

Research Article

Detecting Activation in fMRI Data: An Approach Based on Sparse Representation of BOLD Signal

Blanca Guillen ¹, Jose L. Paredes,² and Rubén Medina³

¹Department of Mathematics and Physics, Bioengineering Group, UNET, San Cristóbal, Venezuela

²Research Center for Biomedical Engineering and Telemedicine, Electrical Engineering Department, ULA, Mérida, Venezuela

³Computer Science Department, Universidad de Cuenca, Cuenca, Ecuador

Correspondence should be addressed to Blanca Guillen; blaguillen@gmail.com

Received 29 August 2017; Accepted 3 January 2018; Published 15 February 2018

Academic Editor: Roberto Fedele

Copyright © 2018 Blanca Guillen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposes a simple yet effective approach for detecting activated voxels in fMRI data by exploiting the inherent sparsity property of the BOLD signal in temporal and spatial domains. In the time domain, the approach combines the General Linear Model (GLM) with a Least Absolute Deviation (LAD) based regression method regularized by the pseudonorm ℓ_0 to promote sparsity in the parameter vector of the model. In the spatial domain, detection of activated regions is based on thresholding the spatial map of estimated parameters associated with a particular stimulus. The threshold is calculated by exploiting the sparseness of the BOLD signal in the spatial domain assuming a Laplacian distribution model. The proposed approach is validated using synthetic and real fMRI data. For synthetic data, results show that the proposed approach is able to detect most activated voxels without any false activation. For real data, the method is evaluated through comparison with the SPM software. Results indicate that this approach can effectively find activated regions that are similar to those found by SPM, but using a much simpler approach. This study may lead to the development of robust spatial approaches to further simplifying the complexity of classical schemes.

1. Introduction

The medical imaging modality known as functional Magnetic Resonance Imaging (fMRI) using blood oxygen level-dependent (BOLD) contrast is a noninvasive technique widely accepted as a standard tool for localizing brain activity [1]. During the course of an fMRI study, a series of brain images is acquired by repeatedly scanning the subject's brain while he/she is performing a set of tasks or is exposed to a particular stimulus. A statistical analysis is carried out to analyze the images and detect which voxels are activated by a particular stimulation or task.

Since its development in the early 1990s, a variety of univariate and multivariate methods for analyzing fMRI data have been developed [2]. Among these, the most popular are the univariate approaches based on the General Linear Model (GLM) [1–3] and the multivariate approach based on independent component analysis (ICA) [2]. Traditionally, ordinary least squares (OLS) has been used as the primary approach to solve the inverse problem induced by the GLM

owing to its simplicity and optimality under the maximum likelihood (ML) principle when the background noise is modelled as a white Gaussian noise [4]. However, findings contradicting the Gaussian assumption have been reported by Hanson and Bly [5]. Furthermore, statistical tests with real fMRI data have revealed that the empirical distribution of fMRI data has heavier-than-Gaussian tail distribution [5, 6]. Further, Luo and Nichols [7] have found that the BOLD signal often contains outliers. Under the heavy tail distribution model, it is well known that OLS fails in producing a reliable estimate for the GLM regression parameters. On the other hand, an important study conducted by Daubechies et al. [8] showed that the most critical factor for the success rate of the ICA algorithm in determining neural activity is the sparsity of the components rather than its independence.

In addition to these drawbacks, there are also a variety of research studies that have pointed out the existence of a common underlying principle involved in sensory information processing referred to as “sparse coding” [9]. According to this principle, sensory information is represented by a

relatively small number of simultaneously active neurons out of a large population [9]. Literature has reported two notions of sparseness: (1) “lifetime sparseness” refers to the sporadic activity of a single neuron over time, going from near silence to a burst of spikes in response to only a small subset of stimuli (stimulus selectivity); and (2) “population sparseness” refers to the activation of a small fraction of neurons of a population in a given time window in response to a single stimulus [10, 11]. Therefore, it is natural and well-justified to explore sparse representations methods to describe fMRI signals of the brain.

