Prior knowledge regularization in statistical medical image tasks

Alessandro Crimi[1], Jon Sporring[1], Marleen de Bruijne[2], Martin Lillholm[3] , Mads Nielsen[1,3]

[1] DIKU, University of Copenhagen, Denmark
[2] Erasmus University Medical Center Rotterdam, The Netherlands,
[3] Nordic Bioscience, Denmark.
ac@nordicbioscience.com,

Abstract. The estimation of the covariance matrix is a pivotal step in several statistical tasks. In particular, the estimation becomes challenging for high dimensional representations of data when few samples are available. Using the standard Maximum Likelihood estimation (MLE) when the number of samples are lower than the dimension of the data can lead to incorrect estimation e.g. of the covariance matrix and subsequent unreliable results of statistical tasks. This limitation is normally solved by the well-known Tikhonov regularization adding partially an identity matrix; here we discuss a Bayesian approach for regularizing the covariance matrix using prior knowledge. Our method is evaluated for reconstructing and modeling vertebra and cartilage shapes from a lower dimensional representation and a conditional model. For these central problems, the proposed methodology outperforms the traditional MLE method and the Tikhonov regularization.

1 Introduction

The covariance matrix estimation is required before performing Principal Component Analysis (PCA), Factor Analysis (FA), regressions and several statistical tasks. If the number of samples is small compared to the dimensionality of the data, then the covariance estimation is poor. For solving this problem, several methods were proposed, the most well-known is the Tikhonov regularization [1] where the covariance matrix is boosted by an identity matrix using a regularization parameter (also called the mixing parameter). The regularization parameter is selected to maximize the expected accuracy of the shrunken estimator, using, e.g. cross validation. The resulting covariance matrix can be shown to outperform the standard MLE estimation. The overall aim of this paper is to improve the accuracy of medical diagnosis of diseases, such as Osteoporosis (OP), Osteoarthritis (OA) and Atherosclerosis. This paper is organized as follows: the next section introduces the point distribution model and the main regularization methods, then we analyze two different kinds of priors, the final part contains experiments about vertebrae and cartilage shape reconstruction and a conditional model for aorta location.

2 Background: Statistical shape model

Statistical shape analysis is a geometrical analysis where a set of shapes, often represented with vectors, are measured statistically to describe geometrical properties from similar shapes [2]. In particular, with a Point Distribution Model [2], a shape $\boldsymbol{x} \in \mathbb{R}^d$ relies on a set of p labeled landmark points $\boldsymbol{x} = [\boldsymbol{p}_1^T, \boldsymbol{p}_2^T, \ldots, \boldsymbol{p}_p^T]^T$ with $\boldsymbol{p}_i \in \mathbb{R}^2$ or $\boldsymbol{p}_i \in \mathbb{R}^3$.

Principal component analysis (PCA) is a relevant tool for studying correlations of movement between groups of landmarks among the training set population. It is an orthogonal projection of the data onto a lower dimensional linear subspace, such that the variance of the projected data is maximized. For the data matrix $\mathbf{X} = [(\mathbf{x}_1 - \mu)|(\mathbf{x}_2 - \mu)|\ldots|(\mathbf{x}_n - \mu)]$, where $\mu = E[\mathbf{x}]$ is the mean of the aligned training examples, the PCA coordinates $\mathbf{Y} = [\mathbf{y}_1|\mathbf{y}_2|\ldots|\mathbf{y}_n]$ are given by:

$$\boldsymbol{Y} = \mathbf{V}^T\mathbf{X}. \tag{1}$$

In the previous expression $\boldsymbol{V}$ is the matrix of the eigenvectors on column form obtained with Singular Value Decomposition (SVD) [3] of the estimated covariance matrix $\hat{\boldsymbol{\Sigma}} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$. Summarizing, statistical shape analysis is performed as follows: given n training shapes a linear model is obtained by first aligning them to a common coordinate system using Procrustes analysis [4], which produces the data matrix $\boldsymbol{X}$. This aligned training set forms a cloud in the d dimensional space, which can be considered a sample from a probability density. Estimating the covariance matrix $\hat{\boldsymbol{\Sigma}}$ and calculating its eigenvectors $\boldsymbol{V}$ and eigenvalues $\boldsymbol{\Lambda}$ gives the linear shape model

$$\tilde{\boldsymbol{x}} = \boldsymbol{\mu} + \boldsymbol{V}_t\boldsymbol{b}, \tag{2}$$

where $\mathbf{V}_t \in \Re^{d\times t}$, whose columns are the eigenvectors corresponding to the t largest eigenvalues, and $\boldsymbol{b} \in \Re^t$ the vector of shape parameters. The performance of such models can depend on the error of estimation of the covariance matrix $\boldsymbol{\Sigma}$ for the PCA.

