Statistics of Anatomic Geometry

Summarizing shape variability and hypothesis testing

Ian Dryden (University of Nottingham, UK)

Ian.Dryden@Nottingham.ac.uk

http://www.maths.nott.ac.uk/∼ild

MICCAI Tutorial, 26 October 2005

1. HIGH-LEVEL BAYESIAN IMAGE ANALYSIS

(following Grenander/Miller et al, Pizer et al., Cootes/Taylor et al,....)

X - co-ordinates of a template

Z - vector of pixels/voxels in an image

- Bayes Theorem:

$$P(X|Z) \propto P(X)P(Z|X)$$

- Bayesian inference can be carried out by sampling from the posterior of the template co-ordinates or obtaining an estimate (e.g. posterior mean, posterior maximum) of the template.

- Further inference on fitted templates, such as hypothesis tests, is also of great practical interest.

- We shall mainly docus on reduction in dimensionality, coping with invariances, the use of linear (tangent) spaces and hypothesis testing.

DATA

Consider n data matrices $X_i, i = 1, \ldots, n$ to be available, each a $k \times m$ matrix. Often this might be a training dataset, and we wish to build our statistical models from such training data.

The data might be

- Landmark co-ordinates for k landmarks in m dimensions (usually $m = 2$ or $m = 3$).
- Image gray levels from images of size $r \times c$.
- A discrete set of points on a function derived from an object (e.g. polar co-ordinates of an outline/surface)
- A discrete set of points on a function derived from an image (e.g. intensity histogram)
- Combinations of the above.

We work with the p-vector $\mathrm{vec}(X)$ [the stacked columns of X].

2. DIMENSION REDUCTION

$X_i, i = 1, \ldots, n$: $k \times m$ matrices

Frequently k is quite large. However, even for small landmark datasets we have trouble visualising in km dimensions ($k \geq 3, m \geq 2$), and models in the co-ordinate space need a large number of paramaters.

Various methods are available for dimension reduction, and for representing the main features of variability. We concentrate on:

- Principal components analysis (PCA)
- Canonical correlation analysis (CCA)
- Independent components analysis (ICA)

2.1 Principal Components Analysis (PCA) (e.g. see Mardia et al., 1979, pp.213).

Sample covariance matrix:

$$S = \frac{1}{n-1} \sum_{i=1}^{n} \mathrm{vec}(X_i - \bar{X}) \{\mathrm{vec}(X_i - \bar{X})\}^T,$$

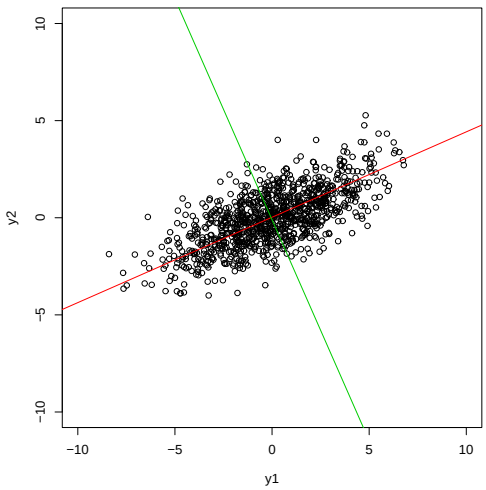
where $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is a p-vector, and S is a $p \times p$ positive definite symmetric matrix.

- Aim: Find orthonormal projections of the original coordinates X_i which each maximise the variance, subject to being orthogonal to previous eigenvectors, i.e.

$\max(a_i^T S a_i)$ subject to $a_i^T a_i = 1$ and $a_i^T a_j = 0, j = 1, \ldots, i-1$.

- Solution: obtain the eigenvectors $\hat{\gamma}_j$ and corresponding eigenvalues $\hat{\lambda}_j$ of S, $j = 1, \ldots, q = \min(n, p-1)$, where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_q$.

Simple example:



Geometrically the PCs are the principal axes of the point cloud. This is **sample** PCA, using an estimate of the population covariance matrix. Alternatively we could use the population covariance matrix (if known).

