Resistant GPA algorithms based on the M and LMS estimation

Geehong Hyun[a], Bo-Hui Lee[b], Yong-Seok Choi[1,b]

[a]Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, USA;
[b]Department of Statistics, Pusan National University, Korea

Abstract

Procrustes analysis is a useful technique useful to measure, compare shape differences and estimate a mean shape for objects; however it is based on a least squares criterion and is affected by some outliers. Therefore, we propose two generalized Procrustes analysis methods based on M-estimation and least median of squares estimation that are resistant to object outliers. In addition, two algorithms are given for practical implementation. A simulation study and some examples are used to examine and compared the performances of the algorithms with the least square method. Moreover since these resistant GPA methods are available for higher dimensions, we need some methods to visualize the objects and mean shape effectively. Also since we have concentrated on resistant fitting methods without considering shape distributions, we wish to shape analysis not be sensitive to particular model.

Keywords: generalized Procrustes analysis, least median of squares estimator, least squares, M-estimation, mean shape, resistant, shape analysis

1. Introduction

The shape is all geometrical information that remains when location, size and rotational effects are filtered from an object (Kendall, 1984). Two objects have the same shape if they can be translated, scaled and rotated to each other such that they match exactly. Procrustes analysis is an important tool for estimating a mean shape as well as measuring, describing and comparing the shapes of objects in many disciplines.

When several objects are fitted using Procrustes analysis the method has been called generalized Procrustes analysis (GPA) (Gower, 1975); however, the method has been called ordinary Procrustes analysis (OPA) when a single objects is fitted to one other. Since the methods use a least squares (LS) technique, they are not resistant to landmarks or object outliers. We consider procedures that are not sensitive to these outliers. Verboon and Heiser (1992) proposed the method using M-estimator in OPA, in this paper we produce the theorem for estimating parameters on high dimension and Dryden and Walker (1999) adapted the least median of squares (LMS) estimator.

In GPA, we propose two algorithms, calling them M GPA and LMS GPA algorithms, having resistant property in low concentration (which implies the data with large variability) as well as high concentration data (which implies the data with small variability) with several object outliers. We then compare these methods with LS method by visualizing the estimated mean shapes. A simulation and some examples is used to demonstrate the performances of the algorithms explicitly.

---

2. Procrustes analysis

Procrustes analysis is a very useful tools to analyze landmark data and involves fitting the objects using similarity transformations to be as close as possible by minimizing the sum of squared Euclidean distances. There are two techniques are two techniques in Procrustes analysis. One is OPA, which is used for fitting and comparing two configurations. Another is GPA for obtaining a mean shape and to explore the structure of shape variability in a dataset of more than two configurations.

Consider two centered $k \times m$ matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ of coordinates from k landmarks in m dimensions. The estimation of $\boldsymbol{\alpha}, \boldsymbol{\Gamma}$, and β by OPA method based on LS (LS OPA) is carried out by minimizing the squared Euclidean distance

$$O(\mathbf{X}_1, \mathbf{X}_2) = \left\| \mathbf{X}_2 - \beta \mathbf{X}_1 \boldsymbol{\Gamma} - \mathbf{1}_k \boldsymbol{\alpha}^T \right\|^2, \tag{2.1}$$

where $\|\mathbf{A}\| = \sqrt{\operatorname{tr}(\mathbf{A}^T \mathbf{A})}$, $\boldsymbol{\alpha}$ is an $m \times 1$ location, $\boldsymbol{\Gamma}$ is an $m \times m$ rotation such that $|\boldsymbol{\Gamma}| = \mathbf{1}$ and $\beta > 0$ is a scale parameter (Dryden and Mardia, 2016).

Consider GPA method by LS (LS GPA) with $n \geq 2$ configurations of k landmarks in $m \geq 2$ dimensions. Let $k \times m$ configurations $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be centered and scaled. We wish to minimize following objective function,

$$G(\mathbf{X}_1, \ldots, \mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left\| \left(\beta_i \mathbf{X}_i \boldsymbol{\Gamma}_i + 1_k \boldsymbol{\alpha}_i^T \right) - \left(\beta_j \mathbf{X}_j \boldsymbol{\Gamma}_j + 1_k \boldsymbol{\alpha}_j^T \right) \right\|^2$$

$$= \sum_{i=1}^{n} \left\| \left(\beta_i \mathbf{X}_i \boldsymbol{\Gamma}_i + \mathbf{1}_k \boldsymbol{\alpha}_i^T \right) - \boldsymbol{\mu} \right\|^2, \tag{2.2}$$

subject to a constraint on the centroid size of the mean shape $\boldsymbol{\mu}$,

$$C(\boldsymbol{\mu}) = \|\mathbf{C}\boldsymbol{\mu}\| = 1,$$

where $\boldsymbol{\alpha}_i$ is location, $\boldsymbol{\Gamma}_i$ is rotation such that $|\boldsymbol{\Gamma}| = 1$, $\beta_i > 0$ is scale, $i = 1, \ldots, n$, $C(\boldsymbol{\mu})$ is the centroid size, and $\mathbf{C}$ is the $k \times k$ centering matrix.

Then the solution to the minimization of equation (2.2) over $\boldsymbol{\mu}$ is given by (Dryden and Mardia, 2016)

$$\hat{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{\beta}_i \mathbf{X}_i \hat{\boldsymbol{\Gamma}}_i + \mathbf{1}_k \hat{\boldsymbol{\alpha}}_i^T \right). \tag{2.3}$$

An explicit eigenvector solution to the mean shape for two dimensional data can be found as the complex eigenvector corresponding to the largest eigenvalue of the complex sum of squares and products matrix (Kent, 1994). The Procrustes mean shape $\hat{\boldsymbol{\mu}}$ has to be found iteratively for $m \geq 3$ dimensional data. The GPA algorithm of Gower (1975), which is modified by Ten Berge (1977), is useful for a practical implementation when the objects are in $m > 2$ dimensions.

3. Resistant Procrustes analysis

3.1. M and LMS OPA

The LS OPA method in our the fitting problem is not resistant since it is badly affected by landmark outliers; therefore, we need some resistant fitting methods. These methods have been studied for

appropriate fit of OPA in shape analysis. In particular, Verboon and Heiser (1992) suggested a resistant fit based on M-estimator. However, the procedure may not work well when a scale parameter is introduced and location is not considered. We would like to consider the fitting under the full set of Euclidean similarity transformations in our application. Dryden and Walker (1999) also adapt the LMS of Rousseeuw (1984) for shape analysis (LMS OPA).

