

A Sparse and Spatially Constrained Generative Regression Model for fMRI Data Analysis

Vangelis P. Oikonomou, Konstantinos Blekas*, and Loukas Astrakas

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 regression model that constitutes a well-known probabilistic approach. By treating regression coefficients as random variables, we can apply an enhanced Gibbs distribution function that captures spatial constraints and at the same time allows sparse representation of fMRI time series. The proposed scheme is described as a maximum *a posteriori* approach, where the known expectation maximization algorithm is applied offering closed-form update equations for the model parameters. We have demonstrated that our method produces improved performance and functional activation detection capabilities in both simulated data and real applications.

Index Terms—Expectation maximization (EM) algorithm, functional magnetic resonance imaging (fMRI) analysis, general linear regression model (GLM), Markov random field (MRF), relevance vector machine (RVM).

I. INTRODUCTION

FUNCTIONAL magnetic resonance imaging (fMRI) measures the tiny metabolic changes that take place in an active part of the brain. It is a common diagnostic method for 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. fMRI is based on the increase in blood flow to the local vasculature that accompanies neural activity of the brain [1], [2]. 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. Because the latter has paramagnetic properties, the intensity of the fMRI images increases in the activated areas according to the paradigm. fMRI detects changes of deoxyhemoglobin levels and generates blood oxygen level-dependent (BOLD) signals related to the activation of neurons.

The objective of the fMRI data analysis is to detect the weak BOLD signal from the noisy data and determine the activated regions of the brain. It consists of two stages: preprocessing and statistical analysis. The first stage is usually carried out in vari-

ous steps, such as realignment, segmentation, motion correction, spatial normalization, spatial smoothing, etc. [2]. Model-based statistical analysis is most commonly achieved through the use of the general linear regression model (GLM) [3], [4]. At the end of the learning process, the statistical activation map is drawn based on *t*- or *F*-statistics, displaying the activation areas and the importance of each voxel.

A significant drawback of the basic GLM approach is that spatial and temporal properties of fMRI data are not taken into account. However, the fMRI data are biologically generated by structures that involve spatial properties, since adjacent voxels tend to have similar activation level [5]. Moreover, the produced activation maps contain many small activation islands and so there is a need for spatial regularization. Another desirable property is to handle temporal correlations derived from neural, physiological, and physical sources [6] and have a mechanism that can automatically address the model order. The latter is a very important issue in many model-based applications including regression. If the order of the regression model is too large, it may overfit the observations and does not generalize well, while if it is too small it might miss trends in the data [7].

Within the literature, there are several methods that incorporate either spatial correlations or sparse properties into the estimation procedure. Spatial characteristics of fMRI are usually taken into account by spatially smoothing the data with a fixed Gaussian kernel, as a preprocessing step [2], [8]. Other approaches have been also proposed that elaborate denoising techniques applied before the statistical analysis (see, e.g., [8]–[11]). fMRI data have also been analyzed by multivariate methods such as spatial independent component analysis (ICA) [12], [13]. Under the Bayesian framework, spatial dependences in fMRI have been modeled through Gaussian Markov random fields (MRFs) priors applied either to temporal and spatial components of the signal or to the noise process, using Markov chain Monte Carlo (MCMC) to compute posterior inference [14], [15]. Also, Gaussian spatial priors have been placed over the regression coefficients, as well as on autoregressive (AR) coefficients of the noise process [6], [8], [16], [17]. A special type of Gaussian MRF has been applied in [18], where the covariance matrix is assumed to be known based on the distance over pixel coordinates, while learning is made through variational expectation maximization (EM). Alternatively, spatial smoothing can be obtained either through Laplace- or Cauchy-type priors [6], [19], or as a postprocessing step working over the resulting statistical map [10], [20], [21].

On the other hand, methods with sparse properties have been applied to fMRI data analysis. These include sparse regression models over the linear coefficients of GLM [22]–[24], either the coefficients of spatio-temporal AR models [15], or the weights

Manuscript received March 31, 2010; revised August 24, 2010 and November 26, 2010; accepted December 12, 2010. Date of publication January 6, 2011; date of current version December 21, 2011. Asterisk indicates corresponding author.

V. P. Oikonomou is with the Department of Computer Science, University of Ioannina, Ioannina 45110, Greece (e-mail: voikonom@cs.uoi.gr).

*K. Blekas is with the Department of Computer Science, University of Ioannina, Ioannina 45110, Greece (e-mail: kblekas@cs.uoi.gr).

L. Astrakas is with the Medical School, University of Ioannina, Ioannina 45110, Greece (e-mail: astrakas@uoi.gr).

Digital Object Identifier 10.1109/TBME.2010.2104321

on the space domain of images [25]. Moreover, other sparse models have been proposed using of elastic nets [26], or by converting the estimation problem into a linear programming problem [27]. Training of the above methods is performed either by Markov Chain Monte Carlo (MCMC), or through Variational Bayes (VB) framework.

However, only a few works have studied the impact of both spatial and sparse properties on the quality of fMRI data modeling. One such methodology has been presented in [8] that addresses both spatial and sparse capabilities in a hierarchical framework. In particular, the image of the regression coefficients is first decomposed using wavelets, and then a sparse prior is applied over the wavelet coefficients. An alternative approach has been presented in [16] where the regression coefficients are indirectly spatially smoothed using an Ising prior over their indicator variables. Finally, a recent work is described in [17], that applies a multivariate Laplacian prior over coefficients written as a scale mixture followed by a spatial distribution on an auxiliary variable, that allows a spatiotemporal smoothing of data.

In this paper, we propose an advanced Bayesian framework that simultaneously employs both spatial and sparse properties in a more systematic way. The contribution of this paper is twofold. First, we provide directly the regression coefficients with the desired two properties by considering an enhanced prior distribution. Additionally, we manage to establish an efficient learning procedure, based on the EM algorithm, that derives closed-form update equations. More specifically, the general-purpose GLM is used for fMRI time series modeling. The key aspect of our method is the enhanced exploitation of the MRFs by using an effective Gibbs potential function. Traditionally, Gibbs distribution is used for addressing only spatial correlations. In our study, we present a modification of the potential function that, apart from spatial, is able to simultaneously impose sparseness based on the relevance vector machine (RVM) [7]. A maximum *a posteriori* expectation maximization (MAP-EM) algorithm [28] is applied next to train this model providing with update rules in closed form during the M -step. The performance of the proposed methodology is quantitatively and qualitatively evaluated using a variety of simulated and real datasets. Comparison has been made with two types of GLM having only spatial or sparse properties, respectively, and with the spatial ICA.

In Section II, we briefly describe the basic GLM framework and show how we can introduce a Gibbs prior so as to allow spatial correlations. The proposed simultaneous sparse spatial regression model is then presented in Section III together with the MAP-based learning procedure. To assess the performance of the proposed methodology, we present in Section IV numerical experiments with artificial and real fMRI datasets. Finally, in Section V we give conclusion and suggestions for future research.

II. SPATIALLY VARIANT LINEAR REGRESSION MODEL

A. Background

Suppose we are given a set of N fMRI time series $Y = \{\mathbf{y}_1, \dots, \mathbf{y}_N\}$, 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 application of the

GLM) assumes the following functional description:

$$\mathbf{y}_n = \Phi \mathbf{w}_n + \mathbf{e}_n \quad (1)$$

where Φ is the design matrix of size $M \times D$ and \mathbf{w}_n is the vector of the D regression coefficients that are unknown and must be estimated. The last term \mathbf{e}_n is an M -dimensional vector determining the model noise. By assuming temporal correlation structure, it can be modeled as an AR process of order p [15], [29] expressed as

$$\mathbf{e}_n = \mathbf{E}_n \boldsymbol{\xi}_n + \varepsilon_n. \quad (2)$$

The term \mathbf{E}_n is an $M \times p$ matrix containing past error samples, $\boldsymbol{\xi}_n$ is the p -length vector of the AR coefficients, and ε_n is an i.i.d. M -length zero mean Gaussian vector with a precision (inverse variance) λ_n , i.e., $\varepsilon_n \sim \mathcal{N}(0, \lambda_n^{-1} \mathbf{I})$. Alternatively, we can consider the next formulation

$$\Xi_n \mathbf{e}_n = \varepsilon_n \quad (3)$$

where Ξ_n is an $M \times M$ upper diagonal matrix containing the AR coefficients. From this scheme, we obtain the distribution of error as $\mathbf{e}_n \sim \mathcal{N}(0, (\lambda_n \Xi_n^T \Xi_n)^{-1})$. Both versions of the AR noise model are helpful for appropriately setting the likelihood function.