- PC scores:

$$s_{ij} = \hat{\gamma}_j^T \{\mathrm{vec}(X_i) - \mathrm{vec}(\bar{X})\}, \;\; i = 1, \ldots, n, \; j = 1, \ldots, q.$$

Often the first few PC scores provide a useful low dimensional summary of the data.
Percentage of variability explained by first j PC scores is

$$100 \sum_{i=1}^{j} \hat{\lambda}_i / \sum_{i=1}^{q} \hat{\lambda}_i.$$

In our simple example PC score 1 explains 88.6 % of the variability.

- Canonical Variate Analysis - also include group information

HIGH DIMENSIONS $p >> n$. A useful trick for computation, especially needed for images.

Let us write $X = [x_{p1}, \ldots, x_{pn}]$ for the n columns of vectors from a random sample. Now, using the spectral decomposition we have $S = \frac{1}{n} X X^T = \sum_{j=1}^{n} \hat{\omega}_j \hat{\gamma}_j \hat{\gamma}_j^T$. Consider the $n \times n$ matrix $A = \frac{1}{n} X^T X$, and the spectral decomposition is $A = \sum_{j=1}^{n} \delta_j q_j q_j^T$, which can be computed in $O(n^3)$ steps. Now

$$S^2 = \frac{1}{n^2} X X^T X X^T = \sum_{j=1}^{n} \hat{\omega}_j^2 \hat{\gamma}_j \hat{\gamma}_j^T = \frac{1}{n} X A X^T = \sum_{j=1}^{n} \frac{\delta_j}{n} (X q_j)(X q_j)^T.$$

Hence, by equating coefficients, $\hat{\gamma}_j = X q_j / \|X q_j\|, \hat{\omega}_j = \|X q_j\| \sqrt{\delta_j / n}, j = 1, \ldots, n$.
Thus calculating the PCs is practical for huge $p >> n$.

2.2 Canonical correlation analysis (CCA)

• Aim: Find orthonormal projections of the measurements $X_i, i = 1, \ldots, n$ and ANOTHER set of measurements measurements $Y_i, i = 1, \ldots, n$ which maximise the correlation BETWEEN two sets of observations.

Write S_{11}, S_{22} for the sample (non-singular) variance-covariance matrix based on X_i, Y_i respectively, and let S_{12} be the sample correlations between X_i and Y_i. We wish to find $\max(a^T S_{12} b)$ subject to $a^T S_{11} a = 1 = b^T S_{22} b$.

• SOLUTION (e.g. see Mardia et al., 1979, pp.281).

Let $K = S_{11}^{-1/2} S_{12} S_{22}^{-1/2} = (\alpha_1, \ldots, \alpha_q)^T D (\beta_1, \ldots, \beta_q)^T$, where α_i, β_i are the standardized eigenvectors of KK^T and $K^T K$ respectively and $D = \mathrm{diag}(\lambda_1^{1/2}, \ldots, \lambda_q^{1/2})$.

• $a_i = S_{11}^{-1/2} \alpha_i, b_i = S_{22}^{-1/2} \beta_i$ are called the ith canonical correlation (CC) vectors for x and y respectively.

• $a_i^T x$ and $a_i^T y$ are called the ith CC variables (or scores).

• COMPARISON: PCA considers relationships WITHIN groups of variables, but CCA BETWEEN

• **Example** Multiple object analysis - e.g. estimating the features which have strongest correlation between different organs in the body.

2.3 Independent components analysis (ICA) [e.g. Hyvärinen et al, 2001]

- Aim: Find linear combinations of the original coordinates X that are 'most independent' in some sense. In practice, this is achieved by maximising some measure of non-Gaussianity of the variables, to give the ICs $s = W\mathrm{vec}(\mathrm{X})$, where $W = q \times p$ and s is a $q \times 1$ vector.

Common choices: maximum absolute kurtosis, negative entropy, minimum mutual information.

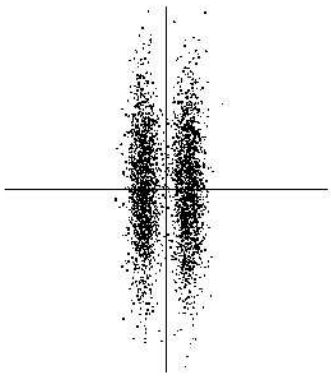
Idea originated from blind-source separation and projection pursuit (looking for 'interesting projections')

- Solution: numerical procedures such as fastICA are commonly available.

