Maximum Likelihood Estimation for General Models in Size and Shape Space

A. Kume

University of Kent

joint work with I.L. Dryden and P. Paine and A.T.A. Wood

Equivalent classes of shapes

$$\mathbf{X}^\dagger = \left(\mathbf{x}_1^\dagger, \mathbf{x}_2^\dagger, ..., \mathbf{x}_{k+1}^\dagger\right) = \left(\begin{array}{cccc} x_{1,1}^\dagger & x_{1,2}^\dagger & \cdots & x_{1,k+1}^\dagger \\ x_{2,1}^\dagger & x_{2,2}^\dagger & \cdots & x_{2,k+1}^\dagger \\ \cdots & \cdots & \cdots & \cdots \\ x_{m,1}^\dagger & x_{m,2}^\dagger & \cdots & x_{m,k+1}^\dagger \end{array} \right).$$

Shape of $\mathbf{X}^\dagger$ is the equivalent class

$$[\mathbf{X}] = \left\{ c\mathbf{R}\mathbf{X}^\dagger + \tau \mid c \in \mathbb{R}^+, \mathbf{R} \in SO(m), \tau \in \mathbb{R}^m \right\}$$

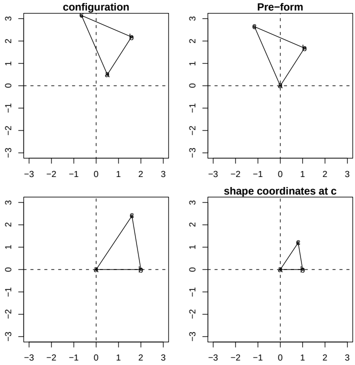
Size-and-shape of $\mathbf{X}^\dagger$ is the equivalent class

$$[\mathbf{X}]^s = \left\{ \mathbf{R}\mathbf{X}^\dagger + \tau \mid \mathbf{R} \in SO(m), \tau \in \mathbb{R}^m \right\}$$

If reflection invariance is required then $SO(m)$ is replaced by $O(m)$

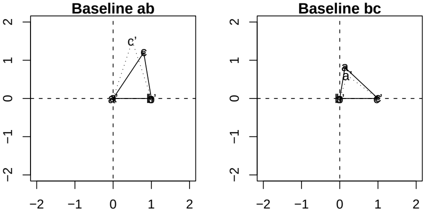
Geometry of shapes

We will focus on the shapes of *labelled* planar configurations.
Bookstein shape coordinates



Shape distances and mean shapes

Bookstein shape coordinates are not appropriate for describing shape differences.



If $\mathbf{X}^c$ and $\mathbf{Y}^c$ are two configurations such that their centres are at $(0, 0)$ and $\|\mathbf{X}^c\| = \|\mathbf{Y}^c\| = 1$, the *partial procrustes* shape distance is obtained as

$$d_P(\mathbf{X}^c, \mathbf{Y}^c) = \inf_{\mathbf{R} \in SO(m)} \|\mathbf{X}^c - \mathbf{R}\mathbf{Y}^c\| = \left\|\mathbf{X}^c - \hat{\mathbf{R}}\mathbf{Y}^c\right\| \quad \hat{\mathbf{R}}\mathbf{Y}^c\mathbf{X}^{c\,t} - \text{symmetric}$$

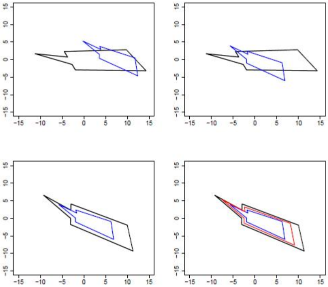
Alternative procrustes distances ρ and d_F are defined as

$$2\sin(\rho/2) = d_P \quad \text{and} \quad d_F = \sin(\rho)$$

Procrustes mean shapes

Partial procrustes mean shape of $\mathbf{X}_1^c, \ldots, \mathbf{X}_n^c$ is defined as

$$\mu^c = \underset{\mathbf{Z}^c}{\text{argmin}} \sum_{i=1}^{n} d_P^2(\mathbf{X}_i^c, \mathbf{Z}^c).$$



