3

Planar Procrustes Analysis

3.1 Introduction

This chapter provides a straightforward introduction to some key concepts when a random sample of planar objects is available. Important aspects of shape analysis are to obtain a measure of distance between shapes, to estimate average shapes from a random sample and to estimate shape variability from a random sample. In this chapter we discuss these ideas by considering the particular case of data in two dimensions. The use of complex notation leads to simple methodology in this important case.

A more comprehensive treatment of shape distances is given in Chapter 4 and a detailed description of Procrustes analysis is given in Chapter 5.

3.2 Shape Distance and Procrustes Matching

Consider two **centred** configurations $y = (y_1, ..., y_k)^{\mathrm{T}}$ and $w = (w_1, ..., w_k)^{\mathrm{T}}$, both in $\mathbb{C}^k$, with $y^* 1_k = 0 = w^* 1_k$, where y^* denotes the transpose of the complex conjugate of y. In order to compare the configurations in shape we need to establish a measure of distance between the two shapes.

A suitable procedure is to match w to y using the similarity transformations and the differences between the fitted and observed y indicate the magnitude of the difference in shape between w and y. Consider the complex regression equation

$$\begin{aligned} y &= (a+ib)1_k + \beta e^{i\theta} w + \epsilon \\ &= [1_k, w]A + \epsilon \\ &= X_D A + \epsilon, \end{aligned} \tag{3.1}$$

where $A = (A_1, A_2)^{\mathrm{T}} = (a + ib, \beta e^{i\theta})^{\mathrm{T}}$ are the 2×1 complex parameters with translation $a+ib$, scale $\beta > 0$ and rotation $0 \leq \theta < 2\pi$; ϵ is a $k \times 1$ complex error vector; and $X_D = [1_k, w]$ is the $k \times 2$ 'design matrix'. To carry out the superimposition we could estimate A by minimizing the least squares objective function, the sum of square errors

$$D^2(y, w) = \epsilon^* \epsilon = (y - X_D A)^* (y - X_D A),$$

where $z^* = (c+id)^* = (c-id)^T$ is the complex conjugate of the transpose of the complex vector z, where $c = \mathrm{Re}(z)$ and $d = \mathrm{Im}(z)$. The full Procrustes superimposition of w

on y is obtained by estimating A with $\hat{A}$, where

$$\hat{A} = (\hat{a}+i\hat{b}, \hat{\beta}e^{i\hat{\theta}})^{\mathrm{T}} = \text{arginf } \epsilon^*\epsilon = \text{arginf}(y-X_DA)^*(y-X_DA).$$

Definition 3.1 *The* **full Procrustes fit (superimposition)** *of w onto y is*

$$w^P = X_D\hat{A} = (\hat{a}+i\hat{b})1_k + \hat{\beta}e^{i\hat{\theta}}w,$$

where $(\hat{\beta}, \hat{\theta}, \hat{a}, \hat{b})$ are chosen to minimize

$$D^2(y,w) = \|y - w\beta e^{i\theta} - (a+ib)1_k\|^2.$$

Remember that we are taking the configurations y and w to be both centred (or ‘centered’ using US spelling of course), i.e. $1_k^T w = 0 = 1_k^T y$.

Result 3.1 *The full Procrustes fit has matching parameters*

$$\hat{a}+i\hat{b} = 0, \tag{3.2}$$

$$\hat{\theta} = \arg(w^* y) = -\arg(y^* w), \tag{3.3}$$

$$\hat{\beta} = (w^* y y^* w)^{1/2}/(w^* w). \tag{3.4}$$

Proof: We wish to minimize (over β, θ, a,b) the expression

$$\begin{aligned} D^2 &= \epsilon^* \epsilon \\ &= \|y - w\beta \mathrm{e}^{i\theta} - (a+ib)1_k\|^2 \qquad (3.5) \\ &= y^* y + \beta^2 w^* w - y^* w \beta \mathrm{e}^{i\theta} - w^* y \beta \mathrm{e}^{-i\theta} + k(a^2+b^2) \end{aligned}$$