Note, we cannot determine the variances of the ICs or the order of the ICs.

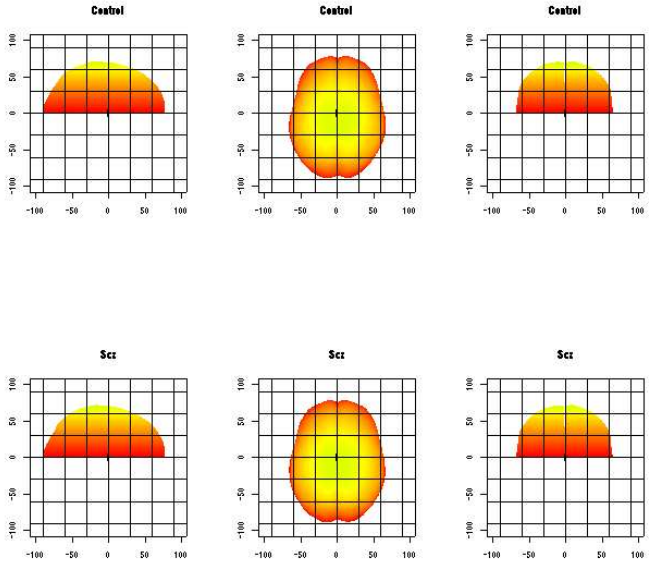
One frequently carries out PCA before ICA to reduce the dimensionality and computation.

- Example: (Hyvärinen and Oja, 2000)

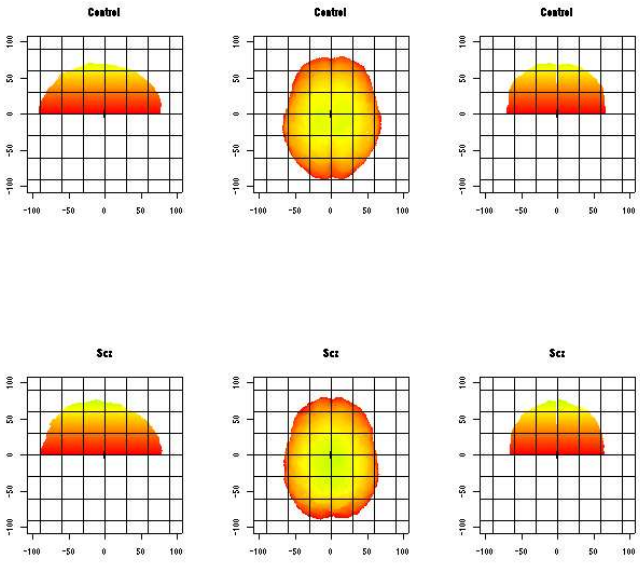


The ICs and PCs are completely different here - 1st PC given by vertical line, single IC would be a horizontal line (highly non-Gaussian and 'most interesting')

