Supplementary Material

for

Regression modelling for size-and-shape data based on a Gaussian model for landmarks

SM.A: Results for Reflection Size-and-Shape

Throughout the main part of the paper we have required that R is a proper rotation, i.e. $R \in SO(m)$. Here we consider the situation where shape is viewed as invariant under reflection, i.e. the reflected version of an object is considered to have the same shape as the original object. Specifically, we consider $R \in O(m)$. Since $O(m) = SO(m) \cup O_-(m)$, where $O_-(m) = \{R \in O(m) : |R| = -1\}$ and $SO(m)$ are isometric, it turns out that Lemma 1 holds as stated, but with $R \in O(m)$, and with $\mathcal{D}(\Delta)$ in (13) and (14) replaced by

$$\mathcal{D}_+(\Delta) = 2^{-m}|\Delta|^{k-m}\prod_{i<j}^{m}(\delta_i^2 - \delta_j^2), \tag{SM1}$$

i.e. the only difference is the factor 2^{-m+1} in (14), which changes to 2^{-m} in (SM1).

The version of Theorem 1 when $R \in O(m)$ is now stated.

Theorem SM1. *Suppose that $X \sim \mathcal{N}_{k\times m}(\mu, I_m \otimes \Sigma)$, where $k \geq m$ and Σ has full rank k. Consider the singular value decomposition $X = U\Delta R^\top$ given by (2). Then the density $f_{1+}(U, \Delta)$ with respect to the measure $(d\Delta)(dU)$, defined via Lemma 1 and equation (SM1), is given by*

$$f_{1+}(U, \Delta; \mu, \Sigma) = \frac{\mathcal{D}_+(\Delta)\mathcal{C}_+(A)}{(2\pi)^{km/2}|\Sigma|^{n/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left(\Delta U^\top \Sigma^{-1} U\Delta + \mu^\top \Sigma^{-1}\mu\right)\right\}, \tag{SM2}$$

where $\mathcal{D}_+(\Delta)$ *is defined in (SM1),* $A = \mu^\top \Sigma^{-1} U \Delta$ *is an* $m \times m$ *matrix and, for given* A*,* $\mathcal{C}_+(A)$ *is defined by*

$$\mathcal{C}_+(A) = \int_{R \in O(m)} \exp\left\{\mathrm{tr}\left(RA^\top\right)\right\}(dR). \tag{SM3}$$

Moreover, the conditional distribution of R *given* U *and* Δ *has density with respect to the unnormalized geometric, or Haar, measure* (dR) *on* $O(m)$ *given by*

$$f_{2+}(R|U,\Delta;\mu,\Sigma) = \frac{1}{\mathcal{C}_+(A)} \exp\left\{\mathrm{tr}\left(RA^\top\right)\right\}. \tag{SM4}$$

The statement of Theorem 2 is unchanged in the reflection size-and-shape case except that the expectation of R_i, conditional on U_i and Δ_i, is calculated using (SM4), the conditional density of R_i over $O(m)$, rather than (17), the conditional density of R_i over $SO(m)$. The relevant formulae for the conditional expectation of the R_i are given in (SM6) and (SM7) below.

The reflection size-and-shape analogue of Lemma A.1 in Appendix A is now stated.

Lemma SM1. *Suppose* $m = 2$ *and define the* 2×2 *matrix* $M = (m_{ij})_{i,j=1}^2$ as in (A2). *Then*

$$\begin{aligned}\bar{R} = E_{\mu,\Sigma}[R|U,\Delta] &= \int_{R \in O(2)} R f_{2+}(R|U,\Delta;\mu,\Sigma)(dR) \\ &= \omega_1 \mathcal{A}(\rho_1) \begin{pmatrix} \cos\alpha_1 & -\sin\alpha_1 \\ \sin\alpha_1 & \cos\alpha_1 \end{pmatrix} + \omega_2 \mathcal{A}(\rho_2) \begin{pmatrix} -\cos\alpha_2 & \sin\alpha_2 \\ \sin\alpha_2 & \cos\alpha_2 \end{pmatrix},\end{aligned}$$

where f_{2+} *is defined in (SM4),*

$$\cos\alpha_1 = (m_{11} + m_{22})/\rho_1, \quad \sin\alpha_1 = (m_{12} - m_{21})/\rho_1$$

$$\rho_1 = \sqrt{(m_{11} + m_{22})^2 + (m_{12} - m_{21})^2}, \quad \rho_2 = \sqrt{(m_{22} - m_{11})^2 + (m_{12} + m_{21})^2},$$

$$\cos\alpha_2 = (m_{22} - m_{11})/\rho_2, \quad \sin\alpha_2 = (m_{12} + m_{21})/\rho_2,$$

$$\omega_j = \mathcal{I}_0(\rho_j)/\{I_0(\rho_1) + I_0(\rho_2)\}, \quad j = 1, 2,$$

and $\mathcal{A}(\rho) = \mathcal{I}_1(\rho)/\mathcal{I}_0(\rho)$*, as previously.*

The analogue of of Proposition A.1 in Appendix A is as follows.

