Statistical shape analysis using Non-Euclidean Metrics

Rasmus Larsen [a],[*] Klaus Baggesen Hilger [a]

[a] *Informatics and Mathematical Modelling, Technical University of Denmark, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, Denmark*

Abstract

The contribution of this paper is the adaptation of data driven methods for non-Euclidean metric decomposition of tangent space shape coordinates. The basic idea is to extend principal component analysis (PCA) to take into account the noise variance at different landmarks and at different shapes. We show examples where these non-Euclidean metric methods allow for easier interpretation by decomposition into meaningful modes of variation. The extensions to PCA are based on adaptation of maximum autocorrelation factors and the minimum noise fraction transform to shape decomposition. A common basis of the methods applied is the assessment of the annotation noise variance at individual landmarks. These assessments are based on local models or repeated annotations by independent operators. We show that the Molgedey-Schuster independent component analysis is equivalent to the maximum autocorrelation factors. Finally, the different subspace methods are compared using a probabilistic formulation based on their ability to represent the data.

Key words: Non-Euclidean metric, maximum autocorrelation factors, independent components, relative warps, minimum noise fractions, active shape models, probabilistic PCA
PACS: , x.x

---

[*]

Email addresses: `rl@imm.dtu.dk` (Rasmus Larsen), `kbh@imm.dtu.dk` (Klaus Baggesen Hilger).
URLs: `www.imm.dtu.dk/ rl` (Rasmus Larsen), `www.imm.dtu.dk/ kbh` (Klaus Baggesen Hilger).

1 Introduction

For the analysis and interpretation of multivariate observations a standard method has been the application of principal component analysis (PCA) to extract latent variables. Cootes et al. applied PCA to the analysis of tangent space shape coordinates (1). For various purposes different procedures for PCA using non-Euclidean metrics have been proposed. The maximum autocorrelation factor (MAF) transform proposed by Switzer (2) defines maximum spatial autocorrelation as the optimality criterion for extracting linear combinations of multispectral images. Contrary to this, PCA seeks linear combinations that exhibit maximum variance. Because imaged phenomena often exhibit some sort of spatial coherence, spatial autocorrelation is often a better optimality criterion than variance. We have previously adapted the MAF transform for analysis of tangent space shape coordinates (3). Green et al. (4) introduced the noise adjusted PCA or the minimum noise fraction (MNF) transformations to decompose multispectral satellite images. The MNF transform is a PCA in a metric space defined by a noise covariance matrix estimated from the data. For image data the noise process covariance is conveniently estimated using spatial filtering. Hilger et al (5) applies the MNF transform to texture modelling in active appearance models (6). Bookstein proposed using bending energy and inverse bending energy as metrics in the tangent space (7). Using the bending energy puts emphasis on the large scale variation, using inverse bending energy puts emphasis on small scale variation.

In this paper we will apply a series of non-Euclidean metric tangent space 2D shape decompositions and compare them in terms of their ability to represent the data using a probabilistic formulation.

2 Methods

In the following two sections we will describe how to use two methods, MAF (2) and MNF (4), for decomposing the tangent space coordinates of a set of shapes into a low-dimensional subspace. The tangent space coordinates are obtained by a generalized Procrustes alignment (8; 9) followed by a projection of the full Procrustes coordinates into the tangent space to the shape space at the full Procrustes mean (e.g. (10)). Let the tangent space coordinates, $\mathbf{x}_i = (x_{i11}, \ldots, x_{i1n}, \ldots, x_{id1}, \ldots, x_{idn})^T$, for shapes $i = 1, \ldots, p$ with $j = 1, \ldots, n$ landmarks in $d \in \{2, 3\}$ dimensions be organised in a $p \times dn$ data matrix $\boldsymbol{X} = [\mathbf{x}_1 \mathbf{x}_2 \ldots \mathbf{x}_p]^T$. Let the Procrustes (sample) mean shape be denoted $\boldsymbol{x}$ and let it be centered on (0,0) in 2D and (0,0,0) in 3D. Further, let the origin of the tangent space coordinate system be the mean shape, then $\boldsymbol{X}$ is doubly centered, i.e. columns as well as rows sum to 0. Additionally, it
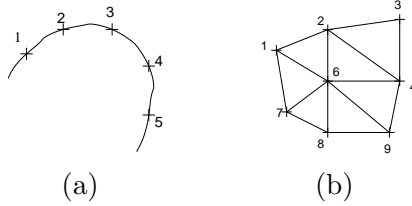
Fig. 1. (a) Landmarks in 2D, landmarks number 2 and 4 and first order neighbours of landmark number 3. (b) Landmarks in 3D, landmarks number {1,2,4,7,8,9} are the first order neighbours of landmark number 6.

is assumed that the landmarks are sampled on curves (in 2D) and surfaces (in 3D) allowing for definition of neighbouring landmarks, i.e. in terms of the order along a curve or through a triangulation of landmarks on a surface (cf. Fig. 1).

