A Sparse Spatial Linear Regression Model for fMRI Data Analysis

Vangelis P. Oikonomou and Konstantinos Blekas

Department of Computer Science, University of Ioannina
P.O. Box 1186, Ioannina 45110 - GREECE
{voikonom,kblekas}@cs.uoi.gr

Abstract. In this study we present an advanced Bayesian framework for the analysis of functional Magnetic Resonance Imaging (fMRI) data that simultaneously employs both spatial and sparse properties. The basic building block of our method is the general linear model (GML) that constitute a well-known probabilistic approach for regression. By treating regression coefficients as random variables, we can apply an appropriate Gibbs distribution function in order to capture spatial constraints of fMRI time series. In the same time, sparse properties are also embedded through a RVM-based sparse prior over coefficients. The proposed scheme is described as a maximum a posteriori (MAP) approach, where the known Expectation Maximization (EM) algorithm is applied offering closed form update equations. We have demonstrated that our method produces improved performance and enhanced functional activation detection in both simulated data and real applications.

1 Introduction

Functional magnetic resonance imaging (fMRI) measures the tiny metabolic changes that take place in an active part of the brain. It is becoming a common diagnostic method of the behavior of a normal, diseased or injured brain, as well as for assessing the potential risks of surgery or other invasive treatments of the brain. Functional MRI is based on the increase in blood flow to the local vasculature that accompanies neural activity of the brain [1]. When neurons are activated, the resulting increased need for oxygen is overcompensated by a large increase in perfusion. As a result, the venous oxyhemoglobin concentration increases and the deoxyhemoglobin concentration decreases. The latter has paramagnetic properties and the intensity of the fMRI images increases in the activated areas. The signal in the activated voxels increases and decreases according to the paradigm. fMRI detects changes of deoxyhemoglobin levels and generates blood oxygen level dependent (BOLD) signals related to the activation of the neurons [1].

The fMRI data analysis consists of two basic stages: preprocessing and statistical analysis. The first stage is usually carried out in four steps: slice timing, motion correction, spatial normalization and spatial smoothing [1]. Statistical analysis can be done using the parametric general linear regression model

S. Konstantopoulos et al. (Eds.): SETN 2010, LNAI 6040, pp. 203–212, 2010.

(GLM) [2] under a Maximum Likelihood (ML) framework for parameter estimation. Sequentially, the t or F statistic is used on order to form a so-called statistical parametric map (SPM) that maps the desired active areas.

A significant drawback of the basic GLM approach is that spatial and temporal properties of fMRI data are not taken into account. However, it is well known that the BOLD signal is constrained spatially due to its physiological nature and preprocessing steps such as realignment and spatial normalization [1]. Within the literature there are several methods that incorporate spatial and temporal correlations into the estimation procedure. A common approach is to apply Gaussian filter smoothing or adaptive thresholding techniques that adjust statistical significance of active regions, according to their size. Alternatively, spatial characteristics of fMRI can be naturally described in a Bayesian framework through the use of Markov Random Fields (MRF) priors [3, 4] and autoregressive (AR) spatio-temporal models [5,6]. The estimation process of most of these works is achieved by either Markov Chain Monte Carlo (MCMC), or Variational Bayes framework. An alternative methodology has been presented in [7], where the image of the regression coefficient is first spatially decomposed using wavelets, and secondly a sparse prior is applied over the wavelet coefficients. Apart from spatial another desired property of analysis is to embody a mechanism that automatically selects the model order. This is a very important issue in many model based applications including regression. If the order of the regressor model is too large it may overfit the observations and does not generalize well. On the other hand, if it is too small it might miss trends in the data. Sparse Bayesian regression offers a solution to the above problem [8,9] by introducing sparse priors on the model parameters.

In this paper we propose a model-based framework that simultaneously employs both spatial and sparse properties in a more systematic way. The basic regression model GLM can be spatially constrained by considering that the regression coefficients follow a Gibbs distribution [10]. By using then a modification of the clique potential function, we can allow the incorporation of sparse properties based on the notion of Relevance Vector Machine (RVM) [8]. A maximum a posteriori expectation maximization algorithm (MAP-EM) [11] is applied next to train this model. This is very efficient since it leads to update rules of model parameters in closed form during the M-step and improves data fitting. The performance of the proposed methodology is evaluated using a variety of simulated and real datasets. Comparison has been made using the typical maximum likelihood (ML) and the spatially variant alone regression model. As the experimental study has showed, the proposed method is more flexible and robust providing with quantitatively and qualitatively superior results.