(remember y and w are centred). Clearly, the minimizing a and b are zero. Let $y^* w = \gamma \mathrm{e}^{i\phi}$ ($\gamma \geq 0$) and then

$$\beta(y^* w \mathrm{e}^{i\theta} + w^* y \mathrm{e}^{-i\theta}) = \beta(\gamma \mathrm{e}^{i(\theta+\phi)} + \gamma \mathrm{e}^{-i(\theta+\phi)}) = 2\beta\gamma\cos(\theta+\phi).$$

So to minimize $\|y - \beta \mathrm{e}^{i\theta} w\|^2$ over θ we need to maximize $2\beta\gamma\cos(\theta+\phi)$. Clearly, a solution for θ is $\hat{\theta} = -\phi =$

$-\arg(y^* w)$. To find the minimizing scale we solve

$$\frac{\partial D^2}{\partial \beta} = 0 = 2\beta w^* w - 2\gamma,$$

where $\gamma = |y^* w|$. Hence,

$$\hat{\beta} = |y^* w|/(w^* w)$$

as required. □

The solution is the standard least squares solution (but with complex variables) and we can write the solution in the familiar form

$$\hat{A} = (\hat{A}_1, \hat{A}_2)^{\mathrm{T}} = (X_D^* X_D)^{-1} X_D^* y \Rightarrow \hat{A}_1 = 0, \quad \hat{A}_2 = w^* y/(w^* w). \tag{3.6}$$

Note that the full Procrustes fit of w onto y is given explicitly by

$$w^P = X_D \hat{A} = \hat{\beta} \mathrm{e}^{i\hat{\theta}} w = w^* y w/(w^* w).$$

The residual vector $r = y - X_D \hat{A}$ is given by

$$r = [I_k - X_D(X_D^* X_D)^{-1} X_D^*]y = (I_k - H_{hat})y$$

where H_{hat} is the ‘hat’ matrix for X_D, i.e.

$$H_{hat} = X_D(X_D^* X_D)^{-1} X_D^*.$$

The minimized value of the objective function is

$$D^2(r, 0) = r^* r = y^* y - (y^* w w^* y)/(w^* w). \qquad (3.7)$$

Now this expression is not symmetric in y and w unless $y^* y = w^* w$. A convenient standardization is to take the configurations to be unit size, i.e.

$$\sqrt{y^* y} = \sqrt{w^* w} = 1.$$

So, if we include standardization, then we obtain a suitable measure of shape distance.

Definition 3.2 *The* **full Procrustes distance** *between* *complex configurations* w *and* y *is given by*

$$\begin{aligned} d_F(w,y) &= \inf_{\beta,\theta,a,b} \left\| \frac{y}{\|y\|} - \frac{w}{\|w\|}\beta e^{i\theta} - 1_k(a+ib) \right\| \\ &= \left\{ 1 - \frac{y^* w w^* y}{w^* w y^* y} \right\}^{1/2}. \qquad (3.8) \end{aligned}$$

The expression for the distance follows from Equation (3.7).

Important point: The term **full** is used because the full set of Euclidean similarity transformations is estimated in the matching (translation, rotation **and** scale), but note that y and w are pre-scaled to unit size.

Note that the full Procrustes fit of w onto y is actually obtained by complex linear regression of y on w.

The term ‘Procrustes’ is used because the above matching operations are identical to those of Procrustes analysis, a commonly used technique for comparing matrices (up to transformations) in multivariate analysis (see Mardia et al., 1979, p.416). In Procrustes analysis the optimal transformation parameters are estimated by minimizing a least squares criterion. The term ‘Procrustes analysis’ was first used by Hurley and Cattell (1962) in factor analysis.