In the following we will consider Q-mode analyses of the matrix $\boldsymbol{X}$. In the case of principal component analysis this is an eigenvalue decomposition of the covariance matrix estimated from observations $\mathbf{z}_{lj} = (x_{1lj}, \ldots, x_{plj})^T$, for $j = 1, \ldots, n$, $l = 1, \ldots, d$. These $\mathbf{z}_{lj}$ are vectors of a landmark coordinate observed over the set of shapes. The maximum likelihood estimator of the covariance matrix is

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{p}\boldsymbol{X}\boldsymbol{X}^T = \boldsymbol{V}\boldsymbol{\Lambda}^2\boldsymbol{V}$$

here $\boldsymbol{\Lambda}^2$ is a diagonal matrix containing the eigenvalues of $\hat{\boldsymbol{\Sigma}}$, and $\boldsymbol{V}$ contains the corresponding conjugate eigenvectors. A point distribution model then consists of retaining the $t \leq \min(nd, p)$ first principal components. Deviations from the Procrustes mean (in tangent space) can then be modelled by

$$\mathbf{x} = \boldsymbol{X}^T\boldsymbol{V}'\boldsymbol{b} \tag{1}$$

where $\boldsymbol{V}'$ is a matrix consisting of the first t columns of $\boldsymbol{V}$, and $\boldsymbol{b}$ defines a set of t parameters of the deformable model.

2.1 Maximum autocorrelation factors

Let the spatial covariance function of a multivariate stochastic variable, $\boldsymbol{Z}_k$, where k denotes spatial position and Δ a spatial shift, be $\Gamma(\Delta) = \text{Cov}\{\boldsymbol{Z}_k, \boldsymbol{Z}_{k+\Delta}\}$. Evidently $\boldsymbol{\Gamma}^T(\Delta) = \boldsymbol{\Gamma}(-\Delta)$. Then by letting the covariance matrix of Z_k be $\boldsymbol{\Sigma}$ and defining the covariance matrix $\boldsymbol{\Sigma}_\Delta = D\{\boldsymbol{Z}_k - \boldsymbol{Z}_{k+\Delta}\}$, we find

$$\boldsymbol{\Sigma}_\Delta = 2\boldsymbol{\Sigma} - \boldsymbol{\Gamma}(\Delta) - \boldsymbol{\Gamma}(-\Delta) \tag{2}$$

$\boldsymbol{\Sigma}_\Delta$ is the covariance matrix of the difference process in lag Δ. We are now able to compute the covariance between a linear combination of the original variables and the shifted variables

3

$$\text{Cov}\{\boldsymbol{w}^T\boldsymbol{Z}_k, \boldsymbol{w}^T\boldsymbol{Z}_{k+\Delta}\} = \boldsymbol{w}^T\boldsymbol{\Gamma}(\Delta)\boldsymbol{w} = \boldsymbol{w}^T\boldsymbol{\Gamma}^T(\Delta)\boldsymbol{w}$$

$$= \frac{1}{2}\boldsymbol{w}^T(\boldsymbol{\Gamma}(\Delta) + \boldsymbol{\Gamma}(-\Delta))\boldsymbol{w}$$

$$= \boldsymbol{w}^T(\boldsymbol{\Sigma} - \frac{1}{2}\boldsymbol{\Sigma}_\Delta)\boldsymbol{w}, \tag{3}$$

Thus the autocorrelation in shift Δ of a linear combination of Z_k is

$$\text{Corr}\{\boldsymbol{w}^T\boldsymbol{Z}_k, \boldsymbol{w}^T\boldsymbol{Z}_{k+\Delta}\} = 1 - \frac{1}{2}\frac{\boldsymbol{w}^T\boldsymbol{\Sigma}_\Delta\boldsymbol{w}}{\boldsymbol{w}^T\boldsymbol{\Sigma}\boldsymbol{w}} = 1 - \frac{1}{2}R(\boldsymbol{w}). \tag{4}$$

In order to maximize this correlation we must minimize the Rayleigh coefficient, $R(\boldsymbol{w})$. This is obtained by choosing as $\boldsymbol{w}$ the conjugate eigenvector corresponding to the least generalized eigenvalue of $\boldsymbol{\Sigma}_\Delta$ wrt. $\boldsymbol{\Sigma}$. The MAF transform is given by the set of conjugate eigenvectors of $\boldsymbol{\Sigma}_\Delta$ wrt. $\boldsymbol{\Sigma}$, $W = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m]$, corresponding to the eigenvalues $\kappa_1 \leq \cdots \leq \kappa_m$ (2). The resulting new variables are ordered so that the first MAF is the linear combination that exhibits maximum autocorrelation. The ith MAF is the linear combination that exhibits the highest autocorrelation subject to it being uncorrelated to the previous MAFs. The autocorrelation of the ith component is $1 - \frac{1}{2}\kappa_i$. We assume first and second order stationarity of the data.