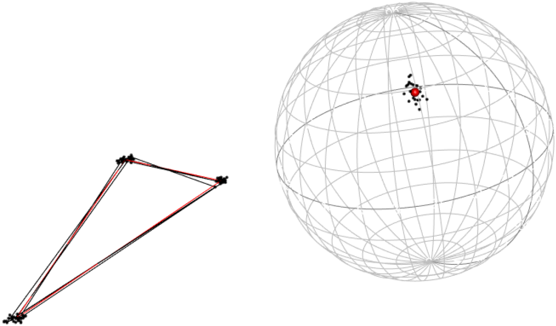


Figure: Landmark space and quotient space

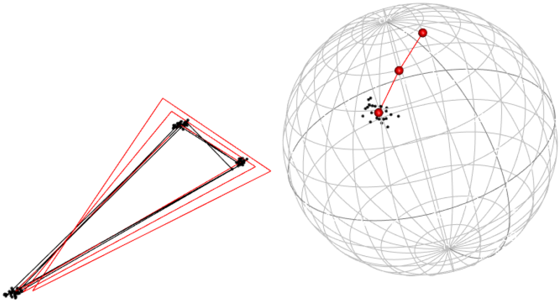


Figure: Marginal regression in (quotient) size-and-shape space

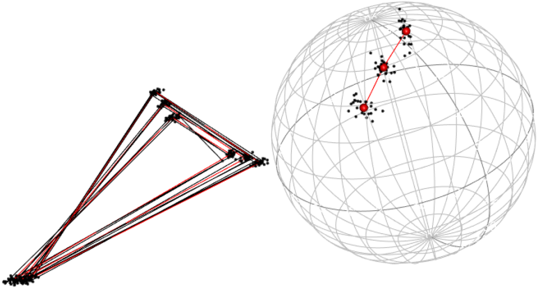


Figure: Marginal regression in size-and-shape space

Human movement data

Landmarks for a particular individual during the pointing action. Each landmark follows a nearly closed curved trajectory.

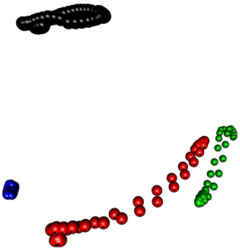
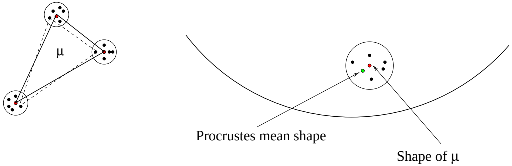


Figure: A representation of the four landmarks: shoulder, elbow, index finger tip and the lower back, during the pointing action while coloured as black, red, green and blue respectively.

Connections between two approaches



$\Sigma^* = \sigma^2 I_{mk}$:

- Procrustes mean shape is a consistent estimator for the shape of μ only for $m = 2$.
- Procrustes tangent space inference is valid for small σ^2.

If $\Sigma^* \neq \sigma^2 I_{mk}$:

- Procrustes mean shape is not a consistent estimator for the shape of μ.
- Approximation: Procrustes mean shape and tangent inference is valid for small values of Σ.
- MLE?

- Translation invariance

$$\mathbf{Y}^\dagger = \mathbf{X}^\dagger + \tau = \left(\mathbf{x}_1^\dagger + \tau, \mathbf{x}_2^\dagger + \tau, ..., \mathbf{x}_{k+1}^\dagger + \tau\right) \quad \tau \in \mathbb{R}^m$$

Standardize translation: $\tau = -\mathbf{x}_1^\dagger$

$$\mathbf{Y}^\dagger = \left(\mathbf{0}, \mathbf{x}_2^\dagger - \mathbf{x}_1^\dagger, ..., \mathbf{x}_{k+1}^\dagger - \mathbf{x}_1^\dagger\right) = (\mathbf{0}, \mathbf{X})$$

Alternatively: $\mathbf{X} = \mathbf{H}\mathbf{X}^\dagger$ where $\mathbf{H}$ is Helmert sub-matrix.
We call $\mathbf{X}$ pre-form.