In the last decade, there has been a growing interest in the use of sparse signal representation techniques for analyzing fMRI data based on the assumption that the components of each voxel’s fMRI signal are sparse and the neural integration of those components is linear. These studies include, among others, sparse representation of nonparametric components of partially linear models [12] and the sparseness of the activity related BOLD signal by using “activelets” [13]. Additionally, methods based on sparse dictionary learning for detecting activated regions in fMRI data have been proposed, including data-driven sparse GLM [14], fast incoherent dictionary learning (FIDL) [15], and sparse representation of whole-brain fMRI signals using an online dictionary learning approach [16]. In parallel, sparse models have been proposed to predict or decode task parameters from individual fMRI activity patterns by using methods such as the elastic net method [17], generalized sparse classifiers [18], or multivariate pattern analysis based on sparse representation models [19, 20].

In the context of sparse regression over the coefficients of the GLM, the solution of the regularized inverse problem induced by the GLM is obtained by using ℓ_1 -regularized least squares (ℓ_1 -LS) based methods [21]. Under this approach, the key assumption is that the underlying distribution of BOLD signal is Gaussian, in contrast with the above-mentioned studies where authors argued that the BOLD activation noise follows heavy tailed gamma distribution [5] and Rician distribution [22]. Besides, the impulsive noise is common in fMRI time series [7]. Thus, there is the need of developing robust regression methods that exploit the sparseness structure of the GLM regression parameters, on one hand, and detract from the effect of the impulsive noise on the coefficient estimations, on the other.

Motivated by recent developments in sparse signal representation and the biological findings of ‘sparse coding’ in the brain, in this paper, we propose a simple yet effective approach based on the sparsity of underlying BOLD signal in fMRI data that exploits both temporal and spatial sparse properties of the fMRI images. This paper has two aims, the first to explore the ability of a new robust regression method named ℓ_0 -LAD (ℓ_0 -regularized Least Absolute Deviation) to solve the inverse problem induced by the GLM. This method is suitable for applications where the underlying contamination follows a statistical model with heavier-than-Gaussian tails [23]. The second aim is to test the suitability of a Laplacian model in the spatial domain. More specifically, in the time domain, each time series related to a voxel is considered as a linear combination of a few elements (column

vectors/atoms) of a suitably designed dictionary of stimuli and confounds. The ℓ_0 -LAD algorithm is used to find whether or not a stimulus is present in the observed signal and, if so how much it contributes to signal formation. Subsequently, in the spatial domain, in order to determine brain activation zones due to a particular stimulus, a statistical map is generated, where each voxel’s value is represented by the value of the estimated parameter associated with the stimulus of interest for that voxel. The activated voxels are obtained by thresholding the statistical map, where the threshold is determined by exploiting the previous knowledge about sparsity of activated regions by assuming that the spatial parameters follows a Laplacian distribution model.

Our rationale in spatial domain is that the set of activated voxels in response to a given task is spatially sparse in the sense that only a reduced number of voxels that comprise the brain volume are activated. In fact, the behavior exhibited by different sets of spatial parameters associated with different stimuli of the *in vivo* PBAIC-2007 fMRI database [24] used in this work confirms what it is expected; that is, for each set, a very large number of spatial parameters have very small values while only a reduced number of parameters have significant nonzero value. The problem here is to decide how large must the spatial parameter of a voxel be to declared it as activated. In order to be able to identify activated voxels, a statistical analysis (described in Appendix) based on qualitative and quantitative tests over spatial parameters of each set was conducted. The results of this analysis suggest the Laplacian distribution as a plausible distribution model for spatial parameters.

The assumption of a Laplacian distribution model for explaining the set of spatial parameters has been proposed [19] in the context of fMRI data analysis based on sparse representation under the framework of predictive modelling. Unlike our approach, this is a multivariate analysis where the prediction task is formulated as a regression problem for which the voxel time series are the predictive variables and the time series of a fixed stimulus task is the response variable. In this context, it seems acceptable to assume a Laplacian model considering that it is common to find in the literature the Laplacian model associated with sparse linear models leading to an ℓ_1 -regularization term under the maximum a posteriori (MAP) principle [4, 25, 26].