One problem now arises, namely, how should we choose Δ. Switzer suggests that we estimate Σ_Δ for a shift in lag 1. Blind source separation by independent component analysis using the Molgedey-Schuster (MS-ICA) algorithm (11) is equivalent to MAF, as is shown in the Appendix. The purpose of this algorithm is to separate independent signals from linear mixings. MS-ICA does this by exploiting differences in autocorrelation structure between the independent signals. Kolenda et al. (12) use an iterative procedure for identifying the optimal lags based on the sum of pairwise absolute differences between the autocorrelations of the estimated independent components. In this study we use Switzers original suggestion. This is based on the assumption that the noise is separated from the interesting latent variables in terms of autocorrelation already in lag 1.

For shape analysis decomposition of the datamatrix $\boldsymbol{X}$ using MAF is carried out in Q-mode. In 2D the difference process covariance matrix $\boldsymbol{\Sigma}_\Delta$ is estimated from the lag 1 difference process of landmark coordinates along the contours of the object as sketched in Fig. 1(a). In 3D we estimate the difference process covariance matrix from the differences between the landmark coordinates and a plane fitted to the landmarks in a kth-order neighbourhood.

2.2 Minimum noise fractions

As before we consider a multivariate stochastic variable, $\boldsymbol{Z}_k$. We assume an additive noise structure $\boldsymbol{Z}_k = \boldsymbol{S}_k + \boldsymbol{N}_k$, where $\boldsymbol{S}_k$ and $\boldsymbol{N}_k$ are uncorrelated signal and noise components, with covariance matrices $\boldsymbol{\Sigma}_S$ and $\boldsymbol{\Sigma}_N$, respectively. Thus $\text{Cov}\{\boldsymbol{Z}_k\} = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_S + \boldsymbol{\Sigma}_N$. By defining the signal-to-noise ratio (SNR) as the ratio of the signal variance and the noise variance we find for a linear combination of $\boldsymbol{Z}_k$

$$\text{SNR} = \frac{V\{\boldsymbol{w}^T \boldsymbol{S}_k\}}{V\{\boldsymbol{w}^T \boldsymbol{N}_k\}} = \frac{\boldsymbol{w}^T \boldsymbol{\Sigma}_S \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{\Sigma}_N \boldsymbol{w}} = \frac{\boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}}{\boldsymbol{w}^T \boldsymbol{\Sigma}_N \boldsymbol{w}} - 1 \qquad (5)$$

So the minimum noise fractions are given by the set of conjugate eigenvectors of $\boldsymbol{\Sigma}$ wrt. $\boldsymbol{\Sigma}_N$, $W = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_m]$, corresponding to the eigenvalues $\kappa_1 \geq \cdots \geq \kappa_m$ (4). The resulting new variables are ordered so that the first MNF is the linear combination that exhibits maximum SNR. The ith MNF is the linear combination that exhibits the highest SNR subject to it being uncorrelated to the previous MNFs. The SNR of the ith component is $\kappa_i - 1$.

The central problem in the calculation of the MNF transformation is the estimation of the noise with the purpose of generating a covariance matrix that approximates $\boldsymbol{\Sigma}_N$. Usually the spatial nature of the data is utilized and the noise is approximated by the difference between the original measurement and a spatially filtered version or a local parametric function (e.g. plane, quadratic function). If repeated measurements are available the noise covariance matrix is readily estimated from these.

If the matrices in Equations (4) and (5) are singular the solution must be found in the affine support of the matrix in the denominator, e.g. by means of a generalized singular value decomposition.

2.3 Relative Warps

Relative warps is a method proposed by Bookstein (7) that utilizes a non-Euclidean metric. The covariance of the data matrix is analysed with respect to the bending matrix of the estimated mean shape. Bookstein proposes to use the bending matrix as a matrix for analysis of large scale variations, and the inverse bending matrtix for analysis of small scale variation. The bending matrix is solely determined from the mean shape and thus holds no information of variation across the dataset.

