On prior information in principal component analysis

Václav Šmídl[†],Miroslav Kárný[‡]

[†] TCD, Department of Electric & Electronic Engineering, TCD,[*] smidl@mee.tcd.ie
[‡] Dept. of Adaptive Systems, UTIA, Czech Republic, school@utia.cas.cz

1 Introduction

Principal component analysis (PCA) is used for data compression, noise reduction, and feature extraction purposes. Its usefulness and many advantages are well known. Performance of PCA depends on the amount and characteristics of noise in the observed data. In data with a low signal to noise ratio (SNR), inhomogeneous, or correlated noise, the performance of PCA can be poor.

The problem has been addressed theoretically in several papers [1, 2] with respect to the properties of noise, and an optimal scaling of data for PCA was defined. The authors concluded that the optimal metric can be derived directly from the known covariance matrix of noise, and suggested particular solutions for specific data. However, it is not easily satisfied in practice. This was the motivation for searching for a more practical approach.

We suggest that the covariance matrix of noise - and thus the optimal metric for PCA - can be estimated. However, it is not a trivial task and the estimates have to de determined iteratively. In order to stabilize the iteration process it is necessary to introduce prior knowledge that is included using the Bayesian paradigm. The performance of the method is demonstrated on simulated dynamic image data.

However, the method is easily applicable to any problem where PCA is successfully used.

2 Problem description and proposed solution

The observed image sequence consists of T images having N pixels each, stored column-wise. The images are assumed to be linear combinations of $r \ll \min(N, T)$ underlying images, P $(N \times r)$, weighted by coefficients, Q $(r \times T)$. The observed data D consist of this combination corrupted by an additive zero mean noise E:

$$D = \mu + E = PQ + E$$

The noise is assumed to contain no outlying realizations so that its distribution can be considered normal. Properties of the noise are thus fully characterized by the covariance matrix $\mathcal{C}$. While noise E has NT elements it's covariance matrix $\mathcal{C}$ is huge $NT \times NT$ matrix. It is much larger than the number NT of available data D and thus a restricted covariance structure has to be considered. Usually, common variance $1/\omega > 0$ is assumed for all pixels and times, where ω is the precision parameter. Then, the model of data O becomes

$$D \sim \mathcal{N}(\mu, I_N \otimes I_T \omega^{-1}) \tag{1}$$

The covariance is $\mathcal{C} = (I_N \otimes I_T)\omega^{-1}$, where I_N is the identity matrix and $\otimes$ is the Kronecker product. The use of the precision, ω, instead of the variance simplifies formal manipulations.

The maximum likelihood estimate of μ, of rank r, minimizes the quadratic form in (1) and thus coincides with the PCA estimate [3].

[*] currently working on MOUMIR project

The results are poor when

1. the covariance $\mathcal{C}$ does not have the assumed structure $\Longrightarrow$ more realistic modelling of the noise.
2. the noise level $1/\omega$ is too high compared to the signal values μ. $\Longrightarrow$ more realistic modelling of the signal

Model of noise covariance. Here, the direct extension $\mathcal{C} = I_N \otimes \Omega^{-1}$ of the classical assumption is considered. Where the precision matrix Ω is matrix of size $T \times T$. Elements on the diagonal models changing variances of noise in time. Non-diagonal elements models the noise correlations.

We search for a joint estimator of μ, Ω. It is a non-trivial task as it can be shown that the joint maximum likelihood estimate of μ and Ω does not exist. Thus, it is impossible to separate signal and noise spaces without additional information. We are using simple prior, stating that we expect uncorrelated noise with the same variance for all time moments

$$\Omega \sim W(\gamma I_T, \gamma w)$$

where W denotes Wishart's distribution with parameters γ, w.

Model of signal. The method was originally developed for analysis of dynamic medical image data. The biological processes are relatively slow and thus the observed adjacent images are supposed usually similar. Formally, we suppose that weights Q are smooth curves. The values $Q_{k(t)}$ of k-th curve $k = 1, \ldots, r$ at time $t = 2, \ldots, T$ are related to the preceding values through the simple time-dependent auto-regression

$$Q_{k(t)} \sim \mathcal{N}(a_{t-1} Q_{k(t-1)}, \beta^{-1}), \tag{2}$$

where the coefficients $a = [a_1 \ldots a_{T-1}]$ - approximating the curve evolution - and the precision β are assumed to be common to all curves. The arbitrariness of the initial values $Q_{k(1)}$ is modelled by the flat normal probability density function (p.d.f.) $Q_{k(1)} \sim \mathcal{N}(0, 1/\varepsilon)$ with a small precision ε. These assumptions, applied to $\mu = PQ$ with orthonormal images P, translate into the prior p.d.f. on μ

$$\mu \sim \kappa \mathcal{N}(0, \beta \Delta \Delta')$$

where Δ is the $(T \times T)$ matrix with the non-zero entries $\Delta_{1,1} = \varepsilon^{0.5}$, $\Delta_{t,t} = 1$, $\Delta_{t-1,t} = -a_{t-1}, t = 2, \ldots, T$ and zeros otherwise. The κ is modification of normalizing factor expressing the restriction of μ to the rank r.