The design matrix Φ contains some explanatory variables (or effects) that describe various experimental factors. Its construction is crucial for the statistical analysis of fMRI data. The number of regressors (columns of the design matrix) depends on the experiment and on the problem formulation in order to address several factors of the fMRI time series, such as long-range correlations and movement effects [6], [22].

In fMRI data analysis, the goal is to find the involvement of experimental factors in the generation process of time series through the estimation of coefficients \mathbf{w}_n . Following (1), we can model the sequence \mathbf{y}_n with a normal distribution

$$p(\mathbf{y}_n | \mathbf{w}_n, \lambda_n, \boldsymbol{\xi}_n) = \mathcal{N}(\Phi \mathbf{w}_n, (\lambda_n \Xi_n^T \Xi_n)^{-1}). \quad (4)$$

Training of the GLM can be viewed as a maximum likelihood (ML) estimation problem for the model parameters $\Theta = \{\mathbf{w}_n, \lambda_n, \boldsymbol{\xi}_n\}_{n=1}^N$. The log-likelihood function can be then written in two equivalent ways by means of the error formulation (2) or (3)

$$L_{\text{ML}}(\Theta) = \sum_{n=1}^N \left\{ \frac{M}{2} \log \lambda_n - \frac{\lambda_n}{2} \|\Xi_n (\mathbf{y}_n - \Phi \mathbf{w}_n)\|^2 \right\} \quad (5)$$

$$L_{\text{ML}}(\Theta) = \sum_{n=1}^N \left\{ \frac{M}{2} \log \lambda_n - \frac{\lambda_n}{2} \|\mathbf{y}_n - \Phi \mathbf{w}_n - \mathbf{E}_n \boldsymbol{\xi}_n\|^2 \right\}. \quad (6)$$

The maximization procedure leads to the following rules that are iteratively applied:

$$\hat{\mathbf{w}}_n = (\mathbf{F}_n \Phi)^{-1} \mathbf{F}_n \mathbf{y}_n \quad (7)$$

$$\hat{\lambda}_n = \frac{M}{\|\Xi_n (\mathbf{y}_n - \Phi \hat{\mathbf{w}}_n)\|^2} \quad (8)$$

$$\hat{\boldsymbol{\xi}}_n = (\mathbf{E}_n^T \mathbf{E}_n)^{-1} \mathbf{E}_n (\mathbf{y}_n - \Phi \hat{\mathbf{w}}_n) \quad (9)$$

where $\mathbf{F}_n = \Phi^T \Xi_n^T \Xi_n$.

B. GLM With MRF-Based Spatial Prior

Let $\mathbf{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_N\}$ be the set of N vector of regression coefficients of N voxels. Spatial properties of the data can be accomplished through the use of MRF over \mathbf{W} . The Gibbs density function takes the following form:

$$p(\mathbf{W}) = Z \exp\{-U(\mathbf{W})\} \quad (10)$$

where Z is a normalization factor. The function $U(\mathbf{W})$ is the energy function given by

$$U(\mathbf{W}) = \frac{1}{2} \sum_{n=1}^N V_{N_n}(\mathbf{w}_n) \quad (11)$$

where $V_{N_n}(\mathbf{w}_n)$ denotes the clique potential function within the neighborhood N_n of n th voxel, written as

$$V_{N_n}(\mathbf{w}_n) = \beta_n \sum_{k \in N_n} \|\mathbf{w}_n - \mathbf{w}_k\|^2 \quad (12)$$

and β_n is the regularization parameter. The neighborhood N_n is the set of voxels that are horizontally, vertically, or diagonally adjacent to the n th voxel, having a cardinality equal to $|N_n|$. Although there is an obvious dependence on neighbors, without loss of generality we can assume the next prior distribution for each voxel

$$p(\mathbf{w}_n | \beta_n) = Z_n \exp\left\{-\frac{1}{2} V_{N_n}(\mathbf{w}_n)\right\} \quad (13)$$

where the first term is the normalization factor formulated as $Z_n \propto \beta_n^{|N_n|/2}$. Finally, a Gamma prior is imposed on the regularization parameter β_n , as well as the noise precision λ_n of the form

$$p(\beta_n) = \text{Gamma}(\beta_n | b_\beta, c_\beta) \propto \beta_n^{c_\beta - 1} e^{-b_\beta \beta_n} \quad (14)$$

$$p(\lambda_n) = \text{Gamma}(\lambda_n | b_\lambda, c_\lambda) \propto \lambda_n^{c_\lambda - 1} e^{-b_\lambda \lambda_n}. \quad (15)$$

The estimation problem is now formulated as a MAP framework, in the sense of maximizing the posterior density of model parameters $\Theta = \{\mathbf{w}_n, \beta_n, \lambda_n, \boldsymbol{\xi}_n\}_{n=1}^N$. The MAP log-likelihood function is then

$$L_{\text{MAP}}(\Theta) = \sum_{n=1}^N \left\{ \frac{M}{2} \log \lambda_n - \frac{\lambda_n}{2} \|\boldsymbol{\Xi}_n(\mathbf{y}_n - \Phi \mathbf{w}_n)\|^2 - \frac{\beta_n}{2} \sum_{k \in N_n} \|\mathbf{w}_n - \mathbf{w}_k\|^2 + \frac{|N_n|}{2} \log \beta_n + G(\beta_n) + G(\lambda_n) \right\} \quad (16)$$

where the function $G()$ has the following form:¹

$$G(x) = c_x \log x - b_x x. \quad (17)$$

By taking the partial derivatives with respect to model parameters, the next updated rules are obtained

$$\hat{\mathbf{w}}_n = (\lambda_n \mathbf{F}_n \Phi + B_n)^{-1} (\lambda_n \mathbf{F}_n \mathbf{y}_n + B \mathbf{W}_n) \quad (18)$$

¹We follow the methodology described in [7] where the maximization is made over a logarithmic scale using $p(\log x) = xp(x)$.

$$\hat{\beta}_n = \frac{|N_n| + 2c_\beta}{\sum_{k \in N_n} \|\hat{\mathbf{w}}_n - \hat{\mathbf{w}}_k\|^2 + 2b_\beta} \quad (19)$$

$$\hat{\lambda}_n = \frac{M + 2c_\lambda}{\|\boldsymbol{\Xi}_n(\mathbf{y}_n - \Phi \hat{\mathbf{w}}_n)\|^2 + 2b_\lambda} \quad (20)$$

where $B_n = \sum_{k \in N_n} (\beta_n + \beta_k) \mathbf{I}$ and $B \mathbf{W}_n = \sum_{k \in N_n} (\beta_n + \beta_k) \mathbf{w}_k$ correspond to the effect of neighbors on n th voxel's regression coefficients. Note that in the aforementioned scheme, the neighborhood N_n can include voxels whose regression coefficients either have $\hat{\mathbf{w}}_k$ or not \mathbf{w}_k been updated. For the AR coefficients ξ_n , the update rule is exactly the same as in the ML case (9). The aforesaid learning scheme can be incorporated into an EM framework [28]. In particular, during the E -step the expected values of the hidden variables \mathbf{w}_n are computed (18), while the adaptation of model parameters is performed in the M -step according to (9), (19), and (20).

