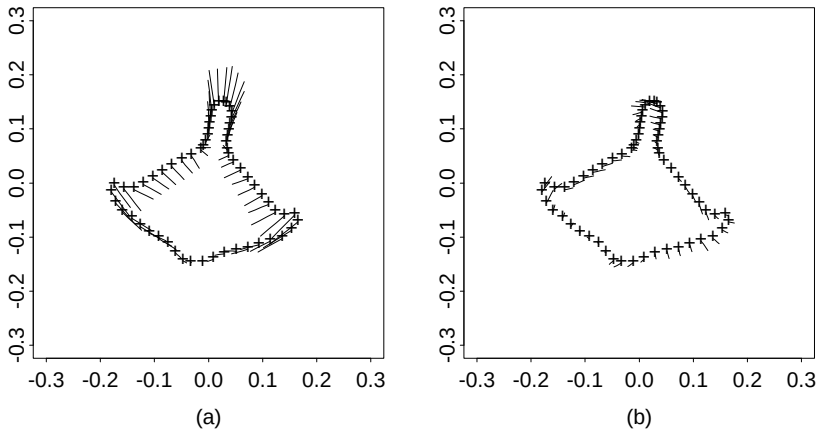

Figure 56 The first (a) and second (b) PCs for the T2 Small vertebra outline data. Each plot shows the full Procustes mean shape with vectors drawn from the mean (+) to an icon which is $c = +6$ standard deviations along each PC from the mean shape.

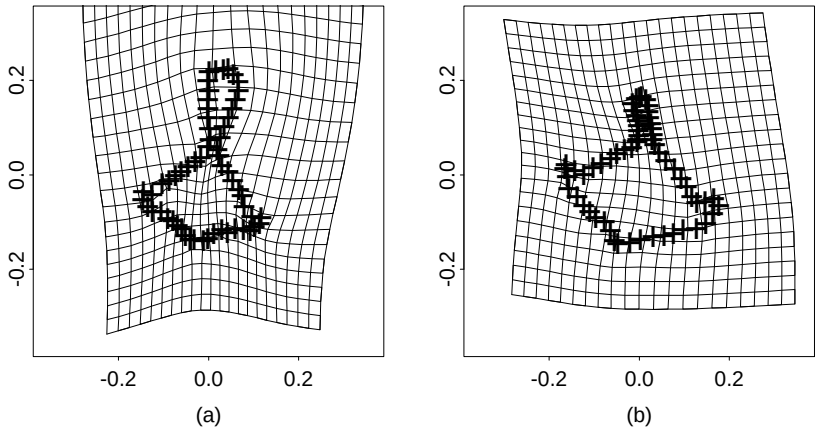

Figure 57 The first (a) and second (b) PCs for the T2 Small vertebra outline data. A square grid is drawn on the mean shape and deformed using a pair of thin-plate splines (see Chapter 10) to an icon $c = 6$ standard deviations along each PC. The plots just show the deformed grid at $c = 6$ for each PC and not the starting grids on the mean.

(using the partial Procrustes coordinates of Equation (1.28)) is evaluated and in Figure 54 we see sequences of shapes evaluated along the first two PCs. Alternative representations are given in Figures 55, 56 and 57. Shapes are evaluated in the tangent space and then projected back using the approximate linear inverse transformation of Equation (4.20) for visualization. The percentages of variability captured by the first two PCs are 65% and 9%, so the first PC is a very strong component here.

The first PC **includes** a contribution from the length of the spinous process (the protrusion on the 'top' of the bone) in contrast to the relative width of the bone. The angle between lines joining landmarks 1 to 5 and 2 to 3 decreases as the height of landmark 4 increases, whereas there is little change in the angles from the lines

joining 1 to 6 and 2 to 6. The second PC includes the effect of asymmetry in the end of the spinous process and asymmetry in the rest of the bone.

Pairwise plots of the elements of the vector $(s_i, d_{Fi}, c_{i1}, c_{i2}, c_{i3})^{\mathrm{T}}$, $i = 1, \ldots, n$, are given in Figure 58, where s_i are the centroid sizes, d_{Fi} are the full Procrustes distances to the mean, and c_{i1}, c_{i2} and c_{i3} are the first three standardized PC scores. There appears to be one bone that is much smaller than the rest and it also appears that there is some correlation between the first PC score and the centroid size of the bones.

An overall measure of shape variability is the root mean square of full Procrustes distance $RMS(d_F)$, which here is 0.07, and the shape variability in the data is quite small.