A very preliminary version of this manuscript was presented at the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society which focused mainly on exploiting the sparsity of the BOLD signal in the time domain whereas in the spatial domain the activation maps were generated by selecting the 300 most significant spatial parameters [27].

This paper is organized as follows. In Section 2, we briefly describe the GLM framework under the assumption that the parameter vector is known to be sparse. The proposed method for detecting active voxels by exploiting sparsity is then presented in Section 3. To assess the performance of the proposed methodology, we present in Section 4 numerical experiments with synthetic and real fMRI datasets. Finally, some concluding remarks are drawn in Section 6 where we give suggestions for future research avenues.

2. General Linear Model With a Sparse Parameter Vector

The GLM for the observed response variable y_j at j th voxel, $j = 1, \dots, N$, is given by

$$y_j = \mathbf{X}\boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j, \quad (1)$$

where $\mathbf{y}_j \in \mathbb{R}^M$, with M the number of experimental scans, $\mathbf{X} \in \mathbb{R}^{M \times L}$ denotes the design matrix which will be described shortly, $\boldsymbol{\beta}_j \in \mathbb{R}^L$ represents the signal strength at the j th voxel with L the number of basic parametric functions, and $\boldsymbol{\varepsilon}_j \in \mathbb{R}^M$ is the noise vector at the j th voxel. This model aims to represent the time series related to a specific voxel as a linear combination of the column vector of the design matrix. Particularly, under the assumption that the parameter vector $\boldsymbol{\beta}$ is sparse and a few column vectors of \mathbf{X} are combined to form the observed data y_j . In the GLM representation, one is interested in finding the contribution of each column vector of \mathbf{X} to the signal formation and from there to decide whether or not the j th voxel has been activated by a particular stimulus. This leads naturally to an L -dimensional inverse problem whose solution we are interested in finding. The most widely used approach for solving the inverse problem is the OLS method which is optimum under the ML principle when the entries of the noise vector are assumed to be independent, identically distributed following a zero mean Gaussian distribution with an unknown variance σ^2 . However, if one knows that the underlying contamination no longer follows a Gaussian noise model and, instead, it is better characterized by a statistical model with heavier-than-Gaussian tails a more robust method must be used to estimate the regression parameters [4]. Furthermore, as mentioned above OLS does not exploit the fact that the parameter vector $\boldsymbol{\beta}$ is sparse.

In this work, the solution to the inverse problem (1) is addressed under the framework of an ℓ_0 -based constrained LAD method. That is, the solution to (1) is given by

$$\begin{aligned} \hat{\boldsymbol{\beta}}_j &= \min_{\boldsymbol{\beta}_j} \|\mathbf{y}_j - \mathbf{X}\boldsymbol{\beta}_j\|_1, \\ \text{subject to } &\|\boldsymbol{\beta}_j\|_0 \leq D, \end{aligned} \quad (2)$$

where $\|\cdot\|_1$ denotes the ℓ_1 -norm, $\|\boldsymbol{\beta}\|_0$ denotes the ℓ_0 pseudonorm that counts the number of nonzero entries in $\boldsymbol{\beta}_j$, and $D \ll L$ is the sparsity level of $\boldsymbol{\beta}_j$ which is unknown. The LAD based methods are optimum under the ML principle when the underlying contamination follows a Laplacian distributed model [23] and, therefore, more robust than OLS. Furthermore, the ℓ_0 -based constraint induces the desirable sparsity in the inverse problem solution. A detailed description of ℓ_0 -LAD method is given in next section.