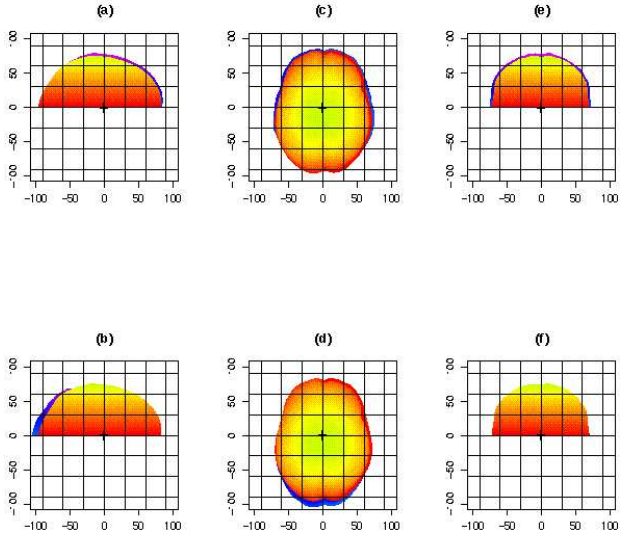
Application: Surface shape analysis



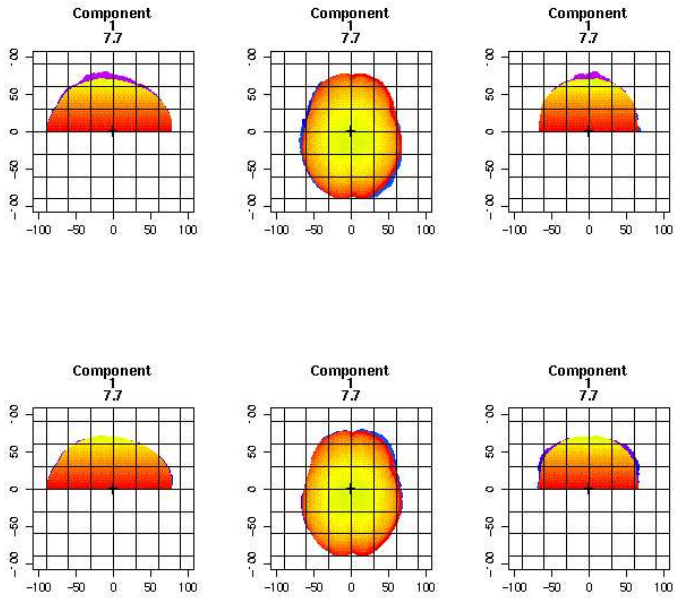
Three orthographic projections of the mean forms for the controls (top row) and patients (bottom row). The columns show (1) Sagittal view, (2) Axial view, (3) Coronal view. The colouring indicates the height above the horizontal AC-PC



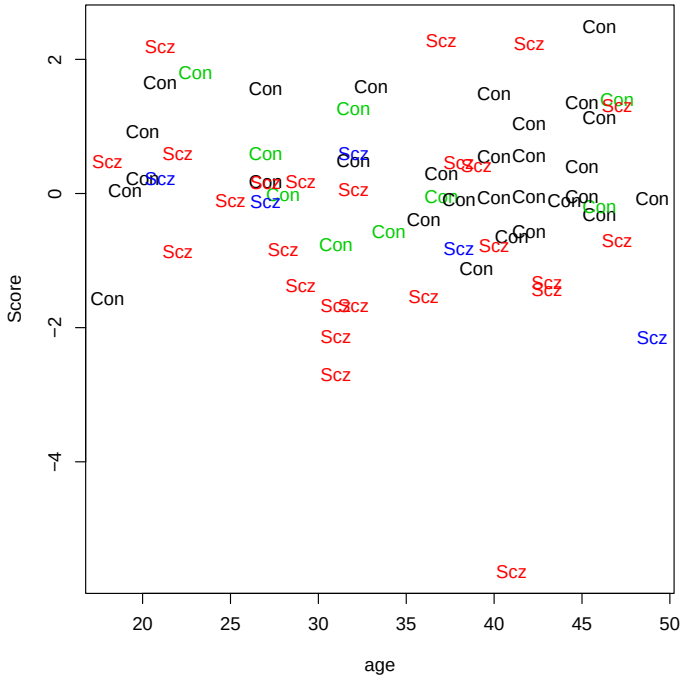
Three orthographic projections of the mean forms for the controls (top row) and patients (bottom row). Each plot shows the pooled mean plus $6\times$ (group mean - pooled mean) in order to exaggerate the differences for easier interpretation.



Plots of the mean cortical form $\pm$ 3 standard deviations along symmetrical PC1. The columns show (1) Sagittal view, (2) Axial view, (3) Coronal view. The rows indicate: (top row) Red/yellow: $\hat{\mu}$ and Blue: $\hat{\mu}$ + 3 sd's along PC1, (bottom row) Red/yellow: $\hat{\mu}$ and Blue: $\hat{\mu}$ - 3 sd's along PC1.



Plots for independent component (IC) which is signifi cantly different between the two groups



The IC scores versus age.

3. INVARIANCES AND PROCRUSTES

In medical imaging and other application areas certain invariances are present which we are NOT INTERESTED in. For example:

- TRANSLATION
- ROTATION
- SCALE
- SHEARS
- NON-AFFINE DEFORMATIONS

Translation

Translation invariance is straightforward to deal with. Consider the case of k landmarks in m dimensions.