Following Minka (13) we use a probabilistic principal component analysis model for choice of dimensionality. Let a multivariate response $\boldsymbol{X}$ of p dimensions be modelled by a linear combinations of a set of basis vectors $\boldsymbol{w}_i$, $i = 1, \ldots, t$ plus noise

$$\boldsymbol{X} = \sum_{i=1}^{t} \boldsymbol{w}_i b_i + \boldsymbol{\mu} + \boldsymbol{\epsilon} = \boldsymbol{W}\boldsymbol{b} + \boldsymbol{\mu} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \in N(\boldsymbol{0}, \boldsymbol{\Sigma_N}) \tag{6}$$

where $\boldsymbol{b}$ has dimension $t < p$. The vector $\boldsymbol{\mu}$ defines the mean of $\mathbf{x}$, while $\boldsymbol{W}$ and $\boldsymbol{\Sigma}_N$ defines its variance. For PCA the noise variance is spherical, i.e. $\boldsymbol{\Sigma}_N = v\boldsymbol{I}_p$. Furthermore, we assume a spherical Gaussian prior density for $\boldsymbol{b}$, $\boldsymbol{b} \in N(\boldsymbol{0}, \boldsymbol{I}_t)$. For this model the maximum likelihood estimators for the model parameters given observations $\mathbf{x}_i$, $i = 1, \ldots, n$ are

$$\hat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \quad \hat{\boldsymbol{W}} = \boldsymbol{U}_t(\boldsymbol{\Lambda}_t - \hat{v}\boldsymbol{I}_t)^{1/2}\boldsymbol{R} \quad \hat{v} = \frac{1}{p-t}\sum_{j=t+1}^{p}\lambda_j$$

Where $\boldsymbol{U}_t$ contains the eigenvectors corresponding to the top t eigenvalues of the maximum likelihood estimate of the dispersion matrix of the observations $\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T$, λ_j is the jth eigenvalue of $\hat{\boldsymbol{\Sigma}}$, the diagonal matrix $\boldsymbol{\Lambda}_t$ contains the corresponding eigenvalues, and $\boldsymbol{R}$ is an arbitrary orthogonal matrix. The likelihood of the data, D, then becomes

$$p(D|\boldsymbol{W}, \boldsymbol{\mu}, \boldsymbol{v}) = $$
$$(2\pi)^{-np/2}|\boldsymbol{W}\boldsymbol{W^T} + \boldsymbol{vI}|^{-n/2}\exp(-\frac{1}{2}tr((\boldsymbol{W}\boldsymbol{W^T} + \boldsymbol{vI})^{-1}\boldsymbol{n}\hat{\boldsymbol{\Sigma}})), \tag{7}$$

Let us instead assume a general unrestricted covariance structure of the noise, which may contain intercorrelated effects. Then it is fairly easily shown that by an initial linear transformation that diagonalises $\boldsymbol{\Sigma}_N$, using the result above, that the maximum likelihood estimate of $\boldsymbol{W}$ consists of the first t minimum noise fraction factors (cf. Eq. (5)).

For a given model the log-likelihood (LL) of the data can be estimated. However, with ever increasing model complexity, better reconstruction of the data is obtained, and thus a corresponding increase in LL is observed. The LL estimates must therefore be penalized by e.g. using the Bayesian information criterion (BIC) or Akaike's "An information criterion" (AIC). BIC and AIC are straightforward to compute, given the probability of the data and the degrees of freedom in the model. In general BIC is more conservative and penalizes harder on model complexity in comparison to AIC. BIC is thus expected to perform better, when the true underlying model is not a member of

the imposed family of models. Let n represent the size of the training data, and let d specify the number of parameters in the imposed model, then BIC and AIC reduces to

$$\text{AIC} = -2\log(p(D|\boldsymbol{W}, \boldsymbol{\mu}, \boldsymbol{\Sigma_N})) + 2d/n$$
$$\text{BIC} = -2\log(p(D|\boldsymbol{W}, \boldsymbol{\mu}, \boldsymbol{\Sigma_N})) + d\log(n).$$

where $d = (t+1)p$.

3 Materials

We demonstrate the properties of the techniques that we propose on a dataset consisting of 2D annotations of the outline of the right and left lung from 115 standard PA chest radiographs. The chest radiographs were randomly selected from a tuberculosis screening program and contained normal as well as abnormal cases. The annotation process was conducted by identification of three anatomical landmarks on each lung outline followed by equidistant distribution of pseudo landmarks along the 3 resulting segments of the outline. In Fig. 2(b) the landmarks used for annotation are shown. Each lung field is annotated independently by two observers - Dr. Bram van Ginneken and Dr. Bart M. ter Haar Romeny. The dataset was supplied to us by Dr. Bram van Ginneken. For further information the reader is refered to the Ph.D. thesis of van Ginneken (14).

4 Results

We intend to use the annotation by two independent observers to estimate the annotation uncertainty. Initially the lung annotations are aligned to a common reference frame by concatenating the annotations of the two observers and performing a generalized Procrustes analysis (GPA) (8; 9). Now we can compute the differences between the two sets of annotations and estimate an inter-observer covariance matrix of the landmark coordinates.