III. SIMULTANEOUS SPARSE AND SPATIAL GLM

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 according to the model. Moreover, an important issue when using the regression model is how to define its order D . The appropriate value of D depends on the shape of data to be fitted, that is models of smaller order lead to underfitting, while large values of D may lead to overfitting. It is well known that both cases may lead to serious deterioration of the fitting performance. The problem can be tackled using the Bayesian regularization method that has been successfully employed in the RVM model [7].

In order to simultaneously cover both spatial and sparse properties, the Gibbs distribution function (10) needs to be reformulated. This can be accomplished by using the following energy function:

$$U(\mathbf{W}) = \frac{1}{2} \sum_{n=1}^N \left\{ V_{N_n}^{(1)}(\mathbf{w}_n) + V_{N_n}^{(2)}(\mathbf{w}_n) \right\}. \quad (21)$$

The first term of the aforesaid function is the sparse term used for describing local relationships of the n th voxel coefficients. This can be expressed as

$$V_{N_n}^{(1)}(\mathbf{w}_n) = \mathbf{w}_n^T \mathbf{A}_n \mathbf{w}_n \quad (22)$$

where \mathbf{A}_n is a diagonal matrix containing the D hyperparameters $\{\alpha_{n1}, \dots, \alpha_{nD}\}$. In addition, a Gamma prior is imposed on them

$$p(\alpha_n) = \prod_{d=1}^D \text{Gamma}(\alpha_{nd} | b_\alpha, c_\alpha) \propto \prod_{d=1}^D \alpha_{nd}^{c_\alpha - 1} e^{-b_\alpha \alpha_{nd}}. \quad (23)$$

In this way, a two-stage hierarchical prior is achieved which is actually a Student- t distribution with heavy tails [7]. Sparsity is obtained since this scheme enforces most hyperparameters α_{nd} to be large; thus, the corresponding coefficients w_{nd} are set to zero and finally eliminated.

The second term of the proposed Gibbs energy function captures the spatial correlation and is responsible for the clique

potential of the n th voxel

$$V_{N_n}^{(2)}(\mathbf{w}_n) = \beta_n \sum_{k \in N_n} z_{nk} \|\mathbf{w}_n - \mathbf{w}_k\|^2. \quad (24)$$

In comparison with the potential function of the basic spatial scheme (12), this formulation provides a variation in the neighbors' contribution to the calculation of the clique potential value, as reflected by the parameters z_{nk} . Experiments have shown that the introduction of such weights can increase the flexibility of spatial modeling and can be proved advantageous in cases around the borders of the activation regions (edges). In the literature, there are other Gibbs potential functions that embody the same desired property around edges, see for example [30]. However, they do not provide closed-form update rules such as in our case.

The prior distribution of the regression coefficients of the n th voxel is then written as

$$p(\mathbf{w}_n | \beta_n, z_n, \alpha_n) = Z_n \exp \left\{ -\frac{1}{2} V_{N_n}^{(1)}(\mathbf{w}_n) - \frac{1}{2} V_{N_n}^{(2)}(\mathbf{w}_n) \right\} \quad (25)$$

where the normalization factor Z_n can be expressed as

$$Z_n \propto \beta_n^{|N_n|/2} \prod_{k \in N_n} z_{nk}^{1/2} \prod_{d=1}^D \alpha_{nd}^{1/2}. \quad (26)$$

We also assume that the regularization parameter β_n , the noise precision λ_n , and the weights z_{nk} are Gamma distributed variables. Based on the aforesaid formulation, the data analysis problem can be treated as a MAP approach for the set of regression model variables $\Theta = \{\mathbf{w}_n, \xi_n, \beta_n, \lambda_n, z_n, \alpha_n\}_{n=1}^N$. The MAP log-likelihood function can be given as

$$\begin{aligned} L_{\text{MAP}}(\Theta) = & \sum_{n=1}^N \left\{ \frac{M}{2} \log \lambda_n - \frac{\lambda_n}{2} \|\Xi_n(\mathbf{y}_n - \Phi \mathbf{w}_n)\|^2 \right. \\ & - \frac{1}{2} \mathbf{w}_n^T \mathbf{A}_n \mathbf{w}_n - \frac{\beta_n}{2} \sum_{k \in N_n} z_{nk} \|\mathbf{w}_n - \mathbf{w}_k\|^2 + \frac{|N_n|}{2} \log \beta_n \\ & + \frac{1}{2} \sum_{k \in N_n} \log z_{nk} + \frac{1}{2} \sum_{d=1}^D \log \alpha_{nd} + G(\beta_n) + G(\lambda_n) \\ & \left. + \sum_{k \in N_n} G(z_{nk}) + \sum_{d=1}^D G(\alpha_{nd}) \right\}. \quad (27) \end{aligned}$$

Setting the partial derivatives with respect to regression coefficients equal to zero, we obtain the following update rule:

$$\hat{\mathbf{w}}_n = (\lambda_n \mathbf{F}_n \Phi + BZ_n + \mathbf{A}_n)^{-1} (\lambda_n \mathbf{F}_n \mathbf{y}_n + BZW_n) \quad (28)$$

where

$$\begin{aligned} BZ_n &= \sum_{k \in N_n} (\beta_n z_{nk} + \beta_k z_{kn}) \mathbf{I} \\ BZW_n &= \sum_{k \in N_n} (\beta_n z_{nk} + \beta_k z_{kn}) \hat{\mathbf{w}}_k. \end{aligned}$$

For the rest three model variables $\{\beta_n, z_n, \alpha_n\}$, the next update equations are found

$$\hat{\beta}_n = \frac{|N_n| + 2c_\beta}{\sum_{k \in N_n} z_{nk} \|\hat{\mathbf{w}}_n - \hat{\mathbf{w}}_k\|^2 + 2b_\beta} \quad (29)$$

$$\hat{z}_{nk} = \frac{1 + 2c_z}{\hat{\beta}_n \|\hat{\mathbf{w}}_n - \hat{\mathbf{w}}_k\|^2 + 2b_z} \quad (30)$$

$$\hat{\alpha}_{nd} = \frac{1 + 2c_a}{\hat{w}_{nd}^2 + 2b_a} \quad (31)$$

while the AR coefficients ξ_n and the noise precision λ_n have the same form as previously defined [(9) and (20), respectively]).

Again, the whole procedure can be incorporated into an EM framework by treating the regression coefficients as hidden variables. In this way, their expectation is computed in the E -step (28), while the maximization of the complete-data MAP log-likelihood function is performed during the M -step giving update rules for the model parameters (29)–(31). We call this method SSGLM. Following (27), it is easy to see that when $a_{nd} = 0$ the proposed method is reduced to the previously described spatial GLM scheme (setting also $z_{nk} = 1$) keeping only the spatial component. On the opposite case, when $\beta_n = 0$ or $z_{nk} = 0$ only the sparse part is maintained and this scheme becomes equivalent to the RVM-based sparse regression modeling [7]. In the Appendix, we present an EM-based alternative description of the aforementioned model, where we obtain the marginal distribution of the observations \mathbf{y}_n by integrating out the regression coefficients \mathbf{w}_n and treating them as hidden variables.