Transform from X to $Y = HX$ where H is the $k-1 \times k$ Helmert submatrix (just a particular orthogonal matrix of contrasts without the first row, e.g. see Dryden and Mardia, 1998, p34).

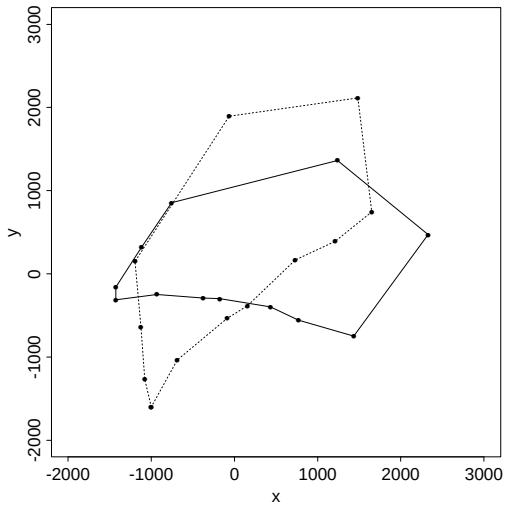
Note that Y is in a $km - m$ Euclidean space and standard multivariate based statistical inference can be carried out in the linear space.

Further invariances, e.g. Euclidean similarity

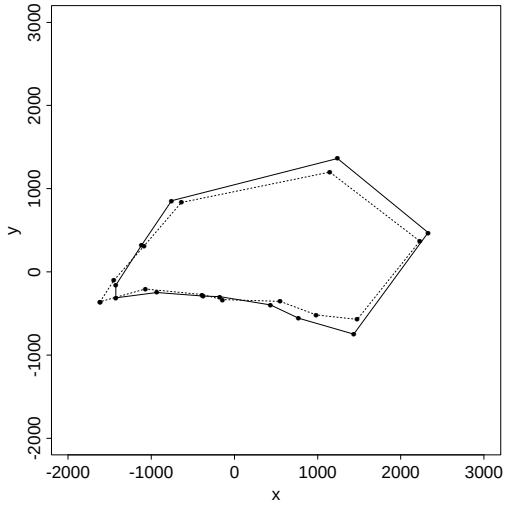
Provided the variability in the data is fairly small then we can also deal with further invariances in a relatively straightforward way.

FIRST REGISTER CONFIGURATIONS, e.g. by PROCRUSTES ANALYSIS

We wish to register Adult (- - - - - -) onto Juvenile (———-)



Procrustes registration of the Adult (- - - - -) onto the juvenile (———)



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

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

Solution:

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

$$\hat{\Gamma} = UV^{\mathrm{T}}$$

where

$$X_2^{\mathrm{T}} X_1 = \|X_1\| \, \|X_2\| V \Lambda U^{\mathrm{T}}, \;\; U, V \in SO(m)$$

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

$$\hat{\beta} = \frac{\mathrm{trace}(X_2^{\mathrm{T}} X_1 \hat{\Gamma})}{\mathrm{trace}(X_1^{\mathrm{T}} X_1)}.$$

The minimized Procrustes sum of squares is:

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

where $\rho(X_1, X_2)$ is the (non-Euclidean) **Riemannian shape distance** (Kendall, 1984).

Procrustes fit $X_1^P = \hat{\beta} X_1 \hat{\Gamma} - 1_k \gamma^{\mathrm{T}}$

Procrustes residual vector $r = X_1^P - X_2$

NB: not symmetric in X_1, X_2.

PERTURBATION MODEL for several shapes:

$$X_i = \beta_i(\mu + E_i)\Gamma_i + 1_k \gamma_i^{\mathrm{T}}$$

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

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

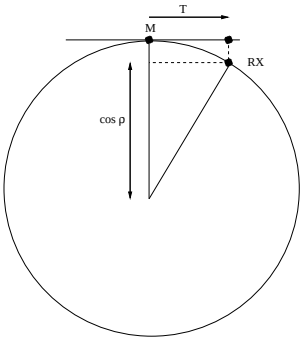
Least squares approach. Iterative algorithm (GPA algorithm, Gower, 1975) needed for $m > 2$ dimensions.