Proposition SM1. *Suppose that the* 3×3 *matrix* $\mu^\top \Sigma^{-1} U^\top \Delta$ in (17) has singular value decomposition

$$\mu^\top \Sigma^{-1} U^\top \Delta = T_1 \Phi T_2^\top, \tag{SM5}$$

where $\Phi = \mathrm{diag}\{\phi_1, \phi_2, \phi_3\}$. Define $\Xi = \mathrm{diag}\{\xi_1, \xi_2, \xi_3, \xi_4\}$. *Then*

$$\bar{R} = E_{\mu,\Sigma}[R|U,\Delta] = \int_{R \in O(3)} R f_2(R|U,\Delta;\Sigma,\mu)(dR) = T_1 \Omega_+ T_2^\top, \tag{SM6}$$

where f_{2+} *is defined in (A4),* $\Omega_+ = diag\{\omega_1, \omega_2, \omega_3\}$ *and, using the notation explained in Proposition A.1*

$$\omega_{j+} = 1 - \frac{\mathcal{C}_{6+}(\Xi_k) + \mathcal{C}_{6+}(\Xi_\ell)}{\pi \mathcal{C}_{4+}(\Xi)}, \quad j,k,\ell \in \{1,2,3\}, \quad j \neq k \neq \ell \neq j,$$

and

$$\mathcal{C}_{q+}(\Lambda) = \mathcal{C}_q(\Lambda) + \mathcal{C}_q(-\Lambda), \tag{SM7}$$

with $\mathcal{C}_q(\Lambda)$ *defined in (A4).*

For the calculation of the Bingham normalizing constants appearing in Proposition SM1, see the comments following Proposition A.1.

SM.B: Proofs

Proof of Lemma 1 and formula (SM1). The version of Lemma 1 with $R \in O(m)$ is given in Theorem 3.1 of Diaz-Garcia et al. (1997). The reason why the factor 2^{-m+1} is present in (14) and the factor 2^{-m} is present in (SM1) is now explained. Using (2), and assuming that the diagonal elements of

$$\Delta = \mathrm{diag}\{\delta_1, \dots, \delta_m\}$$

are distinct and positive, we may write

$$X = \sum_{j=1}^{m} \delta_j u_j r_j^\top = \sum_{j=1}^{m} \delta_j (\epsilon_j u_j)(\epsilon_j r_j^\top),$$

where u_j and r_j are, respectively, column j of U and R, and $\epsilon_j = \pm 1$, $j = 1, \ldots, m$. If we only require $R \in O(m)$ as opposed to $R \in SO(m)$ then, almost everywhere with respect to Lebesgue measure on $\mathbb{R}^{k\times m}$,

$$(U, \Delta, R) \mapsto X \tag{SM1}$$

is a $2^m : 1$ map, hence the factor 2^{-m} in (SM1). On the other hand, if we require that $R \in SO(m)$, then only those sign changes with an even number of the ϵ_j negative will leave the sign-transformed R in $SO(m)$, in which case the map (SM8) is $2^{m-1} : 1$. This is why the factor 2^{-m+1} is present in (14). □

Proof of Theorem 1 and Theorem SM1. Substituting $X = U\Delta R^\top$ from (2) into the exponent of (12) and rearranging,

$$-\frac{1}{2}\text{tr}\left\{(X-\mu)^\top \Sigma^{-1}(X-\mu)\right\} = \text{tr}\left(RA^\top\right) - \frac{1}{2}\text{tr}\left\{\Delta U^\top \Sigma^{-1} U \Delta + \mu^\top \Sigma^{-1}\mu\right\}, \tag{SM9}$$

where $A = \mu^\top \Sigma^{-1} U \Delta$ as before. Using Lemma 1 and (SM9), (15) follows after integrating over $R \in SO(m)$ and using (16). Also, (17) follows because, from (SM9), it is seen that the conditional density with respect to (dR) is proportional to $\exp\{\text{tr}(RA^\top)\}$, with normalizing constant given by (16). The details of the proof of Theorem SM1 are similar, the only differences being that the relevant integral is over $R \in O(m)$, $\mathcal{D}(\Delta)$ in (14) is replaced by $\mathcal{D}_+(\Delta)$ in (SM1), and $\mathcal{C}(.)$ in (15), (16) and (17) is replaced in (SM2) and (SM4) by $\mathcal{C}_+(.)$ defined in (SM3). □

Proof of Theorem 2. Let $\mathcal{Q}^{(r)}(B, \Sigma)$ denote the expectation of the full log-likelihood (19) over the missing data $R_1, \ldots, R_n$ conditional on the observed data $\Delta_1, \ldots, \Delta_n$ and $U_1, \ldots, U_n$, obtained using (3), and with $B = B^{(r)}$ and $\Sigma = \Sigma^{(r)}$. For a given r, we choose

