http://www.stat.sc.edu/~dryden/course/ild-ch-04.pdf

Statistical Shape Analysis

Ian Dryden (University of Nottingham)

Ian.Dryden@Nottingham.ac.uk
http://www.maths.nott.ac.uk/∼ild

3e cycle romand de statistique et probabilités appliqués

Les Diablerets, Switzerland, March 7-10, 2004.

Session I

Dryden and Mardia (1998, chapters 1,2,3,4)

- Introduction
- Motivation and applications
- Size and shape coordinates
- Shape space
- Shape distances.

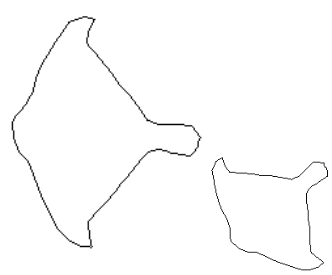
In a wide variety of applications we wish to study the geometrical properties of objects.

We wish to measure, describe and compare the size and shapes of objects

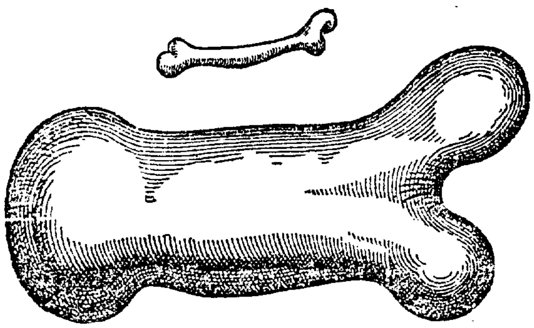
Shape: location, rotation and scale information (similarity transformations) can be removed. [Kendall, 1984]

Size-and-shape: location, rotation (rigid body transformations) can be removed.

An object's shape is invariant under the similarity transformations of translation, scaling and rotation.



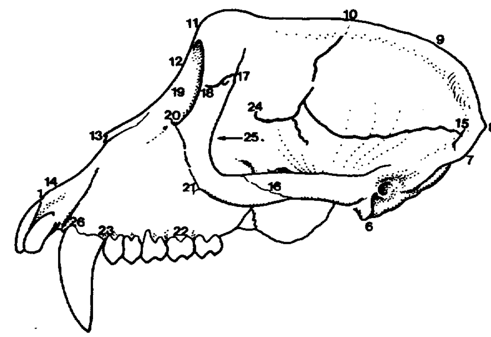
Two mouse second thoracic vertebra (T2 bone) outlines with the same shape.

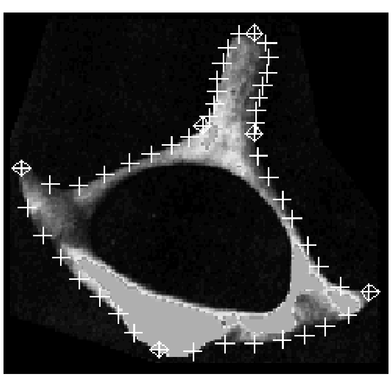


From Galileo (1638) illustrating the differences in shapes of the bones of small and large animals.

• Landmark: point of correspondence on each object that matches between and within populations.

Different types: anatomical (biological), mathematical, pseudo, quasi



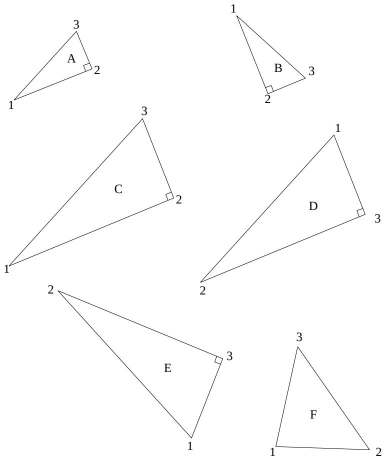


T2 mouse vertebra with six mathematical landmarks (line junctions) and 54 pseudo-landmarks.

• Bookstein (1991)

Type I landmarks (joins of tissues/bones)
Type II landmarks (local properties such as maximal curvatures)
Type III landmarks (extremal points or constructed landmarks)

• Labelled or un-labelled configurations



Six labelled triangles: A, B have the same size and shape; C has the same shape as A, B (but larger size); D has a different shape but its labels can be permuted to give the same shape as A, B, C; triangle E can be reflected to have the same shape as D; triangle F has a different shape from A,B,C,D,E.

Traditional methods

- ratios of distances between landmarks or angles submitted to multivariate analysis

- the full geometry usually if often lost

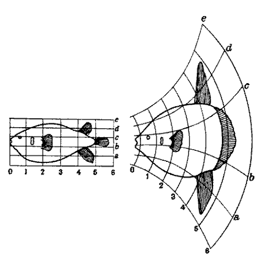
- collinear points?

- interpretation of shape differences in multivariate space?

Geometrical shape analysis

Rather than working with quantities derived from organisms one works with the complete geometrical object itself (up to similarity transformations).

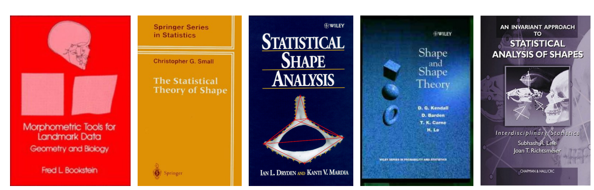
In the spirit of D'Arcy Thompson (1917) who considered the geometric transformations of one species to another

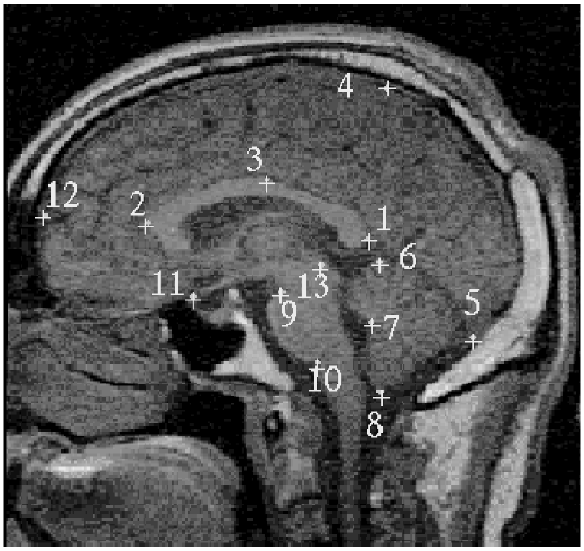


We conside a shape space obtained directly from the landmark coordinates, which retains the geometry of a point configuration at all stages.

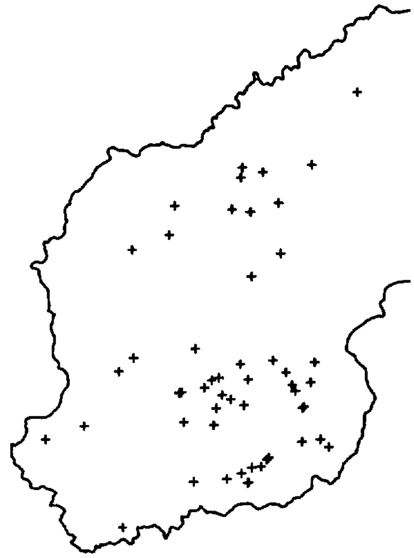
• Pioneers: Fred Bookstein and David Kendall

Summaries of the field are given by Bookstein (1991, Cambridge), Small (1996, Springer), Dryden and Mardia (1998, Wiley), Kendall et al (1999, Wiley), Lele and Richstmeier (2001, Chapman and Hall).

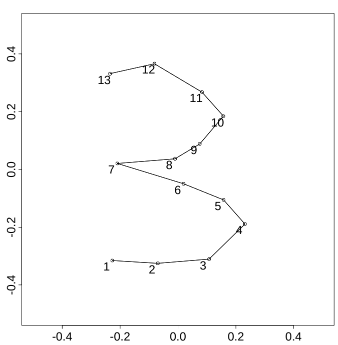




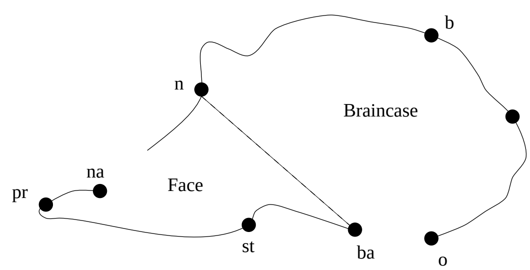
MR brain scan



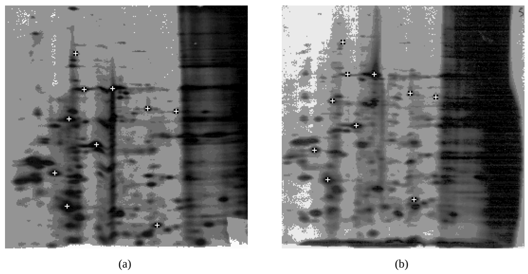
The map of 52 megalithic sites (+) that form the ‘Old Stones of Land’s End’ in Cornwall (from Stoyan et al., 1995).



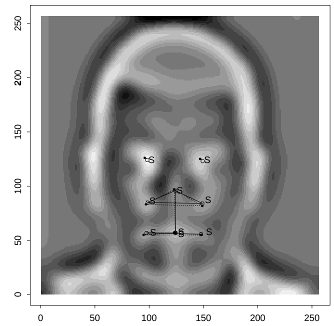
Handwritten digit 3



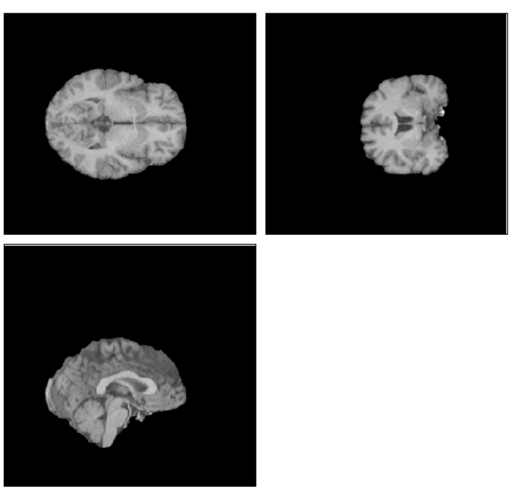
Ape cranium



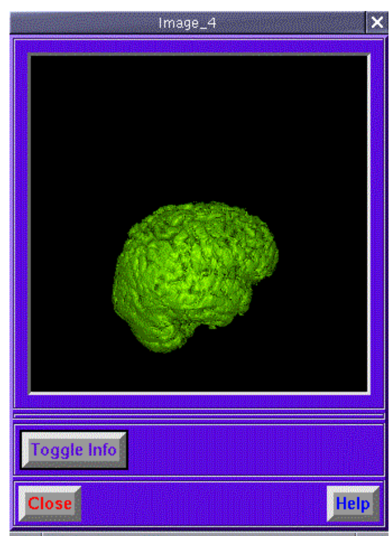
Electrophoretic gel matching



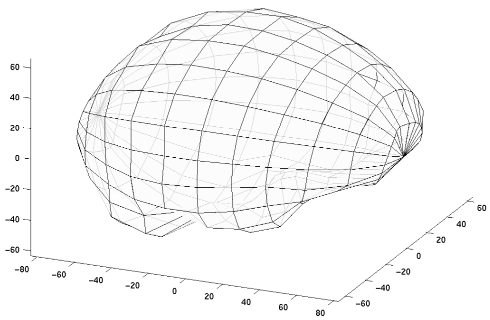
Face recognition



Proton density weighted MR image



Cortical surface extracted from MR scan



203 Pseudo-landmarks on the cortical surface of the brain

OUR FOCUS: k landmarks in m real dimensions

X is a $k \times m$ matrix ($M = \mathbb{R}^{km} \setminus$ coincidence set)

Invariance with respect to Euclidean similarity group (translation, scale and rotation) = $\{\mathbb{R}^m \times \mathbb{R}^+ \times SO(m)\}$

Size....

Any positive real valued function $g(X)$ such that $g(aX) = ag(X)$ for a positive scalar a.

• Centroid size:

$$S(X) = \|CX\| = \sqrt{\sum_{i=1}^{k} \sum_{j=1}^{m} (X_{ij} - \bar{X}_j)^2}$$

where $\bar{X}_j = \frac{1}{k}\sum_{i=1}^{k} X_{ij}$ and

$$C = I_k - \frac{1}{k} 1_k 1_k^T$$

$\|X\| = \sqrt{\text{trace}(X^T X)}$ - Euclidean norm,

I_k - $k \times k$ identity matrix, 1_k - $k \times 1$ vector of ones.

An alternative size measure is the **baseline size**, i.e. the length between landmarks 1 and 2:

$$D_{12}(X) = \|(X)_2 - (X)_1\|.$$

This was used as early as 1907 by Galton for normalizing faces.

Other size measures: square root of area, cube root of volume

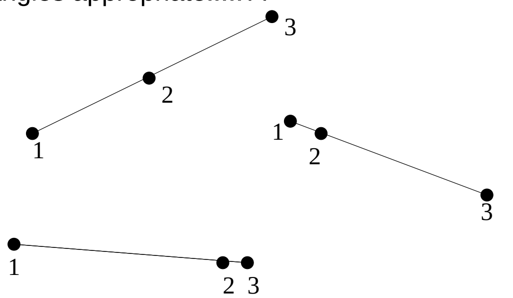
Shape coordinates:

Fixed coordinate system

vs

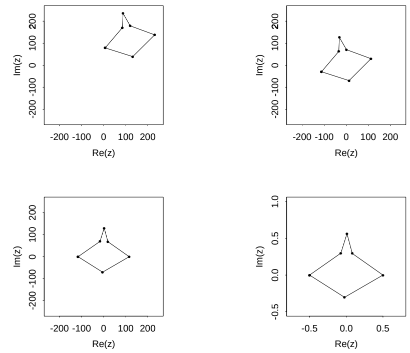
Local Coordinate system

Are angles appropriate.....??