Now we describe resistant ordinary Procrustes analysis using M-estimator (M OPA). Let two matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ denote configurations of k landmarks ($k \geq 3$). Recall that the object $\mathbf{X}_1$ can be fitted onto another object $\mathbf{X}_2$ by minimizing objective function, $O(\alpha, \Gamma, \beta)$, defined as

$$O(\alpha, \Gamma, \beta) = \sum_{i=1}^{k} \rho(\|\mathbf{r}_i\|), \tag{3.1}$$

where residual $\mathbf{r}_i = \mathbf{x}_{(2)i} - \beta \Gamma^T \mathbf{x}_{(1)i} - \alpha$, $\mathbf{x}_{(1)i}$, and $\mathbf{x}_{(2)i}$ are vectors of ith landmark of $\mathbf{X}_1$ and $\mathbf{X}_2$ respectively, and α is a location vector, Γ rotation matrix and β positive scale parameter, and $\rho(\cdot)$ represents a loss function.

Then, Verboon and Heiser (1992) considered the use of Huber's function

$$\rho(\|\mathbf{r}\|) = \begin{cases} \dfrac{\|\mathbf{r}\|^2}{2}, & \|\mathbf{r}\| \leq c, \\[2mm] c\|\mathbf{r}\| - \dfrac{c^2}{2}, & \|\mathbf{r}\| > c \end{cases} \tag{3.2}$$

or biweight function,

$$\rho(\|\mathbf{r}\|) = \begin{cases} c^2 \left[1 - \left(1 - \dfrac{\|\mathbf{r}\|^2}{c^2} \right)^3 \right] / 6, & \|\mathbf{r}\| \leq c, \\[3mm] \dfrac{c^2}{6}, & \|\mathbf{r}\| > c \end{cases} \tag{3.3}$$

as a loss function in (3.1) where c is some tuning constant. For large values of the tuning constant c, objective function reduces to LS OPA.

Theorem 1. *Let two matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ denote configurations of $k \geq 3$ landmarks in $m \geq 2$ dimensions. M OPA estimates of α, Γ, and β can be found out by minimizing the objective function (3.1).*

Then the solutions are written as

$$\hat{\alpha} = \frac{\left(\mathbf{X}_2^T - \hat{\beta} \hat{\Gamma}^T \mathbf{X}_1^T \right) \mathbf{W} \mathbf{1}_k}{\mathbf{1}_k^T \mathbf{W} \mathbf{1}_k}, \tag{3.4}$$

$$\hat{\Gamma} = \mathbf{U} \mathbf{V}^T, \tag{3.5}$$

$$\hat{\beta} = \frac{tr \left[\hat{\Gamma}^T \mathbf{X}_1^T \mathbf{W} \left(\mathbf{I}_k - \frac{\mathbf{1}_k \mathbf{1}_k^T \mathbf{W}}{\mathbf{1}_k^T \mathbf{W} \mathbf{1}_k} \right) \mathbf{X}_2 \right]}{tr \left[\mathbf{X}_1^T \mathbf{W} \left(\mathbf{I}_k - \frac{\mathbf{1}_k \mathbf{1}_k^T \mathbf{W}}{\mathbf{1}_k^T \mathbf{W} \mathbf{1}_k} \right) \mathbf{X}_1 \right]}, \tag{3.6}$$

where $\mathbf{W} = diag(w_1, \ldots, w_k)$ is a diagonal matrix with weight $w_i = \psi(\|\mathbf{r}_i\|) / \|\mathbf{r}_i\|$, $i = 1, \ldots, k$, $\psi(\cdot) = \rho(\cdot)'$, the derivative of $\rho(\cdot)$, and $\mathbf{U}, \mathbf{V} \in SO(m)$ satisfying the singular value decomposition

$$\mathbf{X}_1^T \mathbf{W} \left(\mathbf{I}_k - \frac{\mathbf{1}_k \mathbf{1}_k^T \mathbf{W}}{\mathbf{1}_k^T \mathbf{W} \mathbf{1}_k} \right) \mathbf{X}_2 = \mathbf{U} \mathbf{D}_\lambda \mathbf{V}^T, \tag{3.7}$$

with $\mathbf{D}_\lambda$ *diagonal matrix of singular values.*

The objective function $O(\alpha, \mathbf{\Gamma}, \beta)$ of Theorem 1 can be minimized by an algorithm that is based on the method of iteratively reweighted least squares. There are two steps in the algorithm. First, the weights are calculated by $w_i = \psi(\|\mathbf{r}_i\|) / \|\mathbf{r}_i\|$ with $\rho(\cdot)' = \psi(\cdot)$. We use the weight matrix $\mathbf{W} = \mathbf{1}_k$ as initial values. Next the parameters $\alpha, \mathbf{\Gamma}$, and β are calculated by weighted least squares as given by equations (3.4), (3.5), and (3.6), respectively. We stop the iteration if the conditions $|O^{(s+1)} - O^{(s)}| < \epsilon$ is satisfied, where $O^{(s)}$ is the values of objective function $O(\alpha, \mathbf{\Gamma}, \beta)$ on s^{th} iteration.

Another resistant estimator is the LMS estimator, which has a breakdown of almost 50% (Rousseeuw, 1984) as an alternative to the repeated median. The objective function which Dryden and Walker (1999) have considered, is defined by

$$s^2 (\mathbf{E}) = \text{median} \left(\|(\mathbf{E})_1\|^2, \dots, \|(\mathbf{E})_k\|^2 \right), \tag{3.8}$$

where $\mathbf{E}$ is an error matrix of regression equation

$$\mathbf{X}_2 = g(\mathbf{X}_1) + \mathbf{E},$$

the elements of $g \in G$, a transformation group, are referred to as the similarity transformation parameters $\alpha, \mathbf{\Gamma}$, and β, and $(\mathbf{E})_i$ denotes i^{th} row of $\mathbf{E}$, $i = 1, \dots, k$.

For practical implementation, they considered the exhaustive selection, random selection such as PROGRESS and intelligent subset method.

3.2. M and LMS GPA

Unlike LS method, resistant methods limit the influence of objects outliers and the resulting fit is closer to the true mean shape. Rohlf and Slice (1990) considered GPA fit using the RM technique and gave an algorithm for resistant shape fitting. They compared LS GPA to their methods with wing data from mosquitoes. Er (1998) suggested a robust GPA method via the EM algorithm and compared the method with LS GPA by visualizing the estimates of the mean shape. In this section, we suggest two resistant GPA fitting algorithms based on the M and LMS estimator that are called the M GPA and LMS GPA algorithms.