2.1. The Design Matrix \mathbf{X} . An important component in the GLM is the design matrix \mathbf{X} which in the context of sparse signal representation of fMRI is the dictionary of parametric waveforms, called atoms, in which the time series y_j admits

a sparse representation. In general, the fMRI time series consists of the BOLD signal, noise, and confound components that arise due to hardware and physiological noise (subject's motion, respiration, and heartbeats) [1]. Therefore, it is common practice, to incorporate into \mathbf{X} as much information as possible, such that the model fits to the data as close as possible [3]. For this study, we propose to use a parametric dictionary composed as a union of two subdictionaries: \mathcal{D}_1 and \mathcal{D}_2 . That is, $\mathbf{X} = [\mathcal{D}_1 \mid \mathcal{D}_2]$, where the column vectors of \mathcal{D}_1 are the expected task-related BOLD responses of each stimulus involved in the fMRI experiment given by convolving the stimulus with the model of the hemodynamic response function (HRF). Here the stimulus is assumed to be equivalent to the experimental paradigm, while the HRF is modelled using a canonical HRF [1, 3]. \mathcal{D}_2 , on the other hand, contains a set of confounds modelled, as we will see later, by a low frequency sinusoidal signals.

3. Active Voxel Detection by Exploiting Sparsity

3.1. Parameter Estimation by Solving the ℓ_0 -Regularized LAD. Given the linear model,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (3)$$

where the index labelling the voxel has been removed for simplicity in notation, we want to find the explanatory variables that best suit the model under a certain error criterion. According to (2), we want to locate the column vectors of \mathbf{X} and their contributions such that the data-fitting term $\|\mathbf{X}\boldsymbol{\beta} - \mathbf{y}\|_1$ reaches a minimum subject to the constraint that $\boldsymbol{\beta}$ has just a few nonzero values. This inverse problem can be reformulated as an ℓ_0 -regularized LAD regression problem [28]. This is

$$\min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_1 + \tau \|\boldsymbol{\beta}\|_0, \quad (4)$$

where $\tau > 0$ is the regularization parameter that balances the conflicting objectives of minimizing the data-fitting term while yielding, at the same time, a sparse solution on $\boldsymbol{\beta}$ [4]. Solving this ℓ_0 -regularized LAD problem is NP-hard owing to the sparsity constraint imposed by the ℓ_0 -pseudonorm. In [23], an iterative algorithm has been proposed to solve this optimization problem using a coordinate descent strategy and under the framework of a continuation approach for selecting the regularization parameter τ . Following this approach, the solution to the ℓ_0 -LAD regression problem is achieved by reducing the L -dimensional inverse problem given by (4) to L one-dimensional problems by supposing that all entries of the sparse vector $\boldsymbol{\beta}$ are known but one of them. Therefore, in order to estimate the n th unknown entry of $\boldsymbol{\beta}$, which in turn is the contribution of the n th stimulus/confound in the signal formation, the entries β_j , $j = 1, \dots, L$, $j \neq n$, are treated as known variables taking values estimated in the previous iteration. According to this, the

l_0 -LAD problem reduces to the one-dimensional minimization problem:

$$\hat{\beta}_n = \arg \min_{\beta_n} \sum_{i=1}^M |r_{in} - x_{in}\beta_n| + \tau |\beta_n|_0 + b, \quad (5)$$

where $b = \tau \sum_{j=1, j \neq n}^L |\beta_j|_0$, with $|\beta_j|_0 = 1$ if $\beta_j \neq 0$, otherwise $|\beta_j|_0 = 0$, and r_{in} denotes the i th entry of the n th residual column vector defined as $\mathbf{r}_n = \mathbf{y} - \sum_{j=1, j \neq n}^L \beta_j \mathbf{x}_j$, where $\mathbf{x}_j \in \mathbb{R}^M$ denotes the j th explanatory variable in the design matrix \mathbf{X} or equivalently the j th atom of the dictionary. Note that \mathbf{r}_n is the residual term that remains after subtracting from the observed fMRI signal \mathbf{y} the contributions of all explanatory variables (stimuli and confounds) but the n th one. It was shown in [23] that the solution to the optimization problem (5) can be thought of as a two-stage process: parameter estimation and basis selection. More precisely, in a first stage, an estimation of β_n is found by solving the unregularized optimization problem (5) leading to the weighted median operator as the underlying operation for estimating β_n . That is,

$$\tilde{\beta}_n = \text{MEDIAN} \left(|x_{in}| \diamond \frac{r_{in}}{x_{in}} \right), \quad (6)$$

where $W_i \diamond v_i = \overbrace{v_i, v_i, \dots, v_i}^{W_i \text{ times}}$ is the data replication operation.