$B = B^{(r+1)}$ and $\Sigma = \Sigma^{(r+1)}$ to maximize $\mathcal{Q}^{(r)}(B, \Sigma)$. Recall that $B = [B_1, \ldots, B_p]$ where each B_j is a $k \times m$ parameter matrix. It follows from Theorem 1 and the form of the full log-likelihood (19) that $B^{(r+1)}$ will satisfy

$$\frac{\partial \mathcal{Q}^{(r)}}{\partial B_j}(B, \Sigma) \equiv \Sigma^{-1} \left(\sum_{i=1}^{n} z_{ij} \bar{X}_i^{(r)} - \sum_{i=1}^{n} \sum_{k=1}^{p} z_{ij} z_{ik} B_k \right) = 0_{k,m}, \quad j = 1, \ldots, p, \tag{SM10}$$

where $\bar{X}_1^{(r)}, \ldots, \bar{X}_n^{(r)}$ are defined in (22). Under the assumptions of the theorem, and assuming that Σ has full rank, (SM10) may be written equivalently in the following form, using the Kronecker product:

$$\bar{Y}^{(r)}(Z \otimes I_m) = B^{(r+1)} \left((Z^\top Z) \otimes I_m \right), \tag{SM11}$$

where $\bar{Y}^{(r)}$ and $B^{(r+1)}$ are defined using (24). Next, post-multiply both sides of (SM11) by the matrix $(Z^\top Z)^{-1} \otimes I_m$ to obtain (25). To calculate $\Sigma^{(r+1)}$, we set $\partial \mathcal{Q}^{(r)} / \partial \Sigma = 0_{k,k}$ to obtain

$$\Sigma^{(r+1)} = \frac{1}{mn} \sum_{i=1}^{n} \left\{ X_i X_i^\top - \bar{X}_i^{(r)} \mu_i^{(r+1)\top} - \mu_i^{(r+1)} \bar{X}_i^{(r)\top} + \mu_i^{(r+1)} \mu_i^{(r+1)\top} \right\},$$

where $\mu_i^{(r+1)} = \sum_{j=1}^{p} z_{ij} B_j^{(r+1)}$, $i = 1, \ldots, n$. Moreover, since

$$\mu^{(r+1)} \equiv \left[\mu_1^{(r+1)}, \ldots, \mu_n^{(r+1)} \right] = \bar{Y}^{(r)} \left(P \otimes I_m \right),$$

it follows that

$$\sum_{i=1}^{n} \left(\bar{X}_i^{(r)} \mu_i^{(r+1)\top} + \mu_i^{(r+1)} \bar{X}_i^{(r)\top} - \mu_i^{(r+1)} \mu_i^{(r+1)\top} \right) = \bar{Y}^{(r)} \left(P \otimes I_m \right) \bar{Y}^{(r)\top},$$

which yields (26). Moreover, (27) and (28) follow easily from (25) and (26) respectively, by multiplying out the Kronecker products and using the definition $\bar{Y}^{(r)}$ given in (24). Finally, in the reflection size-and-shape case where $R \in O(m)$, all the details are identical except that the $\bar{R}_i^{(r)}$, $\bar{X}_i^{(r)}$ and $\bar{Y}^{(r)}$ are calculated using the conditional distribution with density (SM4) rather than (17). □

Proof of Lemma A1 and Lemma SM1. Define

$$R_+ \equiv R_+(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \quad R_- \equiv R_-(\theta) = \begin{pmatrix} -\cos\theta & \sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Note that $R_+(\theta)$ is a 2×2 rotation matrix, because $|R_+(\theta)| = 1$, and $R_-(\theta)$ is an orthogonal matrix with $|R_-(\theta)| = -1$. In the case of Lemma A1 we need to find the conditional expectations of $\cos\theta$ and $\sin\theta$ where, conditional on $M = (m_{ij})_{i,j=1}^2$ defined in (A2), θ has a von Mises density on the circle with exponent

$$\mathrm{tr}(R_+ M^\top) = (m_{11} + m_{22})\cos\theta \, + \, (m_{12} - m_{21})\sin\theta. \tag{SM12}$$

Consequently, (A3) in Lemma A1 follows from standard results for trigonometric moments of the von Mises distribution; see e.g. Mardia and Jupp (2000). In the case of Lemma SM1, we need to find the conditional expectation of $\cos\theta$ and $\sin\theta$ where now θ is conditionally distributed according to a mixture of two von Mises densities whose exponents are given by (SM12) and

$$\mathrm{tr}(R_- M^\top) = (m_{22} - m_{11})\cos\theta \, + \, (m_{12} + m_{21})\sin\theta,$$

and mixture proportions ω_1 and ω_2, where ω_1 and ω_2 are defined in Lemma SM1. Using these facts, Lemma SM1 also follows from the standard formulae for the trigonometric moments of the von Mises distribution. $\square$

Proof of Proposition 1 and Proposition SM1. The proof of Proposition 1 depends on three key points. The first is that, due to the existence of a certain 2:1 map from $\mathcal{S}^4$, the unit sphere in $\mathbb{R}^4$, to $SO(3)$, there is a natural identification of the Fisher matrix distributions on $SO(3)$ with the Bingham distributions on $\mathcal{S}^4$; see for example Prentice (1986) and Wood (1993) for detailed discussion. As a consequence of this identification there is a simple relationship between the respective normalization constants: if $\Phi = \mathrm{diag}\{\phi_1, \phi_2, \phi_3\}$ and $\Xi = \mathrm{diag}\{\xi_1, \ldots, \xi_4\}$, where the ξ_i are given by (A7), then

$$\mathcal{C}(\Phi) = \mathcal{C}_4(\Xi)/2, \tag{SM13}$$

where $\mathcal{C}(\Xi)$ is defined in (16) and $\mathcal{C}_4(\Xi)$ is defined in (A4) with $q = 4$.