First we introduce a resistant M GPA algorithm based on the M OPA fit method. Let n matrices $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ be configurations of $k \geq 3$ landmarks in $m \geq 2$ dimensions. Then we can find out the Procrustes mean shape $\hat{\mu}$ by minimizing objective function of the form

$$O(\alpha_i, \mathbf{\Gamma}_i, \beta_i) = \sum_{i=1}^{n} \rho \left(\|\mathbf{R}_i (\mathbf{X}_i, \mu)\| \right), \tag{3.9}$$

subject to $C(\mu) = 1$, where $\rho(\cdot)$ is the loss function like equations (3.2) and (3.3), and $\mathbf{R}_i (\mathbf{X}_i, \mu)$ denotes the residual matrix after OPA fit of $\mathbf{X}_i$ onto the mean shape, that is, $\mathbf{R}_i (\mathbf{X}_i, \mu) = \mu - \beta_i \mathbf{X}_i \mathbf{\Gamma}_i - \mathbf{1}_k \alpha_i^T$, $i = 1, \dots, n$.

The estimate can be found by differentiation the objective function with respect to μ and setting the derivative to zero,

$$\frac{\partial O(\alpha_i, \mathbf{\Gamma}_i, \beta_i)}{\partial \mu} = 0,$$

then the solution can be written as

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^{n} \left(w_i \hat{\beta}_i \mathbf{X}_i \hat{\mathbf{\Gamma}}_i + w_i \mathbf{1}_k \hat{\boldsymbol{\alpha}}_i^T \right)}{\mathbf{1}_n^T \mathbf{W} \mathbf{1}_n}, \tag{3.10}$$

where $\mathbf{W} = \text{diag}(w_1, \ldots, w_n)$ is weight diagonal matrix with $w_i = \psi(\|\mathbf{R}_i\|)/\|\mathbf{R}_i\|$, $i = 1, \ldots, n$, and $\psi(\cdot) = \rho(\cdot)'$, i.e., the derivative of $\rho(\cdot)$.

For our practical implementation, we can adapt M OPA method to GPA procedures to estimate similarity parameters $\boldsymbol{\alpha}_i, \mathbf{\Gamma}_i$, and β_i on the method of iteratively reweighted least squares. Therefore, the parameters are estimated by fitting each object onto the mean shape via M OPA method. After that, the Procrustes mean shape can be obtained from equation (3.10). The algorithm is as follows.

Algorithm 1. (M GPA algorithm) Consider a set of $k \times m$ configuration matrices $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$. For our convenience, we assume that n configurations are centered using Helmert sub-matrix $\mathbf{H}$ and scaled so that $\|\mathbf{X}_i\| = 1$.

Step 1: Fit each $\mathbf{X}_i$ onto $\hat{\boldsymbol{\mu}}_{(i)}$ using M OPA method, where $\hat{\boldsymbol{\mu}}_{(i)}$ is initial Procrustes mean shape except i^{th} object $\mathbf{X}_i$, $i = 1, \ldots, n$.

Step 2: Calculate the residual $\mathbf{R}_i(\mathbf{X}_i, \hat{\boldsymbol{\mu}}_{(i)})$, and the weights $w_i = \psi(\|\mathbf{R}_i\|)/\|\mathbf{R}_i\|$ are obtained, $i = 1, \ldots, n$.

Step 3: Find the estimate of mean shape from equation (3.10).

Step 4: Replace $\hat{\boldsymbol{\mu}}_{(i)}$ with $\hat{\boldsymbol{\mu}}$, repeat steps 1 to 3, until the criterion $|O^{(s+1)} - O^{(s)}| < \epsilon$ is satisfied, where $O^{(s)}$ is the value of objective function $O(\boldsymbol{\alpha}_i, \mathbf{\Gamma}_i, \beta_i)$ on s^{th} iteration.

To achieve resistance, we find the n similarity parameters $(\boldsymbol{\alpha}_i, \mathbf{\Gamma}_i, \beta_i)$ separately using M OPA method, and update the Procrustes mean shape in steps 2 and 3. We do not show that the objective function reduces; however, the algorithm finds the solutions quickly in practice.

Next we develop resistant GPA algorithm based on the LMS estimator. Consider the situation where we have n configuration of k landmarks in m dimensions $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$. Then we can fit the each configuration $\mathbf{X}_i$ onto mean shape. That is,

$$\boldsymbol{\mu} = g_i(\mathbf{X}_i) + \mathbf{E}_i,$$

where the elements of g_i are the similarity transformation parameters $\boldsymbol{\alpha}_i$, $\mathbf{\Gamma}_i$, and β_i, and $\mathbf{E}_i$ is an error matrix, $i = 1, \ldots, n$. Then the LMS objective function can be defined by

$$\text{median} \left(\|\mathbf{E}_1\|^2, \|\mathbf{E}_2\|^2, \ldots, \|\mathbf{E}_n\|^2 \right). \tag{3.11}$$

We can find the estimate of mean shape $\hat{\boldsymbol{\mu}}$ by minimizing the objective function.

As the LMS OPA, this algorithm proceeds by repeatedly drawing subsamples of p different objects. For our implementation we use subsamples chosen by random selection method. The algorithm implemented in PROGRESS (Program for Robust reGRESSion) can require large computation time in our GPA procedures. PROGRESS algorithm was proposed to calculate the parameters by resampling method (Rousseeuw and Leroy, 1987). Therefore, we use a further selection method with the criterion based on the centroid size of randomly chosen subsample. Additionally, we exclude the subsample if the values of centroid size of the subsample are too small or large. The algorithm can be described as follows.

Table 1: Norms of the differences of between true mean and estimated mean shapes

ϵ	High			Low		
	LS	M	LMS	LS	M	LMS
0.1	0.0292	0.0164	0.0274	0.0728	0.0432	0.0643
0.2	0.0365	0.0283	0.0314	0.0785	0.0595	0.0530
0.3	0.0475	0.0347	0.0329	0.0905	0.0678	0.0757
0.4	0.0598	0.0376	0.0582	0.1346	0.0951	0.0522
0.5	0.0505	0.0367	0.0416	0.1350	0.0804	0.0885

LS = least squares; M = M-estimator; LMS = least median of squares.

Algorithm 2. (LMS GPA Algorithm)   Consider a set of $k \times m$ configuration matrices $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$. We assume that n configurations are centered and scaled. Let each subsample be $\mathbf{Y}_j$ for $j = 1, \ldots, n_s$ where $n_s = \binom{n}{p}$ is the total number of subsamples.

