Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging

Ian L. Dryden[1,2*], Alexey Koloydenko[3] and Diwei Zhou[1]

[1] School of Mathematical Sciences, University of Nottingham
[2] Department of Statistics, University of South Carolina
[3] Department of Mathematics, Royal Holloway University of London

1 Introduction

The statistical analysis of covariance matrices occurs in many important applications, e.g. in diffusion tensor imaging or longitudinal data analysis. We consider the situation where it is of interest to estimate an average covariance matrix, describe its anisotropy and to carry out principal geodesic analysis of covariance matrices.

In medical image analysis a particular type of covariance matrix arises in diffusion weighted imaging called a diffusion tensor. The diffusion tensor is a 3×3 covariance matrix which is estimated at each voxel in the brain, and is obtained by fitting a physically-motivated model on measurements from the Fourier transform of the molecule displacement density (Basser et al., 1994). A strongly anisotropic diffusion tensor indicates a strong direction of white matter fibre tracts, and plots of measures of anisotropy are very useful to neurologists. A measure that is very commonly used in diffusion tensor imaging is Fractional Anisotropy

$$FA = \left\{ \frac{k}{k-1} \sum_{i=1}^{k} (\lambda_i - \bar{\lambda})^2 / \sum_{i=1}^{k} \lambda_i^2 \right\}^{1/2}, \quad (1)$$

where $0 \leq FA \leq 1$ and λ_i are the eigenvalues of the diffusion tensor matrix. Note that $FA \approx 1$ if $\lambda_1 >> \lambda_i, i > 1$ (very strong principal axis) and $FA = 0$ for isotropy. In diffusion tensor imaging $k = 3$.

2 Distances between covariance matrices

In applications there are several choices of distances between covariance matrices that one could consider, for example see Table 1. If a sample of covariance matrices is available $S_1, \ldots, S_n$ then we can estimate the population covariance matrix by the Fréchet mean $\hat{\Sigma}$, which minimizes $\sum_{i=1}^{n} \mathrm{dist}(S_i, \Sigma)^2$ with respect to Σ.

Estimators $\hat{\Sigma}_E, \hat{\Sigma}_C, \hat{\Sigma}_H, \hat{\Sigma}_L, \hat{\Sigma}_A$ given in Table 1 are straightforward to compute using arithmetic averages. Note that d_S is obtained by optimal rotation/reflection of $chol(S_2)$ onto $chol(S_1)$ using ordinary Procrustes analysis, where $chol(S)$ is the Cholesky decomposition of S. The Procrustes based estimators $\hat{\Sigma}_S, \hat{\Sigma}_F$ involve the use of the Generalized Procrustes Algorithm, which works well in practice (see Dryden et al., 2009). The Riemannian metric estimator $\hat{\Sigma}_R$ uses a gradient descent algorithm which is guaranteed to converge (e.g. see Pennec et al, 2006). In practice it is similar to the log-Euclidean estimator $\hat{\Sigma}_L$ (Arsigny et al., 2007). Mitteroecker

Name	Notation	Form	Estimator
Euclidean	$d_E(S_1, S_2)$	$\|S_1 - S_2\|$	$\hat{\Sigma}_E$
Log-Euclidean	$d_L(S_1, S_2)$	$\|\log(S_1) - \log(S_2)\|$	$\hat{\Sigma}_L$
Riemannian	$d_R(S_1, S_2)$	$\|\log(S_1^{-1/2} S_2 S_1^{-1/2})\|$	$\hat{\Sigma}_R$
Cholesky	$d_C(S_1, S_2)$	$\|\text{chol}(S_1) - \text{chol}(S_2)\|$	$\hat{\Sigma}_C$
Root Euclidean	$d_H(S_1, S_2)$	$\|S_1^{1/2} - S_2^{1/2}\|$	$\hat{\Sigma}_H$
Procrustes size-and-shape	$d_S(S_1, S_2)$	$\inf_{R \in O(k)} \|\text{chol}(S_1) - \text{chol}(S_2)R\|$	$\hat{\Sigma}_S$
Full Procrustes shape	$d_F(S_1, S_2)$	$\inf_{R \in O(k), \beta \in \mathbf{R}} \left\| \frac{\text{chol}(S_1)}{\|\text{chol}(S_1)\|} - \beta \text{chol}(S_2)R \right\|$	$\hat{\Sigma}_F$
Power Euclidean	$d_A(S_1, S_2)$	$\frac{1}{\alpha}\|S_1^{\alpha} - S_2^{\alpha}\|$	$\hat{\Sigma}_A$

Table 1: Some distances between covariance matrices and notation for the corresponding Fréchet mean estimators.

and Bookstein (2009) have used the Riemannian distance d_R and principal co-ordinate analysis for the study of shape covariance structure in morphometrics.

