SMOOTH PRINCIPAL COMPONENT ANALYSIS WITH APPLICATION TO FUNCTIONAL MAGNETIC RESONANCE IMAGING

Magnus O. Ulfarsson and Victor Solo

Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, MI, 48109
Email: mulfarss@umich.edu, vsolo@umich.edu

ABSTRACT

Multivariate methods such as Principal component analysis (PCA) and Independent component analysis (ICA) have been found to be useful in functional magnetic resonance imaging (fMRI) research. They are often able to decompose the fMRI data so that the researcher can associate their components to some biological processes of interest such as the brain response resulting from a stimulus. In this paper we develop a new smooth version of the PCA derived from a maximum likelihood framework. We are thus led to an unusual use of AIC,BIC namely to choose two (rather than one) parameters simultaneously; the number of principal components and the degree of smoothness. The algorithm is applied to real fMRI data.

1. INTRODUCTION

The term functional Magnetic Resonance Imaging (fMRI) refers to the use of a Magnetic Resonance (MR) scanner to measure (dynamic) functional activity in the brain by means of rapid acquisition of images of brain state. This technology has made it possible to observe brain activity under some stimulus and thus study which brain regions are involved when a subject performs a particular task and therefore better understand human brain organization and function.

fMRI depends on the fact that neuronal activity induces a regional change in the oxygenation level of the blood which can be detected by the MR scanner [1]. This is called blood oxygen level dependency (BOLD) response. The BOLD response is often modelled as a hump like function that reaches maximum in about 5 seconds and then dies out about 10 seconds later [2].

In a typical experiment an MRI scanner is used to record signals that can be used to construct a sequence of brain images with a typical sampling time 2-5 sec while the subject reacts to a stimulus, e.g., auditory, visual, motor, etc. A typical blocked experiment usually consists of two states; the control state and a functional state. The functional state could for example involve finger tapping or a visual fixation on a flickering image and the rest state involves no motor action or visual fixation on non-flickering image. More recently event-related studies [3], where the brain response to brief stimulus is explored, have become more common.

fMRI data acquisition and image reconstruction yields spatiotemporal data. The data is contained in a $T \times M$ matrix $Y = [y_{t,v}]$ where T is the number of time points and M is the number of voxels (3D analogue of pixel). The $T \times 1$ vector y_v will refer to a column vector of Y and represents a voxel v and its associated time series. The $1 \times M$ vector $y_{(t)}^T$ is a row vector of Y and represents a brain image scanned at time t.

Two main paths of analysis are currently used in fMRI research; univoxel analysis and multivoxel analysis. In univoxel analysis [4] response at each voxel is modelled separately as a sum of BOLD response, noise and possible some nuisance signals. Then an inference is made on each voxel to determine if it is activated or not. An activation map is then constructed which is a binary map that can be overlayed on a brain map for interpretation. Multivoxel analysis unlike univoxel analysis do not need information about the stimulus. They seek to find temporal or spatial components that are due to motion, activation etc., that are either unexpected or difficult to model. These components can then be visualized and interpreted or used for modelling purposes. The two most common exploratory methods in fMRI research are Principal component analysis (PCA) [5] that finds temporal/spatial components of maximal variance and Independent component analysis (ICA) [6], [7] that finds spatial/temporal components that are as independent as possible.

Principal component analysis (PCA) is a classical technique [8] that generates an orthonormal basis for the fMRI data such that the first basis vector captures maximal variance, the second basis vector second most variance etc. PCA is a completely data driven method and does not require any information about the experimental design. This could be useful in cases where there is insufficient information about the stimulus i.e., in pharmaceutical experiments. PCA can

also be used for preliminary analysis of the fMRI, for example to find unexpected nuisance signals that could be corrected for. It is usually possible to associate the top few PC to some signals of interest, i.e., drift or stimulus.

But as noted by Mitra [9], due to the orthogonality constraint there is no guarantee that the PCs will correspond to something useful. The stimulus could for example be diffused over many PCs. Another problem with the use of PCs in fMRI research is that they do not recognize the temporal smoothness of the BOLD response.