In section 2 we briefly describe the general linear model and its spatially variant version by setting a Gibbs prior. The proposed simultaneous spatial and sparse regression model is then presented in section 3 and the MAP-based learning procedure. To assess the performance of the proposed methodology we present in section 4 numerical experiments with artificial and real fMRI datasets. Finally, in section 5 we give conclusions and suggestions for future research.

2 A Spatially Variant Generalized Linear Regression Model

Suppose we are given a set of N fMRI time-series $Y = \{\mathbf{y}_1 \ldots, \mathbf{y}_N\}$, where each observation $\mathbf{y}_n$ is a sequence of M values over time, i.e. $\mathbf{y}_n = \{y_{nm}\}_{m=1}^{M}$. The Generalized Linear Model (GLM) assumes that the fMRI time series $\mathbf{y}_n$ are described with the following manner:

$$\mathbf{y}_n = \mathbf{\Phi}\mathbf{w}_n + \mathbf{e}_n \ , \tag{1}$$

where $\mathbf{\Phi}$ is the design matrix of size $M \times D$ and $\mathbf{w}_n$ is the vector of the D regression coefficients which are unknown and must be estimated. Moreover, the last term $\mathbf{e}_n$ in Eq. 1 is a M-dimensional vector determining the error term that is assumed to be Gaussian with zero mean, independent over time with a precision (inverse variance) λ_n, i.e. $\mathbf{e}_n \sim \mathcal{N}(0, \lambda_n^{-1}\mathbf{I})$. The design matrix $\mathbf{\Phi}$ contains some explanatory variables that describes various experimental factors. In block design related experiments it usually has one regressor for the BOLD response plus the mean constant, i.e. it is a two-column matrix. However, we can expand it containing regressors related to other components of the fMRI time series such as drift and movement effects [6].

In fMRI data analysis the goal is to find the involvement of experimental factors in the generation process of time series, that is achieved through the estimation of coefficients $\mathbf{w}_n$. Since $\mathbf{\Phi}\mathbf{w}_n$ is deterministic, we can model the probability density of the sequence $\mathbf{y}_n$ with the normal distribution $p(\mathbf{y}_n|\mathbf{w}_n, \lambda_n) = \mathcal{N}(\mathbf{\Phi}\mathbf{w}_n, \lambda_n^{-1}\mathbf{I})$. Thus, the problem becomes a maximum likelihood (ML) estimation problem for the regression parameters $\Theta = \{\mathbf{w}_n, \lambda_n\}_{n=1}^{N}$. The maximization of the log-likelihood function:

$$L_{ML}(\Theta) = \sum_{n=1}^{N} \log p(\mathbf{y}_n|\mathbf{w}_n, \lambda_n) = \sum_{n=1}^{N} \Big\{ \frac{M}{2} \log \lambda_n - \frac{\lambda_n}{2} \|\mathbf{y}_n - \mathbf{\Phi}\mathbf{w}_n\|^2 \Big\} \ , \tag{2}$$

leads to the following rules:

$$\hat{\mathbf{w}}_n = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{y}_n \ , \ \ \hat{\lambda}_n = \frac{M}{\|\mathbf{y}_n - \mathbf{\Phi}\hat{\mathbf{w}}_n\|^2} \ . \tag{3}$$

After the estimation procedure, we calculate the t-statistic for each voxel for drawing the statistical map and identifying the activation regions.

The fMRI data are biologically generated by structures that involve spatial properties, since adjacent voxels tend to have similar activation level [12]. Moreover, the produced ML-based activation maps contain many small activation islands and so there is a need for spatial regularization. The Bayesian formulation offers a natural platform for automatically incorporating these ideas. We assume that the vector of coefficients $\mathbf{w}_n$ follows the Gibbs density function according to the following form:

$$p(\mathbf{w}_n|\beta_n) \propto \beta_n^{|N_n|} \exp\Big(-\frac{\beta_n}{2} \sum_{k \in N_n} \|\mathbf{w}_n - \mathbf{w}_k\|^2 \Big) \ , \tag{4}$$

where β_n is the regularization parameter. The summation term denotes the cliques potential function within the neighborhood N_n of the n-th voxel, i.e. horizontally, vertically or diagonally adjacent voxels, while the first term $\beta_n^{|N_n|}$ acts as a normalizing factor. In addition, a Gamma prior is imposed on the regularization parameter β_n as well as the noise precision parameter λ_n with Gamma parameters $\{c_\beta, b_\beta\}$ and $\{c_\lambda, b_\lambda\}$, respectively.