For a general $m \times m$ matrix M, let $T_1 \Phi T_2^\top$ denote the restricted singular value decomposition of M in which $T_1, T_2 \in SO(m)$ and $\Phi = \text{diag}\{\phi_1, \ldots \phi_m\}$, where $\phi_1 \geq \cdots \geq \phi_m$, $\phi_{m-1} \geq 0$ and ϕ_m has the same sign as $|M|$; see, for example, Kendall et al. (1999) for this type of singular value decomposition. Then, by applying the transformation $R \mapsto T_1^\top R T_2$, and exchanging the order of differentation and integration to obtain the appropriate moments, it is seen that

$$\begin{aligned}
\bar{R} &= \int_{R \in \mathcal{SO}(m)} R \exp\left\{\text{tr}(RM^\top)\right\}(dR) \\
&= T_1 \left\{ \int_{R \in \mathcal{SO}(m)} \mathbf{R} \exp\left\{\text{tr}(R\Phi)\right\}(dR) \right\} T_2^\top \\
&= T_1 \,\text{diag}\left\{ \nabla_\Phi \int_{R \in \mathcal{SO}(m)} \exp\left\{\text{tr}(R\Phi)(dR)\right\} \right\} T_2^\top \\
&= T_1 \,\text{diag}\{\nabla_\Phi \mathcal{C}(\Phi)\}\, T_2^\top,
\end{aligned}$$

where $\mathcal{C}(.)$ is defined in (16). Here, $m = 3$ so we can make use of (SM13) to obtain

$$\bar{R} = \frac{1}{\mathcal{C}_4(\Xi)} T_1 \text{diag}\{\partial \mathcal{C}_4(\Xi)/\partial \phi_1, \partial \mathcal{C}_4(\Xi)/\partial \phi_2, \partial \mathcal{C}_4(\Xi)/\partial \phi_3\} T_2^\top, \tag{SM14}$$

where $\Xi = \text{diag}\{\xi_1, \xi_2, \xi_3, \xi_4\}$ depends on ϕ_1, ϕ_2 and ϕ_3 through (A7), and $\mathcal{C}_4(\Xi)$ is given by (A4) with $q = 4$.

The final point is that, to calculate the partial derivatives of the Bingham normalization constant in (SM14), we can make use of the results in Kume and Wood (2007). Specifically, Proposition 1 in Kume and Wood (2007) implies that

$$\frac{\partial \mathcal{C}_4(\Xi)}{\partial \xi_j} = \frac{1}{2\pi} \mathcal{C}_6(\Xi_j), \quad j = 1, 2, 3, 4. \tag{SM15}$$

Consequently, using (A7) and (SM15),

$$\begin{aligned}
\frac{\partial \mathcal{C}_4(\Xi)}{\partial \phi_i} &= \frac{\partial \mathcal{C}_4(\Xi)}{\partial \xi_1}\frac{\partial \xi_1}{\partial \phi_i} + \frac{\partial \mathcal{C}_4(\Xi)}{\partial \xi_2}\frac{\partial \xi_2}{\partial \phi_i} + \frac{\partial \mathcal{C}_4(\Xi)}{\partial \xi_3}\frac{\partial \xi_3}{\partial \phi_i} + \frac{\partial \mathcal{C}_4(\Xi)}{\partial \xi_4}\frac{\partial \xi_4}{\partial \phi_i} \\
&= \frac{1}{2\pi}\left\{\mathcal{C}_6(\Xi_i) - \mathcal{C}_6(\Xi_j) - \mathcal{C}_6(\Xi_\ell) + \mathcal{C}_6(\Xi_4)\right\}, \qquad i = 1, 2, 3,
\end{aligned}$$

where $i, j, \ell \in \{1, 2, 3\}$ and $i \neq j \neq \ell \neq i$. Also, since x in (A4) is a unit vector, it follows that

$$\frac{\partial \mathcal{C}_4(\Xi)}{\partial \xi_1} + \frac{\partial \mathcal{C}_4(\Xi)}{\partial \xi_1} + \frac{\partial \mathcal{C}_4(\Xi)}{\partial \xi_1} + \frac{\partial \mathcal{C}_4(\Xi)}{\partial \xi_1} = \mathcal{C}_4(\Xi),$$

and therefore, as a consequence of (SM15),

$$\mathcal{C}_6(\Xi_1) + \mathcal{C}_6(\Xi_2) + \mathcal{C}_6(\Xi_3) + \mathcal{C}_6(\Xi_4) = 2\pi \mathcal{C}_4(\Xi). \tag{SM16}$$

After substituting these results in (SM14), making use of (SM16) and rearranging, we obtain the expression for $\bar{R}$ in Proposition 1, as required.