Step 1: Draw a subsample $\mathbf{Y}_j$ of p different objects $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p$ from n configurations without replacement.

Step 2: Check chosen subsample $\mathbf{Y}_j$ with size criterion, if satisfied, estimate the parameters g_{ij} from fitting $\mathbf{Y}_j$ onto $\boldsymbol{\mu}$ by LS GPA, where $g_{ij}, i = 1, \ldots, n$ is the similarity parameters for Procrustes fit of each $\mathbf{X}_i$ onto $\boldsymbol{\mu}$ in drawing j^{th} subsample, otherwise, go step 1.

Step 3: Determine the corresponding LMS objective function in equation (3.11) with respect to the whole n objects.

Step 4: Repeat steps 1 to 3, and choose least one of among the objective function's values. Then the estimate of mean shape is defined as

$$\hat{\boldsymbol{\mu}} = \arg\min \left[\text{median} \left(\|\mathbf{E}_1\|^2, \ \|\mathbf{E}_2\|^2, \ldots, \|\mathbf{E}_n\|^2 \right) \right].$$

This algorithm requires more $n-p$ residuals in step 3. So we calculate the residuals using similarity parameters corresponding to minimum of p residuals in step 2. In practice, our random selection method is implemented with smaller subsamples than n_s. The algorithm does not find the Procrustes mean shape faster than M GPA; however, its implementation works well.

4. Examples

We investigate the sensitivity of LS and the resistance of M and LMS GPA methods, and compare their performance by visualizing differences between estimated mean shapes.

A random sample of 20 diamonds are simulated. Each landmark is simulated from mixture models of bivariate normal distributions with different covariances with data generated from $(1-\epsilon)BN(\boldsymbol{\mu}, \boldsymbol{\Sigma}_1) + \epsilon BN(\boldsymbol{\mu}, \boldsymbol{\Sigma}_2)$, where $\boldsymbol{\mu}$ is true mean shape and $\boldsymbol{\Sigma}_1 = 0.2\mathbf{I}_2, \boldsymbol{\Sigma}_2 = 1.0\mathbf{I}_2$ in High of Table 1 and $\boldsymbol{\Sigma}_1 = 0.6\mathbf{I}_2, \boldsymbol{\Sigma}_2 = 2.0\mathbf{I}_2$ in Low of Table 1. The simulation results of Table 1 show that two resistant GPA methods have smaller norm values of the residual than LS. As the graphs (b), (c), and (d) of Figure 1 visualize, M and LMS GPA methods also yield more close estimates to the true mean shape than LS result.

Figure 2(a) shows the centered and scaled landmarks for the T2 mouse vertebrae data (with high concentration) with four artificial object outliers. Graph (b) indicates that the LS GPA method produces a Procrustes mean shapes affected by to object outliers. Figure 2(c) and (d) shows that the
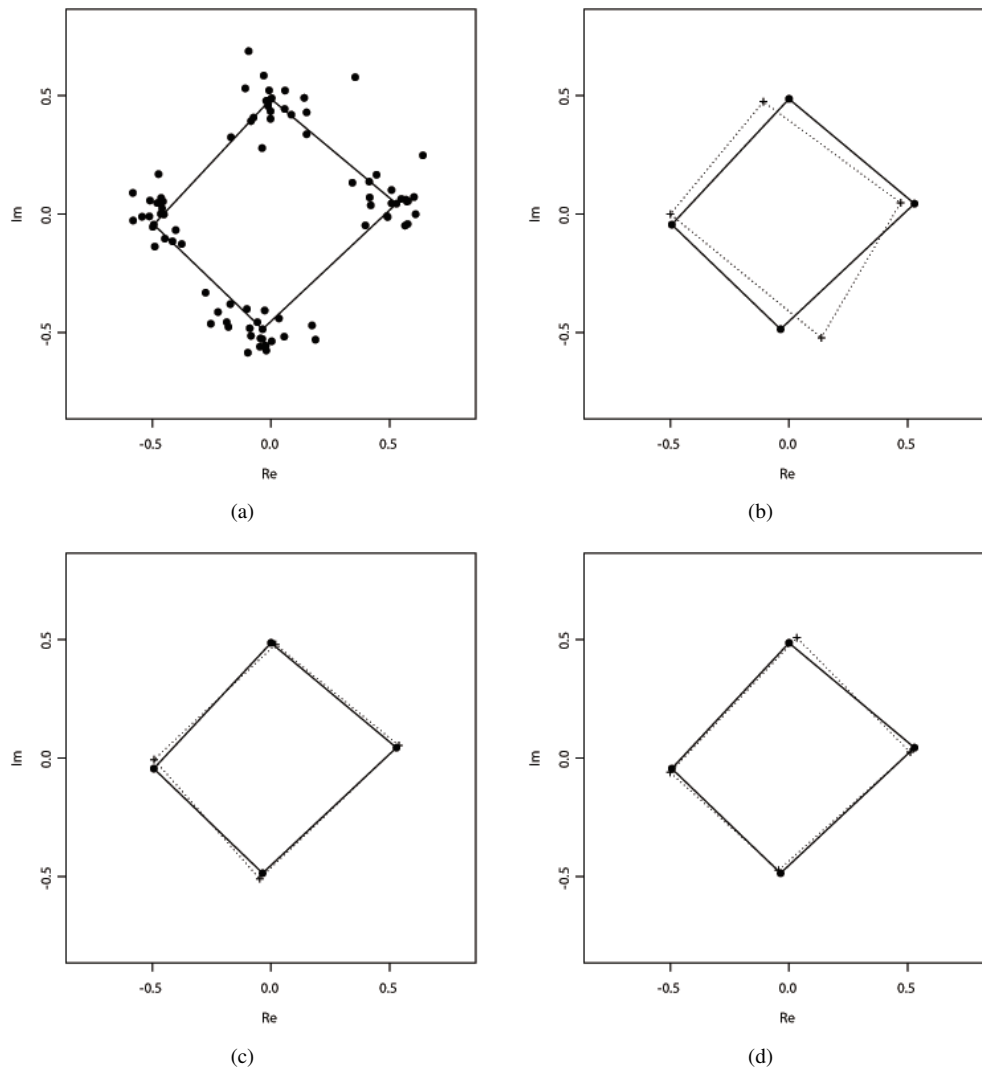
Figure 1: *Procrustes mean shapes displayed with $\epsilon = 0.2$ in low data set of Table 1; (a) the centered and scaled landmarks and true mean shape, (b) LS, (c) M, and (d) LMS, where the solid line is true mean shape and the dotted line is the estimates of mean shape. LS = least squares; M = M-estimator; LMS = least median of squares.*

estimates by M and LMS GPA methods are closer to the true mean shapes; therefore, the resistant algorithms work well for the estimated mean shape of Figure 2.