The estimation problem can now be formulated as a maximum a posteriori (MAP) approach, in the sense of maximizing the posterior of $\Theta = \{\mathbf{w}_n, \beta_n, \lambda_n\}_{n=1}^N$:

$$L_{MAP}(\Theta) = \sum_{n=1}^{N} \Big\{ \log p(\mathbf{y}_n|\mathbf{w}_n, \lambda_n) + \log p(\mathbf{w}_n|\beta_n) + \log p(\beta_n) + \log p(\lambda_n) \Big\} \quad (5)$$

The maximization problem can be easily found that leads to the following updated rules:

$$\hat{\mathbf{w}}_n = (\lambda_n \mathbf{\Phi}^T \mathbf{\Phi} + B_n)^{-1} (\lambda_n \mathbf{\Phi}^T \mathbf{y} + BW_n) \ , \quad (6)$$

$$\hat{\beta}_n = \frac{|N_n| + c_\beta}{\frac{1}{2} \sum_{k \in N_n} \|\hat{\mathbf{w}}_n - \hat{\mathbf{w}}_k\|^2 + b_\beta} \ , \quad (7)$$

$$\hat{\lambda}_n = \frac{M + c_\lambda}{\frac{1}{2} \|\mathbf{y}_n - \mathbf{\Phi} \hat{\mathbf{w}}_n\|^2 + b_\lambda} \ , \quad (8)$$

where $B_n = \sum_{k \in N_n} (\beta_n + \beta_k) \mathbf{I}$ and $BW_n = \sum_{k \in N_n} (\beta_n + \beta_k) \mathbf{w}_k$ that determine the contribution of neighbors inside the clique. Equations 6-8 are applied iteratively until the convergence of the MAP log-likelihood function. The above scheme can be also described within an Expectation-Maximization (EM) framework [11], where the E-step computes the expectation of the hidden variables ($\mathbf{w}_n$) and use them next for updating the model parameters during the M-step. This approach will be referred next as SVGLM.

3 Simultaneous Sparse and Spatial Regression

A desired property of the linear regression model is to offer an automatic mechanism that will zero out the coefficients that are not significant and maintain only large coefficients that are considered significant based on the model. Moreover, an important issue when using the regression model is how to define its order D. The problem can be tackled using the Bayesian regularization method that has been successfully employed in the Relevance Vector Machine (RVM) model [8].

In order to capture both spatial and sparse properties over regression coefficients, the Gibbs distribution function needs to be reformulated. This can be accomplished by using the following Gibbs density function:

$$p(\mathbf{w}_n|\beta_n, z_n, \alpha_n) \propto \left(\beta_n^{|N_n|} \prod_{k \in N_n} z_{nk} \prod_{d=1}^{D} \alpha_{nd}^{1/2} \right) \exp\left(-\frac{1}{2} \Big\{ V_{N_n}^{(1)}(\mathbf{w}_n) + V_{N_n}^{(2)}(\mathbf{w}_n) \Big\} \right) . \quad (9)$$

The first term in the exponential part of this function is the sparse term used for describing local relationships of the n-th voxel coefficients. This is given by:

$$V_{N_n}^{(1)}(\mathbf{W}) = \mathbf{w}_n^T \mathbf{A}_n \mathbf{w}_n \ , \tag{10}$$

where $\mathbf{A}_n$ is a diagonal matrix containing the D elements of the hyperparameter vector $\alpha_n = (\alpha_{n1}, \ldots, \alpha_{nD})^T$. By imposing a Gamma prior over hyperparameters, a two-stage hierarchical prior is achieved, which is actually a Student-t distribution with heavy tails [8]. This scheme enforces most α_{nd} to be large, thus the corresponding coefficients w_{nd} are set zero and finally eliminated.