In Greek mythology Procrustes was the nickname of a robber Damastes, who lived by the road from Eleusis to Athens. He would offer travellers a room for the night and fit them to the bed by stretching them if they were too short or chopping off their limbs if they were too tall. The analogy is rather tenuous but we can regard one configuration as the bed and the other as the person being ‘translated’, ‘rotated’ and possibly ‘rescaled’ so as to fit as close as possible to the bed.

We discuss Procrustes methods in further detail in Chapter 5.

Example 3.1 Consider a juvenile and an adult from the sooty mangabey data of Section 1.2.9. The unregistered outlines are shown in Figure 29. In Figure 30 we see the full Procrustes fit of the adult onto the juvenile and in

Figure 31 we see the full Procrustes fit of the juvenile onto the adult. In matching the juvenile to the adult $\hat{\theta} = 45.5^\circ$ and $\hat{\beta} = 1.131$. We see that the estimate of scale in matching the adult to the juvenile is $\hat{\beta}^R = 0.875$ and the rotation is $\hat{\theta}^R = -45.5^\circ$. Note that $\hat{\beta}^R \neq 1/\hat{\beta}$ because the adult and juvenile are not the same size (the matching is not symmetric). Computing the measure of full Procrustes shape distance we see that $d_F = 0.105$. Further understanding of shape distances is given in Chapter 4 which will help us to interpret this value of the distance. $\square$

This is not the only choice of distance between shapes, and further choices of distance are considered in Section 4.2.2. However, the full Procrustes distance is a natural distance from a statistical point of view, obtained from a

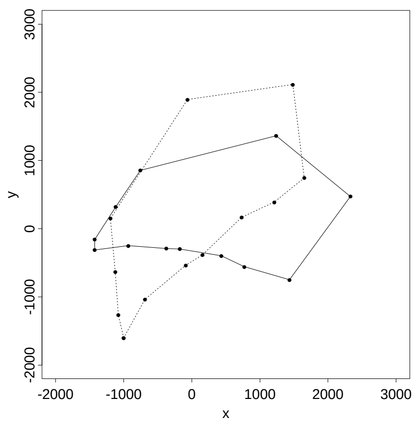

Figure 29 Unregistered sooty mangabeys: juvenile (——) and adult (- - -).

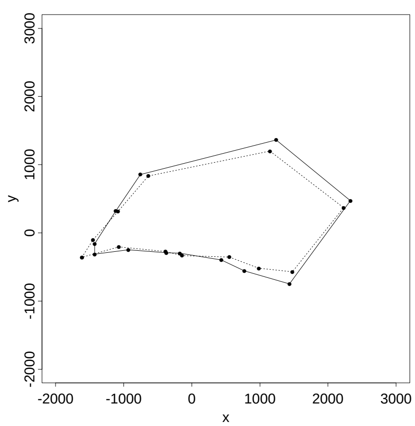

Figure 30 The Procrustes fit of the adult sooty mangabey (- - -) onto the juvenile (——).

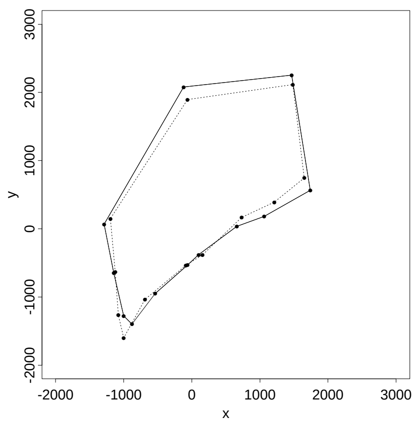

Figure 31 The Procrustes fit of the juvenile sooty mangabey (——) onto the adult (- - -).

least squares criterion and optimizing over the full set of similarity parameters. The squared full Procrustes distance naturally appears exponentiated in the density for many simple probability distributions for shape, as we shall see in Chapter 6.