An estimate of the covariance matrix, $\hat{\boldsymbol{\Sigma}}$, is traditionally obtained as follows. The shapes $\boldsymbol{x}$ are assumed to be distributed according to the normal density,

$$\begin{aligned} &p(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n|\boldsymbol{\Sigma}, \boldsymbol{\mu}) \\ &\quad = \frac{1}{(2\pi)^{\frac{(nd)}{2}}|\boldsymbol{\Sigma}|^{\frac{n}{2}}} exp\left[-\frac{1}{2}\Sigma_{i=1}^n(\boldsymbol{x}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-\mathbf{1}}(\boldsymbol{x}_i - \boldsymbol{\mu})\right], \end{aligned} \tag{3}$$

where $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$. The density of $\boldsymbol{x}$ is called the Likelihood, and its maximum for varying $\boldsymbol{\Sigma}$ is called the Maximum Likelihood Estimate (MLE) [5]. The point of maximum is found to be,

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T, \tag{4}$$

which is slightly biased, but for large n the bias is negligible.

3 Tikhonov regularization

Using the MLE estimation (4), when the dimensionality of the shape space is high, and the available number of samples is small, the resulting matrix may be rank deficient, implying that some eigenvalues have magnitude close to zero, and that the corresponding eigenvectors are arbitrary. This limitation encourages the introduction of more robust covariance matrix estimators [6–10]. Most common is to use a simple form of Tikhonov regularization [1], where nonzero values are added to the diagonal elements of the covariance matrix, e.g.

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{reg}} = \hat{\boldsymbol{\Sigma}} + \lambda \boldsymbol{I}, \tag{5}$$

where λ is a positive regularization parameter. This regularization parameter is not likely to be known in advance, and finding its optimal value can be a cumbersome and computationally heavy task [10]. The analytical methods for covariance estimation [6–9] propose several methods to estimate the regularization parameter using just the sample covariance matrix. However, these methods rely on the estimation of a specific cost function.

4 Gaussian prior regularization

As already showed in [11], starting from the definition of normal distribution in covariance matrix space, and using the Bayes theorem, it is possible to define a covariance matrix as a random variable:

$$p(\boldsymbol{\Sigma}) = (2\pi s^2)^{-\frac{n^2}{2}} \exp\left(-\frac{\|\boldsymbol{\Sigma} - \boldsymbol{B}\|_2^2}{2s^2}\right), \tag{6}$$

where $\boldsymbol{B}$ and s represent the mean and variance of the covariance matrix. The Gaussian prior should ideally be imposed on the space of symmetric, positive definite matrices, as e.g. approximated by the exponential of the norm of the differences of logarithms of matrices [12], but this makes the equation prohibitively complicated, hence we consider (6) as an approximation. The MAP estimate of (6) is found to be a system of third degree polynomials [11], and we seek the solution by the following iterative scheme, estimating the covariance matrix as shown in [11] as

$$\hat{\boldsymbol{\Sigma}}_{t+1} = \frac{1}{n}\left(\boldsymbol{X}\boldsymbol{X}^T - \frac{2\hat{\boldsymbol{\Sigma}}_t(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}} - \boldsymbol{B})^T\hat{\boldsymbol{\Sigma}}_t}{s^2}\right), \tag{7}$$

which we have found to converge for $t \to \infty$, when starting at

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{0}} = \frac{1}{n}\boldsymbol{X}\boldsymbol{X}^T. \tag{8}$$

Following [13], in case of landmark shapes as the data set, we find it useful to define the mean $\boldsymbol{B}$ of the prior distribution of $\hat{\boldsymbol{\Sigma}}$ in the following way:

$$\boldsymbol{B}_{ij} = \exp\left(-\boldsymbol{A}_{\boldsymbol{ij}}\right), \tag{9}$$

where $\boldsymbol{A_{ij}} = \left\| E[\boldsymbol{p}_i] - E[\boldsymbol{p}_j] \right\|_2$, $E[\boldsymbol{p}_i]$ being the i'th component the mean of the aligned training examples.