- Rotation invariance

$$
\begin{aligned}
\mathbf{Y}^\dagger &= (\mathbf{X}^\dagger + \tau)\mathbf{R} \quad \mathbf{R} \in SO(m) \\
&= (\mathbf{0}, \mathbf{XR}) \quad \text{if} \quad \tau = -\mathbf{x}_1^\dagger
\end{aligned}
$$

Standardise Translation and Rotation (**size-and-shape variables**):
Apply the singular-values-decomposition of

$$
\begin{aligned}
\mathbf{X} = \mathbf{R}\Delta\mathbf{O} =& \mathbf{R}\begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \cdots & \cdots & & \\ 0 & 0 & \cdots & \sqrt{\lambda_m} \end{pmatrix}\mathbf{O} \\
=& \mathbf{R} \, diag(\sqrt{\lambda_1}, ..., \sqrt{\lambda_m})\mathbf{O}
\end{aligned}
$$

where $\mathbf{R} \in SO(m), \mathbf{O} \in \mathcal{V}(k, m)$ and $\lambda_1, \lambda_2.., \lambda_m$ are the m eigen values of $\mathbf{XX}^T$. We can use $\Delta\mathbf{O}$ as size-and-shape variables.

Shape distributions

Distribution of $\Delta\mathbf{O}$ if $\mathbf{X} = \mathbf{R}\Delta\mathbf{O} = \sqrt{s}\mathbf{R}\tilde{\Delta}\mathbf{O} \sim \mathcal{N}_{mk}(\mu, I_m \otimes \sigma^2 I_k)$.

$$d\mathbf{X} \propto d\mathbf{R} d\Delta d\mathbf{O} \quad \text{or} \quad d\mathbf{X} \propto s^{km/2-1} d\mathbf{R} d\tilde{\Delta} d\mathbf{O} ds$$

where $d\Delta = \prod_{i=1}^m \lambda_i^{(k-m-1)/2} \prod_{i>j}(\lambda_i - \lambda_j) \prod_{i=1}^m d\lambda_i$, $d\mathbf{R}$ and $d\mathbf{O}$ represent the Haar measures in the respective orthogonal groups $SO(m)$ and $\mathbf{O} \in O(k)$.

$$f_{\mathcal{N}}(\mathbf{X} = \mathbf{R}\Delta\mathbf{O}; \mu, \sigma^2 I_{mk}) \propto e^{-\frac{vec(\mathbf{X}-\mu)vec(\mathbf{X}-\mu)^t}{2\sigma^2}})$$

$$\propto e^{-\frac{s+\|\mu\|^2}{2}} e^{\frac{tr(\mathbf{R}\Delta\mathbf{O}\mu^t)}{\sigma^2}} \quad s = \|\Delta\|^2 = \|\mathbf{X}\|^2$$

$$f(\Delta\mathbf{O}; \mu, \sigma^2 I_{mk}) d\Delta d\mathbf{O} = e^{-\frac{s+\|\mu\|^2}{2}} \int_{SO(m)} e^{\frac{tr(\mathbf{R}\Delta\mathbf{O}\mu^t)}{\sigma^2}} d\mathbf{R} d\Delta d\mathbf{O}$$

$$f(\tilde{\Delta}\mathbf{O}; \mu, \sigma^2 I_{mk}) d\Delta d\mathbf{O} = e^{-\frac{\|\mu\|^2}{2}} \int_{\mathbb{R}^+} s^{km/2-1} e^{-\frac{s}{2\sigma^2}} \int_{SO(m)} e^{\sqrt{s}\frac{tr(\mathbf{R}\tilde{\Delta}\mathbf{O}\mu^t)}{2\sigma^2}} d\mathbf{R}$$