The subject of shape analysis is different from conventional multivariate analysis, because the invariances under similarity transformations lead to a non-Euclidean distance as the suitable measure of distance between shapes. We shall use the full Procrustes distance to assess closeness of shapes and to provide a criterion for estimating a mean shape.

3.3 Estimation of Mean Shape

There are many situations where we wish to obtain an estimate of an average shape. We now consider a method for estimating a population mean shape, which provides a suitable notion of average shape.

Consider the situation where a random sample of configurations $w_1, \ldots, w_n$ is available from the perturbation model

$$w_i = \gamma_i 1_k + \beta_i \mathrm{e}^{i\theta_i}(\mu + \epsilon_i), \quad i = 1, \ldots, n,$$

where $\gamma_i \in \mathbb{C}$ are translation vectors, $\beta_i \in \mathbb{R}^+$ are scale parameters, $0 \leq \theta_i < 2\pi$ are rotations, $\epsilon_i \in \mathbb{C}$ are independent zero mean complex random errors, and μ is the population mean configuration. We can estimate $[\mu]$, the shape of the population mean (**mean shape**), by a variety of methods.

Definition 3.3 *The* **full Procrustes estimate of mean shape** $[\hat{\mu}]$ *is obtained by minimizing (over* μ*) the sum of square full Procrustes distances from each* w_i *to an unknown unit size mean configuration* μ*, i.e.*

$$[\hat{\mu}] = \arg\inf_{\mu} \sum_{i=1}^{n} d_F^2(w_i, \mu).$$

We shall often use the term **full Procrustes mean** instead of 'full Procrustes estimate of mean shape', in order to shorten terminology. It should be remembered throughout that the full Procrustes mean is a sample estimate of mean shape.

Let us assume that the configurations $w_1, \ldots, w_n$ have been centred, so that $w_i^* 1_k = 0$.

Result 3.2 (Kent, 1994) *The* **full Procrustes mean shape** $[\hat{\mu}]$ *can be found as the eigenvector corresponding to the largest eigenvalue of the* **complex sum of squares and products matrix**

$$S = \sum_{i=1}^{n} w_i w_i^* / (w_i^* w_i) = \sum_{i=1}^{n} z_i z_i^*, \tag{3.9}$$

where the $z_i = w_i / \|w_i\|, \quad i = 1, \ldots, n$, *are the pre-shapes.*

Proof: We wish to minimize

$$\sum_{i=1}^{n} d_F^2(w_i, \mu) = \sum_{i=1}^{n} \left\{ 1 - \frac{\mu^* w_i w_i^* \mu}{w_i^* w_i \mu^* \mu} \right\} \tag{3.10}$$

$$= n - \mu^* S \mu / (\mu^* \mu). \tag{3.11}$$

Therefore,

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

Hence, $\hat{\mu}$ is given by the complex eigenvector corresponding to the largest eigenvalue of S (using, for example, Mardia et al., 1979, Equation A.9.11). All rotations of $\hat{\mu}$ are also solutions, but these all correspond to the same shape $[\hat{\mu}]$. $\square$

The eigenvector is unique (up to a rotation) provided there is a single largest eigenvalue of S (which is the case for most practical datasets). We shall see in Section 6.2 that the solution corresponds to the maximum likelihood estimate of modal shape under the complex Bingham model.

The **full Procrustes fits** or **full Procrustes coordinates** of $w_1, \ldots, w_n$ are

$$w_i^P = w_i^* \hat{\mu} w_i / (w_i^* w_i), \; i = 1, \ldots, n, \tag{3.12}$$

where each w_i^P is the full Procrustes fit of w_i onto $\hat{\mu}$. Calculation of the full Procrustes mean shape can also be obtained by taking the arithmetic mean of the full Procrustes coordinates, i.e. $\frac{1}{n}\sum_{i=1}^n w_i^P$ has the same shape as the Procrustes mean shape $[\hat{\mu}]$ (see Result 5.2).