In case of 3D shapes, such as an m-rep model of cartilage, equation (9) can be extended as

$$\boldsymbol{B_{ij}} = \exp\left(-\boldsymbol{A_{ij}}\right) \exp\left(-\boldsymbol{T_{ij}}\right), \tag{10}$$

where $\boldsymbol{T}$ is the matrix with all the angular distance of the normals of the atom points. Since the matrices defined in equations (9) and (10) consider the relationship between variables, are symmetric, by construction their eigenvalues are positive or equal to zero; we can assume that they are covariance matrices.

5 Wishart prior regularization

A shape representation based on a Wishart distribution priors assume independence between points, hence it is a prior of spatial noise and not shape variation where probably a Gaussian prior is more suitable.

Consider shapes consisting of points,

$$\boldsymbol{x} = [\boldsymbol{p}_1^T, \boldsymbol{p}_2^T, \ldots, \boldsymbol{p}_p^T]^T \tag{11}$$

with $\boldsymbol{x} \in \mathbb{R}^d$, a collection of shapes,

$$\boldsymbol{X} = [(\boldsymbol{x}_1 - \ \boldsymbol{\mu})|(\boldsymbol{x}_2 - \ \boldsymbol{\mu})| \ldots |(\boldsymbol{x}_n - \boldsymbol{\mu})], \tag{12}$$

as a $d \times n$ matrix of reals, ignore the stochastic dependency on the mean, and write

$$\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}^T. \tag{13}$$

When the landmarks $\boldsymbol{p}_i - E[\boldsymbol{p}_i]$ are independently and normal distributed as $N(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$, then $\boldsymbol{S}$ is distributed according to the Wishart distribution [5, Chapter 7],

$$p(\boldsymbol{S}|\boldsymbol{\Sigma}, n) = \frac{|\boldsymbol{S}|^{(n-d-1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{S}\right)\right)}{2^{nd/2}\,|\boldsymbol{\Sigma}|^{n/2}\,\Gamma_d\left(\frac{n}{2}\right)}, \tag{14}$$

where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & 0 & \ldots & 0 \\ 0 & \boldsymbol{\Sigma}_2 & & \\ \vdots & & \ddots & \\ 0 & & & \boldsymbol{\Sigma}_p \end{bmatrix}, \tag{15}$$

and Γ_d is the multivariate Gamma function,

$$\Gamma_d(n) = \pi^{d(d-1)/4} \prod_{i=1}^{d} \Gamma\left(n - \frac{1}{2}(i-1)\right). \tag{16}$$

Using the Bayes theorem, we can write

$$p(\boldsymbol{\Sigma}|\boldsymbol{\Psi}, m) = \frac{|\boldsymbol{\Psi}|^{m/2} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Psi}\boldsymbol{\Sigma}^{-1}\right)\right)}{2^{md/2}\,|\boldsymbol{\Sigma}|^{(m+d+1)/2}\,\Gamma_d\left(\frac{m}{2}\right)}, \tag{17}$$

where $\boldsymbol{\Psi}$ and m are fixed as parameters of the density. The inverted Wishart density originates as the density of $\boldsymbol{S}^{-1}$. Since the evidence is independent on $\boldsymbol{\Sigma}$, we find the MAP estimate as:

$$\hat{\boldsymbol{\Sigma}} = \arg\max_{\boldsymbol{\Sigma}} p(\boldsymbol{S}|\boldsymbol{\Sigma}, n)p(\boldsymbol{\Sigma}|\boldsymbol{\Psi}, m) \tag{18a}$$

$$= \arg\max_{\boldsymbol{\Sigma}} \frac{c\,|\boldsymbol{\Psi}|^{m/2}\,\exp\left(-\frac{1}{2}\mathrm{tr}\left(\left(\boldsymbol{S}+\boldsymbol{\Psi}\right)\boldsymbol{\Sigma}^{-1}\right)\right)}{|\boldsymbol{\Sigma}|^{(n+m+d+1)/2}} \tag{18b}$$