The ℓ_0 -regularization term in (5) leads to the second stage of the estimation process where a hard thresholding operator is applied on the estimated value given by (6). That is

$$\hat{\beta}_n = \begin{cases} \tilde{\beta}_n, & \|\mathbf{r}_n\|_1 - \|\mathbf{r}_n - \tilde{\beta}_n \mathbf{x}_n\|_1 > \tau \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

From (7), it can be seen that the n th entry of $\hat{\beta}$ is considered relevant and, hence, a nonzero element if $\tau < \|\mathbf{r}_n\|_1 - \|\mathbf{r}_n - \tilde{\beta}_n \mathbf{x}_n\|_1 \leq |\tilde{\beta}_n| \|\mathbf{x}_n\|_1$. This latter inequality shows that the regularization parameter τ controls whether $\tilde{\beta}_n$ is significant or not based on an estimate of its magnitude. In this work τ is treated as an adaptable parameter whose value changes as the iterative algorithm progresses. That is, $\tau^{(k+1)} = \alpha \tau^{(k)}$, $k = 1, \dots, K$, where $0 < \alpha < 1$, $\tau^{(1)} = \|\mathbf{X}^T \mathbf{y}\|_\infty$ and K is the number of iterations, for further details see [27].

3.2. Detecting Activation Zones by Exploiting a Spatial Sparsity-Induced Model. Once the regression parameters are determined following the approach described above, each voxel is represented by a L -dimensional vector β_j . We next exploit the sparseness of the BOLD signal in the spatial domain by assuming a sparsity-inducing model. Although there are many statistical models that induce sparsity, we use a Laplacian model for reasons that will be further explained below. Let z_j , $j = 1, \dots, N$, be a statistical map related to a particular stimulus. That is,

$$z_j = \mathbf{c}^T \hat{\beta}_j, \quad j = 1, \dots, N, \quad (8)$$

where the vector $\mathbf{c} = [c_1, c_2, \dots, c_L]^T$ is defined according to the stimulus task that we are interested in. For instance, if one is interested in evaluating the presence of the l th stimulus, all the components of \mathbf{c} are set to zero but the l th which is set to one; in this case, $z_j = \hat{\beta}_{lj}$, $j = 1, \dots, N$, leading to a statistical map related to each stimulus. Next, from the statistical map, it is possible to generate activation maps by exploiting the a priori knowledge about sparseness of the activation zone, in the sense that only a reduced number of voxels that conforms the cerebral volume are activated by a particular stimulus. Since the statistical map and the sequence $\{\hat{\beta}_{lj}\}_{j=1}^N$ are equivalent, it is expected that this sequence be sparse. That is, only a few parameters are significant; intuitively, those associated with the voxels localized in the regions are activated by the l th stimulus.

As was mentioned above, the reasons that drive us to use a Laplacian model as a sparsity-inducing model are based on the following considerations. (1) The a priori knowledge about the sparseness of the activated regions suggests that the probability that $z_j \rightarrow 0$ is very high. (2) A statistical analysis, detailed in Appendix, based on graphical tools (histogram, distribution fitting, and QQ-plots) and quantitative values (Akaike information criterion (AIC)) computed over the spatial parameters z_j associated with different stimuli belonging to the PBAIC-2007 fMRI database, has yielded statistical evidence that supports the Laplacian model assumed for \mathbf{z} . Furthermore, this analysis suggests the feasibility of adopting the Exponential Power Distribution (EPD), a more general approach that includes the Laplacian and Gaussian models as particular cases. (3) Results of applying the *normalp* package [29], an R package containing a collection of tools related to EPD, to both synthetic and real spatial data, corroborate the Laplacian distribution as the model more closely related to the empirical distribution of data.