The second term of the exponential part (Eq. 9) captures the sparse property and is responsible for the clique potential of the n^{th} voxel:

$$V_{N_n}^{(2)}(\mathbf{W}) = \beta_n \sum_{k \in N_n} z_{nk} \|\mathbf{w}_n - \mathbf{w}_k\|^2 \ . \tag{11}$$

In comparison with the potential function of the SVGLM method (Eq. 4), here each neighbor contribute with a different weight, as denoted by parameters z_{nk}, to the computation of the clique energy value. The introduction of these weights can increase the flexibility of spatial modeling. As experimentally have shown, this can be proved advantageous in cases around the borders of activation regions (edges). Finally, the first part of Eq. 9 acts as a normalization factor.

We also assume that the regularization parameter β_n, the noise precision λ_n and the weights z_{nk} follow Gamma distribution. Training of the proposed model is therefore converted into a MAP-estimation problem for the set of model parameters $\Theta = \{\theta_n\}_{n=1}^N = \{\mathbf{w}_n, \beta_n, \lambda_n, z_n, \alpha_n\}_{n=1}^N$:

$$L_{MAP}(\Theta) = \sum_{n=1}^{N} \log p(\mathbf{y}_n|\theta_n) + \log\{p(\mathbf{w}_n|\beta_n, z_n, \alpha_n) p(\beta_n) p(\lambda_n) p(z_n) p(\alpha_n)\} \ . \tag{12}$$

By setting the partial derivative equal to zero the following closed form update rule for regression coefficients can be obtained:

$$\hat{\mathbf{w}}_n = (\lambda_n \mathbf{\Phi}^T \mathbf{\Phi} + BZ_n + \mathbf{A}_n)^{-1} (\lambda_n \mathbf{\Phi}^T \mathbf{y}_n + BZW_n) \ , \tag{13}$$

where the matrices BZ_n and BZW_n are: $BZ_n = \beta_n \sum_{k \in N_n} (z_{nk} + z_{kn}) \mathbf{I}$ and $BZW_n = \beta_n \sum_{k \in N_n} (z_{nk} + z_{kn}) \mathbf{w}_k$. For the other model parameters we have:

$$\hat{\beta}_n = \frac{|N_n| + c_\beta}{\frac{1}{2} \sum_{k \in N_n} z_{nk} \|\mathbf{w}_n - \mathbf{w}_k\|^2 + b_\beta} \ , \tag{14}$$

$$\hat{z}_{nk} = \frac{1 + c_z}{\frac{1}{2} \hat{\beta}_n \|\hat{\mathbf{w}}_n - \hat{\mathbf{w}}_k\|^2 + b_z} \ , \tag{15}$$

$$\hat{\alpha}_{nd} = \frac{1 + 2c_a}{\hat{w}_{nd}^2 + 2b_a} \ , \tag{16}$$

while the noise precision λ_n has the same form as previously defined in SVGLM, (Eq. 8). The whole procedure can be integrated in an EM framework, where the

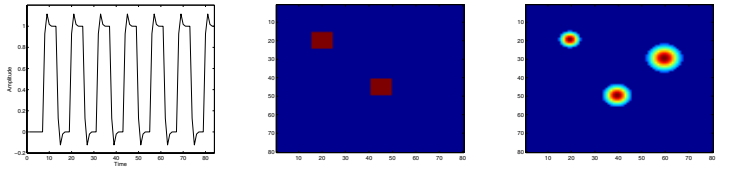
expectation of regression coefficients are computed in the E-step (Eq. 13), and the maximization of the complete-data log-likelihood is performed during the M-step (Eqs 14-16), giving update equations for model parameters. The above scheme is iteratively applied until the convergence of the MAP function. Notice that in the above equations we took into consideration that the weights of n-th voxel occurs two times into the summation term, one as the central voxel, and $|N_n|$ times as a neighbor of different voxels. We call this method SSGLM.

4 Experimental Results

We have tested the proposed method, SSGLM, using various simulated and real datasets. Comparison has been made with the simple ML method and the SVGLM as has been presented in Section 2. The SVGLM and SSGLM have been initialized with the same manner. First, the ML estimates of the regression coefficients $\mathbf{w}_n$ are obtained and use them next for initializing the rest model parameters β_n,λ_n,z_{kn} and a_{np}, according to Eqs. (14)-(16), respectively. During the experiments the parameters of Gamma prior distributions were set $c_\beta = b_\beta = c_z = b_z = 1$, $c_\lambda = b_\lambda = 10^{-8}$ and $b_\alpha = c_\alpha = 10^{-8}$ (making them non-informative as suggested by the RVM methodology [8]).