At this point, we can consider $\boldsymbol{\Psi}$ as independent of $\boldsymbol{\Sigma}$, the solution to (18) is found by differentiation as:

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n+m+d+1}\left(\boldsymbol{S}+\boldsymbol{\Psi}\right). \tag{19}$$

Or we can include $\boldsymbol{\Psi} = s^2\boldsymbol{\Sigma}^{-1}$ in the estimation, in which case the maximization results in system of quadratic equations

$$\hat{\boldsymbol{\Sigma}}^2 = \frac{1}{n+2m+d+1}\left(\boldsymbol{S}\hat{\boldsymbol{\Sigma}} + 2s^2\boldsymbol{I}\right). \tag{20}$$

We solve this iteratively as

$$\hat{\boldsymbol{\Sigma}}_{t+1} = \hat{\boldsymbol{\Sigma}}_t - \delta\left(\hat{\boldsymbol{\Sigma}}_t^2 - \frac{1}{n+2m+d+1}\left(\boldsymbol{S}\hat{\boldsymbol{\Sigma}}_t + 2s^2\boldsymbol{I}\right)\right), \tag{21}$$

where δ is a sufficiently small constant to avoid divergence, n and d are respectively still the number of samples and dimension, m is the only user-specified parameter that defines the density of the Wishart distribution and in our experiments was set to 40 for vertebrae and 50 for cartilages.

We call this approach MAP-PCA, and the prior described in (7) as *Normal prior*, while the equation (19) define the *Inverted Wishart* (IWIS) prior, and the equation (20) the *Uncommitted Inverted Wishart* (UIWIS) prior.

6 Experiments

Using high-resolution (full boundary) vertebral shapes from radiographs may lead to reliable results when detecting osteoporotic fragility fractures [14]. However, manual annotation of full boundaries is time consuming. For studies on osteoarthritis, detailed m-rep shape models of cartilage also lead to accurate results. In this case the problem of building a high-resolution model is, however, the computational cost during the automatic segmentation.

In the following sections we give examples of high-resolution shape models reconstructed from the coarse annotations, using the MAP-PCA estimation with the three priors described above, and compare to the result obtained using MLE and Tikhonov estimation.

Given $\boldsymbol{y}$ be an incomplete or lower dimensional shape vector of dimensionality $l < d$, we can obtain the corresponding higher resolution shape $\boldsymbol{x} \in \mathbb{R}^d$ using a linear mapping $\boldsymbol{L} : \mathbb{R}^d \rightarrow \mathbb{R}^l$:

$$\boldsymbol{y} = \boldsymbol{L}\boldsymbol{x}, \tag{22}$$

For our experiments the matrix $\boldsymbol{L}$ is a sampling matrix connected identical points between $\boldsymbol{x}$ and $\boldsymbol{y}$. Since the system is overdetermined, the solution is easily not uniquely defined. In [15], $\boldsymbol{x}$ is obtained minimizing the functional

$$E(\boldsymbol{x}) = \|\boldsymbol{L}\boldsymbol{x} - \boldsymbol{y}\|_2^2 . \tag{23}$$

Since $\boldsymbol{x}$ belongs to the shape model (2), the functional transforms to:

$$E(\boldsymbol{b}) = \|\boldsymbol{Q}\boldsymbol{b} - \boldsymbol{y}\|_2^2 , \tag{24}$$

where $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{V}_t\boldsymbol{\Lambda}_t$, and $\boldsymbol{V}_t$ are the eigenvectors corresponding to the t principal eigenmodes of the covariance matrix estimated with one of the method described in the previous sections. It can be shown that $E(\boldsymbol{b})$ is minimized by $\boldsymbol{b}^* = \boldsymbol{Q}^+\boldsymbol{y}$ [15], where $\boldsymbol{Q}^+$ is the Moore-Penrose pseudo-inverse [3]. Hence $\boldsymbol{x}$ is estimated by

$$\tilde{\boldsymbol{x}} = \boldsymbol{\mu} + \boldsymbol{V}_t\boldsymbol{\Lambda}_t\boldsymbol{b}^* . \tag{25}$$

Experiments are based on the reconstruction of incomplete data and are divided into two groups: vertebrae shapes and cartilage shapes. The eigenvalues and eigenvectors of the covariance matrix computed from the training set through the MLE and the Bayesian methods.