The **Procrustes residuals** are calculated as

$$r_i = w_i^P - \left(\frac{1}{n} \sum_{i=1}^n w_i^P \right) \;, \;\; i = 1, \ldots, n, \tag{3.13}$$

and the Procrustes residuals are useful for investigating shape variability.

Important point: When several objects are fitted using Procrustes superimposition the method has been called **generalized Procrustes analysis (GPA)** (Gower, 1975), whereas when a single object is fitted to one other, as in Section 3.2, the method has been called **ordinary Procrustes analysis (OPA)**. Note that OPA is not symmetrical in the ordering of the objects, whereas GPA is invariant under re-orderings of the objects.

An alternative equivalent procedure to working with centred configurations would be to work with the Helmertized landmarks Hw_i, where H is the sub-Helmert matrix given in Equation (2.9). This procedure was originally used by Kent (1991, 1992, 1994) and the least squares estimate of shape is the leading eigenvector $\hat{\mu}_1$ of HSH^{T}. Note that $H^{\mathrm{T}}\hat{\mu}_1$ is identical to $\hat{\mu}$, up to an arbitrary rotation.

Definition 3.4 *To obtain an overall measure of shape* *variability we consider the* **root mean square** $RMS(d_F)$ *of full Procrustes distance from each configuration to the full Procrustes mean* $[\hat{\mu}]$,

$$RMS(d_F) = n^{-1} \sum_{i=1}^{n} d_F^2(w_i, \hat{\mu}). \tag{3.14}$$

Example 3.2 In Figure 32 we see the raw digitized data from the female and male gorilla skulls from the dataset described in Section 1.2.2. The landmarks have been recorded by a digitizer to be registered so that $opisthion$ is at the origin and the line from $opisthion$ to $basion$ is horizontal. There are $k = 8$ landmarks in $m = 2$ dimensions. In Figures 33 and 34 we also see full Procrustes fits of the females and males separately. For each sex the landmarks match up quite closely because the shape variability is small. The full Procrustes mean for each sex is found from the dominant eigenvector of the complex sum of squares and products matrix for each sex. In Figure 35 we see the full Procrustes superimposition of the female average shape and the male average shape (by GPA). It is also of interest to assess whether there is a significant average shape difference between the sexes and, if so, to describe the difference. We consider methods for testing for average shape differences in Chapter 7. The full Procrustes distance d_F between the mean shapes is 0.059, and the within-sample $RMS(d_F)$ is 0.044 for females and 0.050 for males. We see later in Section 7.1.2 that the difference in mean shapes between the sexes is statistically significant. $\square$

3.4 Shape Variability

After having obtained an average configuration we often wish to examine the structure of shape variability in a sample. A suitable method is to investigate the shape variability in a linearized space about the average shape (a tangent space). For example, we could consider principal components analysis of the Procrustes residuals (which are approximate tangent coordinates) and we now denote the real vectors of the tangent coordinates as $v_i, i = 1, \ldots, n$. These could be the Procrustes residuals r_i of Equation (3.13) or another choice of tangent coordinates which we introduce in Chapter 4.

Definition 3.5 *Let* S_v *be the sample covariance matrix of* *some tangent coordinates* v_i, *i.e.*

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

where $\bar{v} = \frac{1}{n} \sum v_i$. *The orthonormal eigenvectors of* S_v, *denoted by* γ_j, *are the* **principal components** *of* S_v *with corresponding eigenvalues* $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$, *where* $p = \min(n-1, M)$, *where* $M = 2k - 4$ *is the dimension of the shape space. The* **principal component score** *for the* i*th individual on the* j*th principal component (PC) is given by*

$$s_{ij} = \gamma_j^{\mathrm{T}} (v_i - \bar{v}), \ i = 1, \ldots, n \ ; \ j = 1, \ldots, p,$$

