

Covariance weighted Procrustes analysis

Christopher J. Brignell*, William J. Browne and Ian L. Dryden

University of Nottingham

1 Introduction

Standard Procrustes analysis involves matching shape configurations, through translations, rotations, and possibly scaling, to minimize the Euclidean distance between them (eg. see Dryden and Mardia, 1998). This technique effectively assumes that the coordinates have an isotropic covariance structure. However, there are many cases where one may wish to weight the matching, perhaps to reflect the reliability of the landmarks or to register shapes on only part of the structure. Covariance weighted (CW) Procrustes analysis matches shapes to minimize a Mahalanobis norm, with respect to a symmetric, positive definite matrix Σ .

2 CW Procrustes analysis with known covariance matrix

Suppose we have a shape configuration matrix, X ($k \times m$), containing the coordinates of k landmarks in m dimensions, and a reference shape configuration, μ , then standard ordinary Procrustes analysis (OPA) minimizes the Euclidean distance, $D_{OPA}^2(X, \mu) = \|\mu - \beta X \Gamma - 1_k \gamma^T\|^2$, where $\|Z\|^2 = \text{trace}(Z^T Z)$ is the squared Euclidean norm, Γ is an $m \times m$ special orthogonal rotation matrix, $\beta > 0$ is a scale parameter, 1_k is a $k \times 1$ vector of ones and γ is a $m \times 1$ translation vector. However, partial CW Procrustes analysis seeks to produce non-iterative estimates of Γ and γ , for a general covariance matrix, Σ ($km \times km$), to minimize the squared Mahalanobis norm, $D_p^2(X, \mu; \Sigma) = \|\mu - X \Gamma - 1_k \gamma^T\|_{\Sigma}^2$, following Goodall (1991).

It can be shown that the translation, as a function of rotation, which minimizes $D_p^2(X, \mu; \Sigma)$ is,

$$\hat{\gamma} = [(I_m \otimes 1_k)^T \Sigma^{-1} (I_m \otimes 1_k)]^{-1} (I_m \otimes 1_k)^T \Sigma^{-1} \text{vec}(\mu - X \Gamma).$$

Estimation of the necessary rotation to minimize $D_p^2(X, \mu; \Sigma)$ is complicated. If $m \geq 3$ the problem is non-linear and has to be solved numerically. Let $\theta = (\theta_1, \theta_2, \dots, \theta_r)^T$ be the vector of rotation angles to be estimated where $r = m(m-1)/2$. Then, for an estimate of θ , we can calculate $\Gamma \in SO(m)$, $\hat{\gamma}$ and evaluate $D_p^2(X, \mu; \Sigma)$. Using a suitable non-linear minimization routine, the estimate of θ can be improved to further minimize $D_p^2(X, \mu; \Sigma)$. However, if $m = 2$, an analytic solution is possible.

Let $A = [(I_m \otimes 1_k)^T \Sigma^{-1} (I_m \otimes 1_k)]^{-1} (I_m \otimes 1_k)^T \Sigma^{-1}$, and denote the partitioned submatrices as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad X = [X_1 \quad X_2] \quad \mu = [\mu_1 \quad \mu_2]$$

then it can be shown that,

$$\cos \theta = \frac{S(2\lambda - 2Q) + TR}{(2\lambda - 2P)(2\lambda - 2Q) - R^2} \quad \sin \theta = \frac{T(2\lambda - 2P) + SR}{(2\lambda - 2P)(2\lambda - 2Q) - R^2}$$

where,

$$\begin{aligned}
P &= \left[(X_1 + 1_k \beta_1)^T \quad (X_2 + 1_k \beta_2)^T \right] \Sigma^{-1} \left[(X_1 + 1_k \beta_1)^T \quad (X_2 + 1_k \beta_2)^T \right]^T \\
Q &= \left[(X_2 - 1_k \delta_1)^T \quad -(X_1 + 1_k \delta_2)^T \right] \Sigma^{-1} \left[(X_2 - 1_k \delta_1)^T \quad -(X_1 + 1_k \delta_2)^T \right]^T \\
R &= -2 \left[(X_1 + 1_k \beta_1) \quad (X_2 + 1_k \beta_2)^T \right] \Sigma^{-1} \left[(X_2 - 1_k \delta_1)^T \quad -(X_1 + 1_k \delta_2)^T \right]^T \\
S &= -2 \left[(X_1 + 1_k \beta_1)^T \quad (X_2 + 1_k \beta_2)^T \right] \Sigma^{-1} \left[(\mu_1 - 1_k \alpha_1)^T \quad (\mu_2 - 1_k \alpha_2)^T \right]^T \\
T &= 2 \left[(X_2 - 1_k \delta_1)^T \quad -(X_1 + 1_k \delta_2)^T \right] \Sigma^{-1} \left[(\mu_1 - 1_k \alpha_1)^T \quad (\mu_2 - 1_k \alpha_2)^T \right]^T \\
\alpha_i &= A_{i1} \mu_1 + A_{i2} \mu_2 \quad \beta_i = -A_{i1} X_1 - A_{i2} X_2 \quad \delta_i = A_{i1} X_2 - A_{i2} X_1,
\end{aligned}$$