In the case of Proposition SM1 we have, following Wood (1993),

$$\mathcal{C}_+(M) = \mathcal{C}_+(\Phi) = \frac{1}{2}\{\mathcal{C}_4(\Xi) + \mathcal{C}_4(-\Xi)\} = \frac{1}{2}\mathcal{C}_{4+}(\Xi),$$

where $M = T_1 \Phi T_2^\top$ is obtained using (2), $\mathcal{C}_+(.)$ is defined in (SM3) and $\mathcal{C}_{4+}(.)$ is defined in (SM7) with $q = 4$. Moreover, using similar calculations to those in the proof of Proposition 1 at (SM14),

$$\bar{R} = \frac{1}{\mathcal{C}_{4+}(\Xi)} T_1 \mathrm{diag}\{\partial \mathcal{C}_{4+}(\Xi)/\partial \phi_1, \partial \mathcal{C}_{4+}(\Xi)/\partial \phi_2, \partial \mathcal{C}_{4+}(\Xi)/\partial \phi_3\} T_2^\top,$$

from which Proposition SM1 follows. □

SM.C: Calculation of the observed information

We now write the full parameter vector as

$$\theta = (\mathrm{vec}(B_1)^\top, \ldots, \mathrm{vec}(B_p)^\top, \mathrm{vech}(\Sigma)^\top)^\top, \tag{SM17}$$

where vech is the version of the vectorisation operator which stacks the elements on or below the diagonal, and is appropriate for symmetric matrices. For example, if $A = (a_{ij})_{i,j=1}^3$, then $\mathrm{vech}(A) = (a_{11}, a_{21}, a_{31}, a_{22}, a_{32}, a_{33})^\top$.

It is also assumed that $B_1, \ldots, B_p$ have been standardized as in (7) and (9). Define the observed information

$$H(\theta) = -\frac{\partial^2 \ell_M}{\partial\theta\partial\theta^\top}(\theta), \tag{SM18}$$

and define

$$F(A) = \frac{\partial^2 \log \mathcal{C}}{\partial \mathrm{vec}(A)\partial \mathrm{vec}(A)^\top}(A), \tag{SM19}$$

where $\mathcal{C}(A)$ is defined in (16). Let D_n, L_n and K_{mn} denote, respectively, the duplication matrix, the eliminator matrix and the commutator matrix defined in Magnus & Neudecker (1988). For a general $m \times n$ matrix A, K_{mn} is the unique $(mn) \times (mn)$ matrix such that $\mathrm{vec}(A^\top) = K_{mn}\mathrm{vec}(A)$; and for a general symmetric $n \times n$ matrix A, D_n is the unique $n^2 \times n(n+1)/2$ matrix such that $D_n \mathrm{vech}(A) = \mathrm{vec}(A)$, and L_n is the unique $n(n+1)/2 \times n^2$ matrix such that $\mathrm{vech}(A) = L_n \mathrm{vec}(A)$.

A relatively compact expression for the observed information is given in the following result.

Theorem SM2. *Write $\hat{H} = H(\hat{\theta})$ where $H(\theta)$ is defined in (SM18) and $\hat{\theta}$ is the maximum likelihood estimator of θ, and write $\hat{F}_i = F(\mu_i^\top \Sigma^{-1} U_i \Delta_i)$, where F is defined in (SM19). Then $\hat{H}$ may be written as the block matrix*

$$\hat{H} = \begin{bmatrix} \hat{H}_{11} & \hat{H}_{12} \\ \hat{H}_{21} & \hat{H}_{22} \end{bmatrix}$$

with blocks defined by

$$\hat{H}_{11} = -\frac{\partial^2 \ell_M}{\partial \mathrm{vec}(B)\partial \mathrm{vec}(B)^\top}(\hat{\theta}) = \sum_{i=1}^{n} (z_i \otimes I_{km}) \left(I_m \otimes \hat{\Sigma}^{-1} - \hat{M}_{i1}^\top \hat{F}_i \hat{M}_{i1} \right) \left(z_i^\top \otimes I_{km} \right), \tag{SM20}$$

where

$$\hat{M}_{i1} = \left\{ \left(\Delta_i U_i \hat{\Sigma}^{-1} \right) \otimes I_m \right\} K_{km}, \tag{SM21}$$

B is defined in (20),

$$\hat{H}_{12} = \hat{H}_{21}^\top = -\frac{\partial^2 \ell_M}{\partial \mathrm{vec}(B)\mathrm{vech}(\Sigma)^\top}(\hat{\theta})$$
$$= -\sum_{i=1}^n (z_i \otimes I_{km}) \left\{ \left(\hat{\mu}_i^\top \hat{\Sigma}^{-1}\right) \otimes \hat{\Sigma}^{-1} + \hat{M}_{i1}^\top \hat{F}_i \hat{M}_{i2} \right\} D_k$$

and

$$\hat{M}_{i2} = \left(\Delta_i U_i^\top \hat{\Sigma}^{-1}\right) \otimes \left(\hat{\mu}_i^\top \hat{\Sigma}^{-1}\right);$$

and, finally,

$$\hat{H}_{22} = -\frac{\partial^2 \ell_M}{\partial \mathrm{vech}(\Sigma)\mathrm{vech}(\Sigma)^\top}(\hat{\theta})$$
$$= D_k^\top \left[\sum_{i=1}^n \left\{ \hat{\Sigma}^{-1} \left(U_i \Delta_i^2 U_i^\top + \hat{\mu}_i \hat{\mu}_i^\top - \frac{m}{2}\hat{\Sigma} \right) \hat{\Sigma}^{-1} \right\} \otimes \hat{\Sigma}^{-1} \right.$$
$$\left. + \hat{M}_{i2}^\top \hat{F}_i \hat{M}_{i2} \right] D_k.$$