Particular cases

$$\mathbf{X}_i|\mathbf{z}_i \overset{\text{indep}}{\sim} \mathcal{N}_{k\times m}(\boldsymbol{\mu}_i = \sum_{j=1}^{p} z_{ij}\mathbf{B}_j, \mathbf{I}_m \otimes \boldsymbol{\Sigma}),$$

- IID case , $\mathbf{z}_i = 1$, $\boldsymbol{\mu}_i = \boldsymbol{\mu}$
- $\mathbf{z}_i = (1,0)$ or $\mathbf{z}_i = (0,1)$ regressor for gender

$$\boldsymbol{\mu}_i = z_{i1}\mathbf{B}_1 + z_{i2}\mathbf{B}_2$$

- Polynomial regression $\mathbf{z}_i = (1, t_i, \cdots, t_i^{p-1})$

$$\boldsymbol{\mu}_i = \mathbf{B}_1 + t_i\mathbf{B}_2 + t_i^2\mathbf{B}_3 + \cdots t_i^{p-1}\mathbf{B}_p$$

For $i = 1, \ldots, n$ define

$$\bar{\mathbf{R}}_i^{(r)} = E[\mathbf{R}_i | \mathbf{O}_i, \boldsymbol{\Delta}_i; \mathbf{B}^{(r)}, \boldsymbol{\Sigma}^{(r)}], \qquad (1)$$

where $\mathbf{O}_i$, $\boldsymbol{\Delta}_i$ and $\mathbf{R}_i$ are determined using SVD. Write

$$\bar{\mathbf{X}}_i^{(r)} = \bar{\mathbf{R}}_i \boldsymbol{\Delta}_i \mathbf{O}_i, \quad i = 1, \ldots, n, \qquad (2)$$

and define the $n \times p$ matrix $\mathbf{Z} = (z_{ij})$, the $p \times n$ matrix $\mathbf{A} = (a_{ji})$ and the $n \times n$ matrix $\mathbf{P} = (p_{ij})$ by

$$\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_n]^\top, \quad \mathbf{A} = (\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top \quad \text{and} \quad \mathbf{P} = \mathbf{I}_n - \mathbf{Z}(\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top. \qquad (3)$$

Also, for $r \geq 0$, define the $k \times (mn)$ matrix $\bar{\mathbf{Y}}^{(r)}$ and the $k \times (mp)$ matrix $\mathbf{B}^{(r)}$ by

$$\bar{\mathbf{Y}}^{(r)} = [\bar{\mathbf{X}}_1, \ldots, \bar{\mathbf{X}}_n] \quad \text{and} \quad \mathbf{B}^{(r)} = [\mathbf{B}_1^{(r)}, \ldots, \mathbf{B}_p^{(r)}]. \qquad (4)$$

Theorem 2. *Assume that $n \geq p$ and that $\mathbf{Z}$ in (3) has full rank p. Then, given a starting value for $\mathbf{B}^{(0)}$ as defined in (4), the EM updating rule for calculating the sequence $(\mathbf{B}^{(r)}, \boldsymbol{\Sigma}^{(r)})$ is given by*

$$Vec(\mathbf{B}^{(r+1)}) = (\mathbf{A} \otimes \mathbf{I}_{km})\, Vec(\bar{\mathbf{Y}}^{(r)}) \tag{5}$$

and

$$\boldsymbol{\Sigma}^{(r+1)} = Vec(\bar{\mathbf{Y}}^{(r)})^{\top} (\mathbf{P} \otimes \mathbf{I}_{km})\, Vec(\bar{\mathbf{Y}}^{(r)}). \tag{6}$$