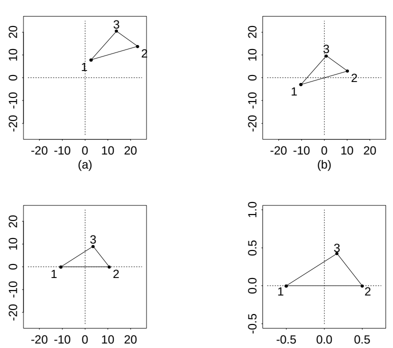


Landmarks: $x_1, x_2, \ldots, x_k \in \mathbb{C}$

- Bookstein shape coordinates (1984,1986) (For two dimensional data)



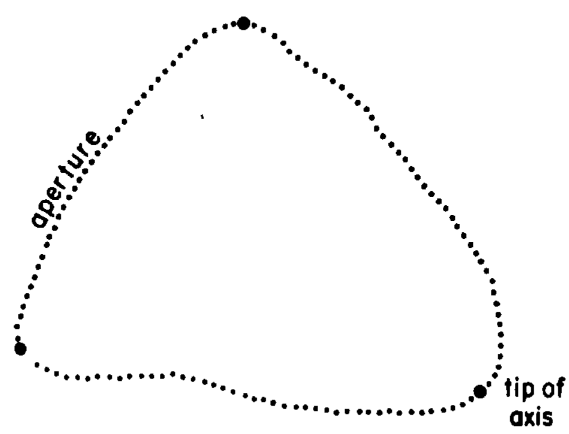
Shape: $u_j^B = \frac{x_j - x_1}{x_2 - x_1} - 0.5, \quad (j = 3, \ldots, k)$



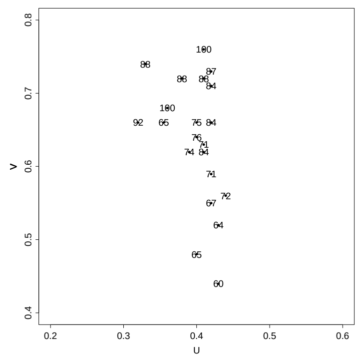
In real co-ordinates:

$$
\begin{aligned}
u_j^B &= -\frac{1}{2} + \{(x_2 - x_1)(x_j - x_1) + (y_2 - y_1)(y_j - y_1)\}/D_{12}^2, \\
v_j^B &= \{(x_2 - x_1)(y_j - y_1) - (y_2 - y_1)(x_j - x_1)\}/D_{12}^2,
\end{aligned}
$$

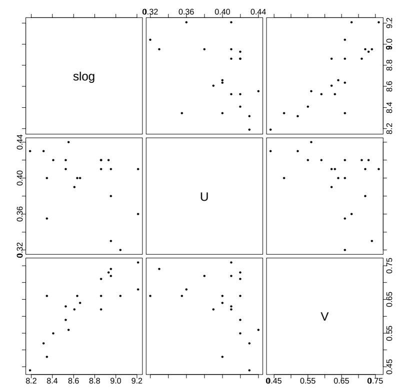
where $j = 3, ..., k$, $D_{12}^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2 > 0$ and $-\infty < u_j^B, v_j^B < \infty$.

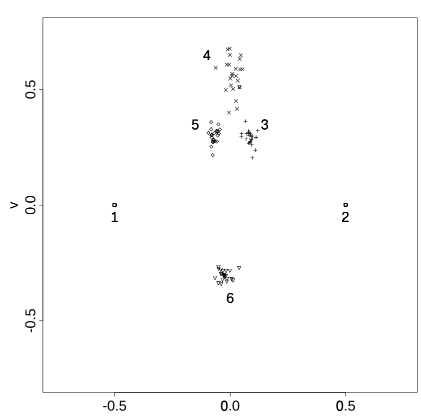


The outline of a microfossil with three landmarks (from Bookstein, 1986).

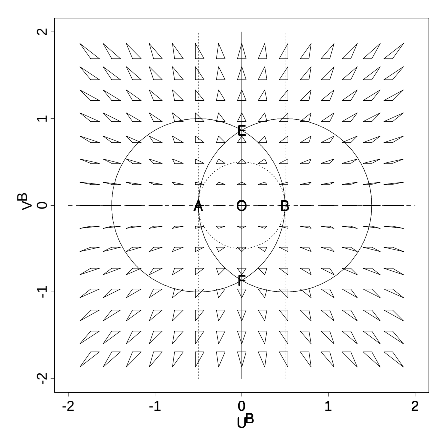


A scatter plot of (U+1/2) for the Bookstein shape variables for some microfossil data. (Bookstein, 1986)





A scatter plot of the Bookstein shape variables for the T2 mouse data.



The shape space of triangles, using Bookstein's coordinates (U^B, V^B). All triangles could be relabelled and reflected to lie in the shaded region.

Kendall's shape coordinates

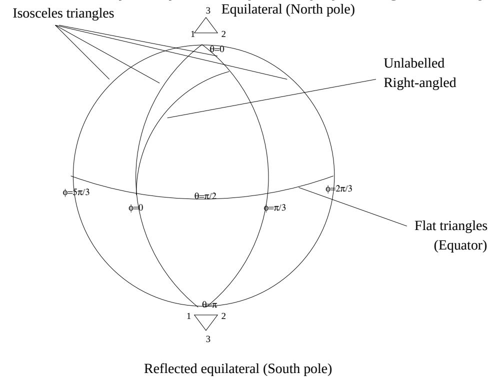
Remove location $z_H = Hz^o = (z_1, ..., z_{k-1})^T$

$$u_j^K + iv_j^K = \frac{z_{j-1}}{z_1} \quad (j = 3, \ldots, k).$$

Simple 1-1 linear correspondence with Booklstein S.V. (equ. 2.11 of book)

For triangles Kendall's SV sends baseline to $-1/\sqrt{3}, 1/\sqrt{3}$

• Kendall's shape sphere (1983) (triangles only)



A mapping from Kendall's shape variables to the sphere is

$$x = \frac{1 - r^2}{2(1 + r^2)} \;,\; y = \frac{u_3^K}{1 + r^2} \;,\; z = \frac{v_3^K}{1 + r^2}$$

and $r^2 = (u_3^K)^2 + (v_3^K)^2$, so that

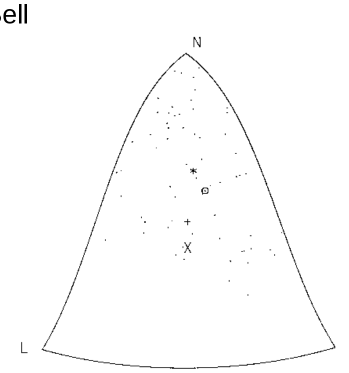
$x^2 + y^2 + z^2 = \frac{1}{4}$.

Kendall's spherical shape shape variables (θ, ϕ) are then given by the usual polar coordinates

$$x = \frac{1}{2}\sin\theta\cos\phi \;,\; y = \frac{1}{2}\sin\theta\sin\phi \;,\; z = \frac{1}{2}\cos\theta,$$

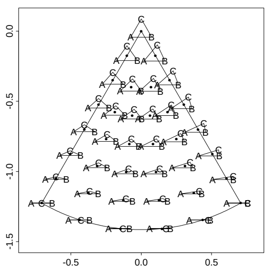
where $0 \leq \theta \leq \pi$ is the angle of latitude and $0 \leq \phi < 2\pi$ is the angle of longitude.

Kendall's Bell



The Schmidt net for 1/12 sphere

$$\xi = 2\sin\left(\frac{\theta}{2}\right) \ , \ \ \psi = \phi \ ; \ 0 \le \xi \le \sqrt{2}, 0 \le \psi < 2\pi.$$



Bookstein coordinates - 3D

Landmarks $X_i = (x_{1i}, x_{2i}, x_{3i})^T$

$$u_j = (u_{1j}, u_{2j}, u_{3j})^T$$

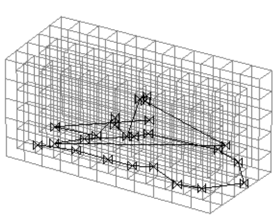
$$= \frac{1}{\|X_2 - X_1\|} A \left(X_j - \frac{(X_1 + X_2)}{2} \right) \ , \ \ j = 3, .., k$$

where A is a 3×3 rotation matrix
(a function of (X_1, X_2, X_3)) and

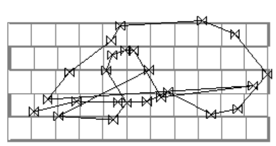
$$X_1 \to (-1/2, 0, 0)^T, \ X_2 \to (1/2, 0, 0)^T,$$

$$X_3 \to u_3 = (u_{13}, u_{23}, 0)^T$$

where $u_{23} \ge 0$, $u_{33} = 0$ and $X_j \to u_j$ for $j = 4, ..., k$.



(a)



(b)

Bookstein 3D coordinates

Goodall-Mardia QR shape coordinates $\ge$ 2 D

Helmertized landmarks $X_H = HX$ ($k \times m$ matrix)

SIZE AND SHAPE (JOINTLY)

$$X_H = T\Gamma \ , \ \ \Gamma \in SO(m),$$

T is lower triangular

SHAPE:

$W = T / \|T\|$

Shape coordinates

1. FILTER OUT TRANSLATION:

a) Shift centroid to origin
b) Take linear orthogonal contrasts, e.g. Helmert contrasts
c) Shift baseline midpoint to origin

2. RE-SCALE:

a) Re-scale to unit centroid size
b) Re-scale to unit area
c) Re-scale to a standard baseline length
d) Re-scale to minimize ‘distance’ to a template

3. REMOVE ROTATION:

a) Rotate baseline to horizontal
b) Rotate to minimize ‘distance’ to a template

Bookstein shape coordinates: 1c/2c/3a
Kendall shape coordinates: 1b/2c/3a
Procrustes shape coordinates: 1a/2d/3b

SHAPE SPACE....Kendall (1984)

1. Remove location (Pre-multiply by Helmert sub-matrix)

$$X_H = HX$$

where jth row of the Helmert sub-matrix H is given by,

$$(h_j, ..., h_j, -jh_j, 0, ..., 0)\ ,\ \ h_j = -\{j(j+1)\}^{-\frac{1}{2}}$$

and the h_j is repeated j times and zero is repeated $k-j-1$ times, $j = 1, ..., k-1$.

Note $C = H^T H$ (centering matrix) so $\|X_C\| = \|X_H\| = S(X)$. (centroid size)

2. Remove size (rescale)

$$Z = \frac{X_H}{S(X)} = \frac{HX}{\|HX\|}.$$

• Z is the PRESHAPE ($\in S^{(k-1)m-1}$)

3. Remove rotation

$$[X] = \{Z\Gamma : \Gamma \in SO(m)\},$$

• $[X]$ is the SHAPE of X.

• Dimensions....

Original configuration: $k \times m$

Centered configuration: $km - m$

Preshape: $km - m - 1$

Shape: $km - m - 1 - m(m-1)/2$

• Shape space is non-Euclidean

SHAPE SPACES

Assume $k \geq m+1$. [k points in m Euclidean dimensions]

$m=1$: Σ_1^k is a unit radius $(k-2)$-sphere.

$m=2$: Σ_2^k is the complex projective space $\mathbb{C}P^{k-2}$.

$m>2$: Σ_m^k has a singularity set $\pi(\mathcal{D}_{m-2})$ of dimension $m-2$ and is NOT a homogeneous space.

For $m>2$ the space spaces Σ_m^{m+1} are topological spheres.

Write $X = U[\Lambda, 0]V$, for the pseudo-singular value decomposition where $U \in SO(m-1), V \in SO(k-1)$, and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_m)$. Let

$$\chi_m^k = \begin{cases} \{X \in S_m^k : \lambda_1 > \ldots > \lambda_{m-1} > |\lambda_m|\} & \text{if } k = m+1 \\ \{X \in S_m^k : \lambda_1 > \ldots > \lambda_{m-1} > \lambda_m\} & \text{if } k > m+1 \end{cases}$$

Le and DG Kendall (1993, Annals of Statistics)

Theorem *On $\pi(\chi_m^k)$, the Riemannian metric can be expressed as*

$$d\rho^2 = \sum_{i=2}^m d\lambda_i^2 + \left(\sum_{i=2}^m \frac{\lambda_i}{\lambda_1} d\lambda_i\right)^2 + \sum_{1\leq i<j\leq m} \frac{(\lambda_i^2-\lambda_j^2)^2}{\lambda_i^2+\lambda_j^2}\phi_{ij}^2$$

$$+\sum_{i=1}^m \sum_{j=m+1}^{k-1} \lambda_i^2 \phi_{ij}^2,$$

where ϕ_{ij} are co-ordinates for $SO(k-1)$.

Planar case: $m=2$ dimensional data

$\Sigma_m^k = S_2^k/SO(2) = \mathbb{C}P^{k-2}$

Helmertized landmarks

$$z_H = Hz^o = (z_1, ..., z_{k-1})^T \in \mathbb{C}^{k-1} \setminus \{0\}$$

Now multiplying by

$\lambda = r\mathrm{e}^{i\omega}, \ \ (r \in \mathbb{R}^+, \omega \in [0, 2\pi))$

rotates and rescales z_H. So,

$$\{\lambda z_H : \lambda \in \mathbb{C} \setminus \{0\}\},$$

is the set representing the SHAPE of z^o. This is a complex line through the origin (but not including it) in $k-1$ dimensions. The union of all such sets is the complex projective space $\mathbb{C}P^{k-2}$

NB: $\mathbb{C}P^{k-2} \equiv S^2$

PLANAR CASE: Procrustes/Riemannian distance

Complex configurations $z^o = (z_1^o, ..., z_k^o)^T$,

$w^o = (w_1^o, ..., w_k^o)^T$

with centroids z_c, w_c.

Shape distance $\rho(z^o, w^o)$ satisfies

$$\cos\rho(z^o, w^o) \ = \ \frac{|\sum_{i=1}^k (z_i^o - z_c)(\overline{w}_i^o - \overline{w}_c)|}{\sqrt{\sum \|z_i^o - z_c\|^2}\sqrt{\sum \|w_i^o - w_c\|^2}}$$

where $\overline{w}_i^o$ means the complex conjugate of w_i^o.

NB $\cos\rho$ is the modulus of the complex correlation between z^o and w^o.

• $k=3$: ρ is the great circle distance on $S^2(1/2)$.

Complex configurations $z^o = (z_1^o, ..., z_k^o)^T$

Bookstein co-ordinates:

$$w_j^B = \frac{z_j^o - z_1^o}{z_2^o - z_1^o} - 0.5, \quad (j = 3, \ldots, k)$$

Kendall co-ordinates:

$$w_j^K = z_{j-1}/z_j, \quad (j = 3, \ldots, k)$$

where $(z_1, \ldots, z_{k-1})^T = Hz^o$

- Linear relationship:

$$\mathbf{w}^K = \sqrt{2} H_1 \mathbf{w}^B$$

where H_1 is lower right $(k-2) \times (k-2)$ partition of H.

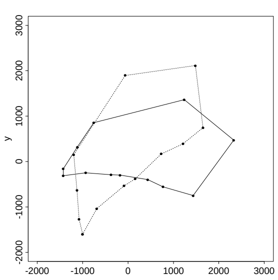
For $k = 3$: $w_3^B = (\sqrt{3}/2) w_3^K$.

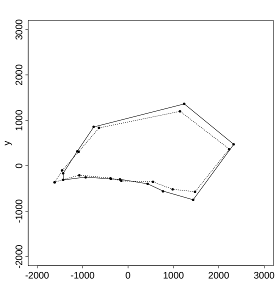
Session II

- Procrustes analysis
- Tangent coordinates
- Shape variability
- Shape models
- Tangent space inference
- Shapes package.