This paper presents a smooth version of the PCA that takes into account the temporal smoothness of the BOLD response detected by the fMRI scanner. The main aspects of our work are: 1) We use maximum likelihood (ML) framework formulate the smooth PCA. This is an extension of what was introduced by Tipping et al [10]. 2) Basis expansion is used to impose the smoothness; Consequently the computation is much simpler. 3) The ML framework allows use of the AIC and the BIC criteria to control the smoothness (select the number of basis functions) and to select the number of principal components. Note that the AIC/BIC criteria is used in a unusual way, i.e. to choose two tuning parameters simultaneously. Another example of this is [11]. Earlier work on this smoothness idea include a work [12] by the second author that introduced Functional Data Analysis [13] to the fMRI problem. More recent work includes Viviani et al. [14] who uses functional PCA. In functional PCA regularization is used to constrain the PCs to be smooth.

The paper is organized as follows. In Section 2 the formulation of the conventional PCA via the SVD is discussed. In Section 3 we discuss smooth PCA. Section 4 discusses the problem of model selection and develops an AIC and BIC criteria for our problem and finally in Section 5 conclusions are drawn.

2. PCA

Principal component analysis can be performed by applying singular value decomposition (SVD) on the data. A SVD of a mean-corrected data matrix Y is given by.

$$Y \approx P\Delta Q^T \tag{1}$$

where P is a $T \times r$ matrix of eigen-timeseries (principal components) where $P^TP = I_r$, and Q is a $M \times r$ matrix of eigen-images where $Q^TQ = I_r$, and $\Delta = \text{diag}(\sqrt{\lambda_1}, ..., \sqrt{\lambda_r})$. The $P\Delta Q^T$ is a rank r approximation to Y in the Frobenius norm.

3. SMOOTH PCA

The model for the Smooth PCA is given by

$$y_v = \mu + \Phi B u_v + \epsilon_v, \quad v = 1, ..., M, \tag{2}$$

where Φ is a $T \times m$ matrix of pre-specified basis functions, B is a $m \times r$ matrix, $B^TB = \Delta_B$, Δ_B is diagonal, $u_v \sim N(0, I_r)$ and $\epsilon_v \sim N(0, \sigma^2 I_T)$. The model induces a distribution on the observed data, i.e., $y_v \sim N(\mu, C)$, where $C = \Phi BB^T\Phi^T + \sigma^2 I_T$. The log-likelihood is given by

$$l(Y;\theta) = -\frac{M}{2} trace(C^{-1}S_y) - \frac{M}{2}\log|C| \tag{3}$$

where $S_y = \frac{1}{M}\sum_{v=1}^{M}(y_v - \mu)(y_v - \mu)^T$ is the sample covariance and θ is a vector of the parameters to be estimated, which are σ^2 and the elements of B. The likelihood is maximized when

$$\hat{B} = K_r(D_r - \sigma^2 I_m)^{1/2} \tag{4}$$

where K_r is the $m \times r$ matrix of unit eigenvectors of $S_\Phi = (\Phi^T\Phi)^{-1/2}\Phi^T S_y \Phi(\Phi^T\Phi)^{-1/2}$ and D_r is a $r \times r$ diagonal matrix that contains the corresponding eigenvalues. When $B = \hat{B}$ the maximum likelihood for σ^2 is given by

$$\hat{\sigma}^2 = \frac{trace(S_y) - trace(D_r)}{T - r} \tag{5}$$

Note that when $m = T$ the columns of ΦB are proportional to the columns of P, i.e., the principal components. The columns of ΦB are referred to as the smooth principal components or the smooth eigen-timeseries. To finish the model fitting we need to choose the number of basis functions m and the number of PCs r.

4. MODEL SELECTION

In this paper we explore the Akaike's AIC criteria [15] and the BIC criteria [16] for model selection. However, in this case the usage is unusual because we need to choose two tuning parameters, i.e., the number of basis functions and the number of principal components. Therefore the AIC and the BIC are two-dimensional functions. The AIC is given by

$$AIC(m, r) = -2l(Y;\hat{\theta}) + 2dim(\hat{\theta}), \tag{6}$$

where $\hat{\theta}$ is the maximum likelihood estimate of the parameter vector θ, and $dim(\hat{\theta})$ is the number of free parameters given by

$$dim(\hat{\theta}) = mr - r(r-1)/2 + r + 1 \tag{7}$$

The BIC has a $\log(T)$ in place of the 2 in the second term of the AIC. We select m and r that give the smallest AIC or BIC value.