Moreover, the updating rules (5) and (6) are equivalent to

$$\mathbf{B}_j^{(r+1)} = \sum_{i=1}^{n} a_{ji} \bar{\mathbf{X}}_i, \qquad j = 1, \ldots, p, \tag{7}$$

and

$$\boldsymbol{\Sigma}^{(r+1)} = \frac{1}{mn} \left\{ \left(\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{X}_i \right) - \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} \bar{\mathbf{X}}_i^{\top} \bar{\mathbf{X}}_j \right\}, \tag{8}$$

where the a_{ji} and p_{ij} are, respectively, the components of the matrices $\mathbf{A}$ and $\mathbf{P}$ defined in (3)

MLE using EM in IID case $\mathbf{X}_i \sim \mathcal{N}(\boldsymbol{\mu}, I_m \otimes \boldsymbol{\Sigma})$

$$(\boldsymbol{\Delta}_i, \mathbf{O}_i) \quad \text{have density} \quad f_1(\boldsymbol{\Delta}, \mathbf{O} | \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\boldsymbol{\mu}^{(r+1)} = \frac{1}{n} \sum_{i=1}^{n} \bar{\mathbf{R}}_i \boldsymbol{\Delta}_i \mathbf{O}_i \quad \bar{\mathbf{R}}_i = E(\mathbf{R} | \boldsymbol{\Delta}_i, \mathbf{O}_i; \boldsymbol{\mu}^{(r)}, \boldsymbol{\Sigma}^{(r)})$$

$$\boldsymbol{\Sigma}^{(r+1)} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{O}_i^T \boldsymbol{\Delta}_i^2 \mathbf{O}_i - {\boldsymbol{\mu}^{(r+1)}}^T \boldsymbol{\mu}^{(r+1)}$$

Expected rotation $\bar{\mathbf{R}} = \frac{1}{\mathcal{C}(\mathbf{M})} \int\limits_{SO(m)} \mathbf{R} e^{tr(\mathbf{R}\mathbf{M})} d\mathbf{R}$

$$\bar{\mathbf{R}} = \mathbf{U}_2 diag\left(\nabla_{\boldsymbol{\Phi}} \log \int_{SO(m)} e^{tr(\mathbf{R}\boldsymbol{\Phi})} d\mathbf{R}\right) \mathbf{U}_1^t \quad \mathbf{M} = \mathbf{U}_1 \boldsymbol{\Phi} \mathbf{U}_2^t$$

and

$$E(\mathbf{X}|\boldsymbol{\Delta}, \mathbf{O}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \bar{\mathbf{R}}(\mathbf{M})\boldsymbol{\Delta}\mathbf{O} \quad \text{with} \quad \mathbf{M} = \boldsymbol{\Delta}\mathbf{O}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^T$$

m=2

$$\bar{\mathbf{R}} = \frac{I_1(\phi_1 + \phi_2)}{I_0(\phi_1 + \phi_2)}\mathbf{U}_2\mathbf{U}_1^t$$

m=3

$$\bar{\mathbf{R}} = \mathbf{U}_2\boldsymbol{\Omega}\mathbf{U}_1^t$$

$$\boldsymbol{\Omega} = \mathbf{I}_3 - \begin{pmatrix} \frac{\mathcal{C}_6(\boldsymbol{\xi}_2)+\mathcal{C}_6(\boldsymbol{\xi}_3)}{\pi\mathcal{C}_4(\boldsymbol{\xi})} & 0 & 0 \\ 0 & \frac{\mathcal{C}_6(\boldsymbol{\xi}_1)+\mathcal{C}_6(\boldsymbol{\xi}_3)}{\pi\mathcal{C}_4(\boldsymbol{\xi})} & 0 \\ 0 & 0 & \frac{\mathcal{C}_6(\boldsymbol{\xi}_1)+\mathcal{C}_6(\boldsymbol{\xi}_2)}{\pi\mathcal{C}_4(\boldsymbol{\xi})} \end{pmatrix} \quad (9)$$