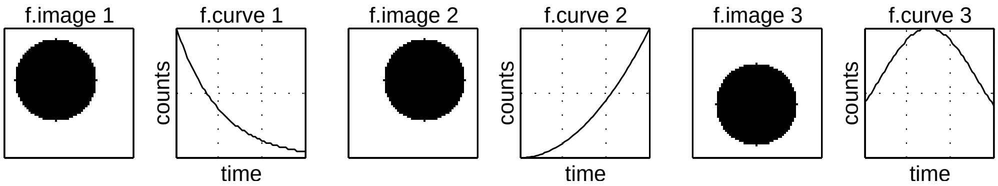
Estimation algorithm. The observation and noise models, together with the chosen prior distribution on unknown parameters $\Theta = (\mu, \Omega, a_1, \ldots, a_{T-1}, \beta, \varepsilon)$ determine the posterior p.d.f. of parameters given by the observations O. The MAP estimate of Θ maximizes the likelihood function given by Bayes rule.

$$f(\Theta|O) \sim f(O|\Theta) f(\Theta)$$

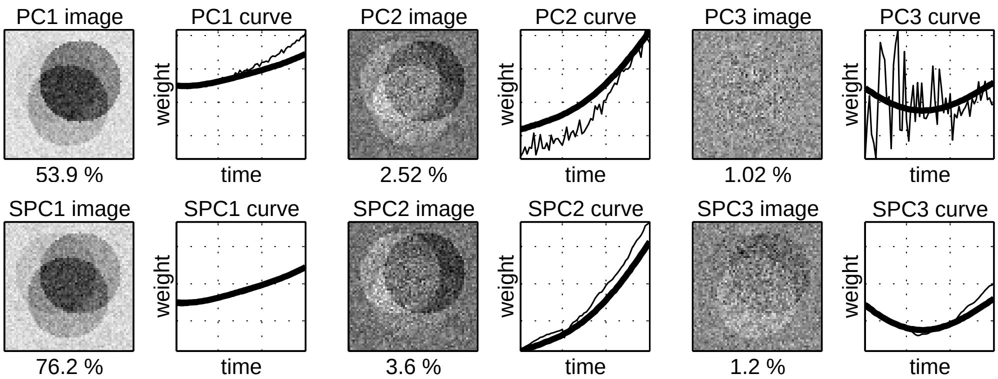
Maximization complexity stems mainly from the restricted rank of the mean value μ. The estimates of μ requires the evaluation of eigenvalues and eigenvectors. The considered dimensions ($T \sim 10^2$) allows to use only numerical evaluation. This makes iterative search inevitable.

3 Experiments

Algorithm of smoothed PCA (SPCA) - that utilizes both prior information of noise and signal - was implemented in Matlab and its performance evaluated in experiments with simulated data. The mathematical phantom consisted of 60 images of size 64×64, ($T = 60, N = 4096$). Each image was a linear combination of three factor images with circular structures.



The figure also includes the curves simulating intensity changes with time. A flat background and uncorrelated Gaussian noise with a high variance was added to the simulated images. PCA of simulated data should recognize three underlying dynamic components. The comparison of first three most significant principal components produced by PCA (PCs) and SPCA (SPCs) are displayed in the following figures



It is remarkable that SPCA algorithm was able to estimate all three underlying images. This data has low SNR and thus the prior information about smoothness was essential for improvement.

Extended information about evaluation of the results will be available in the final paper.

4 Discussion and conclusions

Preliminary experiments with simulated and real medical data have shown that in comparison with PCA, SPCA is able to improve the separation of the signal from noise. The prior information used in the proposed method is rather general. In addition, alternative prior information, better suited to a specific problem, can be chosen and the methodology proposed in this paper can still be used with benefit. The method can be further developed to support an estimation of the number of significant factors and to benefit from similar prior information applied also to the images of principal components. Formally, these extensions are relatively straightforward. However, the increase in complexity of calculations is significant and approximations have to be found in order to make the solution feasible.

Acknowledgements

The work has been partially supported by the following grants: Austro-Czech project Kontakt II-16 (ME-228), GACR 102/99/1564, IGA MZCR NN5382-3/99 and NIH no.AA11653

References

[1] Anderson TW. Estimating linear statistical relationships. Ann Statist 1984; 12:1-45.

[2] Fine J, Pousse A. Asymptotic study of the multivariate functional model. Application to the metric choice in principal component analysis. Statistics 1992; 23: 63-83.

[3] Golub GH, VanLoan CF. Matrix Computations. Baltimore, J Hopkins Univ Press, 1989.