These intercorrelations of landmark coordinates may be visualized by showing the correlation matrix of the coordinates of all landmarks across the dataset. However, this representation is not invariant with respect to rotation of the frame of reference. Such a rotation will shift the correlations between the x and y coordinates. Also, it is convenient to study the correlations in terms of landmarks. In order to express the correlation between a univariate quantity and a set of covariates we use the multiple correlation coefficient, and to
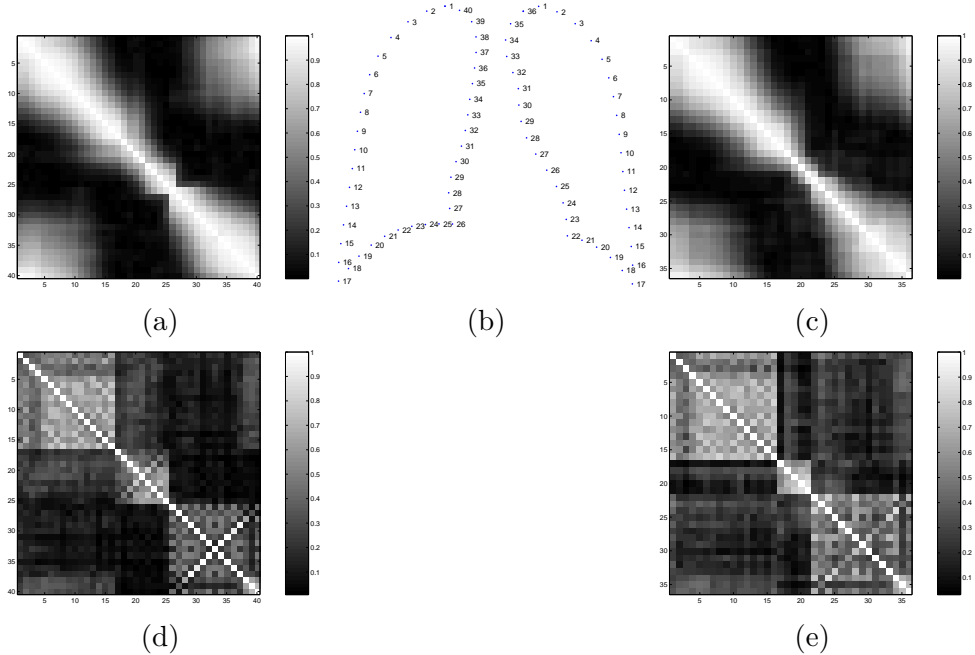
Fig. 2. Landmarks of the left and right lung. Landmark numbers are shown in the middle. The right lung is annotated by 40 landmarks, and the left lung by 36. The anatomical landmarks on the right field are points 1, 17, and 26, on the left field the anatomical landmarks are points 1, 17, and 22. (a),(c) Inter-observer difference canonical correlations between landmarks for the right and left lungs. (d),(e) Inter-neighbour landmark difference canonical correlations between landmark for the right and left lung.

express the correlation between two sets of multivariate quantities (i.e. point coordinates) we may use the canonical correlation. The canonical correlation is determined by use of a canonical correlation analysis (15). This procedure rotates the scattering of two sets of variables (in casu landmarks i and j across the dataset) individually, such that maximum correlation is obtained along a projection axis. This correlation is the canonical correlation.

In Fig. 2 we see these correlations for the right and left lung. The inter lung correlations are neglible. For both set of lungs we see a high degree of correlation along the landmarks of the curved top outline of the lungs. For both lungs landmark 1 is the top point. Again for both lungs there is no or little correlation across the two anatomical landmarks that delimit the bottom segment of the outlines.