Next, consider the 30 handwritten digit 3 data with low concentration and eight artificial object outliers. Graph (b) of Figure 3 shows that the LS GPA method is affected by the object outliers. Graph (c) and (d) that M and LMS GPA algorithms give better Procrustes mean shapes than LS GPA despite the configuration data set having a low concentration and some object outliers. Table 2 shows how the weights by M GPA fit control these object outliers effectively when the objects which have a relatively larger residual than the others with smaller weights ($w_i = 0.131$).
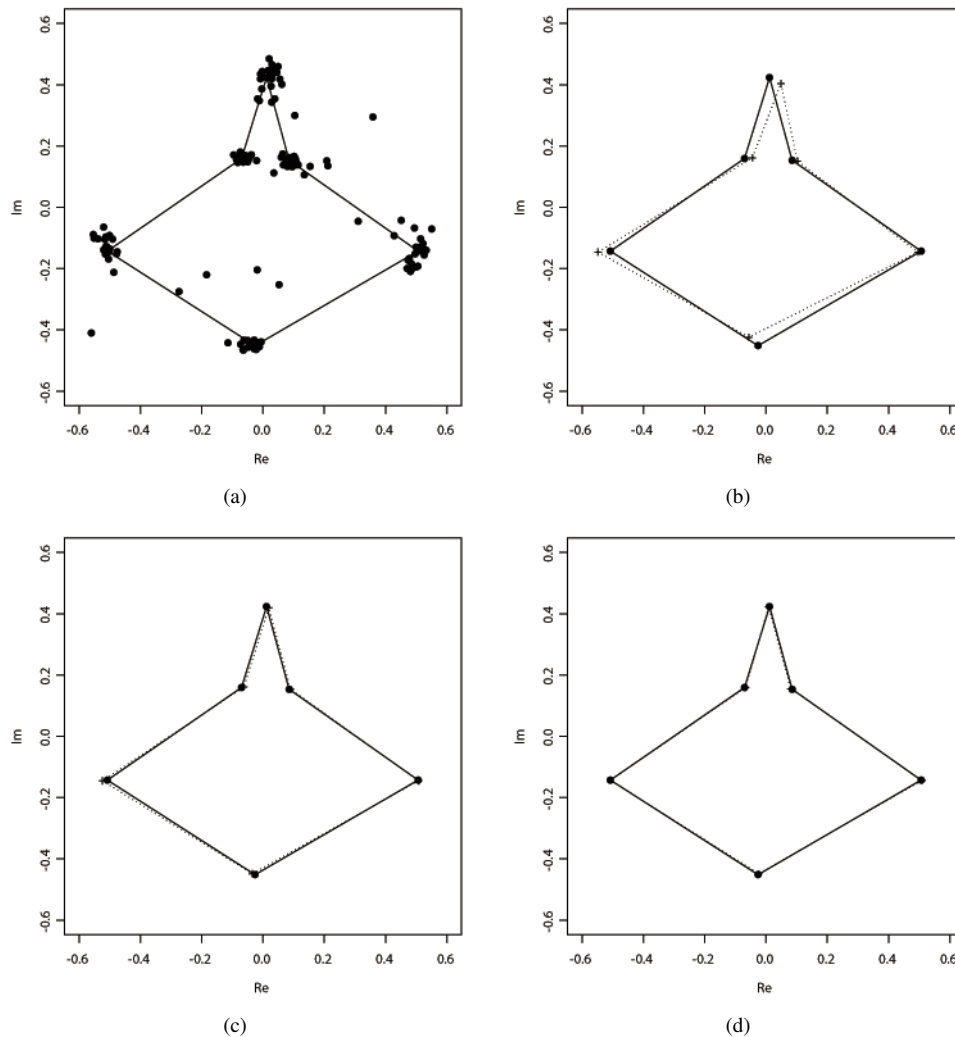
Figure 2: *Procrustes mean shapes for 23 T2 mouse vertebrae data; (a) landmarks for data with 4 object outliers, (b) LS, (c) M, and (d) LMS, where the solid line is Procrustes mean shapes for the data without object outliers and dotted line is that for the data with object outliers. LS = least squares; M = M-estimator; LMS = least median of squares.*

5. Concluding remarks

We have proposed two resistant GPA fitting methods to estimate mean shape avoiding influence of object outliers. M GPA algorithm has found the Procrustes mean shape quickly; however, the LMS GPA algorithm takes more time to compute than M GPA. In LMS GPA, the size criterion has reduced its iteration number effectively; however, the algorithm was somewhat expensive to compute as random selection method.