Remark SM1. Estimated standard errors of the components of $\hat{\theta}$ are given by the diagonal elements of $\hat{H}^{-1}$.

Remark SM2. In view of the standardization in Section 2.3, which we assume is performed here, the rows and columns of $\hat{H}$ which involve derivatives with respect to components of B_1 which have been set to zero (i.e. those elements of B_1 for which the column number is greater than the row number) should be deleted from $\hat{H}$ before calculating the inverse.

Proof of Theorem SM2. The proof of this theorem makes heavy use of results given in Magnus and Neudecker (1998). Write $A_i = \mu_i^\top \Sigma^{-1} U_i \Delta_i$ where, as previously, U_i and Δ_i are obtained from the singular value decomposition of the configuration matrix X_i, i.e. $X_i = U_i \Delta_i R_i^\top$. Then, using the chain rule for vector functions,

$$\frac{\partial \log \mathcal{C}(A_i)}{\partial \mathrm{vec}(B)} = \frac{\partial \mathrm{vec}(\mu_i)^\top}{\partial \mathrm{vec}(B)} \frac{\partial \mathrm{vec}(A_i)^\top}{\partial \mathrm{vec}(\mu_i)} \frac{\partial \log \mathcal{C}(A_i)}{\partial \mathrm{vec}(A_i)}.$$

Since $\mu_i = \sum_{j=1}^{p} z_{ij} B_j$, it follows that

$$\begin{aligned}\text{vec}\left(\mu_i^\top\right) &= \sum_{j=1}^{p} z_{ij}\text{vec}(B_j)^\top \\ &= \text{vec}(B)^\top \left(z_i \otimes I_{km}\right),\end{aligned}$$

and therefore $\partial\text{vec}(\mu_i)^\top / \partial\text{vec}(B) = z_i \otimes I_{km}$. Also, using the standard result that $\text{vec}(ABC) = (C^\top \otimes A)\text{vec}(B)$, and the fact that $\text{vec}(\mu_i^\top) = K_{km}\text{vec}(\mu_i)$, it follows that

$$\begin{aligned}\frac{\partial\text{vec}(A_i)^\top}{\partial\text{vec}(\mu_i)} &= \frac{\partial}{\partial\text{vec}(\mu_i)}\left[\text{vec}(\mu_i^\top)^\top\left\{\left(\Sigma^{-1}U_i\Delta_i\right)\otimes I_m\right\}\right] \\ &= \frac{\partial}{\partial\text{vec}(\mu_i)}\left[\text{vec}(\mu_i)^\top K_{km}^\top\left\{\left(\Sigma^{-1}U_i\Delta_i\right)\otimes I_m\right\}\right] \\ &= K_{km}^\top\left\{\left(\Sigma^{-1}U_i\Delta_i\right)\otimes I_m\right\},\end{aligned}$$

so

$$\frac{\partial\log\mathcal{C}(A_i)}{\partial\text{vec}(B)} = (z_i \otimes I_{km})K_{km}^\top\left\{\left(\Sigma^{-1}U_i\Delta_i\right)\otimes I_m\right\}\frac{\partial\log\mathcal{C}(A_i)}{\partial\text{vec}(A_i)}.$$

Consequently,

$$\begin{aligned}\frac{\partial^2\log\mathcal{C}(A_i)}{\partial\text{vec}(B)\partial\text{vec}(B)^\top} &= \frac{\partial\text{vec}(\mu_i)^\top}{\partial\text{vec}(B)}\frac{\partial\text{vec}(A_i)^\top}{\partial\text{vec}(\mu_i)}\frac{\partial^2\log\mathcal{C}(A_i)}{\partial\text{vec}(A_i)\partial\text{vec}(A_i)^\top} \\ &\qquad\times\frac{\partial\text{vec}(A_i)}{\partial\text{vec}(\mu_i)^\top}\frac{\partial\text{vec}(\mu_i)}{\partial\text{vec}(B)^\top} \\ &= \left(z_i \otimes I_{km}\right) M_{i1}^\top F(A_i) M_{i1}\left(z_i^\top \otimes I_{km}\right).\end{aligned}$$

Similar calculations show that

$$\frac{1}{2}\frac{\partial^2\text{tr}\left(\mu_i^\top\Sigma^{-1}\mu_i\right)}{\partial\text{vec}(B)\partial\text{vec}(B)^\top} = \left(z_i \otimes I_{km}\right)\left(I_m \otimes \Sigma^{-1}\right)\left(z_i^\top \otimes I_{km}\right).$$