4. PROCRUSTES TANGENT SPACE

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

$$T = X\hat{\Gamma} - \cos\rho M$$

where $0 < \rho \leq \pi/2$ is the Riemannian distance between the shapes of M and X, and $\hat{\Gamma}$ 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.

- Alternatively we can use the Procrustes residuals $X^P = \hat{\beta}X\hat{\Gamma} - M$ which are APPROXIMATE tangent co-ordinates.
- For practical purposes standard multivariate statistical techniques in tangent space are good approximations to non-Euclidean shape methods, provided the data are not too highly dispersed.
- Let us write $v_i, i = 1, \ldots, n$ for the Procrustes tangent co-ordinate vectors from a random sample of data.

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

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

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:

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

- PC summary of the data in the tangent space is

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

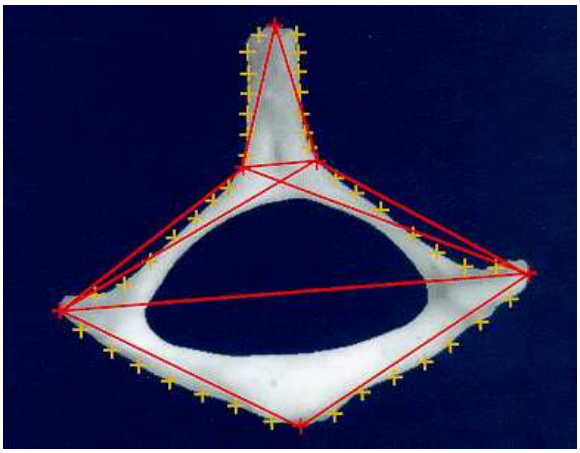
for $i = 1, \ldots, n$.

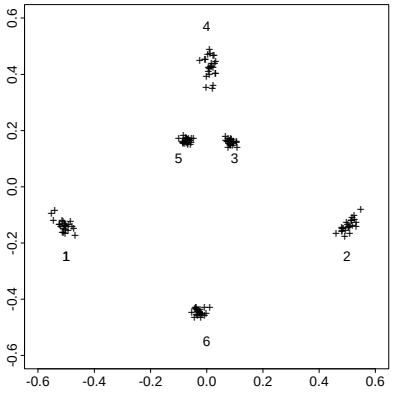
- Standardized PC scores:

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

When $c_{ij} \sim N(0, 1)$ independently, these models are known as Point Distribution Models or Active Shape Models (Cootes et al., 1994).