and λ is the real root less than $\frac{1}{2} \left(P + Q - \sqrt{(P - Q)^2 + R^2} \right)$ of the quartic equation,

$$\begin{aligned}
&16\lambda^4 - 32(P + Q)\lambda^3 + [16(P^2 + Q^2) + 64PQ - 4(S^2 + T^2) - 8R^2]\lambda^2 \\
&+ [8R^2(P + Q) - 32PQ(P + Q) + 8(QS^2 + PT^2 - STR)]\lambda \\
&+ 16P^2Q^2 + R^4 - R^2(S^2 + T^2) + 4RST(P + Q) - 4P^2T^2 - 4Q^2S^2 - 8PQR^2 = 0.
\end{aligned}$$

This can be extended to the case of full ordinary CW Procrustes matching where scaling is included. The solutions are simplified if the covariance matrix is of the form $I_m \otimes \Sigma_k$, where I_m is the m dimensional identity matrix and Σ_k ($k \times k$) is a covariance matrix describing the variability between the landmarks, and are equivalent to the method given by Goodall (1991) of using the Cholesky decomposition of Σ_k^{-1} . Further, if Σ_k is taken to be the identity matrix as well, then CW Procrustes yields the same solutions as standard Procrustes analysis.

Example: Suppose we have a shape generated from a multivariate normal distribution with mean, μ , and covariance matrix, $\Sigma = \Sigma_m \otimes \Sigma_k$, where

$$\mu = \begin{bmatrix} 0 & 5 \\ 10 & 0 \\ 0 & -5 \\ -10 & 0 \end{bmatrix} \quad \Sigma_m = \begin{bmatrix} 0.001 & 0 \\ 0 & 1 \end{bmatrix} \quad \Sigma_k = \begin{bmatrix} 0.01 & 0 & 0 & 0 \\ 0 & 10 & 0 & -9.999 \\ 0 & 0 & 0.01 & 0 \\ 0 & -9.999 & 0 & 10 \end{bmatrix}$$

Figure 1 illustrates the difference of registering the shape to μ using standard Procrustes and CW Procrustes using Σ .

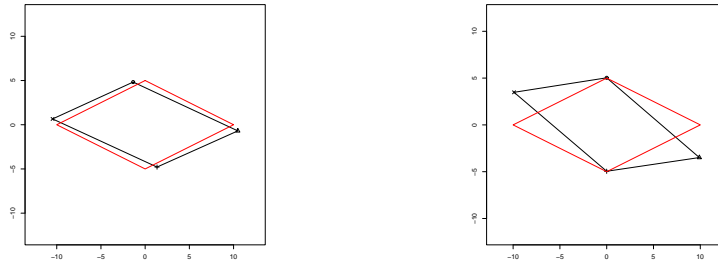


Figure 1: Matching a shape to a template using standard OPA (left) and CW OPA (right).

Consider now the general case where $n \geq 2$ configuration matrices are available, X_1, X_2, \dots, X_n . The method of partial generalized Procrustes analysis (GPA) involves translating and rotating the configurations relative to each other so as to minimize a total sum of squares, under the

model, $X_i = (\mu + E_i)\Gamma_i + 1_k\gamma_i^T$, 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 Γ_i and γ_i are nuisance parameters for rotation and translation. The method of partial CW GPA minimizes a quantity proportional to the sum of squared Mahalanobis norms of pairwise differences,

$$G_p(X_1, \dots, X_n; \Sigma) = \frac{1}{n} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \|(X_i\Gamma_i + 1_k\gamma_i^T) - (X_j\Gamma_j + 1_k\gamma_j^T)\|_{\Sigma}^2,$$

as opposed to the sum of squared Euclidean norms in standard GPA. Full covariance weighted GPA is similarly defined with scaling involved, subject to a constraint on the size of the mean shape. Minimizing $D_p^2(X_i, \bar{X}; \Sigma)$ for each X_i , where \bar{X} is the arithmetic mean of the shapes, is equivalent to minimizing $G_p(X_1, \dots, X_n; \Sigma)$. This suggests the following algorithm:

1. Standard GPA on X_1, \dots, X_n to remove large variations. Calculate \bar{X}_0 .
2. CW OPA to minimize $D_p^2(X_i, \bar{X}; \Sigma)$ for $i = 1, \dots, n$. Calculate \bar{X} .
3. Standard OPA to minimize $D_{OPA}^2(\bar{X}, \bar{X}_0)$. Apply the rotation and translation to each X_i .
4. Repeat steps 2 and 3 until $G_p(X_1, \dots, X_n; \Sigma)$ cannot be reduced.