and a PC summary of the data in the tangent space is

$$v_i = \bar{v} + \sum_{j=1}^{p} s_{ij} \gamma_j, \tag{3.15}$$

for $i = 1, \ldots, n$. *The* **standardized PC scores** *are*

$$c_{ij} = s_{ij} / \lambda_j^{1/2}, \ i = 1, \ldots, n; j = 1, \ldots, p.$$

Note that we usually choose the pole such that $\bar{v} \approx 0$ or $\bar{v} = 0$. The effect of the jth PC can be seen by examining Equation (3.15) for various values of the standardized PC score, and in particular we examine

$$v = \bar{v} + c\lambda_j^{1/2}\gamma_j, \quad j = 1, \ldots, p, \tag{3.16}$$

for a range of values of the standardized PC score c and then project back into configuration space (by adding on the Procrustes mean if using Procrustes residuals).

Suitable values that would cover the full range of the data are $c \in [-3, 3]$ as we would have approximately $c \sim N(0, 1)$ under a multivariate normal model for the tangent coordinates (and hence 99.7% of the variability in the range $c \in [-3, 3]$). Sometimes we wish to exaggerate the effect of a PC by considering c over a wider range, e.g. $c \in [-6, 6]$.

The percentage of variability captured by the jth PC ($j = 1, \ldots, p$) is

$$\frac{100\lambda_j}{\sum_{j=1}^{p} \lambda_j}.$$

Example 3.3 Consider the mouse vertebrae data described in Section 1.2.1. There are $k = 6$ landmarks in $m = 2$ dimensions. The analysis here is similar to Kent (1994). In Figure 36 we have a plot of the Procrustes mean shape obtained from the dominant eigenvector of the complex sum of squares and products matrix. The Procrustes mean shape is centred, with unit size, and rotated so that the line joining the two farthest apart landmarks is horizontal.

Hence, the mean shape has coordinates

$$(-0.51-0.14i, 0.51-0.14i, 0.09+0.15i, 0.01+0.42i, -0.07+0.16i, -0.03-0.45i)^{\mathrm{T}}.$$

In order to examine the structure of variability we examine the eigenstructure of the sample covariance matrix S_v of the Procrustes residuals. The square roots of the eigenvalues of S_v are

$$0.054, 0.020, 0.018, 0.017, 0.011, 0.010, 0.008, 0.005, 0, 0, 0, 0.$$

Hence, the first two principal components explain 69% and 10% of the variability, respectively. The last four zero eigenvalues are expected due to the four constraints for location, rotation and scale. For each principal component, shapes at 6 standard deviations away from the mean are calculated. In Figure 36 we see the mean shape with these unit vectors drawn for the first two PCs. The vectors of the first and second PCs in the figure are given by

$$(0.11 - 0.11i, -0.11 - 0.12i, -0.04, -0.01 + 0.22i, 0.04 - 0.01i, 0.01 + 0.01i)^{\mathrm{T}},$$

$$(0.05 + 0.03i, 0.04 + 0.01i, -0.01, -0.06 + 0.01i, 0.03 - 0.02i, -0.06 - 0.03i)^{\mathrm{T}}.$$

There appears to be a high dependence between certain landmarks, as indicated by the fact that the first PC explains such a large proportion of the variability. The first PC involves a shift downwards and inwards for landmarks 1 and 2, balanced by an upwards movement for landmark 4. ‘At the same time landmarks 3 and 5 move inwards slightly whereas there is little movement in landmark 6. The second PC is not symmetric. If we had chosen to display the PCs in a different manner (e.g. relative to landmarks 1 and 2 as in Bookstein coordinates), then our interpretation would be different. In particular we display the PCs in Figure 37 where the mean and a figure at

3 standard deviations along the PC have been registered relative to landmarks 1 and 2. Our interpretation would be that the first PC includes the movement of landmarks 3, 4, 5 and 6 upwards relative to points 1 and 2. Landmark 4 shows the largest movement, followed by 3 and 5 together and landmark 6 shows the smallest movement. Both of these interpretations are correct, as they are describing the same PCs. □