A. Statistics

After finishing with the estimation procedure, the BOLD contrast can be displayed by calculating the quantity $\mathbf{c}^T \hat{\mathbf{w}}_n$ for each voxel, where \mathbf{c} is the (given) associated contrast vector that specifies particular differences over the coefficients. Two other statistical maps can be also drawn for showing the brain activity according to two statistics: the t -statistics and the posterior distribution, based on the classical and the Bayesian approach, respectively [31]. In the classical approach, the statistical parametric map (SPM) is created by calculating the t -statistic for each voxel

$$t_n = \frac{\mathbf{c}^T \hat{\mathbf{w}}_n}{\sqrt{\mathbf{c}^T \mathbf{C}_{\hat{\mathbf{w}}_n} \mathbf{c}}} \quad (32)$$

where $\mathbf{C}_{\hat{\mathbf{w}}_n}$ is the covariance of the regression coefficients $\hat{\mathbf{w}}_n$ that is calculated according to (38) (assuming a Gaussian distribution). Alternatively, under the Bayesian framework a mapping of the brain is created according to the posterior probability value

$$pp_n = 1 - \Psi \left(\frac{\gamma - \mathbf{c}^T \hat{\mathbf{w}}_n}{\sqrt{\mathbf{c}^T \mathbf{C}_{\hat{\mathbf{w}}_n} \mathbf{c}}} \right) \quad (33)$$

that a voxel is activated or its effect size γ is greater than a threshold. In this form, the function $\Psi(\cdot)$ is the normal cumulative distribution function. These values are then used for obtaining the posterior probability map [32]. During our

experimental study, we present only activation maps produced by SPM.

IV. EXPERIMENTAL RESULTS

We have tested the proposed method (SSGLM) using various simulated and real datasets. In an attempt to examine the contribution of both spatial and sparse properties, we have made comparison with two versions of the GLM: the SeGLM that uses only the spatial constraints as derived by the potential function of (24) and the SpGLM that captures only the potential function related to the sparse properties of the model (22). We have also compared our method with the spatial ICA [12], where we have followed the method described in [13] for estimating the number of independent components (ICs).² Note that during studying simulated datasets where the ground truth was available, the resulting statistical map of ICA was obtained from the independent component that presents the largest correlation with the ground truth. All GLM-based methods were initialized by first calculating the ML estimates of the regression coefficients \mathbf{w}_n (7), and used them next for setting the rest model parameters ξ_n , λ_n , β_n , z_{kn} , and a_{np} , according to (9), (20), and (29)–(31), respectively.³ Also, all Gamma parameters were set equal to 0.5, except those of the error precision and the sparse responsible variables that were set as $b_\lambda = c_\lambda = b_\alpha = c_\alpha = 10^{-6}$ making them noninformative.

A. Experiments With Simulated Data

The simulated datasets were created according to the following generation mechanism. We used a design matrix with two prespecified regressors: the first one, shown in Fig. 1(a), was responsible for the BOLD signal \mathbf{s} of length $M = 84$ and it was selected from a real experiment found on the SPM package while the second one being a constant of ones. Then, we constructed a binary image showing the activated and nonactivated areas, where its pixel intensities correspond to the value of the first coefficient (w_{n1}). Two such simulated images were used for this purpose: the first image [see Fig. 1(b)] was created by a statistical map of a real fMRI example after ML-based estimating the regression coefficients and setting a threshold to the t -values (Dataset 1). The second image [see Fig. 1(c)] was taken from [33] and created by sampling from an MRF model using a Gibbs sampler (Dataset 2). The second coefficient w_{n2} had a constant value equal to 100. The time series data (\mathbf{y}_n) were calculated by using the generative equation of GLM (1) with an additive Gaussian noise of various SNR levels, calculated as

$$\text{SNR} = 10 \log \frac{\mathbf{s}^T \mathbf{s}}{M(1/\lambda_n)}. \quad (34)$$

The noise was constructed according to an AR model of order $p = 3$, whose coefficients took the values $\xi_n = (1, -0.8, 0.6, -0.4)^T$, as proposed in [29].

Two evaluation criteria were used.

²We have used the fast ICA software with its default settings downloaded from <http://www.cis.hut.fi/projects/ica/fastica/>.

³Due to relation between β_n and z_{kn} , we used (19) instead of (29).

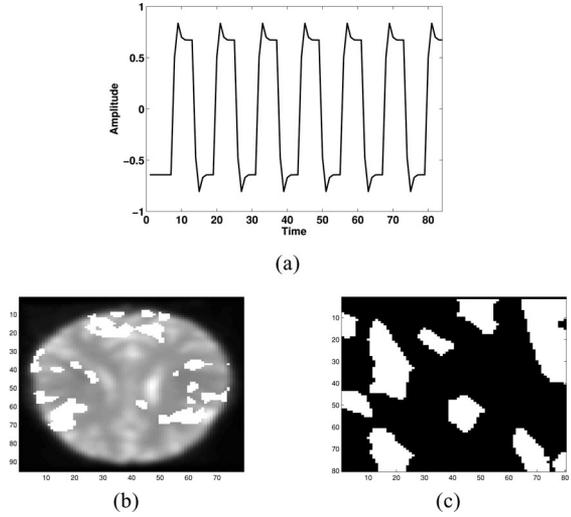


Fig. 1. (a) BOLD signal \mathbf{s} . (b), (c) Two images with random shaped activated areas used for generating simulated data.

- 1) The area under curve (AUC) of the receiver operating curve (ROC) based on t -statistic calculations that reflects the ability of the method to detect the real activations, while minimizing the detections of false activations. ROC curves were generated by considering a voxel to be active if its effect size is greater than a predefined threshold. In our experiments, the aforementioned threshold varied from the minimum to the maximum value of the t -statistic as calculated by each method.
- 2) The normalized mean square error (NMSE) that measures the quality of the curve fitting procedure

$$\text{NMSE} = \frac{\sum_{n=1}^N (\hat{w}_{n1} - \omega_{n1})^2}{\sum_{n=1}^N \omega_{n1}^2} \quad (35)$$

where \hat{w}_{n1} and ω_{n1} are the estimated and the true regression coefficients responsible for the BOLD signal.

Also, we have selected a design matrix (Φ) with 12 columns ($D = 12$): one for the BOLD signal, one column with ones for the constant term, and the first ten basis functions from the discrete cosine transform (DCT) that corresponds to low-pass frequency components of fMRI time series [2]. For every noise realization (SNR value), we performed 50 different runs of each method and calculated statistics (mean value and standard deviation) of the AUC and NMSE measurements. Note that for the spatial ICA method only the AUC measurement was calculated, since the regression coefficients could not be estimated. Furthermore, the resulting statistical map of ICA was obtained from the independent component that had the highest correlation with the ground truth signal.

The diagrams in Fig. 2 present the comparative performance results using error bars for the two activation images in Fig. 1. As is obvious, the proposed method improves not only the activation detection ability (AUC), but also the quality of fitting process (NMSE). Also, the error bars indicate that the observed differences between the methods are statistically significant. Between both versions of GLM, the SeGLM approach, that incorporates

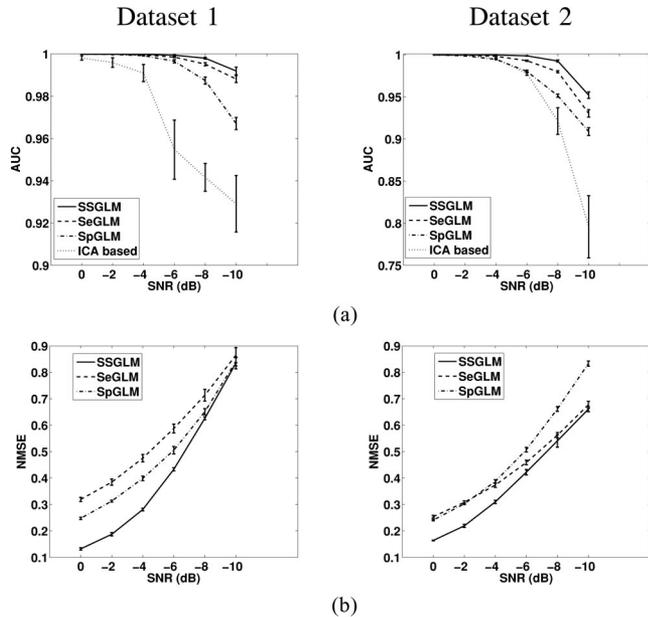


Fig. 2. Plots of mean values of AUC and NMSE with error bars for various SNR values in the case of two simulated images in Fig. 1.

spatial properties, had a tendency toward better discriminative ability. This was expected, since the activated areas in both simulated images in Fig. 1 are not smoothed. On the other hand, the sparse capability during modeling of the SpGLM method causes into improving curve fitting procedure and leads to better NMSE values. This behavior can be further explained by presenting in Fig. 3 some examples of the produced activation maps in the case of the first simulated image [see Fig. 1 (b)] for SNR value -10 dB and two false positive rates (FPR) 0.01 and 0.1. As can be observed, the SeGLM method tends to overestimate the activated areas and produces larger than the true ones. On the other hand, the SpGLM method fails to produce spatially connected areas and detects many small false positive islands. The proposed method SSGLM achieves a balance between spatially distributed modeling and sparse smoothing that improves model accuracy and efficiency. Another useful observation extracted from the activation maps in Fig. 3 is that our approach shows a much reduced dependence on the threshold choice of t -values since it maintains a stable behavior for various values of FPR. Finally, the spatial ICA gave the worst performance and only in large SNR values it yielded results comparable to those obtained by other approaches.