4.1 Experiments with Simulated Data

The simulated datasets used in our experiments were created using the following generation mechanism. We applied a design matrix ($\mathbf{\Phi}$) of size $M \times 2$ with two pre-specified regressors, the first one captures the BOLD signal (Fig. 1 (a)), and the second one being a constant with ones. Then, we constructed an image with the activation regions that corresponds to the value of the first coefficient (w_{n1}), while the second coefficient w_{n2} had a constant value equal to 100. In our study we have used two such images of size 80×80 with different shape of activation areas, rectangular (Fig. 1(b)) and circular (Fig. 1 (c)), respectively. The time series data ($\mathbf{y}_n$) were finally produced by using the generative equation of GLM (Eq. 1) with an additive white Gaussian noise of various signal-to-noise-ratio (SNR) levels, where we performed 50 runs and computed their mean



(a) (b) (c)

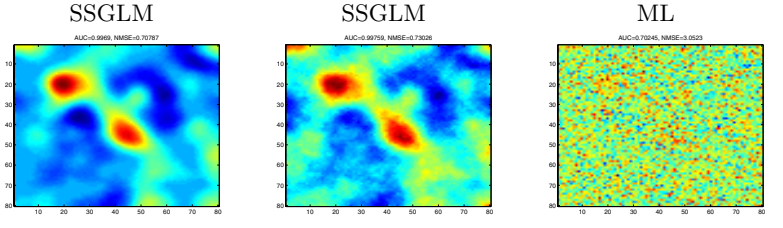
Fig. 1. Simulated data generative features: (a) Bold signal, (b) rectangular and (c) circular shape image of true activated areas

performance. Evaluation has done using two criteria: 1) The Area Under Curve (AUC) of the Receiver Operating Curve (ROC) based on t-statistic calculations and 2) the normalized mean square error (NMSE), between the estimated and the true coefficients responsible for the BOLD signal.

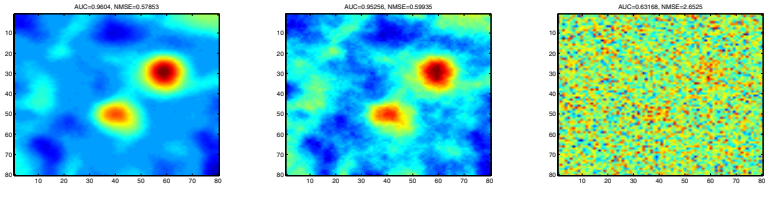
We present in Table 1 the comparative performance results in terms of the above two criteria for several SNR values in the case of rectangular and circular activation regions, respectively. As it is obvious, the proposed spatial sparse model (SSGLM) improves functional activation detection quality, especially for lower values of examined SNR values. In all cases both MAP-based approaches perform significantly better than the simple ML method. Figure 2 presents the mapping results of a typical run in the case of SNR =-20 dB. As it is obvious the proposed SSGLM approach manages to construct much smoother maps of brain activity than the spatial SVGLM model. That is interesting to observe is that SVGLM method has the tendency to overestimate the activation areas and

Table 1. Comparative results for simulated data in various noisy environments

	circular areas						*rectangular areas*					
	AUC			*NMSE*			*AUC*			*NMSE*		
SNR	SSGLM	SVGLM	ML	SSGLM	SVGLM	ML	SSGLM	SVGLM	ML	SSGLM	SVGLM	ML
0	0.999	0.999	0.999	0.118	0.177	0.294	0.998	0.995	0.980	0.129	0.170	0.255
-5	0.998	0.999	0.929	0.551	0.464	0.933	0.998	0.992	0.819	0.415	0.318	0.812
-10	0.998	0.998	0.795	0.704	0.633	1.642	0.995	0.991	0.712	0.541	0.478	1.439
-15	0.986	0.988	0.674	0.807	0.802	2.948	0.978	0.972	0.624	0.641	0.665	2.554
-20	0.920	0.914	0.600	0.993	1.084	5.214	0.898	0.883	0.570	0.855	0.971	4.579
-30	0.763	0.724	0.558	1.748	1.854	9.257	0.747	0.716	0.536	1.437	1.641	8.074



(a) rectangular activated areas



(b) circular activated areas

Fig. 2. An example of the statistical map produced by three comparative methods for two kind of activity (a) rectangular and (b) circular. The SNR value is -20 dB.