The AIC is an unbiased estimate of the mean Kullback-Liebler distance between modelled density and the estimated density. Note this is only true if the modelled density belongs to the same family of probability densities as the density that truly generated the data. But as pointed out in [17]

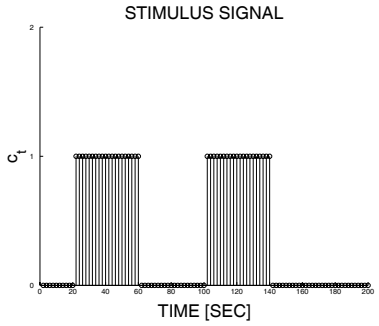

Fig. 1. The stimulus signal

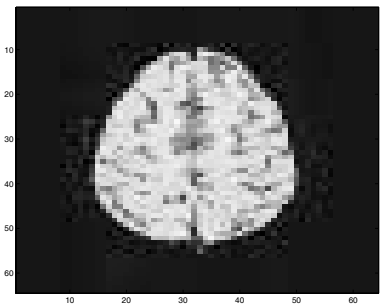

Fig. 2. An example of fMRI image slice.

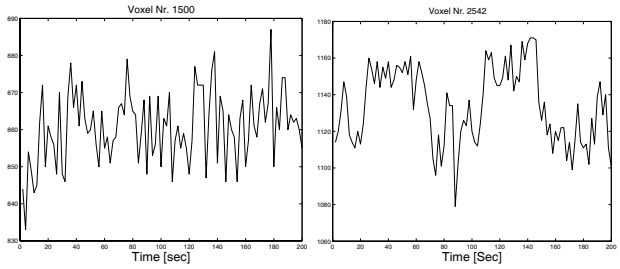

Fig. 3. Examples of fMRI timeseries

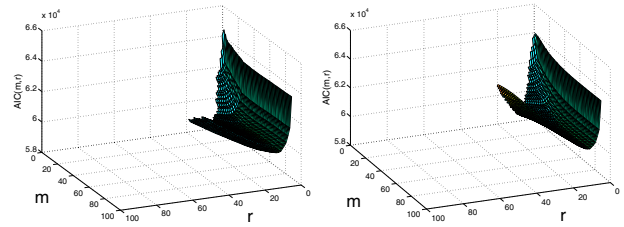

Fig. 4. Left: The AIC statistics (Fourier basis) Right: The AIC statistics (B-splines basis)

page 22, the AIC performs well in practice given that the modelled density and the true density are not grossly different.

5. RESULTS

An fMRI experiment was performed where a human subject performed right hand sequential finger-thumb opposition according to a stimulus signal displayed on Fig. 1. When the signal was high the subject performed finger-thumb opposition when it was low the subject rested. This data set is available on the AFNI homepage.

The data was obtained from a 3T MRI scanner with 2 sec TR. The original data consisted of $T = 100$ time points and $M = 4096 = 64^2$ voxels. The conventional spatial smoothing and temporal filtering were not done. A spatial plot of the data that shows a cross-section of the brain at time $t = 50$ is given on Fig. 2. The timeseries associated with two voxels are displayed on Fig. 3.

Ninety seven voxels associated with the motor cortex were selected for further processing by regressing a parametric model of the BOLD signal [18] to the voxel timeseries and select the most significant voxels (Bonferroni corrected p-value=0.05). This approximately selects the voxels associated with the motor cortex. If all the voxels were used for analysis the model selection criteria picked the number of basis functions r close to 100. This is because most of the voxels outside the motor cortex contain mostly noise. Therefore the smoothness does not help.

The model 2 was fitted for two cases; firstly where Φ was a basis of m Fourier functions and secondly where Φ was a basis of m B-splines functions where the knot locations were uniformly chosen over $[0, T-1]$. The AIC was calculated for $m = 0, ..., 100$ and $r < m$. For the Fourier basis the values that give minimum AIC are $r = 9$ and $m = 49$. The minimum BIC values are $r = 7$ and $m = 28$. For the B-spline basis the values that give minimum AIC are $r = 8$ and $m = 69$. The minimum BIC values are $r = 6$ and $m = 43$. The two-dimensional AIC plots for the Fourier and the B-splines are given on Fig. 4. The plots for the BIC are similar and not shown. Fig. 5 shows the two first smooth eigen-timeseries corresponding the the minimum BIC ($r = 7$,$m = 28$) for Fourier basis. The result for B-splines is similar. We see that the first smooth eigen-timeseries shows the effect of the stimulus signal. Compared to the unsmooth PCA ($m = 100$) we get a signal that looks much more like the expected BOLD response with significant computational savings. Fig. 6 shows a spatial plot on the smooth eigen-timeseries regressed on the fMRI data. The highest activation are in the primary motor cortex and the supplementary motor cortex. The second smooth eigen-timeseries shows the effect of drift. The last 5 smooth eigen-timeseries were not as easily interpreted and are not shown.