The inter-observer covariance matrix defines one sensible metric to use when decomposing the shape variability. This would put less emphasis of landmarks with high annotation variance and more emphasis on landmarks with low annotation variance, and result in a MNF transform. As an alternative to assessing the interobserver differences we may consider the covariance of the difference of neighbouring landmarks. The correlation structure of these are
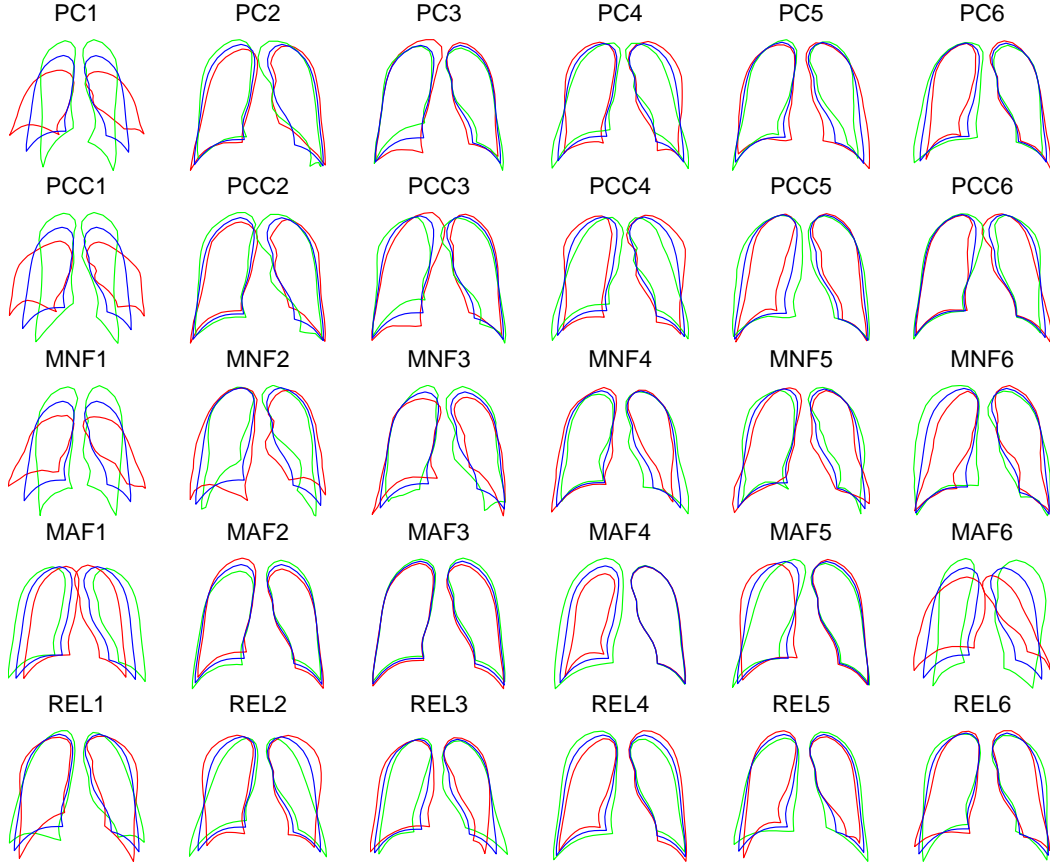
Fig. 3. The 6 most important principal components (PC), principal components on a standardized dataset (PCC), minimum noise fractions (MNF), maximum auto-correlation factors (MAF), and relative warps (REL). The blue curve is the mean shape, and the green and red curves represent ±5 standard deviations as observed in the training set.

also shown in Fig. 2. Here the partitioning of landmarks in three segments for each lung is more pronounced. Using this covariance as metric corresponds to the MAF transform.

In Fig. 3 the 6 most important PCs, PC's on a standardized dataset (PCC) MNFs, MAFs, and relative warps (REL) are shown. The relative warps use the bending matrix of the estimated mean shape as metric and should thus emphasize large scale variations. The noise covariance for the MNF transform is based on the differences between the annotations by the independent operators. The PCs and PCCs are fairly similar, but the MNFs, MAFs, and RELs are different. The latter three all represent uses of metrics that are significantly different from the Euclidean one. The first MNF is an aspect ratio variation, and the following 5 MNF's seems to be a mix of the first PCs. The first MAF is also an aspect ratio variation, and the following MAF's also have evident large scale interpretations. In particular, MAF4 is the relative size of the lungs. The relative warps also give various large scale variations but they

are not as easily interpretable as the MAFs.

In order to further investigate and quantitatively compare the 5 different ways of obtaining a low dimensional shape model we will make an evaluation based on the probabilistic PCA model (13). For each of the 5 methods the loglikelihood, BIC and AIC are computed for all model sizes. The results are shown in Fig. 4. As expected the LL increases with increasing model complexity, i.e. larger models reconstruct the data better. This analysis is therefore inconclusive. More information can be derived from the BIC and AIC plots These regularize the LL by a punishment on the number of free parameters used in the models. The BIC punishes this harder with increasing number of observations than AIC. The analyses confirm the wellknown facts that AIC will tend to choose too complex models if the correct model is not in the model class. In this case the AIC does not reach a minimum for model sizes below 50.

Using BIC we see from the Fig. 4 that for PC the optimal model size is 30. MAF and MNF both have optimal model sizes of 21. Furthermore, MAF as well as MNF perform better at 21 modes than PC at 30 modes in terms of representing the data. The reason that MAF and MNF perform better than PC using the PCA likelihood in Eq. (7) is that the noise covariance is not diagonal as we can see in Fig. 2. MAF and MNF both transform into spaces where estimates of the noise are diagonal. Relative warps and principal components on standardized data are both inferior to PC. MAF and MNF components are not orthogonal in a metric space. However, we have neglected small off-diagonal correlations between MAFs and MNFs when computing AIC, BIC and LL.