The reconstruction error of vertebra and cartilage shapes between a high resolution shape and a reconstructed version of the same from a lower dimensional version is computed for all the p points of the boundary using the expression:

$$E_{full} = \frac{1}{n}\sum_{i=1}^{p} \left\|\boldsymbol{p}_{i,reconst.} - \boldsymbol{p}_{i,orig.}\right\|_2 \tag{26}$$

The performances of MAP-PCA and MLE methods are compared using mean reconstruction error over all the test shapes for different number of principal eigenmodes.

6.1 Reconstruction of vertebra shapes

For clinical studies on osteoporosis for fracture quantification, we use a 6 points representation of a vertebra, due to the fact that three points in the lower border of the vertebra and 3 points in the upper part of the vertebra can describe the heights measure defined by Genant [16] that is the gold standard for fracture quantification. In order to perform more sophisticated shape analysis, a full contour is needed. Therefore, an extrapolation of the full contour from the six points is a useful initialization for a segmentation algorithm like [17]. During the experiments, the shapes in the training set were the full boundaries made of 52 points, while for the test shapes only 6 out of these 52 points were used. Using the shape model, (25), the corresponding high-resolution vertebral boundary was reconstructed from a low-dimensional test shape.

6.2 Reconstruction of cartilage shapes

We tested the improvement of our covariance estimation also for a medial atoms representation of tibial knee-cartilage, in order to produce a higher resolution m-rep model. We have performed this evaluation by removing atoms from the models and by measuring, how well the interpolation allows for reconstruction of the original model. The cartilage data set is composed of 620 knee MRI scans from 159 subjects including both left and right knees and baseline and follow-up scans from a longitudinal 21-month study. The dimensions of the scans are 256 x 256 pixels with around 110 slices. The population includes healthy and diseased knees with varying degree of OA from both men and women at ages from 21 to 78. The test sets were made to be from 10 and 20 samples, and the relative training sets to be the remaining 610 and 600 samples in a leave-one patient-out fashion.

For each knee, we have a three-dimensional m-rep model of the medial tibial cartilage compartment estimated from a fully automatic segmentation [18, 19]. Figure 1 illustrates a cartilage shape model is illustrated for a cropped knee MRI. We produced a low-resolution lattice by removing nodes from the m-rep

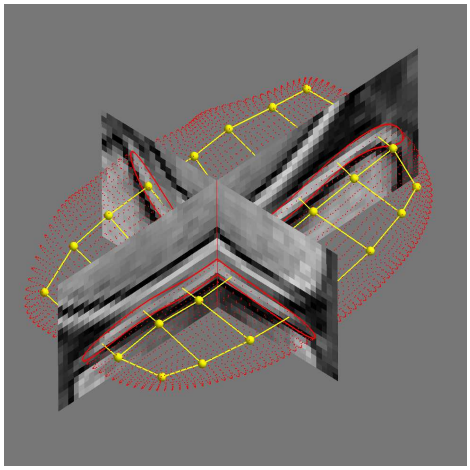

Fig. 1. Sagittal slice from Turb3D T1 MRI. The contours are the manual outlines of femoral and tibial cartilage performed by a radiologist.

representation of the cartilage, and calculated the mean reconstruction error between the original and the reconstructed 3D shapes using (26). To represent a medial lattice for a cartilage sheet, $p = 32$ points were used, hence the dimension is $d = 96$. During the experiments 24 points were removed randomly, and only 8 were left, so the reduced dimensions is $l = 24$.

7 Results

A typical vertebral shape of 52 points (104 dimensions) is depicted in Figure 2(a), while the reduced and reconstructed shape is depicted in Figure 2(b). Figure 2(c)