We need methods that can visualize the objects and mean shape effectively since the OPA and GPA methods are available for higher dimensions. Multivariate techniques such as principle components
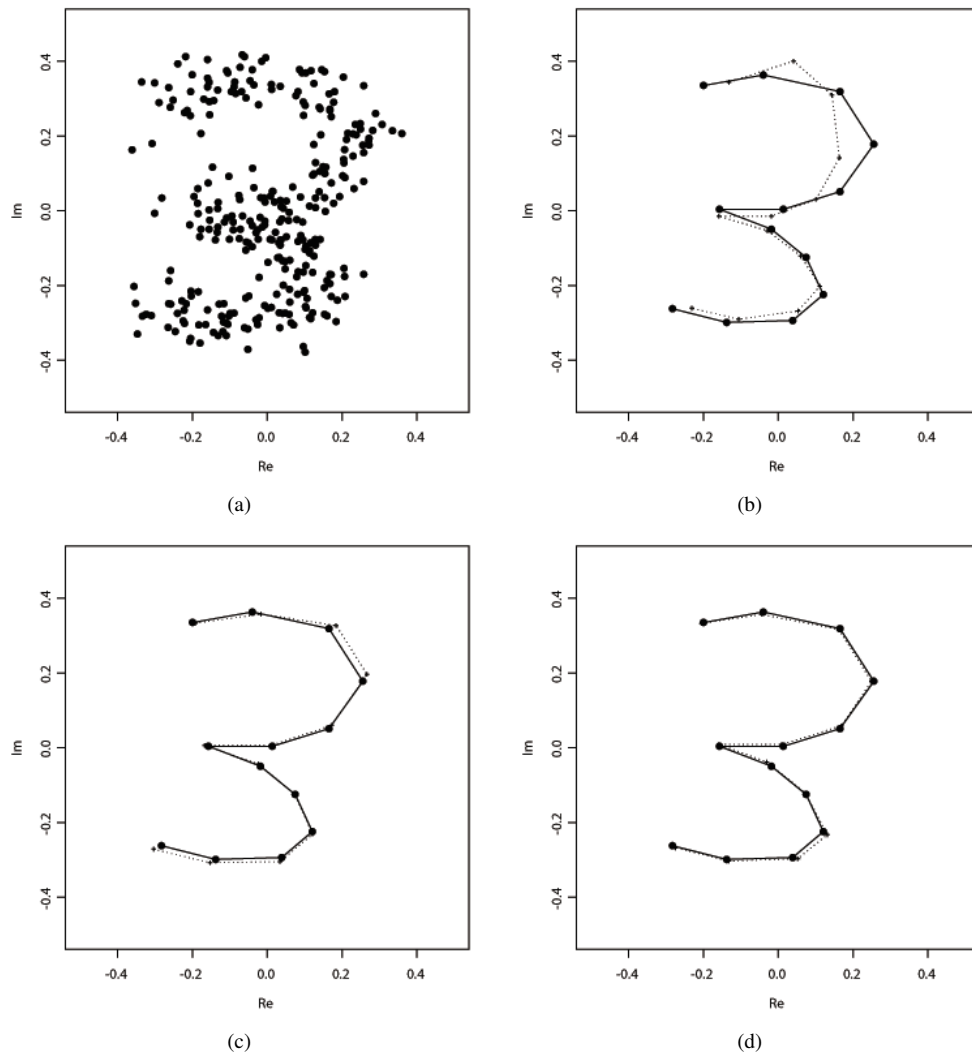
Figure 3: *Procrustes mean shapes for 30 handwritten digit 3 data; (a) landmarks for data with 8 object outliers, (b) LS, (c) M, and (d) LMS, where the solid line is Procrustes mean shapes for the data without object outliers and a dotted line is for the data with object outliers. LS = least squares; M = M-estimator; LMS = least median of squares.*

analysis can be an appropriate method. Resistant fitting methods have been studied without considering shape distributions; however, it is important that the shape analysis should not be sensitive to a particular model such as the complex Bingham distribution or the complex angular central Gaussian distribution. These robustness problem is an issue for further research. An improved method reducing computational time is needed since the LMS GPA algorithm was somewhat expensive to compute as random selection method. We often wish to examine the structure of shape variability after obtaining a mean shape; therefore, a suitable method is needed to investigate the shape variability in a tangent space.

Table 2: Norms of residuals and weights after LS, M, and LMS GPA fits

LS		M				LMS	
$\|\mathbf{R}_i\|$		$\|\mathbf{R}_i\|$		w_i		$\|\mathbf{R}_i\|$	
0.507	0.275	1.283	0.327	0.131	0.757	0.900	0.279
0.325	0.258	0.594	0.279	0.484	0.726	0.600	0.269
0.240	0.507	0.453	1.283	0.562	0.131	0.202	0.844
0.278	0.406	0.359	0.545	0.805	0.694	0.233	0.297
0.430	0.298	0.415	0.325	0.646	0.766	0.330	0.286
0.507	0.203	1.283	0.426	0.131	0.576	0.941	0.174
0.327	0.281	0.267	0.322	0.799	0.806	0.279	0.427
0.507	0.302	1.283	0.228	0.131	0.782	1.011	0.182
0.264	0.350	0.331	0.473	0.669	0.598	0.374	0.263
0.507	0.206	1.283	0.246	0.131	0.792	1.011	0.339
0.241	0.305	0.393	0.389	0.772	0.636	0.289	0.185
0.507	0.507	1.283	1.283	0.131	0.131	1.011	1.011
0.393	0.255	0.430	0.287	0.619	0.776	0.242	0.231
0.264	0.507	0.539	1.283	0.485	0.131	0.168	1.011
0.287	0.284	0.264	0.275	0.733	0.784	0.365	0.203

LS = least squares; M = M-estimator; LMS = least median of squares; GPA = generalized Procrustes analysis.

Appendix: The proof of Theorem 1

Proof: The solutions can be found by differentiating objective function of equation (3.1) in Section 3 with respect to $\alpha, \mathbf{\Gamma}, \beta$ and setting these derivatives to zero. First, for location parameter α,

$$\frac{\partial O(\alpha, \mathbf{\Gamma}, \beta)}{\partial \alpha} = 0,$$

it can be expressed as follows,

$$\sum_{i=1}^{k} \left(\psi(\|\mathbf{r}_i\|) \frac{1}{2\|\mathbf{r}_i\|} \frac{\partial}{\partial \alpha} \operatorname{tr} \left(\mathbf{r}_i^T \mathbf{r}_i \right) \right) = 0, \tag{A.1}$$

where

$$\frac{\partial}{\partial \alpha} \operatorname{tr} \left(\mathbf{r}_i^T \mathbf{r}_i \right) = -2\mathbf{x}_{(2)i} + 2\beta \mathbf{\Gamma}^T \mathbf{x}_{(1)i} + 2\alpha.$$

Then equation (A.1) can be written as

$$\hat{\alpha} = \frac{\sum_{i=1}^{k} \left(w_i \mathbf{x}_{(2)i} - \beta w_i \mathbf{\Gamma}^T \mathbf{x}_{(1)i} \right)}{\sum_{i=1}^{k} w_i},$$

also, describing in matrix notation, we can obtain

$$\hat{\alpha} = \frac{\left(\mathbf{X}_2^T - \beta \mathbf{\Gamma}^T \mathbf{X}_1^T \right) \mathbf{W} \mathbf{1}_k}{\mathbf{1}_k^T \mathbf{W} \mathbf{1}_k}.$$

Next, for rotation matrix $\mathbf{\Gamma}$, the solution can be found by minimizing the Lagrangian function

$$L = \sum_{i=1}^{k} \rho(\|\mathbf{r}_i\|) + \operatorname{tr} \left[\frac{1}{2} \mathbf{\Lambda} \left(\mathbf{\Gamma} \mathbf{\Gamma}^T - \mathbf{I}_m \right) \right],$$

where $(1/2)\mathbf{\Lambda}$ is a symmetric $m \times m$ matrix of Lagrange multipliers. Differentiating the Lagrangian function with respect to $\mathbf{\Gamma}$, we have

$$\frac{\partial L}{\partial \mathbf{\Gamma}} = \sum_{i=1}^{k} \left(\psi(\|\mathbf{r}_i\|) \frac{1}{2\|\mathbf{r}_i\|} \frac{\partial}{\partial \mathbf{\Gamma}} \mathrm{tr}\left(\mathbf{r}_i^T \mathbf{r}_i\right) \right) + \mathbf{\Lambda}\mathbf{\Gamma} = 0. \tag{A.2}$$