PROCRUSTES ANALYSIS

Juvenile (———) Adult (- - - - - -)





Register adult onto juvenile

PLANAR PROCRUSTES ANALYSIS

Two centred configurations $y = (y_1, ..., y_k)^\top$ and $w = (w_1, ..., w_k)^\top$, both in $\mathbb{C}^k$, with $y^* 1_k = 0 = w^* 1_k$,

[y^* - transpose of the complex conjugate of y]

Match w onto y using complex linear regression

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

$X_D = [1_k, w]$ - 'design' matrix
$A = (a + ib, \beta e^{i\theta})^\top$ - similarity transformation parameters

Procrustes match = least squares

Minimize the sum of square errors

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

Full Procrustes fit (superimposition) of w on y

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

where

$\hat{A} = (X_D^* X_D)^{-1} X_D^* y,$
i.e.

$$\begin{aligned} \hat{a} + i\hat{b} &= 0, \\ \hat{\theta} &= \arg(w^* y) = -\arg(y^* w), \\ \hat{\beta} &= (w^* y y^* w)^{1/2}/(w^* w). \end{aligned}$$

Procrustes fit $w^P = w^* y w/(w^* w)$

Procrustes residual vector $r = y - w^P$

Minimized objective function
$D^2(r, 0) = y^* y - (y^* w w^* y)/(w^* w)$
(not symmetric unless $y^* y = w^* w$)

Initially standardize to unit centroid size....

Full Procrustes distance:

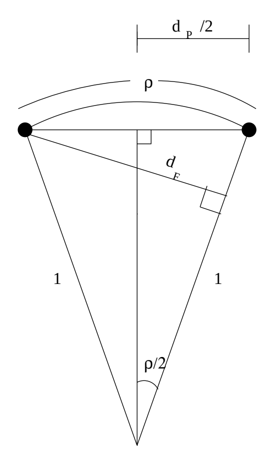
$$\begin{aligned} d_F(w,y) &= \inf_{\beta,\theta,a,b} \left\| \frac{y}{\|y\|} - \frac{w}{\|w\|} \beta e^{i\theta} - a - ib \right\| \\ &= \left\{ 1 - \frac{y^* w w^* y}{w^* w y^* y} \right\}^{1/2}. \end{aligned}$$

FULL Procrustes distance d_F - full set of similarity transformations used in matching

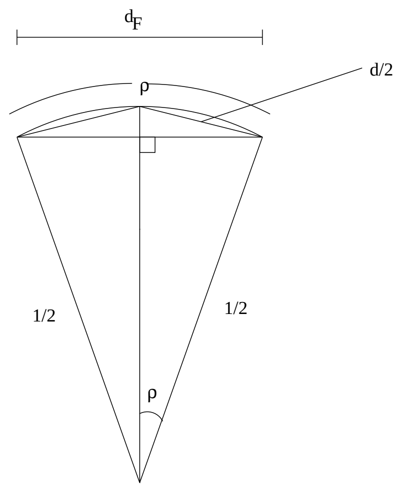
PARTIAL Procrustes distance d_P - matching over translation and rotation ONLY

For fairly similar shapes they are very similar,
as $d_F = d_P + O(d_P^3) = \rho + O(\rho^3)$

In this course for simplicity we shall concentrate on FULL Procrustes matching.

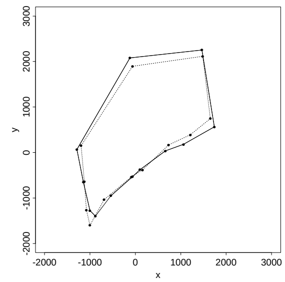


Section of the pre-shape sphere



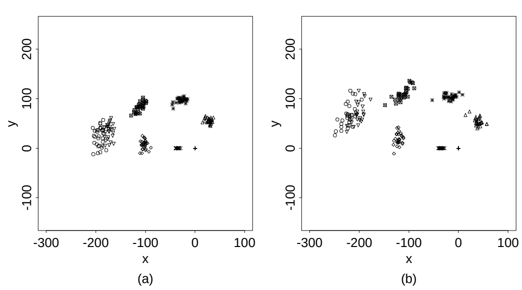
Section of the SHAPE SPHERE FOR TRIANGLES, illustrating the relationship between d_F, d_P and ρ

Procrustes residuals from the match of w onto y are different from y onto w



JUV to ADULT (above): $\hat{\theta} = 45.5^{\rm o}$, $\hat{\beta} = 1.131$.
ADULT to JUV: $\hat{\theta}^R = -45.5^{\rm o}$, $\hat{\beta}^R = 0.875 \neq 1/1.131$

Female (left) and Male (right) gorilla skulls



Mean shape? Shape variance/covariance?

CONFIGURATION MODEL

Random sample of n configurations $w_1, \ldots, w_n$ from the perturbation model

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

where $\gamma_i \in \mathbb{C}$ - translations
$\beta_i \in \mathbb{R}^+$ - scales
$0 \leq \theta_i < 2\pi$ - rotations
$\epsilon_i \in \mathbb{C}$ are independent zero mean complex random errors
μ is the population mean configuration.

AIM: to estimate $[\mu]$ - the shape of μ

Procrustes mean:

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

Consider w_i to be centred: $w_i^T 1_k = 0$.

(Kent, 1994) **Procrustes mean shape** $[\widehat{\mu}]$ is the dominant eigenvector of

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

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

Proof We wish to minimize

$$\begin{aligned} \sum_{i=1}^{n} d_F^2(w_i, \mu) &= \sum_{i=1}^{n} \left\{ 1 - \frac{\mu^* w_i w_i^* \mu}{w_i^* w_i \mu^* \mu} \right\} \\ &= n - \mu^* S \mu / (\mu^* \mu). \end{aligned}$$

Therefore,

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

Hence, result follows.

- Procrustes fits: match w_i to $\widehat{\mu}$

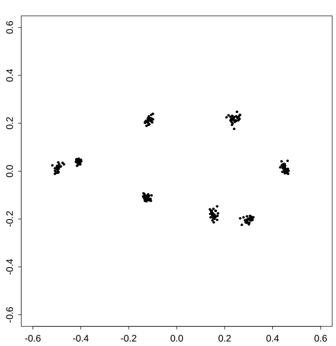
$$w_i^P = w_i^* \widehat{\mu} w_i / (w_i^* w_i), \; i = 1, \ldots, n,$$

NB Arithmetic mean: $\frac{1}{n} \sum_{i=1}^{n} w_i^P$ has same shape as $\widehat{\mu}$.

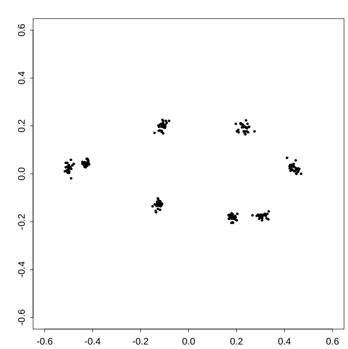
- Procrustes residuals

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

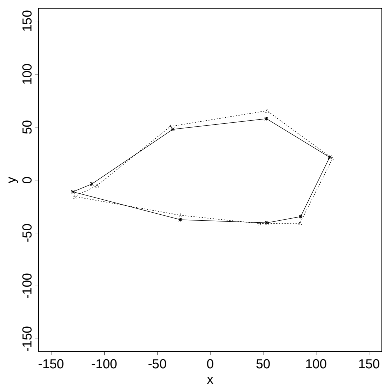
Procrustes fits (Generalized Procrustes analysis)



Female gorillas



Male Gorillas

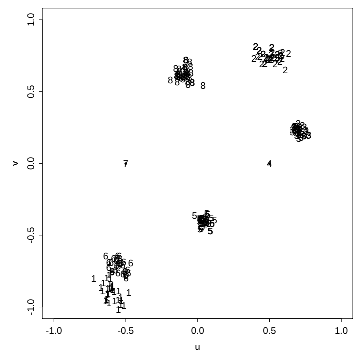


The male (—-) and female (- - -) full Procrustes mean shapes registered by GPA.

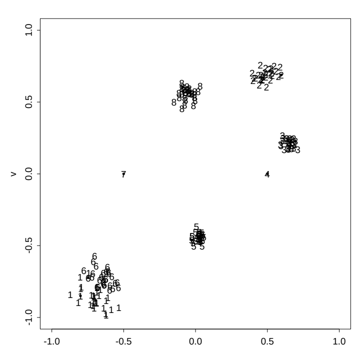
Other mean shape estimates:

• Bookstein mean shape

Take sample mean of Bookstein coordinates U^B



Female Gorillas



Male Gorillas

[In Book chapter 12]

• MDS mean shape (Kent, 1994; Lele 1991)

Obtain average squared Euclidean distance matrix D

let $B = -\frac{1}{2}CDC$ (centred inner product matrix)

Let $f_1, \ldots, f_p$ be the scaled eigenvectors

$MDS_m(D) = [f_1, f_2, \ldots, f_m]$

(invariant under reflections too)

• IMPORTANT: If shape variations small the mean shape estimates are approximately linearly related. i.e. Multivariate normal based inference will be equivalent to first order. (Kent, 1994)

Tangent coordinates

Consider complex landmarks $z^o = (z^o_1, ..., z^o_k)^\top$ with pre-shape

$$z = (z_1, ..., z_{k-1})^\top = Hz^o/\|Hz^o\|.$$

Let γ be a complex pole on the complex pre-shape sphere usually chosen as an average shape.

Let us rotate the configuration by an angle θ to be as close as possible to the pole and then project onto the **tangent plane** at γ, denoted by $T(\gamma)$. Note that $\hat{\theta} = \arg(-\gamma^* z)$ minimizes $\|\gamma - z e^{i\theta}\|^2$.

The **partial Procrustes tangent coordinates** for a planar shape are given by

$$v = e^{i\hat{\theta}}[I_{k-1} - \gamma\gamma^*]z, \; v \in T(\gamma), \qquad (1)$$

where $\hat{\theta} = \arg(-\gamma^* z)$. Partial Procrustes tangent coordinates involve only rotation (and not scaling) to match the pre-shapes.

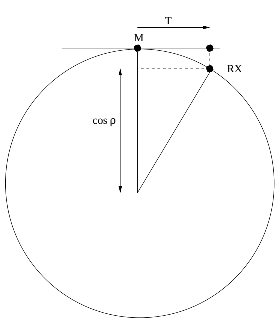
Note that $v^*\gamma = 0$ and so the complex constraint means we can regard the tangent space as a real subspace of $\mathbb{R}^{2k-2}$ of dimension $2k-4$. The matrix $I_{k-1} - \gamma\gamma^*$ is the matrix for complex projection into the space orthogonal to γ. Below we see a section of the shape sphere showing the tangent plane coordinates.

PROCRUSTES TANGENT SPACE

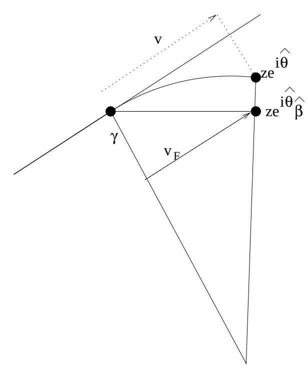
Procrustes tangent co-ordinates T of X at the pole M:

$$T = RX - \cos\rho M$$

where $0 < \rho \leq \pi/2$ is the Riemannian distance between the shapes of M and X, and R is the optimal Procrustes rotation to match X to M.



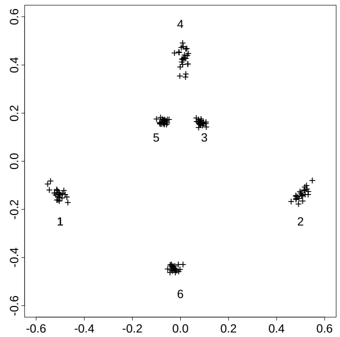
The rays from the origin in Procrustes tangent space correspond to minimal geodesics in shape space.



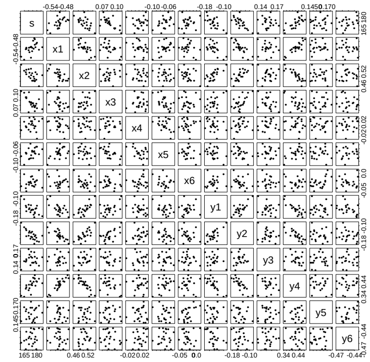
A diagrammatic view of a section of the pre-shape sphere, showing the partial tangent plane coordinates v and the full Procrustes tangent plane coordinates v_F. Note that the inverse projection from v to $ze^{i\hat{\theta}}$ is given by

$$ze^{i\hat{\theta}} = [(1 - v^*v)^{1/2}\gamma + v] \;\; , \;\; z \in \mathbb{C}S^{k-2}. \qquad (2)$$

Hence an icon for partial Procrustes tangent coordinates is given by $X_I = H^\top z$.



Icons for partial Procrustes tangent coordinates for the T2 vertebral data (Small group).



Pairwise scatter plots for centroid size (S) and the (x, y) coordinates of icons for the partial Procrustes tangent coordinates for the T2 vertebral data (Small group).

The Euclidean norm of a point v in the partial Procrustes tangent space is equal to the full Procrustes distance from the original configuration z^o corresponding to v to an icon of the pole $H^\top \gamma$, i.e.

$$\|v\| = d_F(z^o, H^\top \gamma).$$

Important point: This result means that standard multivariate methods in tangent space which involve calculating distances to the pole γ will be equivalent to non-Euclidean shape methods which require the full Procrustes distance to the icon $H^\top \gamma$. Also, if X_1 and X_2 are close in shape, and v_1 and v_2 are the tangent plane coordinates, then

$$\|v_1 - v_2\| \approx d_F(X_1, X_2) \approx \rho(X_1, X_2) \approx d_P(X_1, X_2). \tag{3}$$

For practical purposes this means that standard multivariate statistical techniques in tangent space will be good approximations to non-Euclidean shape methods, provided the data are not too highly dispersed.

Full Procrustes tangent coordinates

An alternative tangent space is obtained by allowing scaling by $\beta > 0$ of the pre-shape z in the matching to the pole γ. In the above section

Shape variability

• Overall measure

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

$RMS(d_F)_{FEMALE} = 0.044$

$RMS(d_F)_{MALE} = 0.050$

• PCA in tangent space to shape space

- PCA of Procrustes residuals $r_i = w_i^P - \widehat{\mu}$
- PCA of Procrustes tangent coordinates v_i
(project r_i so to obtain part that is orthogonal to $\widehat{\mu}$ and its rotations)
- NB for observations close to $\widehat{\mu}$ we have $r_i \approx v_i$

• S_v - sample covariance matrix of some tangent coordinates v_i,

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

where $\bar{v} = \frac{1}{n} \sum v_i$.

γ_j - eigenvectors of S_v: **principal components** (PCs), with eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p \geq 0$

• PC score for the ith individual on the jth PC is:

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

• PC summary of the data in the tangent space is

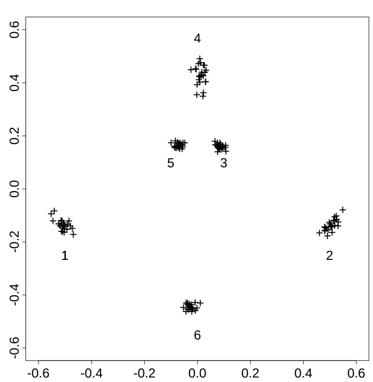
$$v_i = \bar{v} + \sum_{j=1}^{p} s_{ij} \gamma_j,$$

for $i = 1, \ldots, n$.