5   Conclusion

We have demonstrated a series of data driven methods for constructing non-Euclidean metric linear decompositions of the tangent space shape variability in 2D and 3D. We have demonstrated ways of constructing such a metric based on repeated measurements as well as by use of the spatial nature of the outline and surface models considered. It turns out that the MAF and MNF transforms are superior in terms of interpretability for decomposing large scale variation. These methods are tools for determining un-correlated meaningful modes of variation. Furthermore, we have shown that MAF and MNF models can represent shape variations better and using fewer modes than principal components.
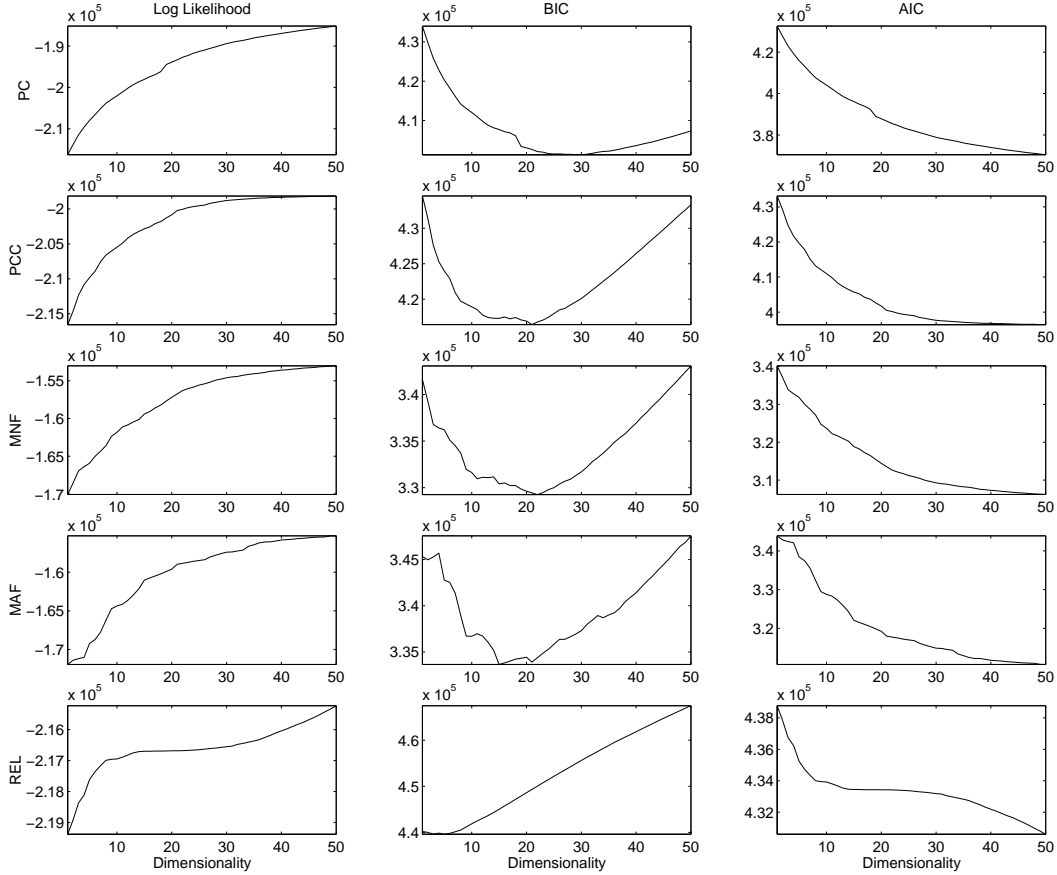
Fig. 4. For all model sizes for the 5 subspace methods applied to the lung dataset: principal components (PC), standardized principal components (PCC), maximum autocorrelation factors (MAF), minimum noise fractions (MNF), and relative warps (REL), the figure shows the loglikelihood, BIC and AIC.

6   Acknowledgements

A   Equivalence of ICA and MAF

It turns out that Molgedey-Schusters algorithm for performing ICA (11) is the same as the MAF analysis (2).

Assuming the linear mixing model of independent component analysis $X = AS$, where X is the $(P \times N)$ data matrix with each row consituting a signal, S is a matrix of the same form as X containing independent signals in the

11

rows, and $\boldsymbol{A}$ is a linear mixing matrix. Furthermore, let $\boldsymbol{X}_\Delta$ and $\boldsymbol{S}_\Delta$ be $\boldsymbol{X}$ and $\boldsymbol{S}$ cyclicly shifted Δ steps rowwise. Then the solution is found by forming

$$\boldsymbol{Q} = \frac{1}{2}\left[\mathbf{X}_\Delta\mathbf{X}^T + \mathbf{X}\mathbf{X}_\Delta^T\right](\mathbf{X}\mathbf{X}^T)^{-1} = \mathbf{A}\left[\frac{1}{2}(\mathbf{S}_\Delta\mathbf{S}^T + \mathbf{S}\mathbf{S}_\Delta^T)(\mathbf{S}\mathbf{S}^T)^{-1}\right]\mathbf{A}^{-1}$$