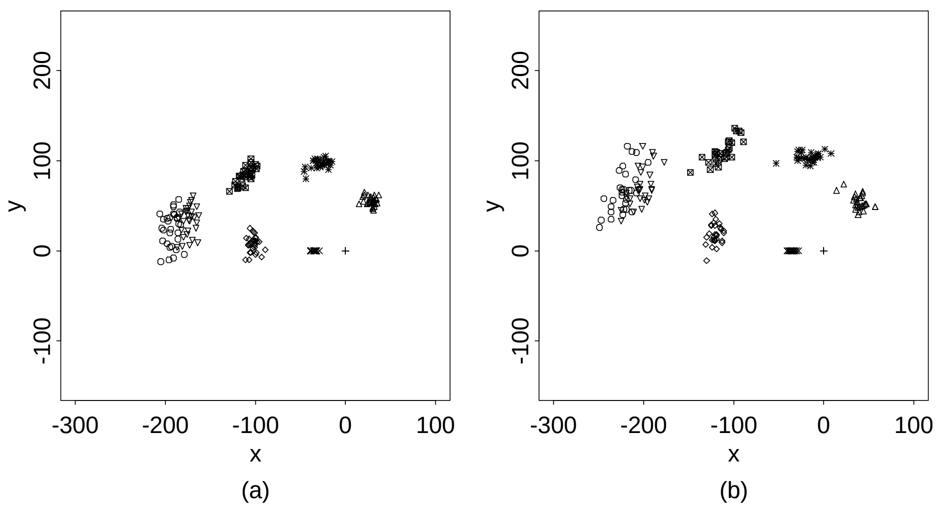

Figure 32 (a) The 30 female gorilla skull landmarks registered in the coordinate system as recorded by a digitizer. (b) The original 29 male gorilla skull landmarks.

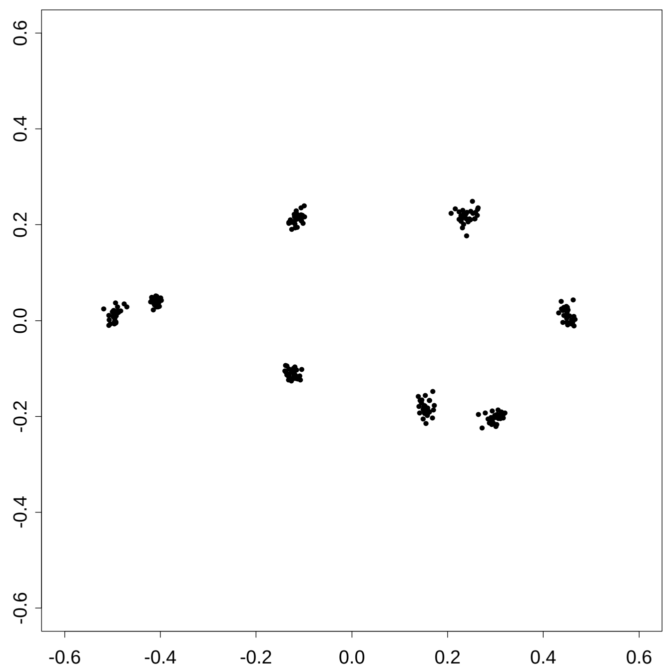

Figure 33 The full Procrustes fits of the female gorilla skulls.