• Standardized PC scores:

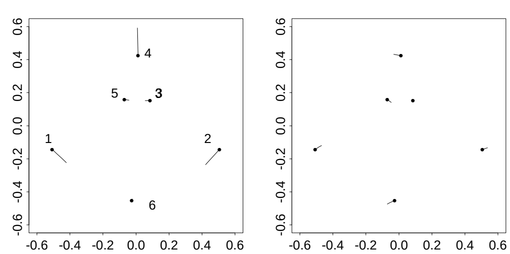
$$c_{ij} = s_{ij} / \lambda_j^{1/2}, \; i = 1, \ldots, n; j = 1, \ldots, p.$$

Mouse vertebra example:



Mouse vertebra example: (PC1 = 69%)

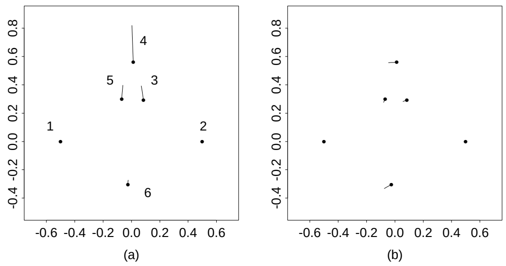
Procrustes registration for display



(a) (b)

Mouse vertebra example: (PC1 = 69%)

Bookstein registration for display

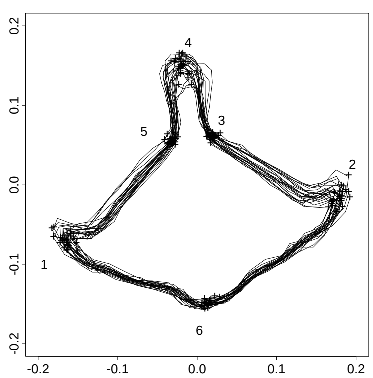


- Important:

If using Bookstein superimposition to calcuate S_v then strong correlations can be induced.....can lead to misleading PCs

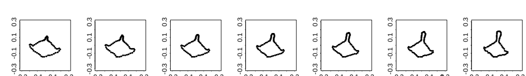
No problem with Procrustes registration, Kent and Mardia (1997)

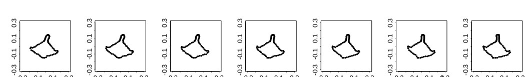
T2 small vertebra outlines



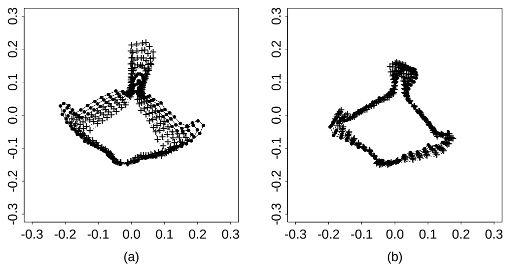
$RMS(d_F) = 0.07$

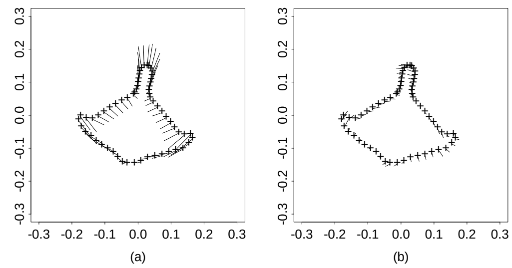
PC1: 65%

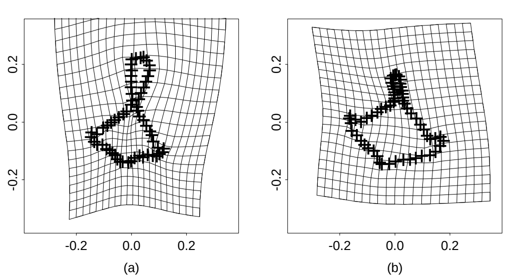




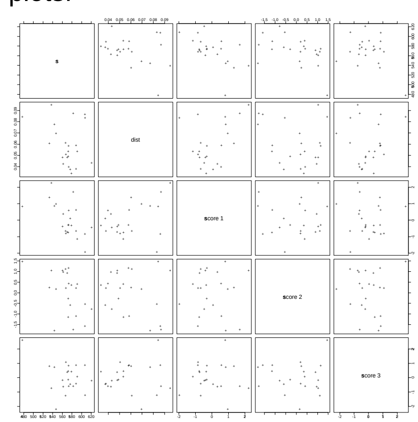
PC2: 9%





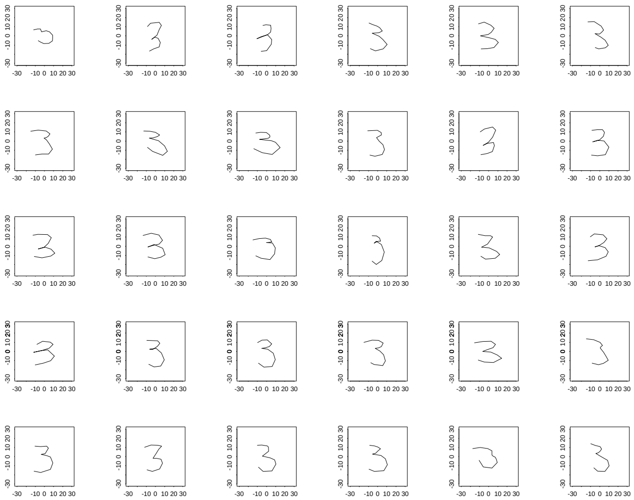


Pairwise plots:



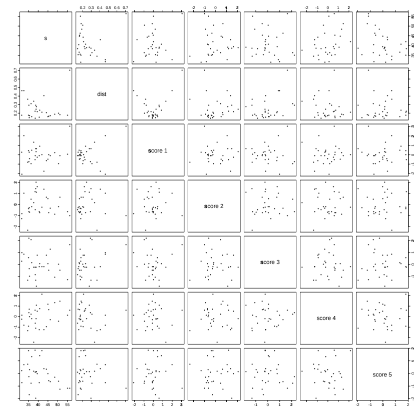
Size, shape distance, PC scores 1, 2, 3

Digit 3 data

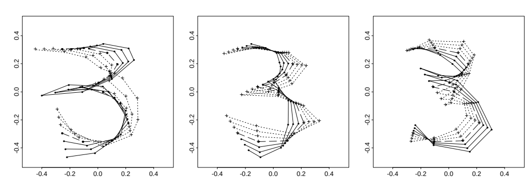


Pairwise plots:

Size, shape distance, PC1: 50%, PC2: 15%, PC3: 13%, PC4: 8%, PC5: 4%



$RMS(d_F) = 0.28$



HIGHER DIMENSIONS

Ordinary Procrustes analysis (match X_1 to X_2 - centred)....Minimize:

$$D^2_{OPA}(X_1, X_2) = \|X_2 - \beta X_1 \Gamma - 1_k \gamma^\top\|^2,$$

Solution:

$$\hat{\gamma} = 0$$

$$\hat{\Gamma} = UV^\top$$

where

$$X_2^\top X_1 = \|X_1\| \, \|X_2\| V \Lambda U^\top, \quad U, V \in SO(m)$$

with Λ a diagonal $m \times m$ matrix. Furthermore,

$$\hat{\beta} = \frac{\text{trace}(X_2^\top X_1 \hat{\Gamma})}{\text{trace}(X_1^\top X_1)},$$

The minimized sum of squares is:

$$OSS(X_1, X_2) = \|X_2\|^2 d_F(X_1, X_2)^2$$

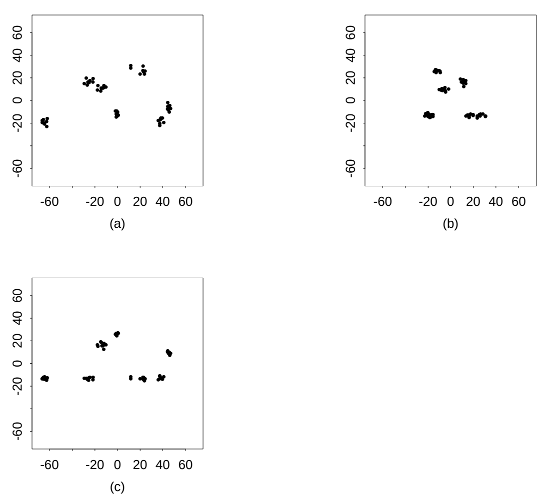
PERTURBATION MODEL:

$$X_i = \beta_i(\mu + E_i)\Gamma_i + 1_k \gamma_i^\top$$

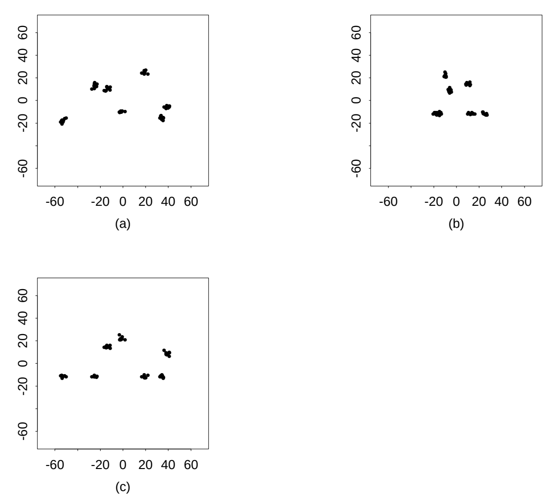
Can estimate the shape of μ by GPA (generalized Procrustes analysis): by minimizing

$$\sum_{i=1}^{n} d_F(X_i, \mu)^2$$

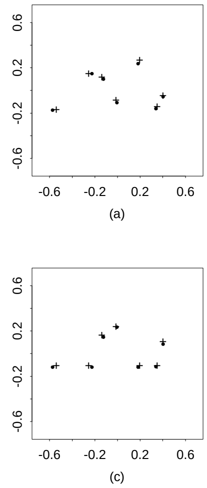
Least squares approach. Iterative algorithm needed for $m > 2$ dimensions



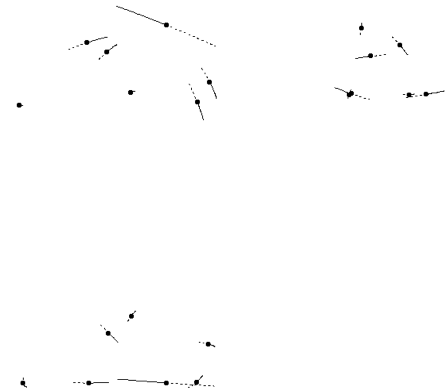
Male macaques



Female macaques

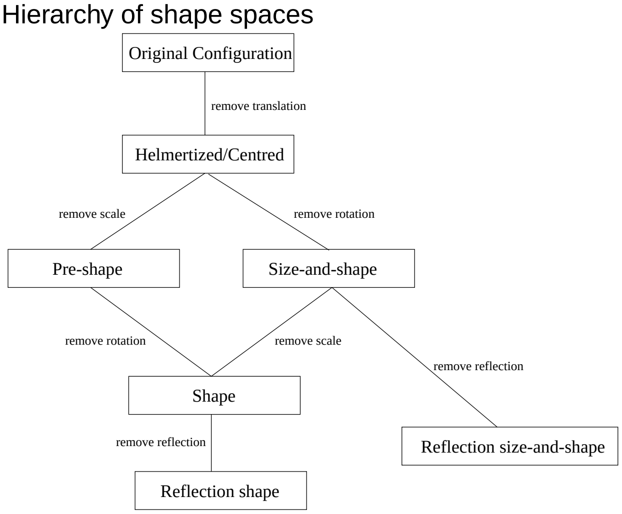


Male (·) Female (+)



PC1 (47%) for Males: +/- 9 s.d.

Hierarchy of shape spaces



Different approaches to inference:

1. Marginal/offset distributions
2. Conditional distributions
3. Directly specified in shape space
4. Distributions in a tangent space
5. Structural models in the tangent space

Preshape distributions (2D)

2D - complex notation: $z = (z_1, \ldots, z_k)^T$ where $1^T z = 0, z^* z = 1$ $[z^* = (\bar{z})^T]$

• complex Bingham (Kent, 1994)

$$f(z) = c(A)\,\mathrm{exp}(z^* A z)$$

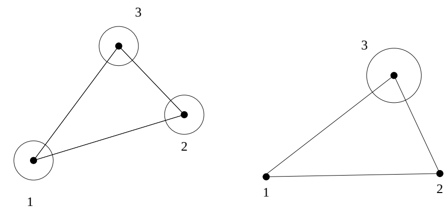
A is Hermitian. NB: $f(z) = f(e^{i\theta} z)$ so suitable for shape analysis.

NB: MLE of modal shape is identical to the PROCRUSTES (least squares) mean

• complex Watson (special case of c. Bingham)

$$f(z) = c(\kappa)\,\mathrm{exp}(\kappa z^* \mu \mu^* z)$$

Shape distributions: offset normal approach



(a) (b)

Mean triangle μ with independent isotropic zero mean normal perturbations with variance σ^2.

Offset normal density (wrt uniform measure) (Mardia and Dryden, 1989; Dryden and Mardia, 1991, 1992)

$$\mathcal{L}_{k-2}(-\kappa(1+\cos 2\rho(X,\mu))\mathrm{exp}(-\kappa(1-\cos 2\rho(X,\mu)))$$

where $\kappa = Size(\mu)^2/(4\sigma^2)$,
$Size(\mu)^2 = \sum |\mu_i - \bar{\mu}|^2$ and
$\mathcal{L}_j(-x) = \sum_{i=0}^{j} \binom{j}{i} \frac{x^i}{i!}$ is the Laguerre polynomial.

Parameters:

$Shape(\mu)$: 2k-4 mean shape parameters
κ: concentration parameter.

DIFFUSIONS AND DISTRIBUTIONS

Diffusion of points in Euclidean shape (WS Kendall):

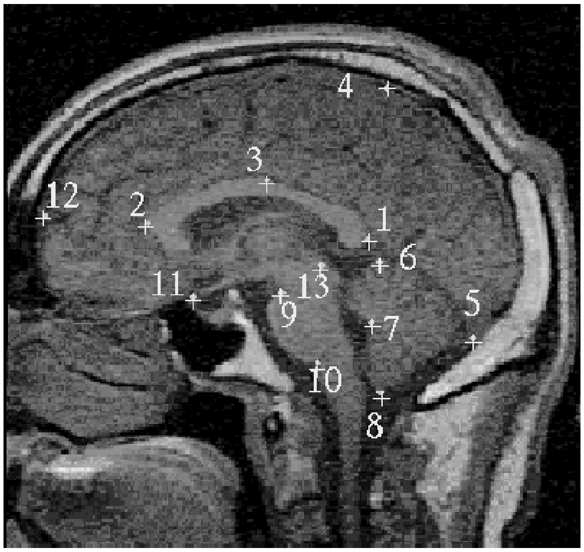
$$dX_i = dB_i - \frac{\kappa}{2} X_i dt \;\;,\;\; i = 1, \ldots, t.$$

Ornstein-Uhlenbeck process for Euclidean points $\rightarrow$ independent size and shape diffusions [with random time change for shape: $d\tau = dt/(\mathrm{size})^2$]. Computer algebra package `Itovsn3` developed through this work.