Due to the independence of the source signals the latter bracketed parenthesis is diagonal. Therefore the mixing matrix can be determined by an eigenvalue decomposition of the matrix $\boldsymbol{Q}$, and the source signals up to a scale factor are estimated by $\mathbf{S} = \mathbf{A}^{-1}\mathbf{X}$. An estimator for the crosscovariance function for a shift Δ is $\frac{1}{N}\mathbf{X}\mathbf{X}_\Delta^T$, and an estimator for the covariance matrix, $\boldsymbol{\Sigma}$, is $\frac{1}{N}\mathbf{X}\mathbf{X}^T$. Therefore using Eq. (2)

$$\boldsymbol{Q} = \frac{1}{2}\left[2\boldsymbol{\Sigma} - \boldsymbol{\Sigma}_\Delta\right]\boldsymbol{\Sigma}^{-1} = \left[\boldsymbol{I} - \frac{1}{2}\boldsymbol{\Sigma}_\Delta\boldsymbol{\Sigma}^{-1}\right]$$

The unity matrix $\boldsymbol{I}$ has no effect on the eigenvectors, so $\boldsymbol{A}$ simply consists of the conjugate eigenvectors of $\boldsymbol{\Sigma}_\Delta$ with respect to $\boldsymbol{\Sigma}$, i.e. the MAF problem given in Eq. (4).

It is easily shown that the MAF transform is invariant to affine transformations. Therefore we may execute a prewhitening beforehand, thus obtaining $\boldsymbol{\Sigma} = \boldsymbol{I}$. Then $\boldsymbol{Q}$ becomes symmetric yielding $\mathbf{A}^{-1} = \mathbf{A}^T$, and the MAF factors become $\mathbf{A}^T\mathbf{X}_{\text{prewhitened}}$, i.e. the independent components.

References

[1] T. F. Cootes, G. J. Taylor, D. H. Cooper, J. Graham, Training models of shape from sets of examples, in: British Machine Vision Conference: Selected Papers 1992, Springer-Verlag, Berlin, 1992, pp. 9–18.

[2] P. Switzer, Min/max autocorrelation factors for multivariate spatial imagery, in: L. Billard (Ed.), Computer Science and Statistics, Elsevier Science Publishers B.V. (North Holland), 1985, pp. 13–16.

[3] R. Larsen, H. Eiriksson, M. B. Stegmann, Q-MAF shape decomposition, in: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2001, 4th International Conference, Utrecht, The Netherlands, Vol. 2208 of Lecture Notes in Computer Science, Springer, 2001, pp. 837–844.

[4] A. A. Green, M. Berman, P. Switzer, M. D. Craig, A transformation for ordering multispectral data in terms of image quality with implications for noise removal, IEEE Transactions on Geoscience and Remote Sensing 26 (1) (1988) 65–74.

[5] K. B. Hilger, M. B. Stegmann, R. Larsen, A noise robust statistical texture model, in: T. Dohi, R. Kikinis (Eds.), Medical Image Computing and Computer-Assisted Intervention - MICCAI 2002, 5th International Con-

ference, Tokyo, Japan, Vol. 2488 of Lecture Note in Computer Science, 2002, pp. 444–451.

[6] T. F. Cootes, G. J. Edwards, C. J. Taylor, Active appearance models, in: Proceedings of the European Conf. On Computer Vision, Springer, 1998, pp. 484–498.

[7] F. L. Bookstein, Morphometric tools for landmark data, Cambridge University Press, 1991, 435 pp.

[8] J. C. Gower, Generalized Procrustes analysis, Psychometrika 40 (1975) 33–50.

[9] C. Goodall, Procrustes methods in the statistical analysis of shape, Journal of the Royal Statistical Society, Series B 53 (2) (1991) 285–339.

[10] I. L. Dryden, K. Mardia, Statistical Shape Analysis, John Wiley & Sons, Chichester, 1998, xx + 347 pp.

[11] L. Molgedey, H. G. Schuster, Separation of a mixture of independent signals using time delayed correlations, Physical Review Letters 72 (23) (1994) 3634–3637.

[12] T. Kolenda, L. K. Hansen, J. Larsen, Signal detection using ica: Application to chat room topic spotting, in: Lee, Jung, Makeig, Sejnowski (Eds.), Proceedings of the 3rd International Conference on Independent Components Analysis and Blind Signal Separation (ICA'2001), San Diego, USA, December 9-13, 2001, pp. 540–545.

[13] T. P. Minka, Automatic choice of dimensionality of PCA, in: T. K. Leen, T. G. Dietterich, V. Tresp (Eds.), Advances in Neural Information Processing Systems 13, MIT Press, 2001, pp. 598–604.

[14] B. van Ginneken, Computer-aided diagnosis in chest radiographs, Ph.D. thesis, Image Sciences Institute, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands, 184 pp. (2001).

[15] H. Hotelling, Relations between two sets of variables, Biometrika 28 (1936) 321–377.