Assuming that the probability distribution of the entries of \mathbf{z} is Laplacian, that is,

$$z_j \sim \mathcal{L}(\eta_l, \kappa_l), \quad j = 1, \dots, N \quad (9)$$

for a particular stimulus, where η_l and κ_l are the localization and scale parameters for the l th stimulus, respectively, it is possible to estimate a threshold θ such that the j th voxel is declared as activated by the l th stimulus if the corresponding statistic z_j defined in (8) exceeds the threshold value. Under this approach, the threshold θ is obtained from the Laplacian cumulative distribution function:

$$F(x) = \begin{cases} 1 - \frac{1}{2} e^{-(x-\eta)/\kappa}, & x \geq \eta; \\ \frac{1}{2} e^{(x-\eta)/\kappa}, & x < \eta, \end{cases} \quad (10)$$

as the solution of $F^{-1}(p)$, for a given probability of false detection p . That is,

$$\begin{aligned} \theta &= F^{-1}(p) \\ &= \eta - \kappa [\text{sgn}(p - 0.5) \ln(1 - 2|p - 0.5|)], \end{aligned} \quad (11)$$

where the parameters η and κ are estimated from samples z_j by using the maximum likelihood estimation (MLE) method [30], that is,

$$\begin{aligned}\hat{\eta} &= \text{median} \{z_1, \dots, z_N\}, \\ \hat{\kappa} &= \frac{1}{N} \sum_{j=1}^N |z_j - \hat{\eta}|.\end{aligned}\quad (12)$$

4. Experiments

Given the critical lack of ground truth, a major difficulty with fMRI research concerns the validation of any analysis of experimental *in vivo* data. For this reason, the proposed method is first evaluated using synthetic fMRI data designed for this purpose following a similar approach to that described in [31]; for further details, see Section 4.1. Secondly, the method is applied to the *in vivo* PBAIC-2007 fMRI database.

In all experiments the dictionary is designed according to the structure $\mathbf{X} = [\mathcal{D}_1 \mid \mathcal{D}_2]$ described in Section 2.1. Although \mathcal{D}_1 and \mathcal{D}_2 vary with each experiment, there is a set of estimated BOLD signals corresponding to 13 different types of stimulus patterns classified as required features in the PBAIC-2007 database that are common to all designed dictionaries. These patterns, identified in the PBAIC-2007 database as follows, Arousal, Dog, Faces, FruitsVegetables, Hits, Instructions, InteriorExterior, SearchFruit, SearchPeople, SearchWeapons, Valence, Velocity, and WeaponsTools, are assembled (in that order) into a dictionary denoted, hereafter, by \mathcal{D}_c .

4.1. Synthetic Data Generation. The synthetic data were generated by blending a simulated activation in nonactivated time series that were extracted from the PBAIC-2007 fMRI database. Thus, knowledge of the ground truth is assured while the noise is representative for real data. According to this, the synthetic time series \mathbf{y}_{syn} at any voxel is given by

$$\mathbf{y}_{\text{syn}} = b\mathbf{s} + \mathbf{y}_{\text{na}}, \quad (13)$$

where $b > 0$ is the activation strength (if $b = 0$ the voxel is a nonactivated one); \mathbf{s} , described below, is the simulated activation time series; and \mathbf{y}_{na} is a nonactivated time series. The activation time series \mathbf{s} , shown in Figure 1(b), was obtained by convolving the canonical hemodynamic response function h used in SPM [32] with a stimulus boxcar function x consisted of 3 randomly placed stimulations that last about 3 s; see Figure 1(a). That is,

$$\mathbf{s}(t) = (x * h)(t), \quad (14)$$

where

$$h(t) = \frac{t^{a_1-1} b_1^{a_1}}{\Gamma(a_1)} e^{-b_1 t} - c \frac{t^{a_2-1} b_2^{a_2}}{\Gamma(a_2)} e^{-b_2 t}. \quad (15)$$

Here, $\Gamma(z)$ denotes the gamma function defined by $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$. In (15), the values for the parameters are $a_1 = 2$, $a_2 = 16$, $b_1 = b_2 = 1$, and $c = 1/16$ [32].