$$\xi_4 = \phi_1 + \phi_2 + \phi_3 \quad \text{and} \quad \xi_i = 2\phi_i - \xi_4 \quad i = 1,2,3 \quad (10)$$

$$\mathcal{C}_4(\boldsymbol{\xi}) = \int_{vv^t=1; v\in\mathbb{R}^4} e^{-v^t\boldsymbol{\xi}v} d_{\mathcal{S}^3}(v) \;\; \text{and} \;\; \mathcal{C}_6(\boldsymbol{\xi}_i) = \int_{vv^t=1; v\in\mathbb{R}^6} e^{-v^t c(\boldsymbol{\xi}, \xi_i, \xi_i) v} d_{\mathcal{S}^5}(v)$$

Evaluate $\mathcal{C}_i$ and $\boldsymbol{\Omega}$ using Saddle point approximation or Holonomic gradient method.

A comparison between HG and SPA

$$\mathbf{\Omega} = diag\left(\nabla_{\mathbf{\Phi}} \log \int_{SO(3)} e^{tr(\mathbf{R\Phi})} d\mathbf{R}\right)$$

$\mathbf{\Phi}$	$\mathbf{\Omega}_{HG}$	$\mathbf{\Omega}_{SPA}$	$\mathbf{\Phi}$	$\mathbf{\Omega}_{HG}$	$\mathbf{\Omega}_{SPA}$	$\mathbf{\Phi}$	$\mathbf{\Omega}_{HG}$	$\mathbf{\Omega}_{SPA}$	$\mathbf{\Phi}$	$\mathbf{\Omega}_{HG}$	$\mathbf{\Omega}_{SPA}$
216	0.99576	0.99576	343	0.99729	0.99729	512	0.99817	0.99817	729	0.99870	0.99870
36	0.98599	0.98604	49	0.98970	0.98975	64	0.99210	0.99216	81	0.99374	0.99381
6	0.98572	0.98577	7	0.98954	0.98960	8	0.99201	0.99207	9	0.99368	0.99375

Isotropic covariance $\Sigma = \sigma^2 \mathbf{I}_k$

EM implementation

1. $$\boldsymbol{\mu}_{r+1} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{U}_{2i}\bar{\mathbf{R}}(\phi_{\boldsymbol{i}}/\boldsymbol{\sigma}_{\boldsymbol{r}}^{\mathbf{2}})\mathbf{U}_{1i}^T\boldsymbol{\Delta}_i\mathbf{O}_i \quad \boldsymbol{\Delta}_i\mathbf{O}_i\boldsymbol{\mu}_r^t = \mathbf{U}_{1i}\phi_i\mathbf{U}_{2i}^T$$

2. $$\sigma_{r+1}^2 = \frac{1}{mk}\sum_{i=1}^{n}\frac{tr(\Delta_i^2)}{n} - tr(\boldsymbol{\mu}_{r+1}\boldsymbol{\mu}_{r+1}^t)$$

We get the Procrustes algorithm if above $\bar{\mathbf{R}}(\phi/\boldsymbol{\sigma}_{\boldsymbol{r}}^{\mathbf{2}}) = \mathbf{I}_m$.

Isotropic covariance and IID model $m = 3$

$$\boldsymbol{\mu} = \begin{pmatrix} 50 & 0 & 0 & 0 \\ 0 & 10 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \mathcal{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_3 \otimes \mathbf{I}_k) \quad k = 4$$

n	$\rho(\hat{\mu}_{proc}, \mu)$	$\rho(\hat{\mu}_{mle}, \mu)$	$\hat{\sigma}_{proc}$	$\hat{\sigma}_{mle}$
1500	0.105	0.021	0.156	0.199
2000	0.102	0.011	0.156	0.199
3000	0.105	0.009	0.157	0.201
3500	0.101	0.011	0.157	0.201
1500	0.205	0.047	0.229	0.299
2000	0.199	0.031	0.229	0.297
3000	0.200	0.029	0.230	0.299
3500	0.201	0.032	0.229	0.299
1500	0.373	0.053	0.368	0.495
2000	0.379	0.077	0.373	0.501
3000	0.375	0.055	0.371	0.498
3500	0.390	0.064	0.369	0.496