To gain more insight into the discriminative capability of our method, a more comprehensive analysis of ROC curves was performed. More specifically, Fig. 4 illustrates the calculated mean ROC curves with error bars in the case of the activation image in Fig. 1(b) for two noisy environments of -6 dB and -10 dB. Note that the FPR on x -axis is shown in logarithmic scale. Obviously, the SSGLM method exhibits better results than the other two methods, SeGLM and SpGLM. For example, in the case of -6 dB [see Fig. 4(a)] and for FPR equal to 0.001, the SSGLM method gives 90% of the detection accuracy, while its peers have 70% and 55%, respectively. Higher values of the

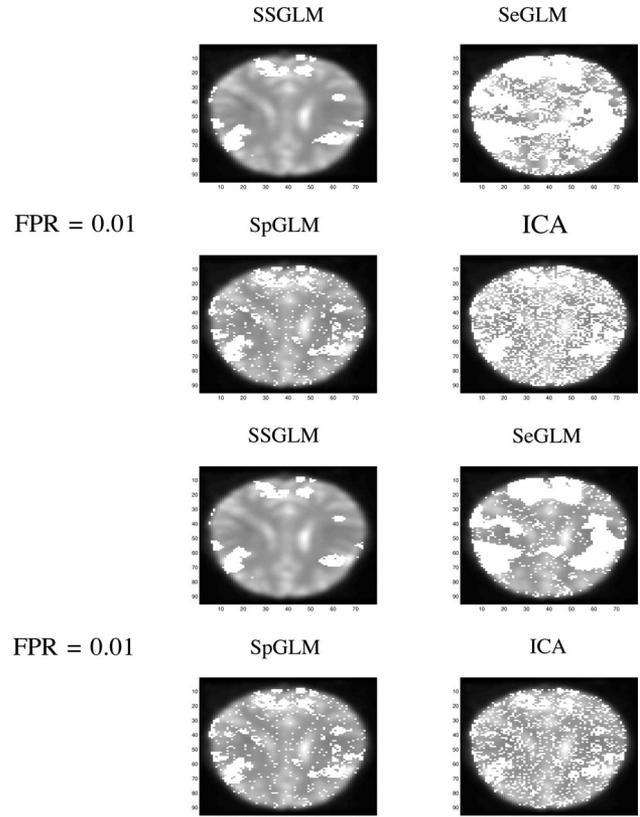


Fig. 3. Resulting activation maps of four comparative methods on SNR -10 dB for two values of FPR.

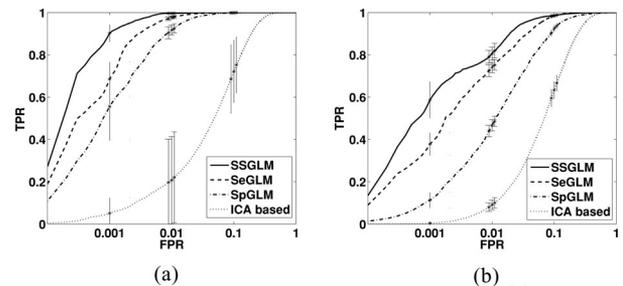


Fig. 4. Mean ROC curves of SSGLM, SeGLM, and SpGLM for (a) -6 dB and (b) -10 dB. FPR is shown in logarithmic scale.

FPR are needed for the SeGLM and the SpGLM methods in order to reach 90%. Likewise, in the case of SNR value -10 dB the SSGLM method shows an accuracy of 60% for FPR equal to 0.001, while the SeGLM and the SpGLM methods have 40% and 10%, respectively. Finally, the performance of the spatial ICA method was poor at small FPR values. As shown in Fig. 4, an accuracy of 90% is achieved when the FPR is much greater than 0.1.

B. Experiments With Real fMRI Data

The proposed approach was also evaluated in a variety of real applications. Since the ground truth is not available in this study, we have selected to visually compare the resulting activation maps by all methods. For any selected dataset, we

followed the standard preprocessing steps of the SPM package, i.e., realignment, segmentation, and spatial normalization, without performing the spatial smoothing step. Data are then scaled by means of their mean value, as described in [8], and finally high-pass filtered using a set of discrete cosine basis functions. During all experiments, we have chosen an AR model of order $p = 3$, suggested in [29]. On the other hand, for evaluating the spatial ICA method we manually selected the independent component (among those estimated from the approach described in [13]) that best meets the design of the experiment and resembles the expected brain activation area.

1) *Block Design fMRI Data*: At first, we have studied a real block design fMRI dataset designed for auditory processing task on a healthy volunteer.⁴ It consisted of 96 acquisitions made in blocks of 6, giving 16 blocks of 42-s duration. The condition for successive blocks alternated between rest and auditory stimulation, starting with rest. Auditory stimulation was performed with bisyllabic words presented binaurally at a rate of 60 words per minute. The functional data start at acquisition 16. The whole brain BOLD/echo-planar imaging (EPI) images were acquired on a modified 2T Siemens MAGNETOM Vision system. A detailed description of it can be found on the SPM manual [34]. After preprocessing, functional images consisted of 68 slices ($79 \times 95 \times 68$, $2 \text{ mm} \times 2 \text{ mm} \times 2 \text{ mm}$ voxels).

We have selected slice 29 of this dataset for making experiments. Twelve regressors ($D = 12$) for the design matrix were used: one for the BOLD response, another one having constant values of ones for modeling the mean brain activity, and finally ten more regressors obtained by DCT basis functions.⁵ Fig. 5(a) illustrates the activation maps of all comparative approaches. The proposed method illustrates clearer an activity around the auditory cortex, as it was the expected. On the other hand, the SeGLM and the spatial ICA methods construct very large and confused activation areas, while the SpGLM has a tendency to yield many small (disconnected) islands.

Another useful observation can be obtained by presenting the statistical maps of three GLM-based methods after setting a threshold value on the calculated t -values.⁶ As shown in Fig. 5, the imposition of a threshold does not affect the performance of the SSGLM method since both activation maps are very similar. This is more apparent in Fig. 6(a) where we plot the estimated size of activated areas (number of voxels) by each method in terms of the threshold value (obtained with a varying significance level from 0.0001 to 0.5). The same observation is made by plotting the calculated t -values of three methods in Fig. 6(b), where the distinction between the activated and nonactivated voxels becomes much easier in the SSGLM method. From a clinical perspective, this is very important since in the standard fMRI analysis the activation boundary varies significantly with the smoothing and the statistical threshold used, and thus complicates clinical decisions [35]. This problem is alleviated by our

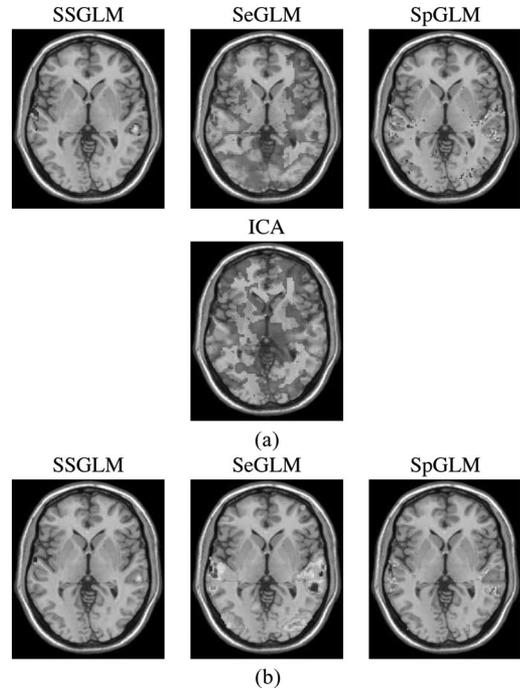


Fig. 5. SPMs of comparative methods based on t -values (a) without and (b) with setting a threshold value.