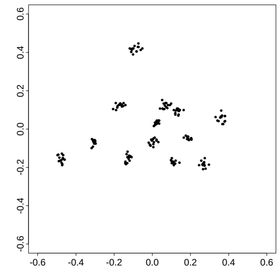
Size and shape, and shape diffusions in Σ_m^k (Le).

Shape density at time t: (from previous slide).

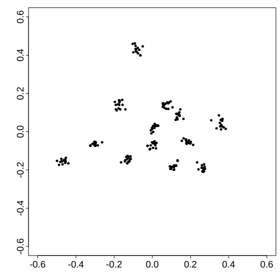
Maximum likelihood based inference



Controls:

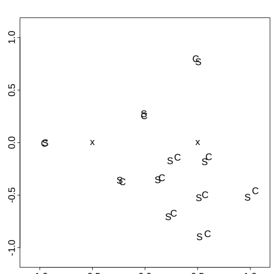


Schizophrenia patients:



Schizophrenia study (Bookstein, 1996;
Dryden and Mardia, 1998)
$k = 13$ landmarks in 2D: $n_1 = 14$ Controls and $n_2 = 14$ Schizophrenia patients

Isotropic offset normal model: independent individuals
Inference: maximum likelihood



LR test $P(\chi^2_{22} > 43.124) = 0.005$.
Monte Carlo permutation test, p-value: 0.038.

INFERENCE: Multivariate normal model in the tangent space (to pooled mean)

Hotelling's T^2 test

$$v_i \sim N(\xi_1, \Sigma) \ \ , \ \ w_j \sim N(\xi_2, \Sigma),$$

$i = 1, \ldots, n_1; j = 1, \ldots, n_2$, all mutually independent and common covariance matrices

$\bar{v}, \bar{w}$ - sample means
S_v, S_w - sample covariance matrices

Mahalanobis distance squared:

$$D^2 = (\bar{v} - \bar{w})^\top S_u^- (\bar{v} - \bar{w}),$$

where $S_u = (n_1 S_v + n_2 S_w)/(n_1 + n_2 - 2)$

Under H_0 equal mean shapes...

$$F = \frac{n_1 n_2 (n_1 + n_2 - M - 1)}{(n_1 + n_2)(n_1 + n_2 - 2) M} D^2$$

$$\sim F_{M, n_1 + n_2 - M - 1}$$

under H_0. [M = dimension of the shape space]

Gorilla (female/male):

$k = 8$ landmarks in $m = 2$ dimensions

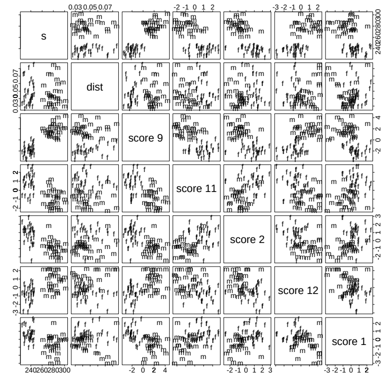
$n_1 = 30, n_2 = 29$

$M = 2k - 4 = 12$

The test statistic is $F = 26.47$
and $P(F_{12,46} > 4.47) = 0.0001$

Pairwise plots:

Size, shape distance, PC scores in direction of mean difference



Goodall's F test:

If $\Sigma \propto I$ then

$$F = \frac{\frac{n_1+n_2-2}{n_1{}^{-1}+n_2{}^{-1}} d_F^2(\hat{\mu}_1, \hat{\mu}_2)}{\sum_{i=1}^{n_1} d_F^2(X_i, \hat{\mu}_1) + \sum_{i=1}^{n_2} d_F^2(Y_i, \hat{\mu}_2)}$$

Under H_0: $F \sim F_{M,(n_1+n_2-2)M}$

• Schizophrenia data:

$k = 13$ landmarks in $m = 2$ dimensions

$n_1 = 14, n_2 = 14$

$M = 2k - 4 = 22$

$F = 1.89$, and $P(F_{22,572} > 1.89) \approx 0.01$

Permutation test: p-value = 0.04

• Hotelling's T^2 test

p-value = 0.66

Comparing several groups: ANOVA

Balanced analysis of variance with independent random samples $(X_{j1}, ..., X_{jn})^T, \ j = 1, ..., n_G$ from n_G groups, each of size n.
Let $\hat{\mu}_j$ be the group full Procrustes means and $\hat{\mu}$ is the overall pooled full Procrustes mean shape. A suitable test statistic is

$$F = n(n-1)n_G \ \frac{\sum_{i=1}^{n_G} d_F^2(\hat{\mu}_i, \hat{\mu})}{(n_G - 1)\sum_{j=1}^{n_G}\sum_{i=1}^{n} d_F^2(X_{ji}, \hat{\mu}_j)}.$$

Under H_0 : equal mean shapes:

$$F \sim F_{(n_G-1)M, n_G(n-1)M}$$

and reject H_0 for large values of the statistic.

Complex Watson inference:

Two independent random samples $z_1, \ldots, z_n$ from $\mathbb{C}W(\mu, \kappa)$ and $y_1, \ldots, y_m$ from $\mathbb{C}W(\nu, \kappa)$. We wish to test between

$$H_0 : [\mu] = [\nu] \text{ and } H_1 : [\mu] \neq [\nu],$$

where $[\mu] = \{e^{i\alpha}\mu : 0 \leq \alpha < 2\pi\}$, (i.e. $[\mu]$ represents the shape corresponding to the modal preshape μ. For large κ it follows that

$$\sum_{i=1}^{n} \sin^2 \rho(z_i, \mu) + \sum_{j=1}^{m} \sin^2 \rho(y_j, \nu) \approx \frac{1}{2\kappa}\chi^2_{(2k-4)(n+m)}$$

and we also have

$$\sum_{i=1}^{n} \sin^2 \rho(z_i, \hat{\mu}) + \sum_{j=1}^{m} \sin^2 \rho(y_j, \hat{\nu}) \approx \frac{1}{2\kappa}\chi^2_{(2k-4)(n+m-2)}$$

By analogy with analysis of variance we can write

$$\sum_{i=1}^{n} \sin^2 \rho(z_i, \hat{\hat{\mu}}) + \sum_{j=1}^{m} \sin^2 \rho(y_j, \hat{\hat{\mu}}) =$$

$$\sum_{i=1}^{n} \sin^2 \rho(z_i, \hat{\mu}) + \sum_{j=1}^{m} \sin^2 \rho(y_j, \hat{\nu}) + B$$

where $\hat{\hat{\mu}}$ is the overall MLE of μ if the two groups are pooled, and B is analogous to the between sum of squares. Since,

$$\sum_{i=1}^{n} \sin^2 \rho(z_i, \hat{\hat{\mu}}) + \sum_{j=1}^{m} \sin^2 \rho(y_j, \hat{\hat{\mu}}) \approx \frac{1}{2\kappa}\chi^2_{(2k-4)(n+m-1)}$$

it follows that

$$B = \sum_{i=1}^{n} \sin^2 \rho(z_i, \hat{\hat{\mu}}) + \sum_{j=1}^{m} \sin^2 \rho(y_j, \hat{\hat{\mu}}) -$$

$$\sum_{i=1}^{n} \sin^2 \rho(z_i, \hat{\mu}) - \sum_{j=1}^{m} \sin^2 \rho(y_j, \hat{\nu}) \approx \frac{1}{2\kappa}\chi^2_{2k-4}.$$

Therefore, under H_0 we have

$$F_2 = \frac{(n+m-2)B}{\sum_{i=1}^{n} \sin^2 \rho(z_i, \hat{\mu}) + \sum_{j=1}^{m} \sin^2 \rho(y_j, \hat{\nu})}$$

$$\approx F_{2k-4,(2k-4)(n+m-2)}$$

and so we reject H_0 for large values of F_2. Using Taylor series expansions for large concentrations

$$B \approx (n^{-1} + m^{-1})^{-1} \sin^2 \rho(\hat{\mu}, \hat{\nu}),$$

and so for large κ the test statistic F_2 is equivalent to the two sample test statistic of Goodall (1991).

Bayesian approach to inference

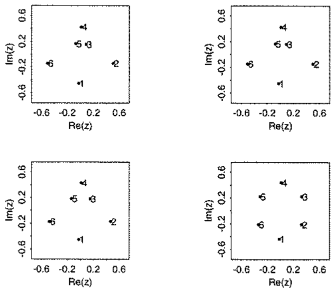
$$\pi(\Theta, \Sigma | u_1, \ldots, u_n) =$$

$$\frac{L(u_1, \ldots, u_n; \Theta, \Sigma)\pi(\Theta, \Sigma)}{\int L(u_1, \ldots, u_n; \Theta, \Sigma)\pi(\Theta, \Sigma) \mathrm{d}\Theta \mathrm{d}\Sigma} \quad .$$

e.g. Data $z_i \sim$ complex Watson(μ, κ known)
Prior $\mu \sim$ complex Bingham (A known)

$$\begin{aligned} \pi(\mu | z_1, \ldots, z_n) &\propto \pi(\mu) L(z_1, \ldots, z_n) \\ &\propto \exp\left\{ \mu^* A \mu + \kappa \sum_{i=1}^{n} z_i^* \mu \mu^* z_i \right\} \\ &= \exp\{\mu^*(\kappa S + A)\mu\}, \end{aligned}$$

Conjugate prior
MAP: dominant eigenvector of $S + A/\kappa$



The smoothed Procrustes mean of the T2 Small data: (Top left) $\lambda = 0$, (top right) $\lambda = 0.1$, (bottom left) $\lambda = 1.0$, (bottom right) $\lambda = 100$.

EDMA (Euclidean distance matrix analysis) [Lele, 91+, Stoyan, 91]

$F(X)$: **form distance matrix** ($k \times k$ matrix of pairs of inter-landmark distances (ILDs))

Estimate $F(\mu)$: population form distance matrix

$(x_j, y_j) \sim N((\mu_j, \nu_j), \sigma^2 I_2), j = 1, \ldots, k$. Then

$$(x_r - x_s)^2 + (y_r - y_s)^2 = D_{rs}^2 \sim \sigma^2 \chi_2^2(\delta_{rs}^2/\sigma^2), \quad (4)$$

$\delta_{rs}^2 = (\mu_r - \mu_s)^2 + (\nu_r - \nu_s)^2$.

Moment estimator

$$\delta_{rs}^4 = \{E(D_{rs}^2)\}^2 - \mathrm{var}(D_{rs}^2).$$

Removes bias.

Estimate of mean reflection size-and-shape

$$[MDS_2(\Delta)] \ , \ (\Delta)_{rs} = \hat{\delta}_{rs},$$

EDMA-I test (Lele and Richtsmeier, 1991)

Form ratio distance matrix

$$D_{ij}(X, Y) = F_{ij}(X)/F_{ij}(Y). \quad (5)$$

Test statistic:

$$T = \max_{i,j} D_{ij}(\hat{\mu}, \hat{\nu}) / \min_{i,j} D_{ij}(\hat{\mu}, \hat{\nu}), \quad (6)$$

Use bootstrap procedures.

EDMA-II (Lele and Cole, 1995)

$\hat{F}_\mu$ and $\hat{F}_\nu$ estimates of average form distance matrix for each group

Scale by group size measure

T = Largest entry in arithmetic difference of scaled matrices

More powerful than EDMA-I

Rao and Suryawanshi (1996)

$G(X)$: **form log-distance matrix**

shape log-distance matrix is

$$G^*(X) = G(X) - \bar{G}1_k 1_k^\top,$$

$$\bar{G} = \frac{2}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (G(X))_{ij}.$$

Average form log-distance matrix is

$$G(\hat{\mu}) = \frac{1}{n} \sum_{i=1}^{n} \log d_i(h_1, h_2),$$

where $d_i(h_1, h_2)$ is the distance between landmarks h_1 and h_2 for the ith object X_i.

Average reflection size-and-shape

$$[MDS_m(\exp(G(\hat{\mu})))].$$

Average reflection shape

$$[MDS_m(\exp(G^*(\hat{\mu})))].$$

For small variations estimates of mean shape or size-and-shape are all very similar...(Kent, 1994)

Distance based (+):

Landmarks not necessarily needed (eg. maximum breadth)

Consistent estimation under general normal models

Distance based (-):

Invariant under reflections

Visualization not straightforward

A choice of metric for averaging needs to be made

• SIZE-AND-SHAPE

Invariance under translation and rotation (not scale)

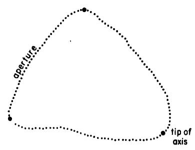
Perturbation model:

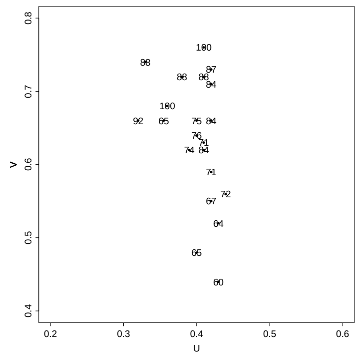
$$X_i = (\mu + E_i)\Gamma_i + 1_k \gamma_i^\top, \quad i = 1, \ldots, n,$$

• ALLOMETRY

The relationship of shape given size

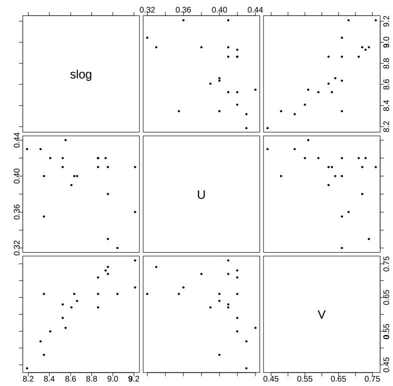
Bookstein's (1996) Microfossils





$V = V^B$ versus $U = U^B + 1/2$

Microfossils:



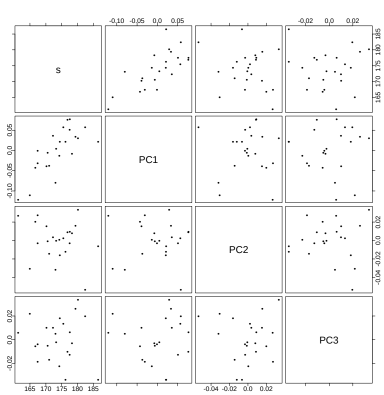
Regression:

$$U = \alpha_1 + \beta_1 \log S, \quad V = \alpha_2 + \beta_2 \log S$$

Tthe fitted values (with standard errors) are
$\widehat{\alpha}_1 = 0.77(0.20)$, $\widehat{\beta}_1 = -0.04(0.02)$,
$\widehat{\alpha}_2 = -1.47(0.32)$ and $\widehat{\beta}_2 = 0.24(0.04)$

Significant linear relationship between $\log S$ and V.

T2 Small mouse vertebrae data



NB: Approx. linear relationship between PC 1 and centroid size.

Shapes package in R:

http://www.cran-r-project.org

Library of shape analysis routines.