□

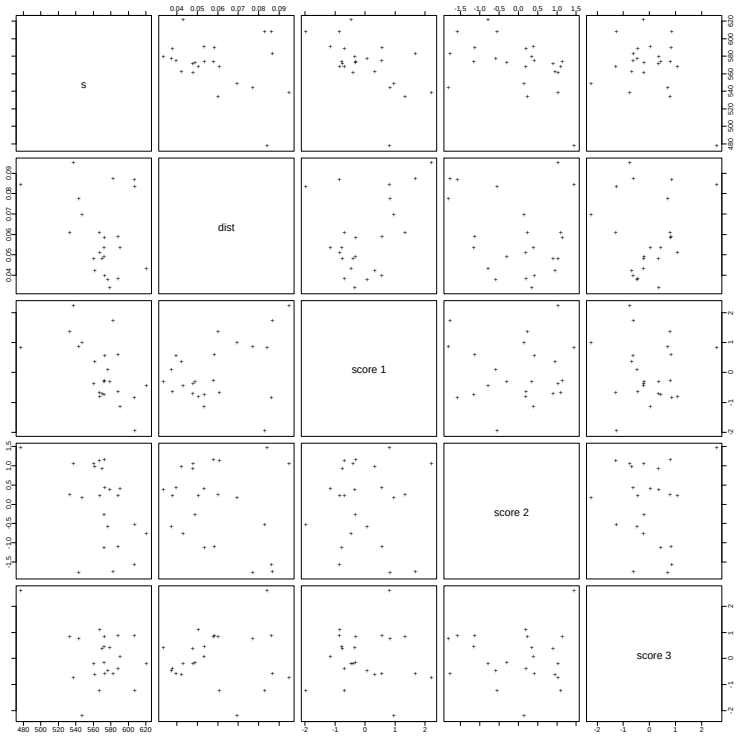

Figure 58 Pairwise plots of $(s_i, d_{Fi}, c_{i1}, c_{i2}, c_{i3})^{\mathrm{T}}$, $i = 1, \ldots, n$, centroid size, full Procrustes distance to the mean shape and the first three standardized PC scores, for the T2 Small outline data.

5.5.2 Point distribution models

Principal components analysis with the full Procrustes coordinates of Equation (3.13) has a particularly simple formulation. Cootes et al. (1992, 1994) use PCA to develop the 'point distribution model' (PDM), which is a PC model for shape and uses Procrustes residuals rather than tangent coordinates. Given n independent configuration matrices $X_1, ..., X_n$ the figures are registered to $X_1^P, \ldots, X_n^P$ by full generalized Procrustes analysis. The estimate of mean shape is taken to be the full Procrustes mean $\hat{\mu}$ which has the same shape as $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i^P$. The sample covariance matrix is

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

and the PCs are the eigenvectors of this matrix, $\gamma_j, j = 1, \ldots, \min(n-1, M)$, with corresponding decreasing eigenvalues λ_j. Note that PCA using this formulation is the same (up to an overall scaling) as using the full Procrustes tangent coordinates of Equation (3.13), with the tangent coordinates pre-multiplied by H^{T}, the transpose of the Helmert sub-matrix.

Visualization of the PCs is carried out as in the previous section. In particular, the structure in the jth PC can be viewed through plots of an icon for mean shape $\hat{\mu}$ with displacement vectors

$$\mathrm{vec}(\hat{\mu}) + c{\lambda_j}^{1/2}\gamma_j$$

for the shapes corresponding to $c \in [-3, 3]$.

Example 5.3 An example of the PDM approach is given

in Figure 59 taken from Cootes et al. (1994), who describe a flexible model for describing shape variability in hands. In Figure 59 the first PC **includes** the spreading of the fingers, the second PC includes movement of the thumb relative to the fingers, and the third PC includes movement of the middle finger. **The shape variability here is complicated because there are multiple effects.** As well as the biological shape variability of hands, the relative positions of the fingers contributes greatly to shape variability here. □

5.5.3 PCA in shape analysis and multivariate analysis

The algebra in geometrical shape analysis is the same as that of conventional PCA in multivariate analysis, since we are assuming our data are sufficiently concentrated that

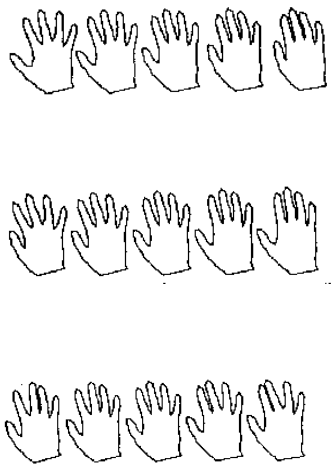

Figure 59 Varying hands: the first three PCs (from Cootes et al., 1994) with values of $c \in \{-2, -1, 0, 1, 2\}$ here.

a tangent plane approximation to shape space can used. However, what is different here is that we project back into configuration space to visualize directly and clearly the effects of each PC.