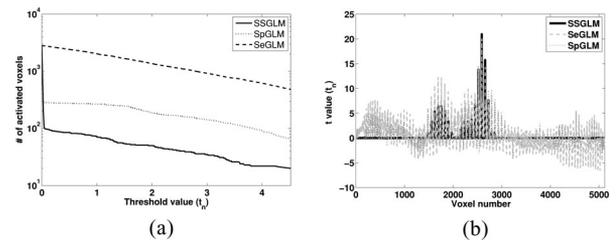


Fig. 6. Plots of (a) the estimated number of activated voxels in terms of threshold value (in logarithmic scale) and (b) the t -values of SPMs as computed by three comparative methods.

methodology which does not require smoothing and produces results that are very insensitive to the threshold choice.

2) *Event-Related Design fMRI Data*: Additional experiments were made by considering event-related cases. At first, we have used a public available dataset found on the SPM web page. This was designed for face recognition using grayscale images of faces, where the subject was asked to make face judgments making key presses. Images were acquired using continuous EPI with $TE = 40 \text{ ms}$, $TR = 2 \text{ s}$, and 24 descending slices (64×64 , $3 \text{ mm} \times 3 \text{ mm}$). More information about these data can be found in [36]. In our study, we have selected slice 18 for making experiments. The contrast vector was set as $\mathbf{c} = [1, 1, 1, 1, 0]$ that describes the response to the presentation of a face image. We have used a design matrix that consists of five ($D = 5$) regressors related to four types of events. In particular, the first four regressors indicate the presence of a face and have been convolved with a “canonical” hemodynamic response function (HRF) while the last one is the constant term. Fig. 7 presents the resulting statistical activation maps. As it is obvious, the most

⁴Obtained from the SPM web page <http://www.fil.ion.ucl.ac.uk/spm/>.

⁵Additional experiments were made with various numbers of DCT basis functions with similar results.

⁶The significance level was set to 0.0001, which gives a threshold value $t_0 = 3.8942$ applied to (32).

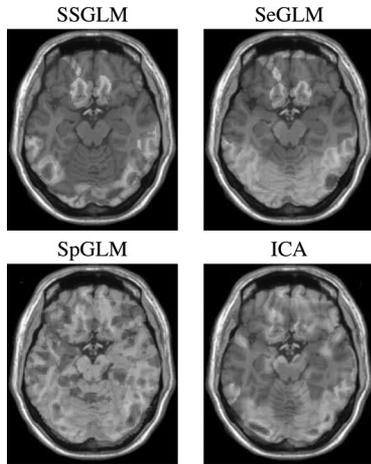


Fig. 7. Estimated statistical maps by four comparative methods in a real visual event-related experiment.

important voxels estimated by the SSGLM method are exclusively observed in the occipital lobe. Comparing with the other methods, the resulting activation areas of whom are to some extent scattered throughout the brain, the proposed approach manages to produce smoother and clearer results.

In the second event-related experiment, we analyzed fMRI data consisted of images acquired from a motor event-related paradigm available at the affiliated University Hospital of Ioannina. During this experiment, patients with restless legs syndrome (RLS) performed random and spontaneous limb movements evoked by sensory leg uneasiness. These movements were used to create the indicator vector in our modeling that was convolved next with the HRF in order to provide the BOLD signal. The motor fMRI dataset was acquired with a clinical magnet 1.5T (Philips Intera) equipped with fast gradients. The imaging protocol consisted of: 1) axial T1-Weighted—a 3-D sequence that provided a high-resolution reference for alignment between subjects. We used a 3-D-volume magnetization-prepared rapid acquisition with gradient echoes (MP-RAGE). Parameters are TR = 25 ms; TE = 4,6 ms; flip angle = 30°; slice thickness = 0.8 mm; acquisition matrix = 256; reconstruction matrix = 256. 2) Multilevel BOLD fMRI using a hand task: we used a gradient echo EPI. Parameters are TR = 2000 ms; TE = 50 ms; flip angle = 40°; acquisition matrix = 64; reconstruction matrix = 128; slice thickness = 5 mm; gap = 0; 160 dynamics. Four dummy scans were performed in the beginning of fMRI to stabilize the magnetization level and allow us to keep all the images. The paradigm consisted of alternate action and resting blocks, 20 s each. During the action epoch the subjects flexed and released continuously at 0.5 Hz rate the fingers of their right hand in unison. The cues for action or rest were announced through headphones using the commands start and stop. The paradigm of 8 action and 8 rest blocks lasted 320 s, and we collected 20 images/slice at TR = 2 s. We restricted motion artifacts by using foam rubber pads and strap across the forehead. Eyes were kept closed, at all times. A detailed description about how these data were acquired can be found in [37]. The design matrix had

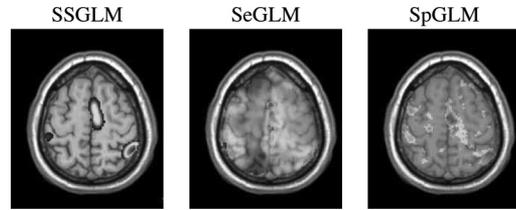


Fig. 8. Resulting statistical maps in the real motor event-related experiment.

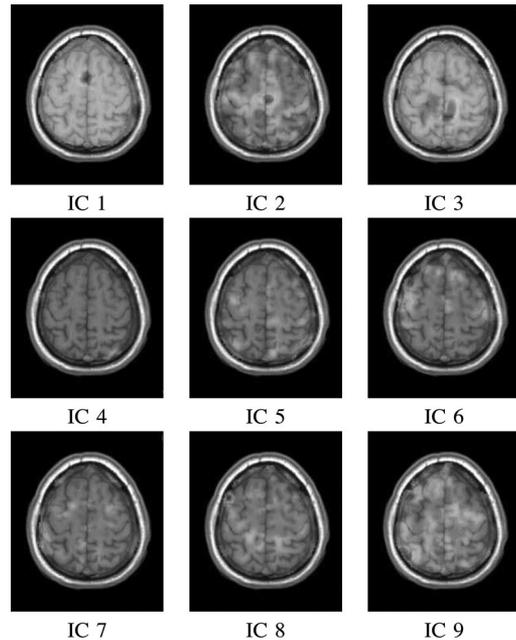


Fig. 9. Nine produced ICs of the motor event-related experiment.

four columns ($D = 4$): the BOLD signal, its time and dispersion derivatives, and a constant.

In our study, we examined the main effect (leg movement versus rest) which means that the contrast vector was set equal to $\mathbf{c} = [1, 0, 0, 0]$. The estimated statistical maps are shown in Fig. 8 for slice 54 of the dataset. Our method produced brain activated regions related to motor function such as: 1) the supplementary motor area; 2) the primary motor areas (precentral gyrus); and 3) the superior parietal lobe. In comparison with the other approaches, it provides contrast maps that appear neurobiologically plausible, where even a simple visual inspection reveals localized maxima in good agreement with the current knowledge of the locations and extent of motor circuitry. In this experiment, the activation map estimated by the spatial ICA was split into many components making the visualization a hard task. This is shown in Fig. 9 that presents nine ICs, each one of which captures only a part of the brain regions related to the motor function.