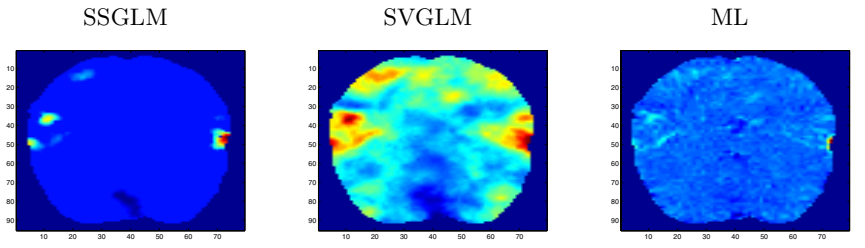


Fig. 3. Maps of the estimated BOLD signal (w_{n1}) obtained by three methods

discover larger regions than their true size. The proposed SSGLM exhibits very clean edges between activated and non - activated areas, and thus visual improvement. Finally, the ML approach completely fails to discover any activation pattern in this experiment.

4.2 Experiments with Real fMRI Data

The proposed approach was also evaluated in real applications. Experiments were made using a block design real fMRI dataset that was downloaded from the SPM web page[1] which was designed for auditory processing task on a healthy volunteer. In our study, we followed the standard preprocessing steps of the statistical parametric mapping package (SPM) manual, which are realignment, segmentation, and spatial normalization, without performing the spatial smoothing step.

We selected the slice 29 of this dataset for making experiments. Figure (3) presents the maps of the BOLD signal (regression coefficients w_{n1}) as estimated by the three comparative approaches SSGLM, SVGLM and ML. As it is obvious the proposed SSGLM approach achieves significantly smoother results, where brain activity is found on the auditory cortex, as it was expected. In addition, produced activation areas are less noisy and very clean in comparison with those produced by the SVGLM which overestimates the brain activity, thus making the decision harder. On the other hand, the resulting map of the ML method is confused without showing any significant distinction between the activated and non activated areas.

Moreover, we find it useful to visually inspect the resulting activation maps obtained by the t-test. In Figure 4 the SPMs of each method are shown, calculated without setting a threshold (Figure 4a), or by using a threshold ($t_0 = 1.6$) on t-value (Figure 4b). Notice that the activation maps of the SSGLM approach are similar in both cases that makes our approach less sensitive to the threshold value. The latter can be more apparent by plotting in Figure 5(a) the estimated size (number of voxels) of activation areas from each method in terms of the threshold value t_0. This behavior can be proved very useful, since there is not need to resort in multiple comparison between t-tests. This can be also viewed in

[1] http://www.fil.ion.ucl.ac.uk/spm/

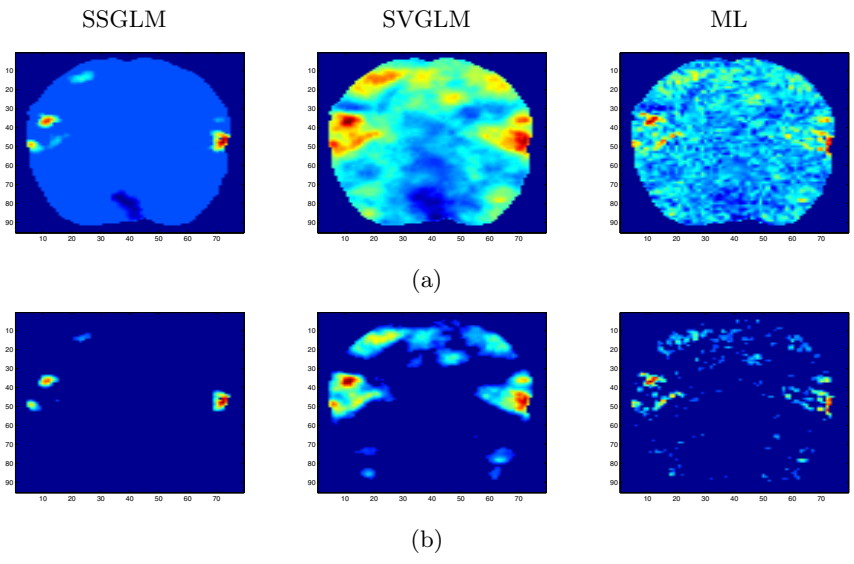


Fig. 4. Statistical parametric maps from the t-statistics (a) without and (b) with a threshold value $t_0 = 1.6$