By substituting

$$\frac{\partial}{\partial \mathbf{\Gamma}} \mathrm{tr}\left(\mathbf{r}_i^T \mathbf{r}_i\right) = -2\beta \mathbf{x}_{(1)i} \mathbf{x}_{(2)i}^T + 2\beta \mathbf{x}_{(1)i} \boldsymbol{\alpha}^T$$

into equation (A.2) we obtain

$$\sum_{i=1}^{k} \left[w_i \beta \left(-\mathbf{x}_{(1)i} \mathbf{x}_{(2)i}^T + \mathbf{x}_{(1)i} \boldsymbol{\alpha}^T \right) \right] + \mathbf{\Lambda}\mathbf{\Gamma} = 0,$$

so in matrix notation,

$$\mathbf{X}_1^T \mathbf{W} \mathbf{X}_2 - \mathbf{X}_1^T \mathbf{W} \mathbf{1}_k \boldsymbol{\alpha}^T = \frac{\mathbf{\Lambda}\mathbf{\Gamma}}{\beta}. \tag{A.3}$$

Substituting the estimate of $\boldsymbol{\alpha}$, the left hand side of equation (A.3) becomes

$$\mathbf{X}_1^T \mathbf{W} \mathbf{X}_2 - \mathbf{X}_1^T \mathbf{W} \frac{\mathbf{1}_k \mathbf{1}_k^T \mathbf{W} (\mathbf{X}_2 - \beta \mathbf{X}_1 \mathbf{\Gamma})}{\mathbf{1}_k^T \mathbf{W} \mathbf{1}_k} = \mathbf{X}_1^T \mathbf{W} \mathbf{X}_2 - \frac{\mathbf{X}_1^T \mathbf{W} \mathbf{1}_k \mathbf{1}_k^T \mathbf{W} \mathbf{X}_2}{\mathbf{1}_k^T \mathbf{W} \mathbf{1}_k} + \frac{\beta \mathbf{X}_1^T \mathbf{W} \mathbf{1}_k \mathbf{1}_k^T \mathbf{W} \mathbf{X}_1 \mathbf{\Gamma}}{\mathbf{1}_k^T \mathbf{W} \mathbf{1}_k}.$$

Thus equation (A.3) can be written as

$$\mathbf{X}_1^T \mathbf{W} \left(\mathbf{I}_k - \frac{\mathbf{1}_k \mathbf{1}_k^T \mathbf{W}}{\mathbf{1}_k^T \mathbf{W} \mathbf{1}_k} \right) \mathbf{X}_2 = \frac{\mathbf{\Lambda}\mathbf{\Gamma}}{\beta} - \frac{\beta \mathbf{X}_1^T \mathbf{W} \mathbf{1}_k \mathbf{1}_k^T \mathbf{W} \mathbf{X}_1 \mathbf{\Gamma}}{\mathbf{1}_k^T \mathbf{W} \mathbf{1}_k}$$

$$= \left(\frac{\mathbf{\Lambda}}{\beta} - \frac{\beta \mathbf{X}_1^T \mathbf{W} \mathbf{1}_k \mathbf{1}_k^T \mathbf{W} \mathbf{X}_1}{\mathbf{1}_k^T \mathbf{W} \mathbf{1}_k} \right) \mathbf{\Gamma}. \tag{A.4}$$

In the right hand side of equation (A.4), put

$$\frac{\mathbf{\Lambda}}{\beta} - \frac{\beta \mathbf{X}_1^T \mathbf{W} \mathbf{1}_k \mathbf{1}_k^T \mathbf{W} \mathbf{X}_1}{\mathbf{1}_k^T \mathbf{W} \mathbf{1}_k} = Q,$$

then,

$$Q\mathbf{\Gamma} (Q\mathbf{\Gamma})^T = \left(\mathbf{U} \mathbf{D}_\lambda \mathbf{V}^T \right) \left(\mathbf{U} \mathbf{D}_\lambda \mathbf{V}^T \right)^T$$

$$= \mathbf{U} \mathbf{D}_\lambda \left(\mathbf{U}^T \mathbf{U} \right) \mathbf{D}_\lambda \mathbf{U}^T$$

$$= \left(\mathbf{U} \mathbf{D}_\lambda \mathbf{U}^T \right) \left(\mathbf{U} \mathbf{D}_\lambda \mathbf{U}^T \right)^T, \tag{A.5}$$

where $\mathbf{U}, \mathbf{V} \in SO(m)$ satisfy the singular value decomposition

$$\mathbf{X}_1^T \mathbf{W} \left(\mathbf{I}_k - \frac{\mathbf{1}_k \mathbf{1}_k^T \mathbf{W}}{\mathbf{1}_k^T \mathbf{W} \mathbf{1}_k} \right) \mathbf{X}_2 = \mathbf{U} \mathbf{D}_\lambda \mathbf{V}^T$$

with diagonal matrix $\mathbf{D}_\lambda$. So, by equation (A.5)

$$\mathbf{Q} = \mathbf{U}\mathbf{D}_\lambda\mathbf{U}^T. \tag{A.6}$$

Hence, by equation (A.4) and (A.6) the estimate of rotation matrix $\hat{\mathbf{\Gamma}}$ follows

$$\mathbf{X}_1^T\mathbf{W}\left(\mathbf{I}_k - \frac{\mathbf{1}_k\mathbf{1}_k^T\mathbf{W}}{\mathbf{1}_k^T\mathbf{W}\mathbf{1}_k}\right)\mathbf{X}_2 = \mathbf{Q}\hat{\mathbf{\Gamma}},$$

$$\mathbf{U}\mathbf{D}_\lambda\mathbf{V}^T = \mathbf{U}\mathbf{D}_\lambda\mathbf{U}^T\hat{\mathbf{\Gamma}},$$

$$\hat{\mathbf{\Gamma}} = \mathbf{U}\mathbf{V}^T.$$

Finally, for scale parameter β, differentiating the objective function with respect to β we obtain

$$\frac{\partial}{\partial\beta}\sum_{i=1}^k \rho\left(\|\mathbf{r}_i\|\right) = \sum_{i=1}^k \left(\psi(\|\mathbf{r}_i\|)\frac{1}{2\|\mathbf{r}_i\|}\frac{\partial}{\partial\beta}\|\mathbf{r}_i\|^2\right) = 0. \tag{A.7}$$