We briefly summarize some of the properties if the distances. All these distances are invariant under simultaneous rotation and reflection of S_1 and S_2, i.e. the distances are unchanged by replacing both S_i by $VS_iV^T,\ V \in O(k), i = 1, 2$. Metrics $d_L(), d_R(), d_F()$ are invariant under simultaneous scaling of $S_i, i = 1, 2$, i.e. replacing both S_i by βS_i. Metric $d_R()$ is also affine invariant, i.e. the distances are unchanged by replacing both S_i by $AS_iA^T, i = 1, 2$ where A is a general $k \times k$ full rank matrix. Metrics $d_L(), d_R()$ have the property that $d(A, I_k) = d(A^{-1}, I_k)$, where I_k is the $k \times k$ identity matrix, and $d_L(), d_R(), d_F()$ are not valid for comparing rank deficient covariance matrices. Finally, there are problems with extrapolation with metric $d_E()$: extrapolate too far and the matrices are no longer positive semi-definite (Arsigny et al., 2007).

An alternative anisotropy measure to FA in (1) is to use the full Procrustes shape distance to isotropy where

$$PA \ = \ \sqrt{\frac{k}{k-1}} d_F(I_k, S) = \left\{ \frac{k}{k-1} \sum_{i=1}^{k} (\sqrt{\lambda}_i - \overline{\sqrt{\lambda}})^2 / \sum_{i=1}^{k} \lambda_i \right\}^{1/2},$$

where $\overline{\sqrt{\lambda}} = \frac{1}{k} \sum \sqrt{\lambda_i}$. We include the scale factor when defining the Procrustes Anisotropy (PA), and so $0 \leq PA \leq 1$, with $PA = 0$ indicating isotropy, and $PA \approx 1$ indicating a very strong principal axis. Another anisotropy measure based on metrics d_L or d_R is the geodesic anisotropy

$$GA = \left\{ \sum_{i=1}^{k} (\log \lambda_i - \overline{\log \lambda})^2 \right\}^{1/2},$$

where $0 \leq GA < \infty$ (Arsigny et al., 2007), which has been used in diffusion tensor analysis in medical imaging with $k = 3$.

In some applications covariance matrices are close to being deficient in rank. For example when FA or PA are equal to 1 then the covariance matrix is of rank 1. The Procrustes metrics can easily deal with deficient rank matrices, which is a strong advantage of the approach. Indeed Kendall's (1984) original motivation for developing his theory of shape was to investigate rank

1 configurations in the context ot detecting 'flat' (collinear) triangles in archaeology. The use of $\hat{\Sigma}_L$ and $\hat{\Sigma}_R$ has strong connections with the use of Bookstein's (1986) hyperbolic shape space and Le and Small's (1999) simplex shape space, and such spaces cannot deal with deficient rank configurations.

3 Applications

3.1 Anisotropy of diffusion tensors

We consider anisotropy of estimated diffusion tensors in the brain obtained from diffusion weighted images (see Dryden et al., 2009). In Figure 1 we see a coronal view of the brain, and the corpus callosum and cingulum can be seen.

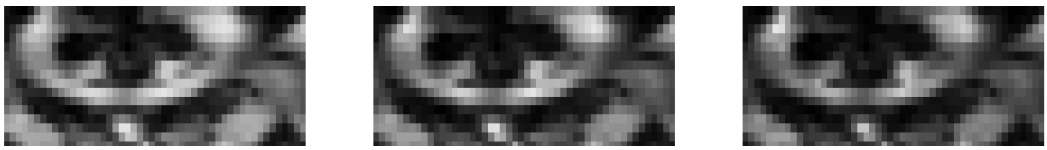

Figure 1: The anisotropy measures (left) FA, (middle) PA, right (GA).

At first sight all three anisotropy measures appear broadly similar. However, the PA image offers more contrast than the FA image in the highly anisotropic region - the corpus callosum. Also, the GA image has rather fewer brighter areas than PA or FA. Due to the improved contrast we believe PA is slightly preferable in this example.