Bookstein data of rat skulls

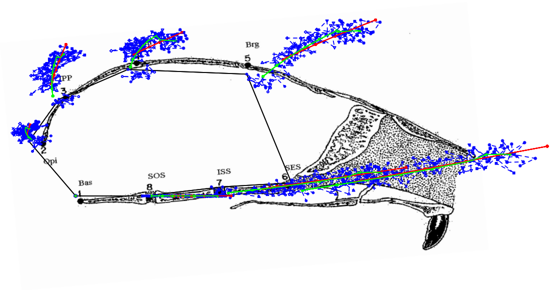


Figure: Linear(red) and quadratic(green) polynomial regression
Reproduced from Bookstein (1991)

Data consists of 18 individual rats observed at 8 different time points when they are 7, 14, 21, 30, 40, 60, 90, and 150 days old.

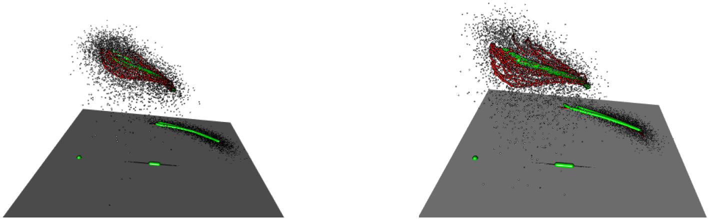


Figure: The fitted polynomial mean paths (qubic-left and quadratic-right) in green, observations are in red; the rotation standartisation is obtained by fixing landmark 1 to the origin, landmark 2 is allowed to vary only along a chosen axis and landmark 3 is varying only in the standardizing plane (the shaded region), landmark 4 is allowed to freely vary in 3-d space. Simulated data from the fitted models are shown in black.

Figure: Quadratic polynomial fit in red and the observations in black.

Figure: Cubic polynomial fit in red and the observations in black.

Concluding remarks

- Models in Landmark Space are relatively easy to interpret in Euclidean space.
- Computationally challenging but the complexity is only due to m-dimensional expectations.
- Guaranteed consistency for our (ML) estimates.
- Flexibility in modelling complex data (more general covariance, regression models).
- The approach could be seen as an adopted version of Procrustes algorithm.

Dryden, I.L. & Mardia, K.V. (1991). General shape distributions in a plane. *Adv. Appl. Prob.* **23**, 259-276.

Dryden, I.L. & Mardia, K.V. (1998). *Statistical Shape Analysis*. John Wiley, Chichester.

Goodall, C. & Mardia, K.V. (1992). The Non-central Bartlett Decomposition and Shape Densities. *J. Mult. Analysis* B **40**, 94-108.

Kume, A. & Wood, A. T. A.(2005): Saddlepoint approximations for the Bingham and Fisher-Bingham normalising constants. *Biometrika*, 92: 465–476.

Kume, A. & Sei, T. (2013). On the explicit form of the Pfaffian of the Fisher-Bingham integral. *Stat. Comp.* 1-12.

Mardia, K. V. & Jupp, P. E. (2000) *Directional Statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester.

Kent J.T. & Mardia K.V. (2001) Shape, Procrustes tangent projections and bilateral symmetry. *Biometrika* **88**, 469-485.

Wood, A. T. A. (1993). Estimation of the concentration parameters of the Fisher matrix distribution on $\mathbf{SO}(3)$ and the Bingham distribution on $S_q, q \geq 2$. *Aus. J. Stat*, **35**, 69-79.