Note that the Procrustes estimates, $\hat{\Gamma}_i$, $\hat{\gamma}_i$ and $\hat{\mu}$, given Σ are also maximum likelihood estimates under the model, $vec(X_i) \sim N_{km}(vec(\mu), \Sigma)$.

Example: Suppose we have 130 shapes generated using μ and Σ from the previous example. Figure 2 illustrates the difference following standard GPA and CW GPA using Σ .

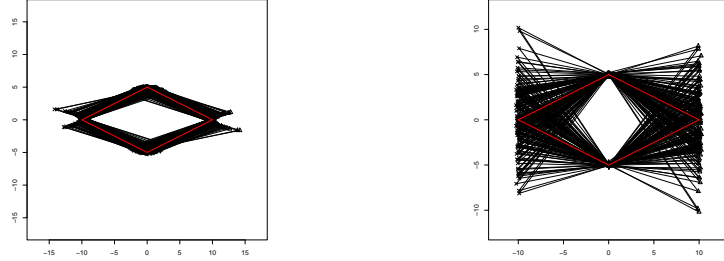


Figure 2: The registration of 130 shapes using standard GPA (left) and CW GPA (right).

3 Estimating the covariance matrix

Suppose the configurations have a multivariate normal distribution, $vec(X_i) \sim N_{km}(vec(\mu), \Sigma)$, then the profile MLE of Σ given $\hat{\Gamma}_i, \hat{\gamma}_i, i = 1, \dots, n$, is

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

where $X_i^P = X_i\hat{\Gamma}_i + 1_k\hat{\gamma}_i^T$ is the partial covariance weighted Procrustes fit of X_i , and $\hat{\Gamma}_i, \hat{\gamma}_i$ are the CW Procrustes estimates of the minimizing rotation and translation. Thus, $\hat{\Sigma}$ can be calculated using the output of standard GPA, or CW GPA, but the estimate depends on the Σ used in the GPA algorithm. Suppose, Σ_0 is an arbitrary estimate of the covariance matrix, then Goodall (1991) proposes the following algorithm to maximize the likelihood for the model:

1. CW GPA with $\Sigma = \Sigma_0$.

2. Estimate $\hat{\Sigma}$.
3. CW GPA with $\Sigma = \hat{\Sigma}$.
4. Iteratively cycle between (2) and (3) until convergence.

However, this algorithm is not guaranteed to converge, and use of different starting points may produce different estimates. Therefore, different starting points are necessary to establish the maximum of the likelihood. Further, $\hat{\Sigma}$ is singular, forcing the use of a generalized inverse and making it impossible to estimate the transformation parameters. This problem can be resolved by adding a small constant to the eigenvalues of the covariance matrix, which can be thought of as a prior specification of variability in the space of the final three eigenvectors. To illustrate this algorithm, $N = 1000$, Monte Carlo samples of the model in the example were generated. The estimates of μ_i and Σ_i for each sample were calculated following standard GPA and CW GPA with unknown Σ and the root mean square error and bias assessed.

	GPA	CWGPA
$\sqrt{(\frac{1}{N} \sum_{i=1}^N \ \hat{\mu}_i - \mu\ ^2)}$	0.697	0.174
$\sqrt{(\frac{1}{N} \sum_{i=1}^N \ \hat{\Sigma}_i - \Sigma\ ^2)}$	19.707	2.558
$E(\hat{\mu}) - \mu$	$\begin{bmatrix} 0.001 & -0.145 \\ 0.454 & 0.001 \\ -0.001 & 0.145 \\ -0.454 & -0.001 \end{bmatrix}$	$\begin{bmatrix} 0.001 & -0.001 \\ 0.004 & 0.001 \\ -0.001 & 0.001 \\ -0.004 & -0.001 \end{bmatrix}$

In trying to minimize the Euclidean norm, standard GPA has had to introduce large rotations to match the outer landmarks, at the expense of producing large variability at the inner landmarks. The output of CW GPA provides a good estimate of Σ , which Lele (1993) showed to be problematic in similar examples using other Procrustes based methods. The root mean square error and the bias are clearly reduced compared to standard Procrustes. Through this relatively simple example, we have been able to demonstrate a generalization of Procrustes techniques, matching using an unknown covariance matrix. We have demonstrated that non-linear methods are not required for $m = 2$. No assumptions are made regarding the initial position of the data. Problems still remain in estimating variability in the directions of rotations and translations, because Procrustes techniques filter this out. However, Bayesian analysis and Markov Chain Monte Carlo simulation can be used to estimate the mean shape and covariance matrix in conjunction with CW Procrustes estimates of the rotation and translation. This also removes the need for multiple start points and replaces the introduction of the small constant with the prior distribution of the covariance matrix.

References

- Dryden, I.L. and Mardia, K.V. (1998). *Statistical Shape Analysis*. Wiley, Chichester.
- Goodall, C. (1991). Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society, Series B*, **53**, 285-339.
- Lele, S. (1993). Euclidean distance matrix analysis (EDMA): Estimation of mean form and mean form difference. *Mathematical Geology*, **25**, 573-602.