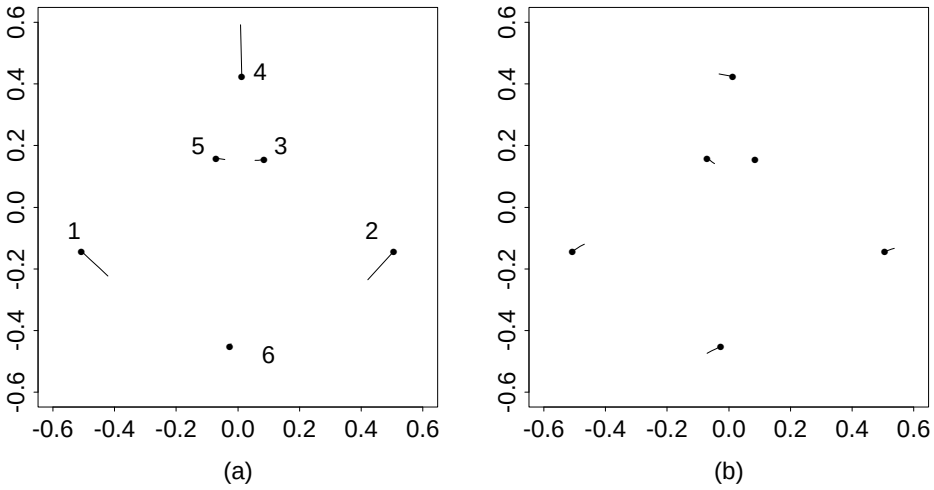
Application: T2 Mouse vertebra example



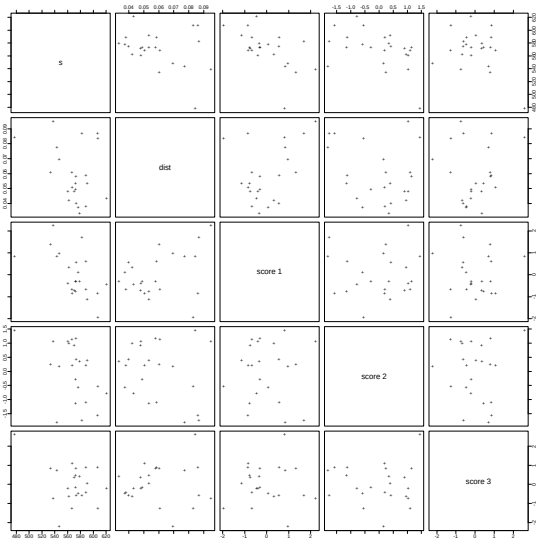


Mouse vertebra example: (PC1 = 69%)

Procrustes registration for display



Pairwise plots:



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

Likewise we can consider Canonical correlation analysis and Independent components analysis.

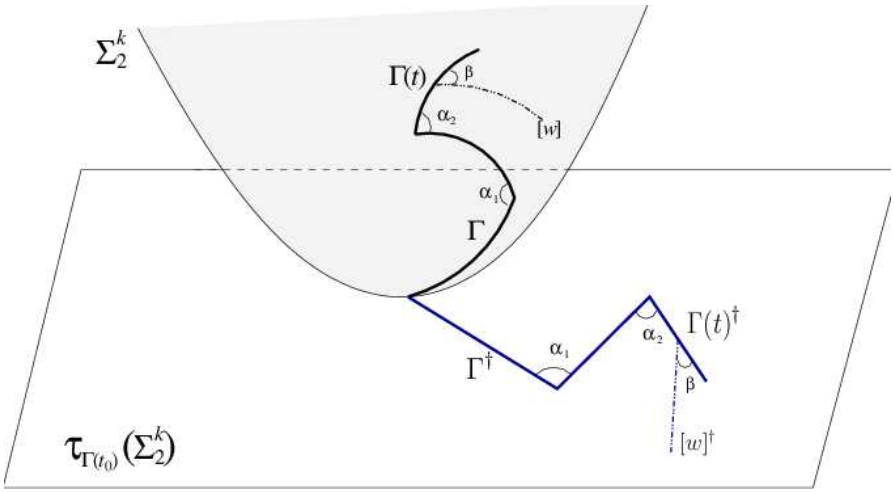
Some potential problems...

- High-dimensional data with low sample sizes - large confidence regions
- No physical constraints (overlap etc.) or physics... Deformation models (Miller, Joshi et al...)
- All based on small variability, and linearised spaces. So, if there IS large movement then one should use the full geometry of the shape space (e.g. see Fletcher...)

Shape space splines

We have also carried out some statistical work on shape space curve fitting for fitting largely dispersed shapes (ILD, Kume and Le, 2005; Evans, ILD and Le, 2005)

Piecewise geodesic shape splines and other shape space curves.



5. INFERENCE - many possibilities...

- Using maximum likelihood or Bayesian inference assuming certain shape distributions.
- Multivariate normal model in the tangent space (to pooled mean)

TWO INDEPENDENT SAMPLE TEST: 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:

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

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] Model assumptions need to hold closely. Lots of parameters to estimate.

Preferred procedure PERMUTATION TEST (Dryden and Mardia, 1993; Bookstein, 1997) or BOOTSTRAP TEST (Amaral et al., 2005).

- Our Bootstrap test has demonstrably better performance for unequal covariance matrices (Behrens-Fisher problem)

ISOTROPY BASED TEST Goodall's F test:

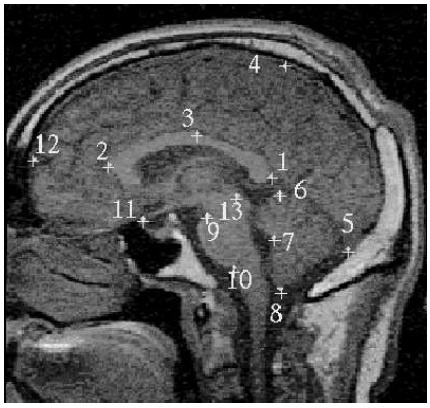
If $\Sigma \propto I$ then

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

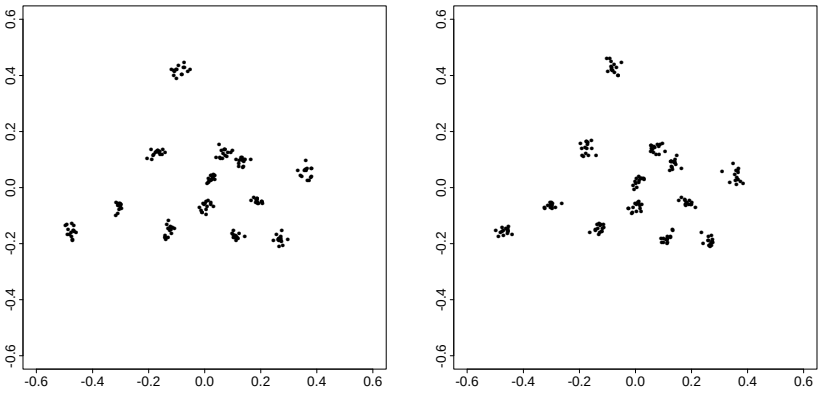
Under H_0: $F \sim F_{M, (n_1 + n_2 - 2)M}$. Very restrictive model.

Again PERMUTATION test or BOOTSTRAP test preferred in practice.

- Challenging example: Schizophrenia data (Bookstein 1997)



Controls and patients



Schizophrenia landmark data example:

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

$n_1 = 14, n_2 = 14$ (rather small)

$M = 2k - 4 = 22$

$F = 1.89$, 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

Hotelling T^2 test has little power here due to small samples.

- But Permutation and Bootstrap tests using the Goodall statistic have good properties.

6. IMAGE AND HISTOGRAM STATISTICS

Note that dimension reduction can also be applied to Z using grey levels, features, functions (e.g. histograms) or combinations of these.

Generative models for grey levels can be constructed from first few PCs - just like the PDMs we have independent Gaussian distributions for the grey level PC scores - Active Appearance Models (Cootes, Taylor et al.). Also, can use CCA or ICA etc.

A further application: intensity histograms (Broadhurst, UNC).

Selected References

Amaral, G. J. A., Dryden, I. L. and Wood, A. T. A. (2005). Pivotal bootstrap methods for k-sample problems in directional statistics and shape analysis, *Technical report, Division of Statistics, University of Nottingham*.

Bookstein, F. L. (1997). Shape and the information in medical images: a decade of the morphometric synthesis, *Computer Vision and Image Understanding*, **66**, 97-118.

Cootes, T. F., Taylor, C. J., Cooper, D. H. and Graham, J., (1992). Training models of shape from sets of examples, *BMVC*,Springer-Verlag, Berlin, 9-18.

Cootes, T. F., Taylor, C. J., Cooper, D. H. and Graham, J., (1994). Image search using flexible shape models generated from sets of examples, *Statistics and Images: Vol. 2*, editor K. V. Mardia,Carfax, Oxford, pages 111-139.

Dryden, I. L. and Mardia, K. V., (1993). Multivariate shape analysis, *Sankhyā*,**A,55**, 460-480.

Dryden, I.L. and Mardia, K.V. (1998). *Statistical shape analysis*, Wiley, Chichester.

Grenander, U. (1994). *General Pattern Theory*, Clarendon Press, Oxford.

Grenander, U. and Miller, M. I. (1994). Representations of knowledge in complex systems (with discussion), *Journal of the Royal Statistical Society, Series B*,**56**, 549-603 .

Fritsch, D., Pizer, S., Yu, L. , Johnson, V. and Chaney, E. (1997). Localization and Segmentation of Medical Image Objects using Deformable Shape Loci, in *International Conference on Information Processing in Medical Imaging (IPMI)*, pp. 127-140.

Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent Components Analysis*, Wiley.

Hyvärinen, A. and Oja, E. (2000). Independent Components Analysis: Algorithms and applications. *Neural Networks*, **13**, 411–430.

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

Kume, A., Dryden, I. L. and Le, H., (2005). Shape space smoothing splines for planar landmark data, *Technical report, Division of Statistics, University of Nottingham*.

Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). *Multivariate Analysis*. Academic Press, London.