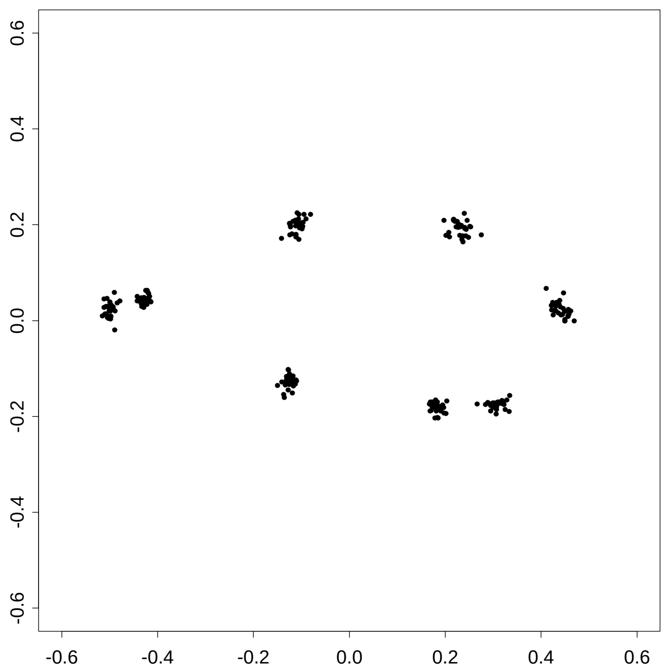

Figure 34 The full Procrustes fits of the male gorilla skulls.

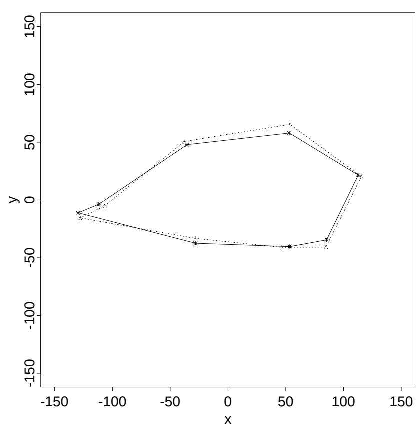

Figure 35 The male (—-) and female (- - -) full Procrustes mean shapes registered by GPA.

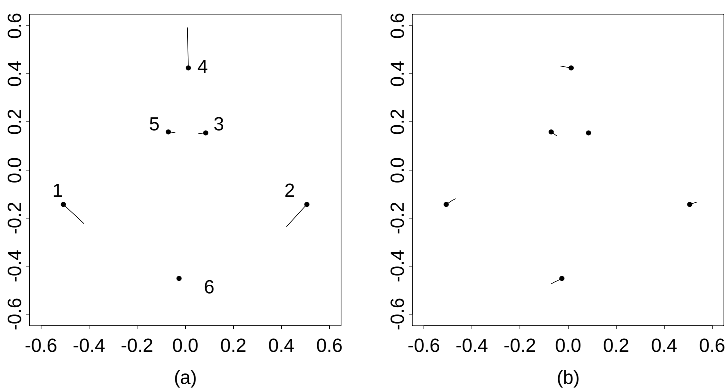

Figure 36 The Procrustes mean shape of the T2 vertebra data (landmarks at the dots) and vectors to 6 standard deviations along the first and second principal components. The first PC is shown in (a) and the second PC is shown in (b).

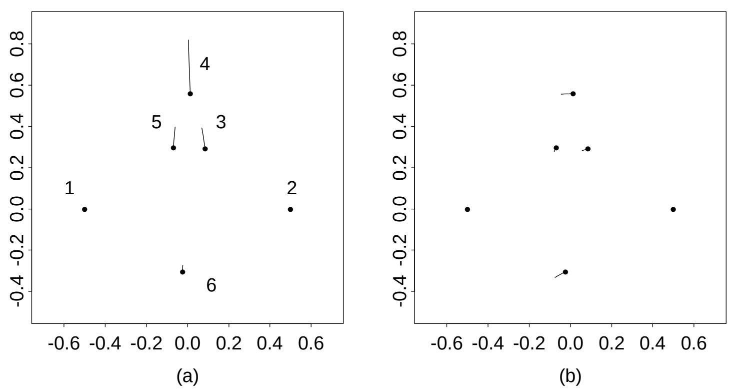

Figure 37 The mean shape with vectors to a figure at 6 standard deviations along the first PC (a) and second PC (b), both the same PCs as in Figure 36 but the icons are registered on a common baseline 1, 2.