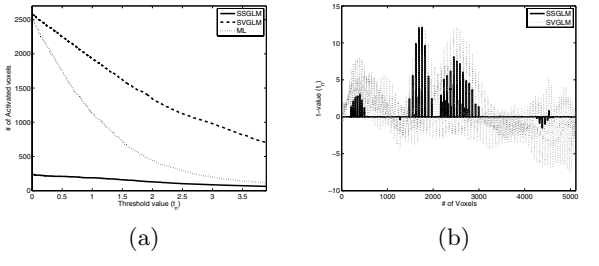


Fig. 5. (a) Plots of the estimated number of activated voxels in terms of threshold value used for producing the SPMs. (b) Plots of the t-values as computed by comparative methods SSGLM (thick line) and SVGLM (thin line).

Figure 5(b) where we plot the calculated t-values of the SSGLM and the SVGLM methods. The distinction between the activated and non activated areas is much more apparent in the case of SSGLM plot.

5 Conclusions

In this work we present an advanced regression model for fMRI time series analysis by incorporating both spatial correlations and sparse capabilities. This is done by using an appropriate prior over the regression coefficients based on the MRF and the RVM schemes. Training is achieved through a maximum a posteriori (MAP) framework that allows the EM algorithm to be effectively used for

estimating the model parameters. This has the advantage of establishing update rules in closed form during the M-step and thus data fitting is computationally efficient. Experiments on artificial and real datasets have demonstrated the ability of the proposed approach to improve the detection performance by providing cleaner and more accurate estimates. We are planning to make experiments with extended kernel design matrix and also to improve its specification by an adaption mechanism, as well as to examine the appropriateness of other types of sparse priors [9].

References

1. Frackowiak, R.S.J., Ashburner, J.T., Penny, W.D., Zeki, S., Friston, K.J., Frith, C.D., Dolan, R.J., Price, C.J.: Human Brain Function, 2nd edn. Elsevier Science, USA (2004)
2. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, Heidelberg (2007)
3. Descombes, X., Kruggel, F., von Cramon, D.Y.: fMRI signal restoration using a spatio-temporal Markov Random Field preserving transitions. NeuroImage 8, 340–349 (1998)
4. Gossl, C., Auer, D.P., Fahrmeir, L.: Bayesian spatiotemporal inference in functional magnetic resonance imaging. Biometrics 57, 554–562 (2001)
5. Woolrich, M.W., Jenkinson, M., Brady, J.M., Smith, S.M.: Fully bayesian spatio-temporal modeling of fmri data. IEEE Transactions on Medical Imaging 23(2), 213–231 (2004)
6. Penny, W.D., Trujillo-Barreto, N.J., Friston, K.J.: Bayesian fmri time series analysis with spatial priors. NeuroImage 24, 350–362 (2005)
7. Flandin, G., Penny, W.: Bayesian fmri data analysis with sparse spatial basis function priors. NeuroImage 34, 1108–1125 (2007)
8. Tipping, M.E.: Sparse Bayesian Learning and the Relevance Vector Machine. Journal of Machine Learning Research 1, 211–244 (2001)
9. Seeger, M.: Bayesian Inference and Optimal Design for the Sparse Linear Model. Journal of Machine Learning Research 9, 759–813 (2008)
10. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. on Pattern Analysis and Machine Intelligence 6, 721–741 (1984)
11. Dempster, A., Laird A., Rubin D.: Maximum likelihood from incomplete data via the em algorithm. Journal of Royal Statistical Society, Series B 39, 1–38 (1977)
12. Harrison, L.M., Penny, W., Daunizeau, J., Friston, K.J.: Diffusion-based spatial priors for functional magnetic resonance images. NeuroImage 41(2), 408–423 (2008)