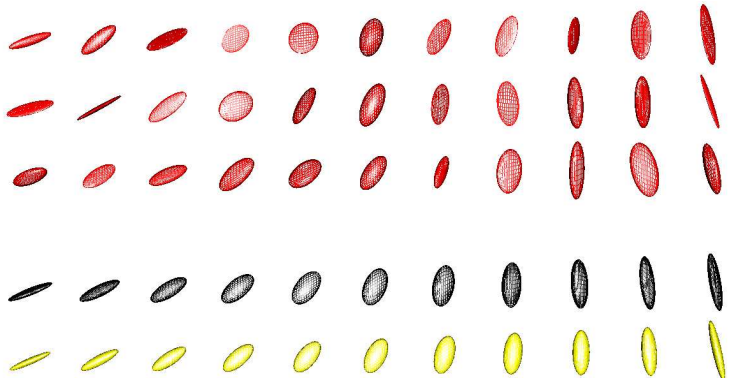

Figure 2: Principal geodesic analysis for covariance matrices. The true geodesic path is given in the penultimate row (black). We then add noise in the three initial rows (red). Then we estimate the mean and find the first principal component (yellow), displayed in the bottom row. (Colour figures will be included in the proceedings available online).

3.2 Principal geodesics of covariance matrices

We consider now an example estimating the principal geodesics of the covariance matrices $S_1, \ldots, S_n$ using the Procrustes size-and-shape metric d_S (see Dryden et al., 2009). In Figure 2, we consider a true geodesic path (black) and evaluate 11 equally spaced covariance matrices along this path. We then add i.i.d. Gaussian noise in the tangent space for three separate

realisations of noisy paths (in red). The overall mean $\hat{\Sigma}_S$ is computed based on all the data ($n = 33$), and then the Procrustes size-and-shape tangent space co-ordinates are obtained based on the Cholesky decompositions of the covariance matrices. The first principal component loadings are computed and projected back to give an estimated minimal geodesic in the covariance matrix space. We plot this path in yellow by displaying 11 covariance matrices along the path. It can be seen that the estimated principal geodesic is very similar to the true geodesic path here. Other extensions include curve fitting through paths of covariance matrices using polynomials and geodesics (e.g. see Evans et al., 2009, for some examples of shape curves).

4 Conclusions

Methodology for estimation and inference in the space of covariance matrices has application in many areas, including diffusion tensor imaging, structural tensor analysis in computer vision, and modelling longitudinal data with Bayesian and random effect models. There are many choices of metric available, each with its advantages. The use of the Procrustes size-and-shape metric d_S is particularly appropriate when the covariance matrices are close to being deficient in rank.

References

Arsigny, V., Fillard, P., Pennec, A. and Ayache, N. (2007). Geometric Means in a Novel Vector Space Structure on Symmetric Positive-Definite Matrices, *SIAM Journal on Matrix Analysis and Applications*, **29**, 328–347.

Basser, P. J., Mattiello, J., and Le Bihan, D. (1994). Estimation of the effective self-diffusion tensor from the NMR spin echo. *J Magn Reson B.*, **103**, 247–254.

Bookstein, F. L. (1986). Size and shape spaces for landmark data in two dimensions (with discussion). *Statistical Science*, **1**, 181–242.

Dryden, I. L., Koloydenko, A., and Zhou, D. (2009). Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *Annals of Applied Statistics*. To appear.

Evans, K., Dryden, I. L., and Le, H. (2009). Shape curves and geodesic modelling. Technical report, Division of Statistics, University of Nottingham. Submitted for publication.

Fillard, P., Arsigny, V., Pennec, X., and Ayache, N. (2007). Clinical DT-MRI estimation, smoothing and fiber tracking with log-Euclidean metrics. *IEEE Transactions on Medical Imaging*, **26**, 1472–1482.

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

Le, H. and Small, C. G. (1999). Multidimensional scaling of simplex shapes. *Pattern Recognition*, **32**, 1601–1613.

Mitteroecker, P. and Bookstein, F. L. (2009). The ontogenetic trajectory of the phenotypic covariance matrix, with examples from craniofacial shape in rats and humans. *Evolution*, **63**, 727–737.

Pennec, X., Fillard, P. and Ayache, N. (2006). A Riemannian Framework for Tensor Computing, *Int. J. Comput. Vision*, **66**, 41–66.