Also see:

http://www.maths.nott.ac.uk/∼ild/shapes

Session III:

- Deformations
- Shape in images
- Temporal shape
- Shape Regression
- Discussion

DEFORMATIONS AND THIN-PLATE SPLINES

The thin-plate spline is the most natural interpolant in two dimensions because it minimizes the amount of bending in transforming between two configurations, which can also be considered a roughness penalty. The theory of which was developed by Duchon (1976) and Meinguet (1979). Consider the (2×1) landmarks $t_j, j = 1, \ldots, k$, on the first figure mapped exactly into $y_i, i = 1, \ldots, k$, on the second figure, i.e. there are $2k$ interpolation constraints,

$$(y_j)_r = \Phi_r(t_j), \quad r = 1, 2, \quad j = 1, \ldots, k, \qquad (7)$$

and we write $\Phi(t_j) = (\Phi_1(t_j), \Phi_2(t_j))^\top,\ j = 1, \ldots, k,$ for the two dimensional deformation. Let

$$T = [\, t_1 \quad t_2 \quad \ldots \quad t_k \,]^\top \ , \ Y = [\, y_1 \quad y_2 \quad \ldots \quad y_k \,]^\top$$

so that T and Y are both $(k \times 2)$ matrices.

A **pair of thin-plate splines (PTPS)** is given by the bivariate function

$$\begin{aligned} \Phi(t) &= (\Phi_1(t), \Phi_2(t))^\top \\ &= c + At + W^\top s(t), \end{aligned} \qquad (8)$$

where t is (2×1), $s(t) = (\sigma(t - t_1), ..., \sigma(t - t_k))^\top$, $(k \times 1)$ and

$$\sigma(h) = \begin{cases} \|h\|^2 \log(\|h\|), & \|h\| > 0, \\ 0, & \|h\| = 0. \end{cases} \qquad (9)$$

The $2k + 6$ parameters of the mapping are $c(2 \times 1), A(2 \times 2)$ and $W(k \times 2)$. There are $2k$ interpolation constraints in Equation (7), and we introduce six more constraints in order for the bending energy in Equation (14) below to be defined:

$$1_k^\top W = 0, \quad T^\top W = 0. \qquad (10)$$

The pair of thin-plate splines which satisfy the constraints of Equation (10) are called **natural thin-plate splines**. Equations (7) and (10) can be re-written in matrix form

$$\begin{bmatrix} S & 1_k & T \\ 1_k^\top & 0 & 0 \\ T^\top & 0 & 0 \end{bmatrix} \begin{bmatrix} W \\ c^\top \\ A^\top \end{bmatrix} = \begin{bmatrix} Y \\ 0 \\ 0 \end{bmatrix}, \qquad (11)$$

where $(S)_{ij} = \sigma(t_i - t_j)$ and 1_k is the k-vector of ones. The matrix

$$\Gamma = \begin{bmatrix} S & 1_k & T \\ 1_k^\top & 0 & 0 \\ T^\top & 0 & 0 \end{bmatrix}$$

is symmetric positive definite and so the inverse exists, provided the inverse of S exists. Hence,

$$\begin{bmatrix} W \\ c^\top \\ A^\top \end{bmatrix} = \begin{bmatrix} S & 1_k & T \\ 1_k^\top & 0 & 0 \\ T^\top & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} Y \\ 0 \\ 0 \end{bmatrix} = \Gamma^{-1} \begin{bmatrix} Y \\ 0 \\ 0 \end{bmatrix},$$

say. Writing the partition of Γ^{-1} as

$$\Gamma^{-1} = \begin{bmatrix} \Gamma^{11} & \Gamma^{12} \\ \Gamma^{21} & \Gamma^{22} \end{bmatrix},$$

where Γ^{11} is $k \times k$, it follows that

$$\begin{aligned} W &= \Gamma^{11} Y \\ \begin{bmatrix} c^\top \\ A^\top \end{bmatrix} &= [\hat\beta_1, \hat\beta_2] = \Gamma^{21} Y, \end{aligned} \qquad (12)$$

giving the parameter values for the mapping. If S^{-1}

exists, then we have

$$\begin{aligned}\Gamma^{11} &= S^{-1} - S^{-1}Q(Q^\top S^{-1}Q)^{-1}Q^\top S^{-1},\\ \Gamma^{21} &= (Q^\top S^{-1}Q)^{-1}Q^\top S^{-1} = (\Gamma^{12})^\top, \quad (13)\\ \Gamma^{22} &= -(Q^\top S^{-1}Q)^{-1},\end{aligned}$$

where $Q = [1_k, T]$, using for example Rao (1973,p39).

Using Equations (12) and (13) we see that $\hat{\beta}_1$ and $\hat{\beta}_2$ are generalized least squares estimators, and

$$\mathrm{cov}((\hat{\beta}_1, \hat{\beta}_2)^\top) = -\Gamma^{22}.$$

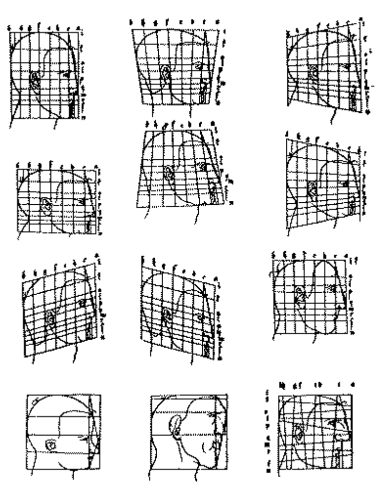
Mardia et al. (1991) gave the expressions for the case when S is singular.

The $k \times k$ matrix B_e is called the **bending energy matrix** where

$$B_e = \Gamma^{11}. \quad (14)$$

There are three constraints on the bending energy matrix

$$1_k^\top B_e = 0, \quad T^\top B_e = 0$$



Early transformation grids of human profiles

and so the rank of the bending energy matrix is $k - 3$.

It can be proved that the transformation of Equation (8) minimizes the total bending energy of all possible interpolating functions mapping from T to Y, where the total bending energy is given by

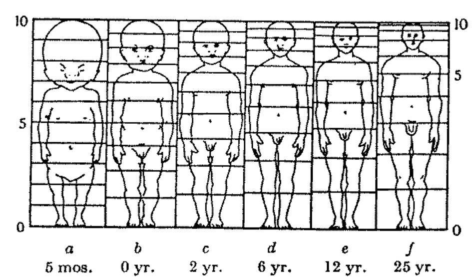
$$J(\Phi) =$$

$$\sum_{j=1}^{2} \int\int_{\mathbb{R}^2} \left(\frac{\partial^2 \Phi_j}{\partial x^2}\right)^2 + 2\left(\frac{\partial^2 \Phi_j}{\partial x \partial y}\right)^2 + \left(\frac{\partial^2 \Phi_j}{\partial y^2}\right)^2 \mathrm{d}x\mathrm{d}y. \quad (15)$$

A simple proof is given by Kent and Mardia (1994a). The minimized total bending energy is given by,

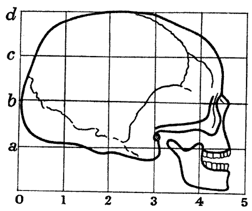
$$J(\Phi) = \mathrm{trace}(W^\top S W) = \mathrm{trace}(Y^\top \Gamma^{11} Y). \quad (16)$$

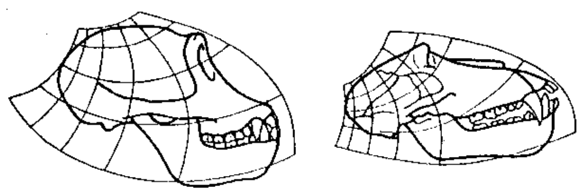
In calculating a deformation grid we do not want to see any more bending locally than is necessary and also do not want to see bending where there are no data.



Early transformation grids modelling six stages through life (from Medawar, 1944).

Following from the original ideas of D'Arcy Thompson (1917) we can produce similar transformation grids, using a pair of thin-plate splines for the deformation from configuration matrices T to Y.





(a) (b)

A regular square grid is drawn over the first figure and at each point where two lines on the grid meet t_i the corresponding position in the second figure is calculated using a pair of thin-plate splines transformation $y_i = \Phi(t_i), i = 1, \ldots, n_g$, where n_g is the number of junctions or crossing points on the grid. The junction points are joined with lines in the same order as in the first figure, to give a deformed grid over the second figure. The pair of thin-plate splines can be used to produce a transformation grid, say from a regular square grid on the first figure to a deformed grid on the second figure. The resulting interpolant produces transformation grids that 'bend' as little as possible. We can think of each square in the deformation as being deformed into a quadrilateral (with four shape parameters). The PTPS minimizes the local variation of these small quadrilaterals with respect to their neighbours.

Consider describing the square to kite transformation which was considered by Bookstein (1989) and Mardia and Goodall (1993). Given $k = 4$ points in $m = 2$ dimensions the matrices T and Y are given by

$$T = \begin{bmatrix} 0 & 1 \\ -1 & 0 \\ 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad Y = \begin{bmatrix} 0 & 0.75 \\ -1 & 0.25 \\ 0 & -1.25 \\ 1 & 0.25 \end{bmatrix}.$$

We have here

$$S = \begin{bmatrix} 0 & a & b & a \\ a & 0 & a & b \\ b & a & 0 & a \\ a & b & a & 0 \end{bmatrix},$$

where $a = \sigma(\sqrt{2}) = 0.6931$ and $b = \sigma(2) = 2.7726$. In this case, the bending energy matrix is

$$B_e = \Gamma^{11} = 0.1803 \begin{bmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{bmatrix}.$$

It is found that

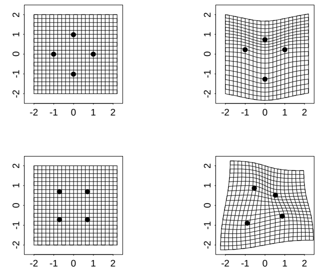
$$W^{\top} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -0.1803 & 0.1803 & -0.1803 & 0.1803 \end{bmatrix}, \tag{17}$$

$$c = 0, \; A = I_2$$

and so the pair of thin-plate splines is given by $\Phi(t) = (\Phi_1(t), \Phi_2(t))^{\top}$, where

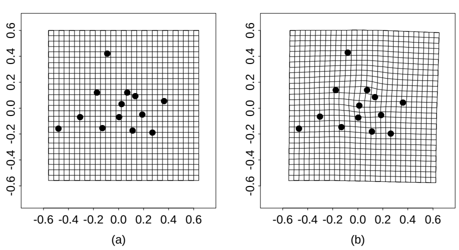
$$\begin{aligned} \Phi_1(t) &= t[1], \qquad (18) \\ \Phi_2(t) &= t[2] + 0.1803 \sum_{j=1}^{4} (-1)^j \sigma(\|t - t_j\|). \end{aligned}$$

Note that Equation (18) is as expected, because there is no change in the $t[1]$ direction. The affine part of the deformation is the identity transformation.



Transformation grids for the square (left column) to kite (right column) (after Bookstein, 1989). In the second row the same figures as in the first row have been rotated by 45^o and the deformed grid does look different, even though the transformation is the same.

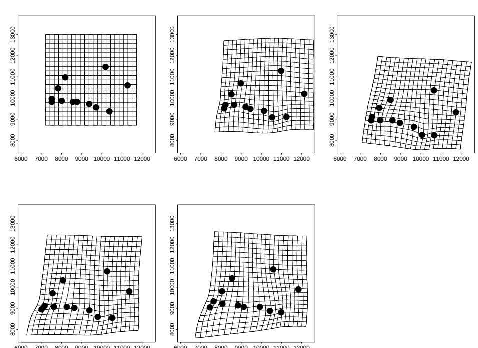
We consider Thompson-like grids for this example (above). A regular square grid is placed on the first figure and deformed into the curved grid on the kite figure. We see that the top and bottom most points are moved downwards with respect to the other two points. If the regular grid is drawn



A thin-plate spline transformation grid between the control mean shape estimate and the schizophrenia mean shape estimate.

(left) We see a square grid drawn on the estimate of mean shape for the Control group in the schizophrenia study. Here there are $n_g = 30 \times 29 = 870$ junctions and there are $k = 13$ landmarks. (right) we see the schizophrenia mean shape estimate and the grid of new points obtained from the PTPS transformation. It is quite clear that there is a shape change in the centre of the brain, around landmarks 1, 9 and 13.

on the first figure at a different orientation, then the deformed grid does appear to be different, even though the transformation is the same. This effect is seen in the Figure where both figures have been rotated clockwise by 45^o in the second row.



A series of grids showing the shape changes in the skull of some sooty mangabey monkeys

Bookstein (1989, 1991)'s principal and partial warps are useful for decomposing the thin-plate spline transformations into a series of large scale and small scale components.

Consider the pair of thin-plate splines transformation from $t \in \mathbb{R}^2$ to $y \in \mathbb{R}^2$, which interpolates the k points T to Y ($k \times 2$) matrices. An eigen-decomposition of the $k \times k$ bending energy matrix B_e of Equation (14) has non-zero eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{k-3}$ with corresponding eigenvectors $\gamma_1, \gamma_2, \ldots, \gamma_{k-3}$. The eigenvectors $\gamma_1, \gamma_2, \ldots, \gamma_{k-3}$ are called the **principal warp eigenvectors** and the eigenvalues are called the **bending energies**. The functions,

$$P_j(t) = \gamma_j^\top s(t) \ \ , \ \ j = 1, \ldots, k-3,$$

are the **principal warps**, where $s(t) = (\sigma(t-t_1), \ldots, \sigma(t-t_k))^\top$.

Here we have labelled the eigenvalues and eigenvectors in this order (with λ_1 as the smallest eigenvalue corresponding to the first principal warp) to follow Bookstein's (1996b) labelling of the order of the warps. The principal warps do not depend on the second figure Y. The principal warps will be used to construct an orthogonal basis for re-expressing the thin-plate spline transformations. The principal warp deformations are univariate functions of two dimensional t, and so could be displayed as surfaces above the plane or as contour maps. Alternatively one could plot the transformation grids from t to $y = t + (c_1 P_j(t), c_2 P_j(t))^\top$ for each j, for particular values of c_1 and c_2. Note that the principal warps are orthonormal.

The **partial warps** are defined as the set of $k-3$ bivariate functions $R_j(t), j = 1, ..., k-3$, where

$$R_j(t) = Y^\top \lambda_j \gamma_j P_j(t) = Y^\top \lambda_j \gamma_j \gamma_j^\top s(t).$$

The jth **partial warp scores** for Y (from T) are defined as

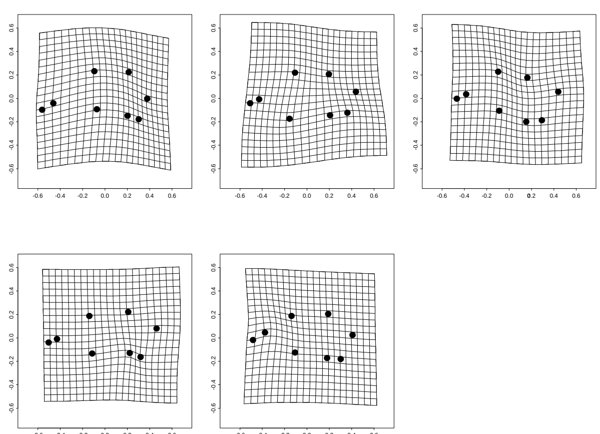
$$(p_{j1}, p_{j2})^\top = Y^\top \gamma_j \ , \ \ j = 1, \ldots, k-3,$$

and so there are two scores for each partial warp.