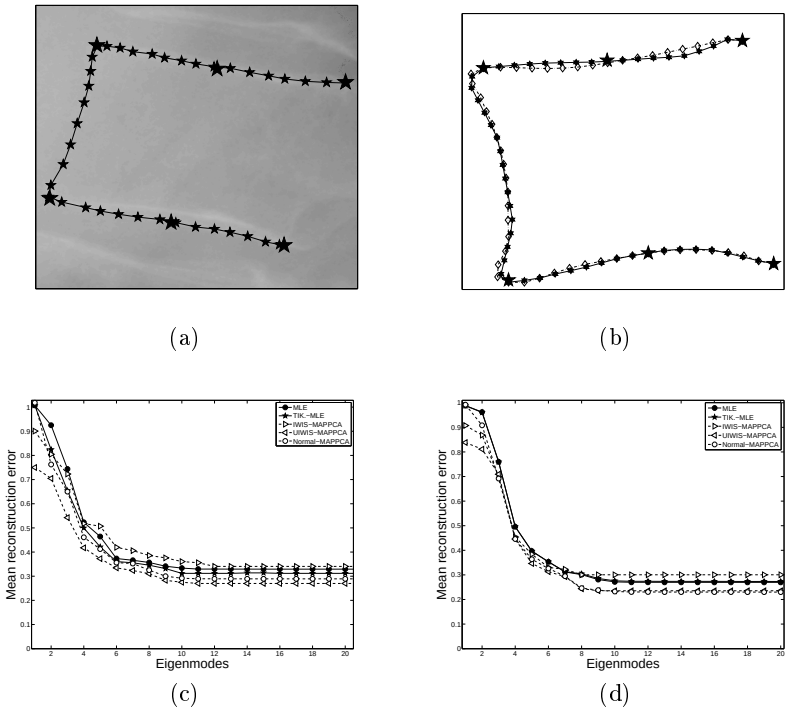

Fig. 2. (a) An image of a vertebra with the shape annotation. (b) An original (continuous line) and reconstructed (dashed line) shape annotation. The shape's 52 points are reconstructed using only the 6 points depicted as the big stars. The vertebra reconstruction error for different number of training shapes: (c) 20 and (d) 40. Here the results with the IWIS and UIWIS are for noisy test shapes with $s = 1$ and $m = 40$ as parameters of the equation (19) and (20). Instead the Normal prior uses $s = 2$ and the Tikhonov regularization parameter is 0.4665. Due to no further changes of the curves, only the first 20 eigenmodes are depicted.

and 2(d) shows the different mean reconstruction errors obtained using (26) for two different sample sizes and when varying t, the number of eigenvalues included in the reconstruction. The vertebra experiments demonstrate that the MAP-PCA method generally improves the reconstruction especially for few samples and a small number of eigenvalues. It seems the Wishart prior performs relatively better than the Normal prior.

For m-rep modeling of tibial knee-cartilage, the result using the first 20 eigenmodes, is shown in Figure 3(a)(b)(c). The mean reconstruction error using the discussed covariance estimators is shown in Figure 3(d)(e). The cartilage shapes show a considerable improvement, when using the Normal prior, and an almost as good improvement, when using the Uncommitted Inverted Wishart prior when using only 10 training shapes.

8 Future experiments

Doing further investigation about regression, we also obtained early good results with a conditional shape model for finding the aorta given the spine. Where the spine model is a collection of the 6 points representation of four lumbar vertebrae, while the aorta model is a sequence of aorta-wall points. For each wall there are 9 points equidistantly sampled along aorta boundaries, hence no sliding along boundaries is necessary to model. An aorta shape can be predicted using an equation similar to (25) where the unknown points of the previous experiments are the aorta points $\mathbf{x_1}$: Figure 4 illustrates the typical result of our method with the lumbar vertebrae (thin continuous line), the annotated aorta (thick continuous line) and the predicted aorta (dashed line).

$$\mathbf{x_1} = \mu_\mathbf{1} + \mathbf{\Sigma_{12}}\mathbf{\Sigma_{22}^{-1}}(\mathbf{S_2} - \mu_\mathbf{2}) \tag{27}$$

and where $\mu_\mathbf{1}$ and $\mu_\mathbf{2}$ are the mean shape of all the aortas and the mean shape of all the spines respectively. $\mathbf{\Sigma_{11}}$ represents the covariance of the aortas, $\mathbf{\Sigma_{22}}$ the covariance of spines, and $\mathbf{\Sigma_{12}}$ and $\mathbf{\Sigma_{21}}$ are the cross-covariance. Here the need is to regularize the covariance matrix $\mathbf{\Sigma_{22}}$ of vertebra.