Consequently,

$$
\begin{aligned}
H(\theta) &= -\frac{\partial^2 \ell_M(\theta)}{\partial \mathrm{vec}(B) \partial \mathrm{vec}(B)^\top} \\
&= -\sum_{i=1}^{n} \frac{\partial^2}{\partial \mathrm{vec}(B) \partial \mathrm{vec}(B)^\top} \log f_1(U_i, \Delta_i; B, \Sigma) \\
&= \sum_{i=1}^{n} \frac{\partial^2}{\partial \mathrm{vec}(B) \partial \mathrm{vec}(B)^\top} \left\{ \frac{1}{2} \mathrm{tr} \left(\mu_i^\top \Sigma^{-1} \mu_i \right) - \log \mathcal{C}(\mu_i \Sigma^{-1} U_i \Delta_i) \right\} \\
&= \sum_{i=1}^{n} (z_i \otimes I_{km}) \left(I_m \otimes \Sigma^{-1} - M_{i1}^\top F(A_i) M_{i1} \right),
\end{aligned}
$$

and (SM20) follows after substituting $B = \hat{B}$ and $\Sigma = \hat{\Sigma}$.

When differentiating with respect to $\mathrm{vech}(\Sigma)$, the following facts are useful:

$$
\frac{\partial \log |\Sigma|}{\partial \mathrm{vech}(\Sigma)} = D_k^\top \mathrm{vec}(\Sigma^{-1}); \qquad \frac{\partial \mathrm{vec}(\Sigma^{-1})}{\partial \mathrm{vech}(\Sigma)^\top} = -\left(\Sigma^{-1} \otimes \Sigma^{-1} \right) D_k;
$$

and for a constant matrix C of compatible dimension,

$$
\frac{\partial^2 \mathrm{tr} \left(C^\top \Sigma^{-1} C \right)}{\partial \mathrm{vech}(\Sigma) \partial \mathrm{vech}(\Sigma)^\top} = 2 D_k^\top \left\{ \left(\Sigma^{-1} C C^\top \Sigma^{-1} \right) \otimes \Sigma^{-1} \right\} D_k.
$$

Using the above expressions, the proof of the remainder of Theorem SM2 is straightforward and is omitted. □

We now consider the calculation of the observed information specified in Theorem SM2 in the case $m = 3$; the case $m = 2$ follows directly from standard results for derivatives of the Bessel function $\mathcal{I}_0$ and we omit the details. The only step not yet explained is how to calculate $F(A)$ in (SM19) when $m = 3$. Following the definition (A5), define

$$
\Xi_{ij} = \mathrm{diag}\{\xi_1, \xi_2, \xi_3, \xi_4, \xi_i, \xi_i, \xi_j, \xi_j\}.
$$

We have the following result whose proof we omit.

Proposition SM2. *Suppose A has singular value decomposition $T_1 \Phi T_2^\top$ as in Proposition 1. Then the function $F(A)$ has the following expression in terms of the Bingham normalising constant:*

$$F(A) = \left(T_2 \otimes T_1^\top\right) \left\{E\left[\mathrm{vec}(\check{R})\mathrm{vec}(\check{R})^\top\right] - \mathrm{vec}(\Omega)\mathrm{vec}(\Omega)^\top\right\} \left(T_2^\top \otimes T_1\right),$$

where the density of $\check{R} = (\check{R}_{uv})_{u,v=1}^3$ is f_2 in (17) with parameter matrix $\Phi = diag(\phi_1, \phi_2, \phi_3)$ rather than A, and Ω is defined in Proposition 1. Moreover, the elements of $E[vec(\check{R})vec(\check{R})^\top]$ may be calculated as follows:

$$E[\check{R}_{ii}^2] = 1 + \frac{1}{\mathcal{C}_4(\Xi)} \left[\frac{3\mathcal{C}_8(\Xi_{jj}) + 3\mathcal{C}_8(\Xi_{\ell\ell}) + 2\mathcal{C}_8(\Xi_{j\ell})}{\pi^2} - \frac{2}{\pi}\left\{\mathcal{C}_6(\Xi_j) + \mathcal{C}_6(\Xi_\ell)\right\}\right];$$

for $i \neq j$,

$$E[\check{R}_{ii}\check{R}_{jj}] = 1 + \frac{1}{\mathcal{C}_4(\Xi)}\left\{\frac{3\mathcal{C}_8(\Xi_{\ell\ell}) + \mathcal{C}_8(\Xi_{ij}) + \mathcal{C}_8(\Xi_{i\ell}) + \mathcal{C}_8(\Xi_{j\ell})}{\pi^2} - \frac{\mathcal{C}_6(\Xi_i) + \mathcal{C}_6(\Xi_j) + 2\mathcal{C}_6(\Xi_\ell)}{\pi}\right\},$$

$$E[\check{R}_{ij}^2] = \frac{\mathcal{C}_8(\Xi_{ij}) + \mathcal{C}_8(\Xi_{\ell m})}{\pi^2 \mathcal{C}_4(\Xi)}$$