V. CONCLUSION

In fMRI data analysis, the spatial extension of the hemodynamic response in a neighborhood of voxels introduces a significant weakness for the detection process of the activated areas. Moreover, the presence of temporal correlations

deteriorates the performance. In this paper, we present an advanced method to tackle these two problems by efficiently incorporating both spatial correlations and sparse properties. This is done by using an appropriate prior over the regression coefficients based on MRFs modeling and RVMS. Training of the proposed model is achieved through an EM-MAP framework providing update rules in closed form. Experiments on artificial and real datasets have demonstrated the ability of the method to improve the detection performance and robustness, especially in noisy environments, with enhanced estimation accuracy. Our method showed a significantly reduced sensitivity to the threshold value of the produced statistical map without needing to make multiple comparisons. Our future research is focused to three directions: 1) to examine the appropriateness of other types of sparse priors [38]; 2) to try alternative potential functions of the Gibbs distribution; and 3) to assume a Student- t distribution instead of Gaussian for modeling the excitation noise aiming to achieve more robust statistical inference and handle more efficiently outlying observations [39].

APPENDIX

In the previous analysis, the regression coefficients have been treated as model parameters. However, following the RVM methodology [7], we can integrate them out and obtain a reduced model with less parameters $\Theta = \{\beta_n, z_{nk}, \alpha_{nd}, \lambda_n, \xi_n\}_{n=1}^N$. The marginal log-likelihood for each voxel n can be obtained by the following integration:

$$\begin{aligned} & \log p(\mathbf{y}_n | \beta_n, z_{nk}, \alpha_{nd}, \xi_n) \\ &= \log \int p(\mathbf{y}_n | \mathbf{w}_n, \lambda_n, \xi_n) p(\mathbf{w}_n | \beta_n, z_{nk}, \alpha_{nd}) d\mathbf{w}_n. \end{aligned} \quad (36)$$

Since both densities are known (4) and (25), we can easily found the marginal log-likelihood

$$\begin{aligned} & \log p(\mathbf{y}_n | \beta_n, z_{nk}, \alpha_{nd}, \xi_n) \propto \frac{|N_n|}{2} \log \beta_n \\ & - \frac{1}{2} \sum_{k \in N_n} \log z_{nk} - \frac{1}{2} \sum_{d=1}^D \log \alpha_{nd} - \frac{1}{2} \log |\mathbf{S}_n| \\ & + \frac{M}{2} \log \lambda_n - \frac{1}{2} \left\{ \mathbf{m}_n \mathbf{S}_n \mathbf{m}_n - \lambda_n \mathbf{y}_n^T \Xi_n^T \Xi_n \mathbf{y}_n \right. \\ & \left. - \beta_n \sum_{k \in N_n} z_{nk} \mathbf{w}_k^T \mathbf{w}_k \right\} \end{aligned} \quad (37)$$

where

$$\mathbf{S}_n = (\lambda_n \mathbf{F}_n \Phi + BZ_n + \mathbf{A}_n)^{-1} \quad (38)$$

$$\mathbf{m}_n = \mathbf{S}_n (\lambda_n \mathbf{F}_n \mathbf{y}_n + BZW_n). \quad (39)$$

Therefore, the MAP log-likelihood function is written as

$$\begin{aligned} L_{\text{MAP}}(\Theta) = & \sum_{n=1}^N \left\{ \log p(\mathbf{y}_n | \beta_n, z_n, \alpha_n, \xi_n) + G(\beta_n) \right. \\ & \left. + G(\lambda_n) + \sum_{k \in N_n} G(z_{nk}) + \sum_{d=1}^D G(\alpha_{nd}) \right\}. \end{aligned} \quad (40)$$

Following the MAP-EM framework, the regression coefficients \mathbf{w}_n are treated as hidden variables where their expectation is calculated at the E -step (28). During the M -step, the expectation of the complete data MAP log-likelihood function (27) is maximized, where the expectation is made with respect to the posterior distribution of regression coefficients \mathbf{w}_n . Notice here that this posterior can be considered as Gaussian with mean \mathbf{m}_n and covariance \mathbf{S}_n . By setting the partial derivatives with respect to model parameters equal to zero, the next update rules are obtained

$$\hat{\beta}_n = \frac{|N_n| + 2c_\beta}{\sum_{k \in N_n} \hat{z}_{nk} \mathcal{E}_{\mathbf{w}_n | \mathbf{y}_n, \theta_n} \{ \|\mathbf{w}_n - \mathbf{w}_k\|^2 \} + 2c_\beta} \quad (41)$$

$$\hat{z}_{nk} = \frac{1 + 2c_z}{\hat{\beta}_n \mathcal{E}_{\mathbf{w}_n | \mathbf{y}_n, \theta_n} \{ \|\mathbf{w}_n - \mathbf{w}_k\|^2 \} + 2b_z} \quad (42)$$

$$\hat{\alpha}_{nd} = \frac{1 + 2c_a}{\mathcal{E}_{\mathbf{w}_n | \mathbf{y}_n, \theta_n} \{ w_{nd}^2 \} + 2b_a} \quad (43)$$

$$\hat{\lambda}_n = \frac{M + 2c_\lambda}{\mathcal{E}_{\mathbf{w}_n | \mathbf{y}_n, \theta_n} \{ \|\Xi_n(\mathbf{y}_n - \Phi \mathbf{w}_n)\|^2 \} + 2b_\lambda} \quad (44)$$

$$\hat{\xi}_n = (\mathbf{E}_n^T \mathbf{E}_n)^{-1} \mathbf{E}_n \mathcal{E}_{\mathbf{w}_n | \mathbf{y}_n, \theta_n} \{ \|\Xi_n(\mathbf{y}_n - \Phi \mathbf{w}_n)\|^2 \} \quad (45)$$

which are iteratively applied. The aforesaid expectations can be easily calculated using

$$\mathcal{E}_{\mathbf{w}_n | \mathbf{y}_n, \theta_n} \{ \mathbf{w}_n \} = \mathbf{m}_n, \quad \mathcal{E}_{\mathbf{w}_n | \mathbf{y}_n, \theta_n} \{ \mathbf{w} \mathbf{w}_n^T \} = \mathbf{S}_n + \mathbf{m}_n \mathbf{m}_n^T. \quad (46)$$

ACKNOWLEDGMENT

This paper is dedicated to the memory of our teacher, friend, and colleague Prof. N. P. Galatsanos, who contributed significantly to the research and preparation of this manuscript through helpful advice and comments, but unfortunately recently passed away as we were finalizing revisions. He was a great mentor who continues to provide inspiration to all of his former colleagues.

REFERENCES

- [1] P. Jezzard, P. M. Matthews, and S. M. Smith, *Functional MRI: An Introduction to Methods*. New York: Oxford Univ. Press, 2001.
- [2] R. Frackowiak, J. Ashburner, W. Penny, S. Zeki, K. Friston, C. Frith, R. Dolan, and C. Price, *Human Brain Function*, 2nd ed. New York: Elsevier, 2004.
- [3] K. J. Friston, "Analysis of fMRI time series revisited," *NeuroImage*, vol. 2, pp. 45–53, 1995.
- [4] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, 2007.
- [5] L. Harrison, W. Penny, J. Daunizeau, and K. Friston, "Diffusion-based spatial priors for functional magnetic resonance images," *NeuroImage*, vol. 41, no. 2, pp. 408–423, 2008.
- [6] W. D. Penny, N. J. Trujillo-Barreto, and K. J. Friston, "Bayesian fMRI time series analysis with spatial priors," *NeuroImage*, vol. 24, pp. 350–362, Jan. 2005.
- [7] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.
- [8] G. Flandin and W. Penny, "Bayesian fMRI data analysis with sparse spatial basis function priors," *NeuroImage*, vol. 34, pp. 1108–1125, 2007.
- [9] X. Descombes, F. Kruggel, and von D.Y. Cramon, "fMRI signal restoration using a spatio-temporal Markov random field preserving transitions," *NeuroImage*, vol. 8, pp. 340–349, 1998.
- [10] N. Hartvig and J. Jensen, "Spatial mixture modeling of fMRI data," *Human Brain Mapp.*, vol. 11, no. 4, pp. 233–248, 2000.