9 Conclusions

Efficient estimation of covariance matrices is an important task for statistical shape analysis. In this paper we discuss a novel method called MAP-PCA for estimate the covariance matrix in case of small sample size. The matrix obtained with MAP-PCA generally outperforms the traditional MLE method with and without Thikhonov regularization. In addition, the experiments show that the choice of a suitable prior leads to better results. In particular, the Wishart priors assume statistically independent points, and therefore it performs best for variations due to noise and not shape variation. The Normal prior is not limited to zero off-block-diagonal elements and may be used to steer the estimate towards preferred shape variations. Our conclusion is that the choice of prior is related to various factors, by the number of samples available to the shape variance, and that this method is a valid substitute to the Thikhonov regularization where the result is based on the search of the optimal mixing parameter.

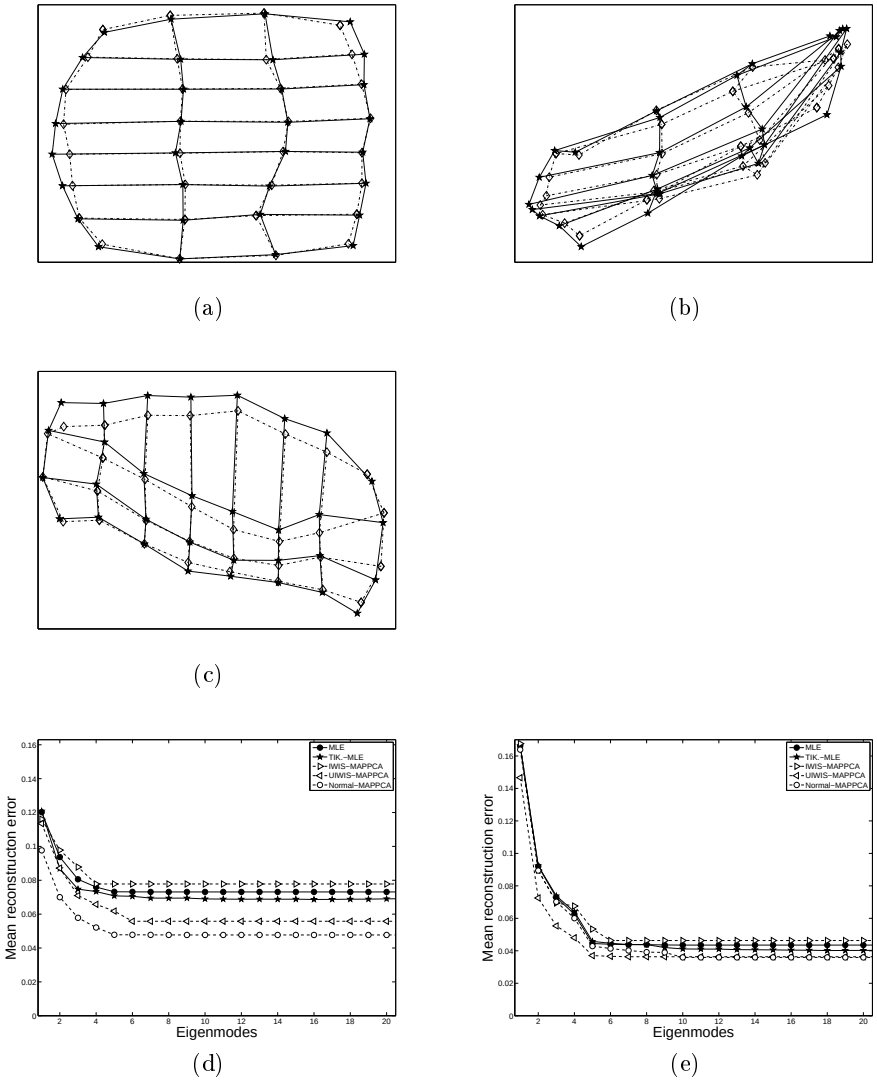

Fig. 3. Overlay of the original and the reconstructed cartilage shape: (a) projection on the plane XY, (b) projection on the plane XZ, and (c) projection on the plane YZ. The continuous line is the original shape and the dashed one is the reconstruction. Here the removed points are chosen randomly. The reconstruction error for different number of training shapes: (d) 10 and (e) 20. Here the results are for noisy test shapes cf. Figure 3, and using $s = 1$ and $m = 50$ in the equation (20). Instead the Normal prior uses $s = 4$ and the Tikhonov regularization parameter is 0.6443. Due to no further changes of the curves, only the first 20 eigenmodes are depicted.