Likewise, in order to reduce the number of voxels to be analyzed, instead of trying to model the cerebral cortex a synthetic volume was designed. This volume (depicted in Figure 2) consists of 4 slices of size 20×20 voxels, each one with activated regions of different shapes (white-colored regions). There are altogether 108 activated voxels that represent 6.75% of the total number of voxels in the volume.

The synthetic fMRI database was generated from 32000 nonactivated time series in the following steps. First, the 32000 nonactivated time series were divided into 20 sets of 1600 time series each. Then, each set of 1600 time series was inserted into the synthetic volume. Finally, for each one of the 108 selected voxels (white regions) the synthetic activation was added to their BOLD signals, while remaining voxels retain their original real BOLD signals. Using 4 different levels of activation strength ($b = 1, 2, 3, 4$) in (13) 4 groups (labelled as g_1, g_2, g_3, g_4) of synthetic data were generated. That is, synthetic activation time series within each group have the same activation strength. In total, 80 datasets were analyzed. Note that since the background time series are directly extracted from *in vivo* fMRI data the contrast between activation and background will not be exactly the same, even for time series within the same group. In order to have a quantitative measure of the difference between groups the signal-to-noise ratio (SNR) is determined for each group (see Table 1), which is defined as the average of the SNR_i , $i = 1, \dots, 20$. Here the i th SNR is calculated as in [31]: $\text{SNR} = \text{std}(bs)/\text{std}(\mathbf{y}_{\text{na}})$ with $\text{std}(\cdot)$ being the standard deviation.

Regarding the design of the dictionary, in this experiment $\mathcal{D}_1 = [\mathbf{s} \mid \mathcal{D}_c]$, that is, the concatenation of the stimulus of interest \mathbf{s} with the 13 stimuli is described above, while \mathcal{D}_2 is a parametric dictionary whose atoms are the DCT functions:

$$\phi_{\omega_k, 0}[m] = a_k \cos(\omega_k m), \quad k = 0, \dots, K-1, \quad (16)$$

where $a_0 = 1/\sqrt{M}$, $a_k = \sqrt{2/M}$, $k \geq 1$, $\omega_k = k\pi/M$, $m = (2n+1)/2$, $n = 0, \dots, M-1$; and $K = (2 \cdot M \cdot TR)/128$ denotes the number of DCT basis functions. Here, TR is the repetition time and it is set to 1.75 s, and M is the number of fMRI scans.

4.2. In Vivo fMRI Data. The PBAIC-2007 fMRI dataset was collected with a Siemens 3T Allegra scanner using an EpiBOLD sequence, with imaging parameters TR and TE being set to 1.75 s and 25 ms, respectively. Subjects were engaged in a Virtual Reality task during which they had to perform a number of tasks in a hypothetical neighborhood. The tasks included, among others, the acquisition of pictures of neighbors with piercing, the gathering of specific objects (e.g., fruits, weapons) and the avoidance of a growling dog (for a more detailed description of the experiment, see [33]). Three subjects' data were available in the competition (labelled: subject 1, subject 13, and subject 14). Each subject's data consisted of 3 runs with a time duration of approximately 20 minutes each (704 volumes in each run). Each run also included 24 time series of features describing the subjects experiences over 704 TRs each, which had been convolved