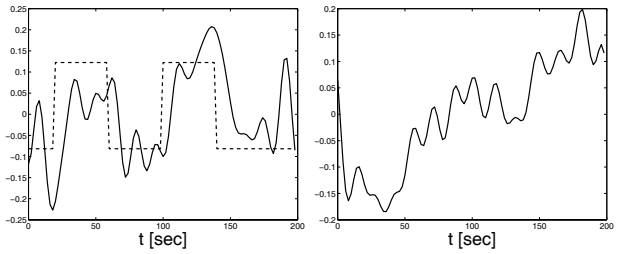

Fig. 5. Left: The first smooth eigen-timeseries and the stimulus signal. Right: The second smooth eigen-timeseries

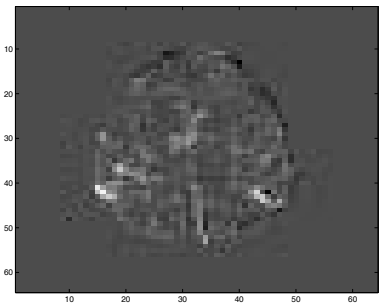

Fig. 6. The first smooth eigen-timeseries regressed on the fMRI data

6. CONCLUSION

We have presented a smooth PCA method that is able to exploit the temporal smoothness of the BOLD signal by restricting the fMRI data to a smooth subspace of basis functions. The AIC and the BIC criteria were used to select two tuning parameters; the number of basis functions and the number of eigen-timeseries, and they show that a smooth PCA is preferred. Work is underway extending this smoothness idea to Independent component analysis (ICA).

7. REFERENCES

[1] S. Ogawa, D.W. Tank, R. Menon, J.M. Ellerman, S.G. Kim, and H. Merkle. Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA*, 89:5951–5955, 1992.

[2] R.B. Buxton. *Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques*. Cambridge University Press, Cambridge, England, 2002.

[3] B.R. Rosen, R.L. Buckner, and A.M. Dale. Event-related functional MRI: Past, present and future. *Proceedings of the National Academy of Sciences*, 95(31):773–780, Feb 1998.

[4] K.J. Friston, P. Jezzard, and R. Turner. Analysis of functional MRI series. *Human Brain Mapping*, pages 153–171, 1994.

[5] K. J. Friston, C. D. Frith, P. F. Liddle, and R. S. Frackowiak. Functional Connectivity: The principal component analysis of large data sets. *J. Cereb. Blood Flow Metab.*, 13:5–14, 1993.

[6] V.D. Calhoun, T. Adali, J. Hansen, J. Larsen, and J.J. Pekar. ICA of functional MRI data: an overview. *4th International Symposium on Independent Component Analysis and Blind Signal Separation*, pages 281–288, April 2003.

[7] M.J. McKeown, S. Makeig, Brown G.G., T.P. Jung, Bell A.J. Kinderman, S.S., and T.J. Sejnowski. Analysis of fMRI data by blind separation into independent spatial component. *Human Brain Mapping*, 6:160–188, 1998.

[8] I.T. Jolliffe. *Principal component analysis*. Springer Verlag, New York, NY, 1986.

[9] P.P. Mitra and B Pesaran. Analysis of dynamic brain imaging data. *Biophysical Journal*, 76:691–708, 1999.

[10] M.E. Tipping and C.M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B*, 61:611–622, 1999.

[11] V. Solo. Signals in coloured noise: Joint nonparametric estimation of signal and of noise spectrum. In *Proc. ICASSP03 Hong Kong*, volume VI, pages 302–305, 2003.

[12] V. Solo, P. Purdon, and E. Brown. Spatio-temporal signal processing for multi-subject functional MRI studies. In *Proc. ICASSP01 Salt Lake City*, 2001.

[13] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer Verlag, New York, NY, first edition, 1997.

[14] R. Viviani, G. Gron, and M. Spitzer. Functional principal component analysis of fmri data. *Human Brain Mapping*, 24:109–129, 2005.

[15] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Automat. Contr*, 1974.

[16] G. Schwartz. Estimating the dimension of a model. *Ann. Stat.*, 6:461–464, 1978.

[17] H. Linhart and W. Zucchini. *Model Selection*. Wiley, New York, NY, 1986.

[18] Cohen M.S. Parametric analysis of fMRI data using linear system methods. *NeuroImage*, 6:93–103, 1997.