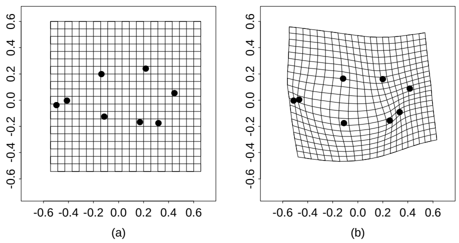
Since

$$W^\top s(t) = \sum_{j=1}^{k-3} R_j(t),$$

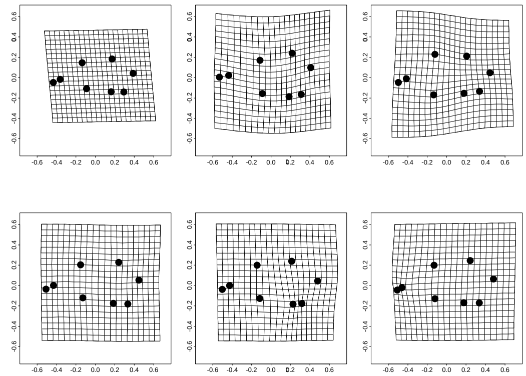
we see that the non-affine part of the pair of thin-plate splines transformation can be decomposed into the sum of the partial warps. The jth partial warp corresponds largely to the movement of the landmarks which are the most highly weighted in the jth principal warp. The jth partial warp scores indicate the contribution of the jth principal warp to the deformation from the source T to the target Y, in each of the Cartesian axes.



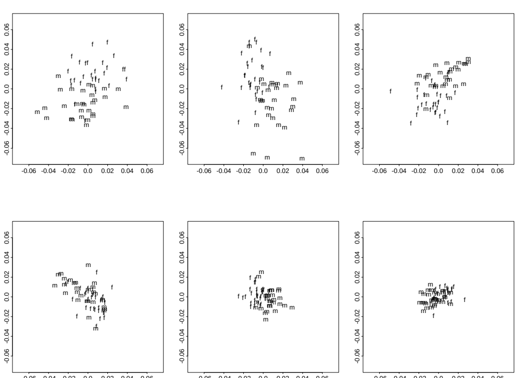
The five principal warps for the the pooled mean shape of the gorillas



A thin-plate spline transformation grid between a female and a male gorilla skull midline.



Affine and partial warps for Gorilla (Female to male mean shapes)



Affine scores and the partial warp scores for female (f) and male (m) gorilla skulls.

RELATIVE WARPS

Principal component analysis with non-Euclidean metrics

Define the **pseudo-metric space** $(\mathbb{R}^p, d_A)$ as the real p-vectors with **pseudo-metric** given by

$$d_A(x_1, x_2) = \sqrt{(x_1 - x_2)^\top A^-(x_1 - x_2)},$$

where x_1 and x_2 are p-vectors, and A^- is a generalized inverse (or inverse if it exists) of the positive semi-definite matrix A.

The Moore–Penrose inverse is a suitable choice of generalized inverse. If A is the population covariance matrix of x_1 and x_2, then $d_A(x_1, x_2)$ is the Mahalanobis distance. The norm of a vector x in the metric space is

$$\|x\|_A = d_A(x, 0) = (x^\top A^- x)^{1/2}.$$

We could carry out statistical inference in the metric space rather than in the usual Euclidean $(A = I)$

space (after Bookstein, 1995, 1996b; Kent, personal communication; Mardia, 1977, 1995). A simple way to proceed is to transform from $x \in (\mathbb{R}^p, d_A)$ to $y \in (A^-)^{1/2}x$ in Euclidean space. For example, consider principal component analysis (PCA) of n centred p-vectors $x_1, \ldots, x_n$ in the metric space. Transforming to $y_i = (A^-)^{1/2}x_i$ the principal component (PC) loadings are the eigenvectors of

$$S_y = \frac{1}{n}\sum_{i=1}^{n} y_i y_i^\top.$$

Denote the eigenvectors (p-vectors) of S_y as $\gamma_{yj}, j = 1, \ldots, p$ (assuming $p \leq n-1$), with corresponding eigenvalues $\lambda_{yj}, j = 1, \ldots, p$. The **principal component scores** for the jth PC on the ith individual are

$$r_{ij} = \gamma_{yj}^\top y_i = (\gamma_{yj}^\top (A^-)^{1/2})x_i,\ i = 1, \ldots, p.$$

So, the (unnormalized) PC loadings on the original data are $(A^-)^{1/2}\gamma_{yj}$ which are the eigenvectors of $(A^-)^{1/2}S_x(A^-)^{1/2}$, where $S_x = \frac{1}{n}\sum_{i=1}^{n} x_i x_i^\top$ (using standard linear algebra, e.g. Mardia et al., 1979, Appendix). The first few PCs in the metric space (with loadings given by the eigenvectors of

$$(A^-)^{1/2}S_x(A^-)^{1/2}$$

) can be useful for interpretation, emphasizing a different aspect of the sample variability than the usual PCA in Euclidean space. If our analysis is carried out in the pseudo-metric space, then we say the our analysis has been carried out *with respect to* A.

If a random sample of shapes is available, then one may wish to examine the structure of the within group variability in the tangent space to shape space. We have already seen PCA with respect to the Euclidean metric, but an alternative is the method of relative warps. Relative warps are PCs with respect to the bending energy or inverse bending energy metrics in the shape tangent space.

Consider a random sample of n shapes represented by Procrustes tangent coordinates $v_1, ..., v_n$ (each is a $2k-2$-vector), where the pole μ is chosen to be an average pre-shape such as from the full Procrustes mean. The sample covariance matrix in the tangent plane is denoted by S_v and the sample covariance matrix of the centred tangent coordinates $x_i = (I_2 \otimes H^\top)v_i, \quad i = 1, \ldots, n$ is denoted by S_c ($2k \times 2k$). In our examples we have used the covariance matrix of the Procrustes fit coordinates. The bending energy matrix is calculated for the average shape B_e and then the tensor product is taken to give $B_2 = I_2 \otimes B_e$, which is a $2k \times 2k$ matrix of rank $2k-6$. We write B_2^- for a generalized inverse of B_2 (e.g. the Moore–Penrose generalized inverse).

We consider PCA in the tangent space with respect to a power of the bending energy matrix, in particular with respect to B_e^α.

Let the non-zero eigenvalues of $(B_2^-)^{\alpha/2}S_c(B_2^-)^{\alpha/2}$ be $l_1, \ldots, l_{2k-6}$ with corresponding eigenvectors $f_1, ..., f_{2k-}$ and

$$(B_2^-)^{\alpha/2} = \sum_{r=1}^{2k-6} \lambda_r^{-\alpha/2}\gamma_r\gamma_r^\top,$$

with $\lambda_1, \ldots, \lambda_{2k-6}$ the eigenvalues of B_2 with corresponding eigenvectors $\gamma_1, \ldots, \gamma_{2k-6}$. The eigenvectors $f_1, ..., f_{2k-6}$ are called the **relative warps**. The **relative warp scores** are

$$a_{ij} = (f_j)^\top (B_2^-)^{\alpha/2}x_i,\ \ j = 1, \ldots, 2k-6,\ \ i = 1, \ldots, n.$$

Important remark: The relative warps and the relative warp scores are useful tools for describing the non-linear shape variation in a dataset. In particular the effect of the jth relative warp can be viewed by plotting

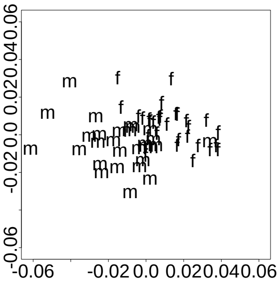
$$H^\top\mu \pm cB_2^{\alpha/2}f_j l_j^{1/2},$$

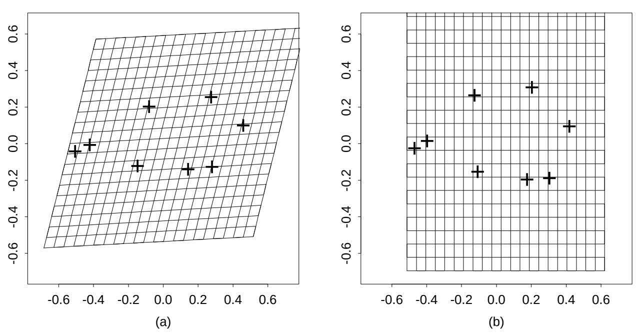
for various values of c, where

$$B_2^{\alpha/2} = \sum_{r=1}^{2k-6} \lambda_r^{\alpha/2}\gamma_r\gamma_r^\top.$$

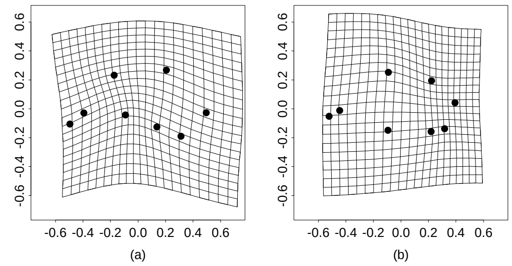
The procedure for PCA with respect to the bending energy requires $\alpha = +1$ and emphasizes **large scale**

variability. PCA with respect to the inverse bending energy requires $\alpha = -1$ and emphasizes **small scale variability**. If $\alpha = 0$, then we take $B_2^0 = I_{2k}$ as the $2k \times 2k$ identity matrix and the procedure is exactly the same as PCA of the Procrustes tangent coordinates. Bookstein (1996b) has called the $\alpha = 0$ case PCA with respect to the Procrustes metric.

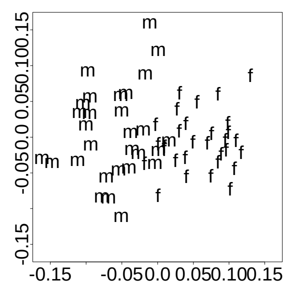


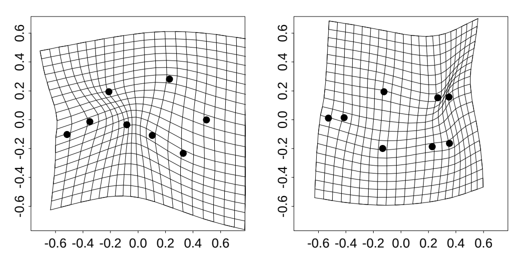


Deformation grids for the two uniform/affine vectors for the gorilla data.

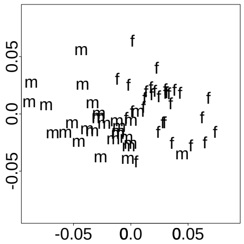


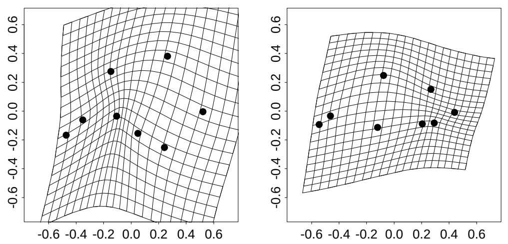
Relative warps: $\alpha = 1$





Relative warps: $\alpha = -1$





Relative warps: $\alpha = 0$

- Prior distribution for configuration

Geometrical object description =
SHAPE + REGISTRATION

where REGISTRATION =
LOCATION, ROTATION and SCALE

- Use training data to estimate any parameters

- Deformable templates:
Grenander and colleagues

- Point distribution models (PDM) Cootes, Taylor, et al.

- Bayesian approach

- Prior model for object shape and registration using SHAPE ANALYSIS

- Likelihood for features measuring goodness-of-fit (feature density)

Bayes Theorem $\rightarrow$ Posterior inference

CASE STUDY:

Object recognition: face images

[example from Mardia, McCulloch, Dryden and Johnson, 1997]

LANDMARKS or FEATURES

• Grey level image $I(x, y)$

• Scale space features (e.g. Val Johnson, Duke; Stephen Pizer et al, UNC Chapel Hill, USA)

Convolution of image with isotropic bivariate Gaussian kernel at a succession of 'scales' (σ)

$$S(x, y; \sigma) =$$

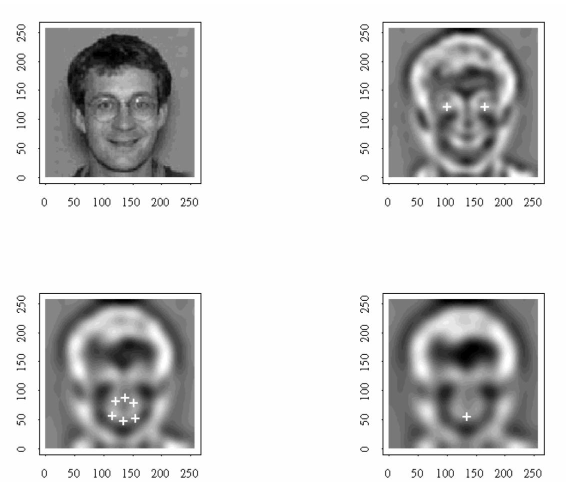
$$\int I(x - h_1, y - h_2) \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}(h_1^2 + h_2^2)} dh_1 dh_2$$

Use 2D FFT

• 'Medialness' : Laplacian of blurred scale space image

$$L_{xx}(\sigma) + L_{yy}(\sigma) = \frac{\partial^2 S(x, y; \sigma)}{\partial x^2} + \frac{\partial^2 S(x, y; \sigma)}{\partial y^2}$$

Pilot study - Face Identification: Choose $k = 9$ landmarks on the medialness image at scales 8, 11, 13



• Feature density (likelihood)

$$L(\text{image}|\text{configuration}) \propto \prod_{i=1}^{k} e^{\frac{1}{2}\kappa_i (L_{x_i x_i} + L_{y_i y_i})}$$

• Johnson et al. (1997) motivate this as mimicking a human observer.

• Features are treated as independent

• High medialness at feature $\rightarrow$ high density

• Treat non-feature grey levels as independent, uniformly distributed (like a human observer ignoring those pixels).