By substituting

$$\frac{\partial}{\partial\beta}\|\mathbf{r}_i\|^2 = -2\left[\text{tr}\left(\mathbf{x}_{(2)i}\mathbf{x}_{(1)i}^T\mathbf{\Gamma}\right)\right] + 2\beta\left[\text{tr}\left(\mathbf{x}_{(1)i}\mathbf{x}_{(1)i}^T\right)\right] + 2\left[\text{tr}\left(\mathbf{x}_{(1)i}\boldsymbol{\alpha}^T\mathbf{\Gamma}^T\right)\right]$$

into equation (A.7), the equation can be expressed as

$$\beta\left[\text{tr}\left(\sum_{i=1}^k w_i\mathbf{x}_{(1)i}\mathbf{x}_{(1)i}^T\right)\right] = \text{tr}\left(\sum_{i=1}^k w_i\mathbf{x}_{(2)i}\mathbf{x}_{(1)i}^T\mathbf{\Gamma}\right) - \text{tr}\left(\sum_{i=1}^k w_i\mathbf{x}_{(1)i}\boldsymbol{\alpha}^T\mathbf{\Gamma}^T\right),$$

also in matrix notation,

$$\beta\left[\text{tr}\left(\mathbf{X}_1^T\mathbf{W}\mathbf{X}_1\right)\right] = \text{tr}\left(\mathbf{X}_2^T\mathbf{W}\mathbf{X}_1\mathbf{\Gamma} - \mathbf{X}_1^T\mathbf{W}\mathbf{1}_k\boldsymbol{\alpha}^T\mathbf{\Gamma}^T\right). \tag{A.8}$$

Substituting each of the estimates of parameters $\boldsymbol{\alpha}$ and $\mathbf{\Gamma}$ into equation (A.8), then it becomes

$$\beta\left[\text{tr}\left(\mathbf{X}_1^T\mathbf{W}\mathbf{X}_1\right)\right] = \text{tr}\left[\mathbf{X}_2^T\mathbf{W}\mathbf{X}_1\mathbf{\Gamma} - \mathbf{X}_1^T\mathbf{W}\left(\frac{\mathbf{1}_k\mathbf{1}_k^T\mathbf{W}\left(\mathbf{X}_2 - \beta\mathbf{X}_1\mathbf{\Gamma}\right)}{\mathbf{1}_k^T\mathbf{W}\mathbf{1}_k}\mathbf{\Gamma}^T\right)\right]$$

$$= \text{tr}\left[\mathbf{X}_2^T\mathbf{W}\mathbf{X}_1^T\mathbf{\Gamma} - \frac{\mathbf{X}_1^T\mathbf{W}\mathbf{1}_k\mathbf{1}_k^T\mathbf{W}\mathbf{X}_2\mathbf{\Gamma}^T}{\mathbf{1}_k^T\mathbf{W}\mathbf{1}_k} + \frac{\beta\mathbf{X}_1^T\mathbf{W}\mathbf{1}_k\mathbf{1}_k^T\mathbf{W}\mathbf{X}_1\mathbf{\Gamma}\mathbf{\Gamma}^T}{\mathbf{1}_k^T\mathbf{W}\mathbf{1}_k^T}\right].$$

Therefore,

$$\beta\left[\text{tr}\left(\mathbf{X}_1^T\mathbf{W}\mathbf{X}_1\right)\right] - \beta\left[\text{tr}\left(\frac{\mathbf{X}_1^T\mathbf{W}\mathbf{1}_k\mathbf{1}_k^T\mathbf{W}\mathbf{X}_1}{\mathbf{1}_k^T\mathbf{W}\mathbf{1}_k}\right)\right] = \text{tr}\left(\mathbf{X}_2^T\mathbf{W}\mathbf{X}_1\mathbf{\Gamma} - \frac{\mathbf{X}_1^T\mathbf{W}\mathbf{1}_k\mathbf{1}_k^T\mathbf{W}\mathbf{X}_2\mathbf{\Gamma}^T}{\mathbf{1}_k^T\mathbf{W}\mathbf{1}_k}\right).$$

Hence, we have

$$\hat{\beta} = \frac{\text{tr}\left[\mathbf{\Gamma}^T\mathbf{X}_1^T\mathbf{W}\left(\mathbf{I}_k - \frac{\mathbf{1}_k\mathbf{1}_k^T\mathbf{W}}{\mathbf{1}_k^T\mathbf{W}\mathbf{1}_k}\right)\mathbf{X}_2\right]}{\text{tr}\left[\mathbf{X}_1^T\mathbf{W}\left(\mathbf{I}_k - \frac{\mathbf{1}_k\mathbf{1}_k^T\mathbf{W}}{\mathbf{1}_k^T\mathbf{W}\mathbf{1}_k}\right)\mathbf{X}_1\right]}.$$

$\square$

References

Dryden IL and Mardia KV (2016). *Statistical Shape Analysis with Applications in R* (2nd ed), John Wiley & Sons, New York.

Dryden IL and Walker G (1999). Highly resistant regression and object matching, *Biometrics*, **55**, 820–825.

Er F (1998). *Robust methods in statistical shape analysis* (Ph.D. thesis), University of Leeds, Leeds, UK.

Goodall C (1991). Procrustes methods in the statistical analysis of shape (with discussion), *Journal of Royal Statistical Society, Series B (Methodological)*, **53**, 285–339.

Gower JC (1975). Generalized Procrustes analysis, *Psychometrika*, **40**, 33–51.

Kendall DG (1984). Shape manifolds, Procrustean metrics, and complex projective spaces, *Bulletin of the London Mathematical Society*, **16**, 81–121.

Kent JT (1994). The complex Bingham distribution and shape analysis, *Journal of Royal Statistical Society, Series B (Methodological)* , **56**, 285–299.

Rohlf FJ and Slice D (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks, *Systematic Zoology*, **39**, 40–59.

Rousseeuw PJ (1984). Least median of squares regression, *Journal of the American Statistical Association*, **79**, 871–880.

Rousseeuw PJ and Leroy AM (1987). *Robust Regression and Outlier Detection*, Wiley, New York.

Siegel AF and Benson RH (1982). A robust comparison of biological shapes, *Biometrics*, **38**, 341–350.

Ten Berge JMF (1977). Orthogonal Procrustes rotation for two or more matrices, *Psychometrika*, **42**, 267–276.

Verboon P and Heiser WJ (1992). Resistant orthogonal Procrustes analysis, *Journal of Classification*, **9**, 237–256.