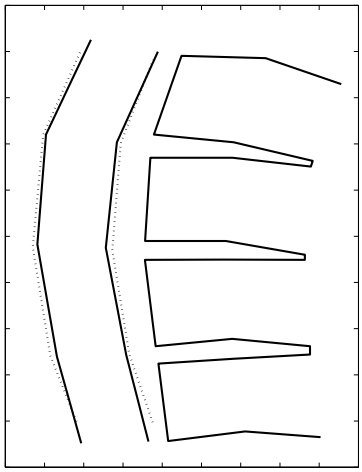

Fig. 4. The typical result of our method: The lumbar vertebrae represented by 6 points(continuous line), the annotated aorta(continuous line) and the predicted aorta (dashed line).

Acknowledgement

We would like to acknowledge the Center for Clinical and Basic Research, Ballerup, Denmark, for providing the annotated vertebra radiographies and the MRI knee scans.

References

1. Tikhonov, A.N., Arsenin, A.N.: Solution of Ill-posed Problems. first edn. Winston & Sons, Washington (1977)
2. Cooper, D., Cootes, T., Taylor, C., Graham, J.: Active shape models, their training and application. Computer Vision and Image Understanding **61** (1995) 38–59
3. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes 3rd Edition: The Art of Scientific Computing. 3rd edn. Cambridge University Press (2007)
4. Bookstein, F.L.: Shape and the information in medical images: A decade of morphometric synthesis. cviu **66**(2) (1997) 97–118
5. Anderson, T.W.: An Introduction to Multivariate Statistical Analysis. 3rd edn. Wiley (2003)
6. Schaefer, J., Strimmer, K.: A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Journal of Statitstical Applications in Genetics and Molecular Biology **4** (2005)

7. Haff, L.R.: Empirical bayes estimation of the multivariate normal covariance matrix. Annals of statistics **8** (1980) 586–597
8. James, W., Stein, C.: Estimation with quadratic loss. In: Proceedings of the fourth Berkely symposium on mathematical statistics and probability. Volume 1. (1961) 361–379
9. Efron, B., Morris, C.: Multivariate empirical bayes and estimation of covariance matrices. Annals of statistics **4** (1976)
10. Friedman, J.H.: Regularized discriminant analysis. Journal of American Statistics Association **84** (1989) 165–175
11. Crimi, A., Ghosh, A., Sporring, J., Nielsen, M.: Bayes estimation of shape-models with application to vertebrae boundaries. In: SPIE Medical Imaging. (2009)
12. Fillard, P., Pennec, X., Arsigny, V., Ayache, N.: Clinical dt-mri estimation, smoothing, and fiber tracking with log-euclidean metrics. IEEE Trans. Med. Imaging **26**(11) (2008) 1472–1482
13. Sporring, J., Jensen, K.H.: Bayes reconstruction of missing teeth. Journal of Mathematical Imaging and Vision **31** (2008) 245–254
14. de Bruijne M.: Shape particle filtering for image segmentation. In: MICCAI, Lecture Notes in Computer Science. (2004) 186–175
15. Blanz, V., Vetter, T.: Reconstructing the complete 3d shape of faces from partial information. Technical Report report of Computer Graphics no. 1, University of Freiburg (2001)
16. Genant, H.K., Jergas, M., Palermo, L., Nevitt, M., Valentin, R.S., Black, D., Cummings, S.R.: Comparison of semiquantitative visual and quantitative morphometric assessment of prevalent and incident vertebral fractures. J Bone Miner Res. **11** (1996) 984–996
17. Iglesias, J., de Bruijne, M.: Semi-automatic segmentation of vertebrae in lateral x-rays using a conditional shape model. Academic Radiology **14** (2007) 1156–1165
18. Dam, E.B., Fletcher, P.T., Pizer, S.M.: Automatic shape model building based on principal geodesic analysis bootstrapping. Medical Image Analysis **12**(2) (2008) 136–151
19. Folkesson, J., Dam, E.B., Olsen, O.F., Pettersen, P., Christiansen, C.: Segmenting articular cartilage automatically using a voxel classification approach. IEEE Transactions on Medical Imaging **26**(1) (2007) 106–115