- [11] O. Friman, M. Borga, P. Lundberg, and H. Knutsson, "Adaptive analysis for fMRI data," *Neuroimage*, vol. 19, no. 3, pp. 837–845, 2003.
- [12] V. Calhoun, T. Adali, L. K. Hansen, J. Larsen, and J. J. Pekar, "Independent component analysis of functional MRI data: An overview," in *Proc. Int. Compon. Anal. (ICA)*, 2003, 2012, pp. 281–288.
- [13] Y. Li, T. Adali, and V. Calhoun, "Estimating the number of independent components for functional magnetic resonance imaging data," *Human Brain Mapp.*, vol. 28, no. 11, pp. 1251–1266, 2007.
- [14] C. Gossel, D. P. Auer, and L. Fahrmeir, "Bayesian spatiotemporal inference in functional magnetic resonance imaging," *Biometrics*, vol. 57, pp. 554–562, 2001.
- [15] M. Woolrich, M. Jenkinson, J. Brady, and S. Smith, "Fully Bayesian spatio-temporal modeling of fMRI data," *IEEE Trans. Med. Imag.*, vol. 23, no. 2, pp. 213–231, Feb. 2004.
- [16] M. Smith and L. Fahrmeir, "Spatial Bayesian variable selection with application to functional magnetic resonance imaging," *J. Amer. Statist. Assoc.*, vol. 102, no. 478, pp. 417–431, 2007.
- [17] M. van Gerven, B. Cseke, F. P. de Lange, and T. Heskes, "Efficient Bayesian multivariate fMRI analysis using a sparsifying spatio-temporal prior," *Neuroimage*, vol. 50, pp. 150–161, 2010.
- [18] W. Ou, "fMRI detection via variational EM approach," MIT OpenCourseWareTech. Rep., 2005. Available: <http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-881-representation-and-modeling-for-image-analysis-spring-2005/projects/>
- [19] L. Fahrmeir, C. Gossel, and A. Hennerfeind, "Robust Spatial Smoothing in functional MRI," in *Exploratory Data Analysis in Empirical Research*, M. Schwaiger and O. Opitz, Eds. New York: Springer-Verlag, 2002, pp. 50–57.
- [20] W. Ou and P. Golland, "From spatial regularization to anatomical priors in fMRI analysis," in *Proc. Inf. Process. Med. Imag.*, 2005, pp. 88–100.
- [21] W. Ou, W. M. Wells, and P. Golland, "Combining spatial priors and anatomical information for fMRI detection," *Med. Image Anal.*, vol. 14, no. 3, pp. 318–331, 2010.
- [22] H. Luo and S. Puthusserypady, "A sparse Bayesian method for determination of flexible design matrix for fMRI data analysis," *IEEE Trans. Circuits Syst. I: Regular Papers*, vol. 52, no. 12, pp. 2699–2706, Dec. 2005.
- [23] H. Luo and S. Puthusserypady, "fMRI data analysis with nonstationary noise models: A Bayesian approach," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 9, pp. 1621–1630, Sep. 2007.
- [24] H. Luo and S. Puthusserypady, "Analysis of fMRI data with drift: Modified general linear model and Bayesian estimator," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 5, pp. 1504–1511, May 2008.
- [25] A. Lukic, M. Wernick, D. Tzikas, X. Chen, A. Likas, N. Galatsanos, Y. Yang, F. Zhao, and S. Strother, "Bayesian kernel methods for analysis of functional neuroimages," *IEEE Trans. Med. Imag.*, vol. 26, no. 12, pp. 1613–1622, Dec. 2007.
- [26] M. Carroll, G. Cecchi, I. Rish, R. Garg, and A. R. Rao, "Prediction and interpretation of distributed neural activity with sparse models," *NeuroImage*, vol. 44, pp. 112–122, 2009.
- [27] Y. Li, P. Namburi, Z. Yu, C. Guan, J. Feng, and Z. Gu, "Voxel selection in fMRI data analysis based on sparse representation," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 10, pp. 2439–2451, Oct. 2009.
- [28] A. Dempster, A. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *J. R. Statist. Soc., Series B*, vol. 39, pp. 1–38, 1977.
- [29] W. Penny, S. Kiebel, and K. Friston, "Variational Bayesian inference for fMRI time series," *NeuroImage*, vol. 19, pp. 727–741, Jul. 2003.
- [30] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, Jul. 1995.
- [31] K. J. Friston and W. Penny, "Posterior probability maps and SPMs," *NeuroImage*, vol. 19, pp. 1240–1249, Jul. 2003.
- [32] K. J. Friston, *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. New York: Academic, 2006.
- [33] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
- [34] *SPM8 Manual*, The FIL Methods Group, Functional Imaging Laboratory, Wellcome Trust Centre for Neuroimaging, Institute of Neurology, London, U.K., 2009.
- [35] E. Vlieger, C. Majoie, S. Leenstra, and G. den Heeten, "Functional magnetic resonance imaging for neurosurgical planning in neurooncology," *Eur. Radiol.*, vol. 17, no. 7, pp. 1143–1153, 2004.
- [36] R. Henson, T. Shallice, M. Gorno-Tempini, and R. Dolan, "Face repetition effects in implicit and explicit memory tests as measured by fMRI," *Cereb. Cortex*, vol. 12, no. 2, pp. 178–186, 2002.
- [37] L. Astrakas, S. Konitsiotis, P. Margariti, S. Tsouli, L. Tzarouhi, and M. I. Argyropoulou, "T2 relaxometry and fMRI of the brain in late-onset restless legs syndrome," *Neurology*, vol. 71, no. 12, pp. 911–916, 2008.
- [38] M. Seeger, "Bayesian inference and optimal design for the sparse linear model," *J. Mach. Learning Res.*, vol. 9, pp. 759–813, 2008.
- [39] K. Lange, R. Little, and J. Taylor, "Robust statistical modeling using the t distribution," *J. Amer. Statist. Assoc.*, vol. 84, pp. 881–896, 1989.



Vangelis P. Oikonomou was born in Karditsa, Greece, in 1976. He received the Dipl., M.Sc., and the Ph.D. degrees in computer science from the University of Ioannina, Janina, Greece, in 2000, 2003, and 2010, respectively.

His research interests include Bayesian methods, medical image processing, and biomedical signal processing.



Konstantinos Blekas received the Diploma degree in electrical engineering in 1993 and the Ph.D. degree in electrical and computer engineering in 1997, both from the National Technical University of Athens, Athens, Greece.

He is currently on the Faculty of the Department of Computer Science, University of Ioannina, Janina, Greece. He has coauthored more than 40 refereed journal and conference articles. His research interests include machine learning and statistical pattern recognition with applications to computer vision,

bioinformatics, and medical imaging.



Loukas Astrakas received the Bachelor's degree in physics from the University of Athens, Athens, Greece, in 1989, and the Ph.D. degree from the National Technical University of Athens, Athens, in 1999.

From 1999 to 2004, he was a Postdoctoral Fellow at the Harvard Medical School, Boston, MA, where he conducted research on brain diseases and burn trauma using magnetic resonance imaging and spectroscopy. During 2006–2008, he worked as a Marie Curie Fellow at the University of Ioannina, Janina, Greece. Since 2009, he has been an Instructor in Medical Physics at the University of Ioannina. His current research interest includes biomedical applications of magnetic resonance.