• Parameters κ_i need to be specified

Registration parameters

- Location

$\mu_x \sim N(\psi_x, \sigma_x^2)$

$\mu_y \sim N(\psi_y, \sigma_y^2)$

- Rotation

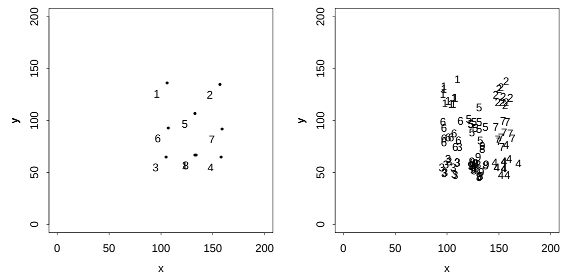
$\theta \sim N(\psi_t, \sigma_t^2)$

- Isotropic scale

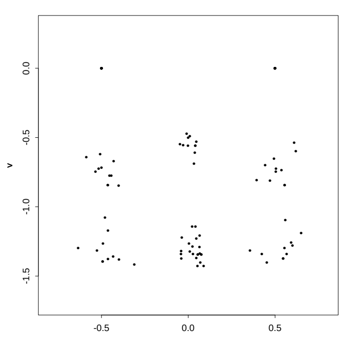
$\beta \sim N(\psi_b, \sigma_b^2)$

Hyperparameters $\psi_x, \sigma_x, \psi_y, \sigma_y, \psi_t, \sigma_t, \psi_b, \sigma_b$ estimated from training data (10 faces)

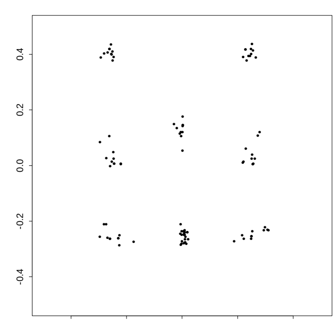
Original raw face data



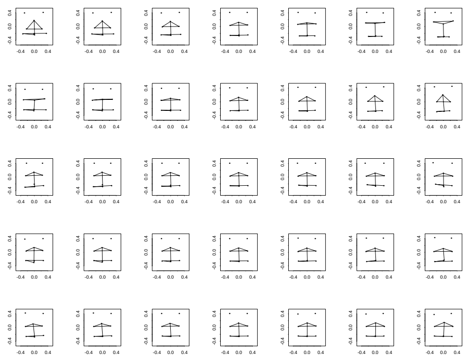
Bookstein registered data



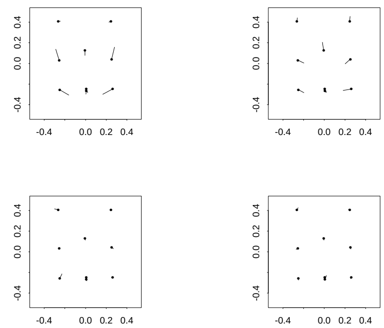
- Least squares Procrustes approach



• First five PCs (explaining 54.4, 29.9, 6.0, 3.7, 2.7% of variability in shape).



• Vector plot from mean to 3 S.D.s for first three PCs



• .... FACE PRIOR

Assume registration and shape independent

Multivariate normal prior model (configuration density):

$$\pi(\text{configuration}) = \pi(\mu_x, \mu_y, \theta, \beta, c_1, \ldots, c_p)$$
$$\propto \mathrm{e}^{-\frac{1}{2}\left(\frac{(\mu_x-\psi_x)^2}{\sigma_x^2}+\frac{(\mu_y-\psi_y)^2}{\sigma_y^2}+\frac{(\theta-\psi_t)^2}{\sigma_t^2}+\frac{(\beta-\psi_b)^2}{\sigma_b^2}+\sum_{i=1}^{p} c_i^2\right)}$$

Bayes theorem $\rightarrow$ Posterior density:

$$\pi(\text{configuration}|\text{image})$$
$$\propto \pi(\text{configuration}) L(\text{image}|\text{configuration})$$

• Draw samples from posterior using MCMC

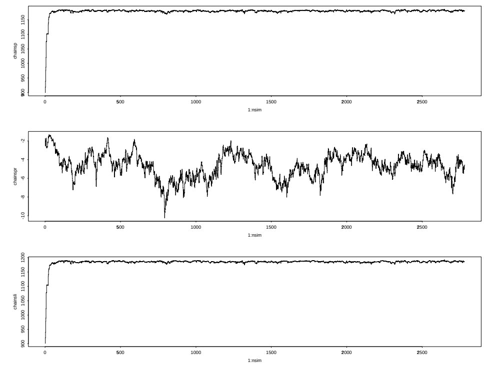
• Object recognition: maximize posterior to obtain most likely configuration given the image

• Straightforward Metropolis-Hastings algorithm

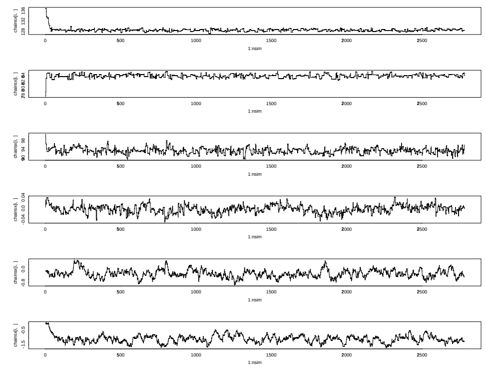
Proposal distribution: independent normal centred on current observation, with varying variance (linearly decreasing over 5 iterations, then jumping back up)

Update each parameter one at a time

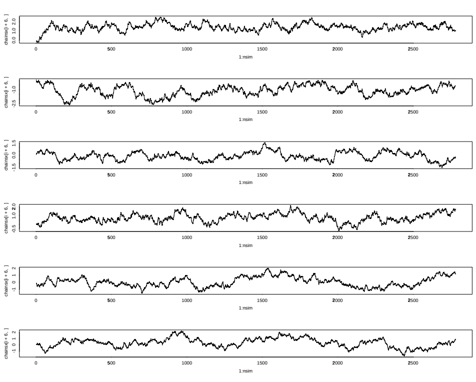
Results: MCMC output for face 2 (in training set): Posterior, Prior, Likelihood



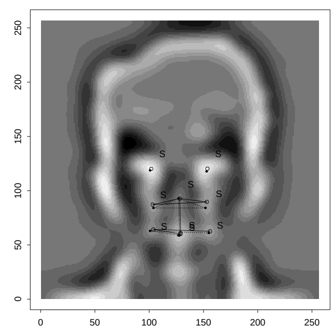
Translations, scale, rotation, PC1, PC2



PC3,...,PC8



MAP estimate overlaid on scale 8



Shape distance to training set.... Procrustes distance ρ and Mahalanobis

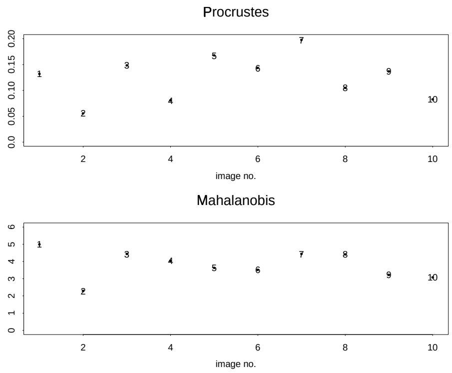


IMAGE REGISTRATION

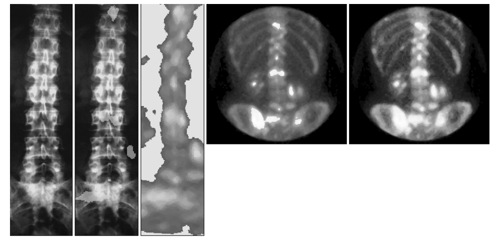
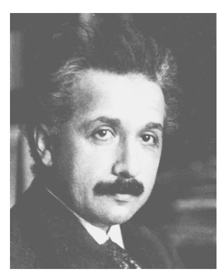


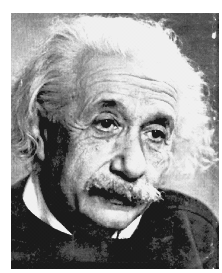
IMAGE AVERAGING

Consider a random sample of images $f_1, \ldots, f_n$ containing landmark configuration $X_1, \ldots, X_n$, from a population mean image f with a population mean configuration μ. We wish to estimate μ and f up to arbitrary Euclidean similarity transformations. The shape of μ can be estimated by the full Procrustes mean of the landmark configurations $X_1, \ldots, X_n$. Let Φ_i^* be the deformation obtained from the estimated mean shape $[\hat{\mu}]$ to the ith configuration. The **average image** has the grey level at pixel location t given by

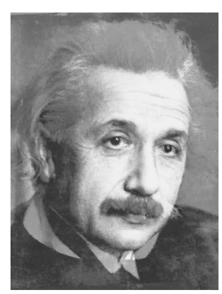
$$\bar{f}(t) = \frac{1}{n} \sum_{i=1}^{n} f_i\{\Phi_i^*(t)\}. \qquad (19)$$



(a)

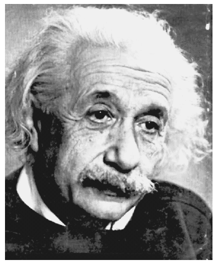


(b)

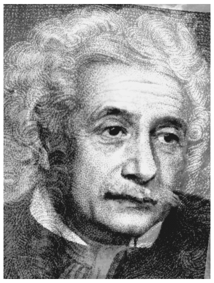


(c)

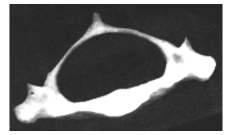
(a)



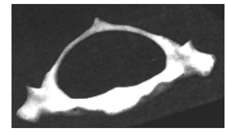
(b)



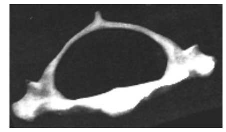
(c)



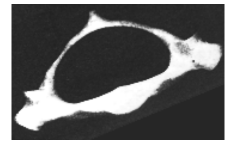
(a)



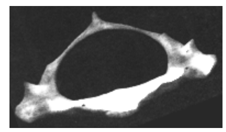
(b)



(c)

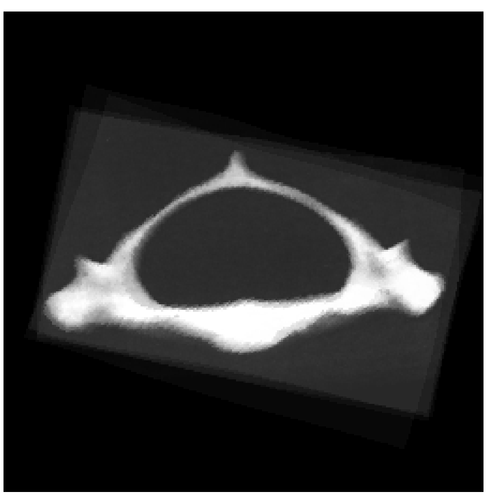


(d)



(e)

Images of five first thoracic (T1) mouse vertebrae.

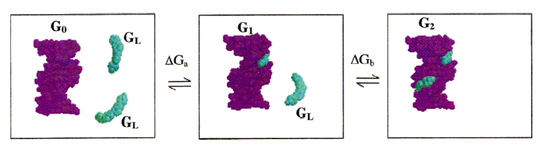


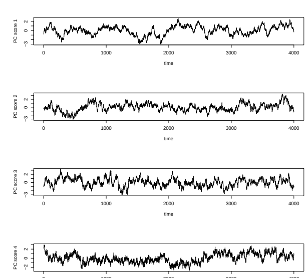
An average T1 vertebra image obtained from five vertebrae images.

SHAPE TEMPORAL MODELS

Stochastic modelling of size and shape of molecules over time: HIGH DIMENSIONAL.

- Practical aim: to estimate entropy. Use tangent space modelling.





• Temporal correlation models for the principal component (PC) scores of size and shape. [AR(2)]
• Non-separable model - different temporal covariance structure for each PC but constant eigenvectors over time.
• Improved entropy estimator based on MLE, interval estimators.
• Properties of estimators under general correlation structures, including long-range dependence.

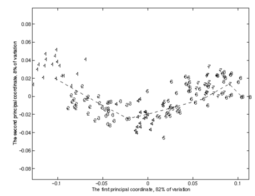
• Temporal shape modelling directly in shape space.

REGRESSION

The minimal geodesic in shape space between the shapes of X and Y where $0 < s_0 = \rho(X, Y)$ [Riemannian distance] is given by:

$$g(s) = \frac{1}{\sin s_0}\left\{X \sin(s_0 - s) + R^T Y \sin s\right\}, \quad 0 \leq s \leq s_0$$

where $R^T Y X^T$ is symmetric (i.e. R^T is the optimal Procrustes rotation of Y on X).



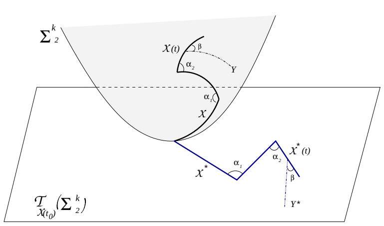
Practical regression models: tangent space regression through origin $\equiv$ fitting geodesics in shape space.

SMOOTHING SPLINES

Smoothing spline fitting through ‘unrolling’ and ‘unwrapping’ the shape space Σ_2^k.

On the Procrustes tangent space at time t_0, the shape space is rolled without slipping or twisting along the continuous piecewise geodesic curve in Σ_2^k. The piecewise linear path in the tangent space is the unrolled path.

A point off the curve is unwrapped onto the tangent plane.



Spline fitting in Σ_2^k: unrolling the spline to the tangent space at t_0 is the corresponding cubic spline fitted to the unwrapped data.

Le (2002, *Bull.London Math.Soc.*), Kume et al. (2003).

Piecewise linear spline $\rightarrow$ piecewise geodesic curve in Σ_2^k.

NONPARAMETRIC INFERENCE

• The full Procrustes mean $\widehat{\mu}$ is a consistent estimator of 'extrinsic mean shape' (Patrangenaru and Bhattacharya, 2003)
• Central limit theorem for $\widehat{\mu}$ and a limiting χ^2 distribution for a pivotal test statistic $\rightarrow$ confidence regions.
• Bootstrap confidence interval for mean shape based on a pivotal statistic - NEEDS CARE in a non-Euclidean space.
• Coverage accuracy of bootstrap confidence region $O(n^{-2})$.

• Bootstrap k sample hypothesis test (not necessary to have equal covariance matrices in each group).
• Need to simulate from the null hypothesis of equal mean shapes, and so the individual samples are moved along a geodesic to the pooled mean without changing the inter-sample shape distances.
• Simulation studies indicate accurate observed significance levels and good power.

DISCUSSION

• At all stages geometrical information always available

• Statistical shape analysis of wide use in many disciplines.

• Great scope for further application in image analysis, e.g. medical imaging.

• Non-landmark - curve - data

Selected References to papers:

DG Kendall (1984,Bull.Lond.Math.Soc), Bookstein (1986,Statistical Science), WS Kendall (1988,Adv.Appl.Probab.), DG Kendall (1989,Statistical Science), Mardia and Dryden (1989, Adv.Appl. Probab; 1989, Biometrika), Dryden and Mardia (1991, Adv.Appl,Probab) Dryden and Mardia (1992, Biometrika) Goodall and Mardia (1993, Annals of Statistics), Le and DG Kendall (1993, Annals of Statistics), Kent (1994, JRSS B), Le (1994, J.Appl.Probab.), Kent and Mardia (1997, JRSS B), Dryden, Faghihi and Taylor (1997, JRSS B), WS Kendall (1998, Adv.Appl.Probab.), Mardia and Dryden (1999, JRSS B), Kent and Mardia (2001, Biometrika), Le (2002, Bull.London.Math.Society), Albert, Le and Small (2003, Biometrika).