and

$$E[\check{R}_{ij}\check{R}_{ji}] = \frac{\mathcal{C}_8(\Xi_{ij}) - \mathcal{C}_8(\Xi_{\ell m})}{\pi^2 \mathcal{C}_4(\Xi)};$$

and for all choices of indices not covered above,

$$E[\check{R}_{ij}\check{R}_{uv}] = 0.$$

SM.D: Further numerical results relating to §4

For the scalar covariance matrix $\Sigma = \sigma^2 I_k$, the differences between the MLE and Procrustes approaches are investigated in §4. Note that for both of these algorithms a right multiplication of X_i's by an element from $O(k)$ transforms the respective solutions in the same way. Hence, if $m = k = 3$, we could then focus without loss of generality only to cases where μ is diagonal. For a given choice of σ and of some diagonal μ, we simulate $n = 20, 50, 100$ and 1000 random perturbations with $\Sigma = \sigma^2 I_k$ and compare the respective quantities for MLE and Procrustes algorithms. We run this procedure 1000 times and the mean values of these quantities are reported in Tables 1, 4 and 5. We notice that these two approaches depart from each other as the mean departs from the multiple of the identity (see the differences for the same σ across the three tables).

n and σ	$\rho_s(\hat{\mu}_p, \mu)$	$\rho_s(\hat{\mu}_{EM}, \mu)$	$\rho(\hat{\mu}_p, \mu)$	$\rho(\hat{\mu}_{EM}, \mu)$	σ_p	σ_{EM}
n= 20 sig= 0.1	0.0101	0.0101	0.0075	0.0075	0.0783	0.0784
n= 50 sig= 0.1	0.0494	0.0494	0.0399	0.0399	0.0747	0.0748
n= 100 sig= 0.1	0.0411	0.0411	0.0256	0.0256	0.0787	0.0787
n= 1000 sig= 0.1	0.0137	0.0136	0.0053	0.0053	0.0813	0.0814
n= 20 sig= 0.3	0.2468	0.1747	0.1519	0.1632	0.2345	0.2965
n= 50 sig= 0.3	0.2213	0.1679	0.1801	0.1682	0.2322	0.2941
n= 100 sig= 0.3	0.1342	0.0404	0.0421	0.0385	0.233	0.295
n= 1000 sig= 0.3	0.1457	0.0246	0.0411	0.0242	0.2369	0.3003
n= 20 sig= 0.8	1.0478	0.7088	0.1547	0.1389	0.4666	0.5979
n= 50 sig= 0.8	1.3456	0.3605	0.5863	0.3677	0.6035	0.8628
n= 100 sig= 0.8	1.0364	0.2633	0.2261	0.2552	0.5806	0.8092
n= 1000 sig= 0.8	1.0349	0.3564	0.2482	0.2637	0.5794	0.8432

Table 4: Details here are the same as those in the caption for Table 1. Choice of mean: $\mu_1 \propto diag(6, 4, 2)$. Note that some of the figures for the EM and Procurstes estimates at $\sigma = 0.1$ are the same to 4 d.p due to the fact that we applied the convergence accuracy of the order 10^{-3} in our EM procedure while the procrustes estimates are its starting values.

n and σ	$\rho_s(\hat{\mu}_p, \mu)$	$\rho_s(\hat{\mu}_{EM}, \mu)$	$\rho(\hat{\mu}_p, \mu)$	$\rho(\hat{\mu}_{EM}, \mu)$	σ_p	σ_{EM}
n= 20 sig= 0.1	0.1039	0.1039	0.0981	0.0981	0.0793	0.0793
n= 50 sig= 0.1	0.0376	0.0375	0.0205	0.0205	0.0796	0.0796
n= 100 sig= 0.1	0.0199	0.0199	0.0194	0.0194	0.0816	0.0817
n= 1000 sig= 0.1	0.0139	0.0139	0.0077	0.0077	0.0816	0.0817
n= 20 sig= 0.3	0.3381	0.3077	0.3267	0.3122	0.2118	0.2668
n= 50 sig= 0.3	0.148	0.0591	0.0597	0.0591	0.2321	0.2925
n= 100 sig= 0.3	0.1852	0.0779	0.0652	0.0674	0.2372	0.2987
n= 1000 sig= 0.3	0.1483	0.0197	0.0164	0.0188	0.2388	0.3014
n= 20 sig= 0.8	1.5146	1.057	0.6288	0.6506	0.565	0.7431
n= 50 sig= 0.8	1.0044	0.5441	0.2987	0.4711	0.5409	0.7368
n= 100 sig= 0.8	1.0563	0.4299	0.1633	0.2993	0.5715	0.7784
n= 1000 sig= 0.8	0.9275	0.6184	0.0505	0.4255	0.5704	0.8453

Table 5: Details here are the same as those in the caption for Table 1. Choice of mean: $\mu_2 \propto diag(1, 1, 1)$. Note that some of the figures for the EM and Procurstes estimates at $\sigma = 0.1$ are the same to 4 d.p due the fact that we applied the convergence accuracy of the order 10^{-3} in our EM procedure while the procrustes estimates are its starting values.