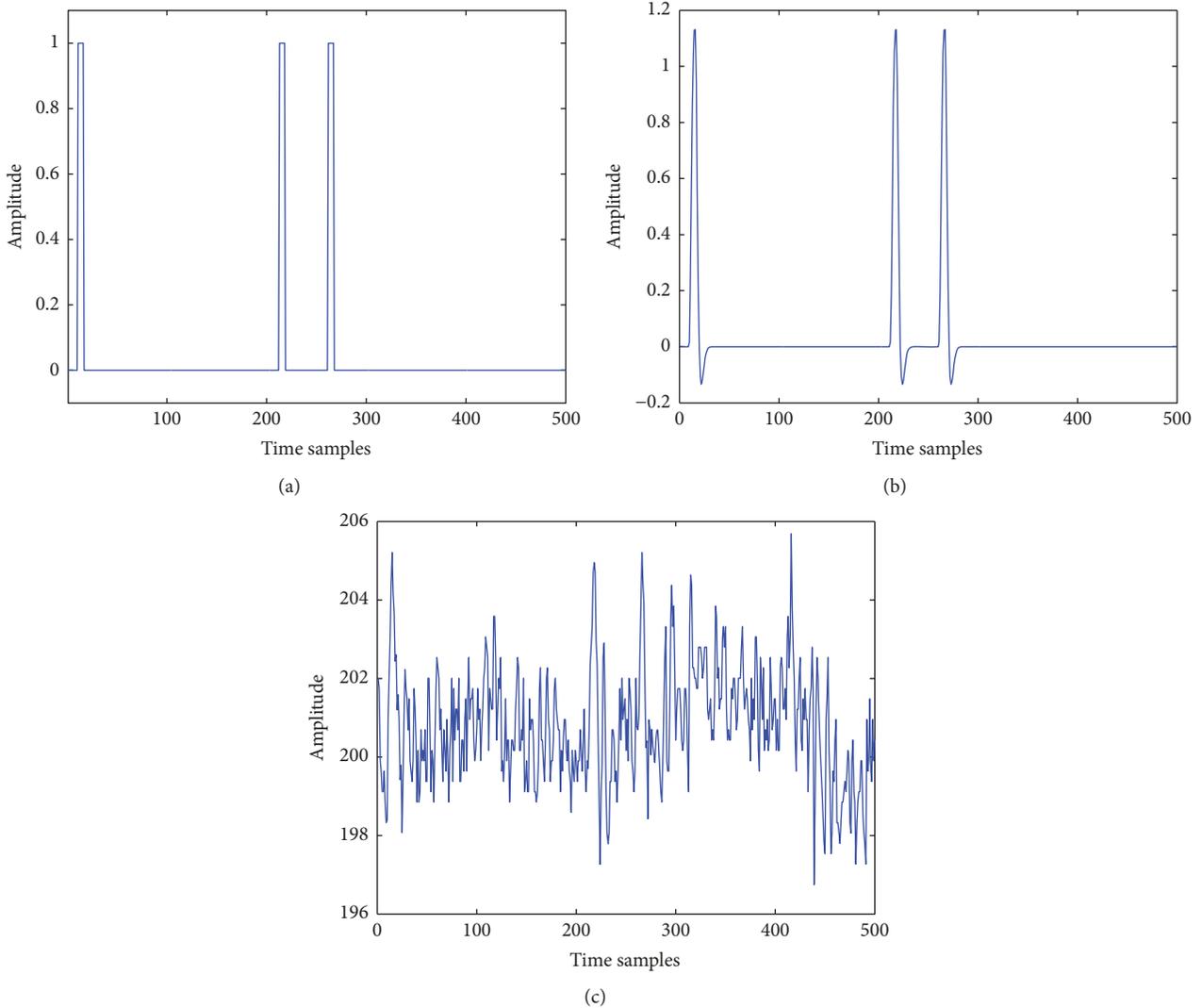


FIGURE 1: (a) Synthetic boxcar signal $x(t)$. (b) Synthetic activation $s(t) = (x * h)(t)$, where h denotes the canonical hemodynamic response function. (c) A synthetic time series y_{syn} generated by using (13) with $b = 4$.



FIGURE 2: A synthetic volume consists of 4 slices. White regions denote the set of voxels that has been activated by the stimulus s .

with the double gamma hemodynamic response filter (HRF) produced by the SPM software [32]. In this study only the 13 features (tasks) classified as required in [33] were considered. For the analysis, fixation periods (the time when the virtual world is off) were excluded from the original dataset leading to a total of 500 volumes to analyze in each run. Each volume contains $64 \times 64 \times 34$ voxels with a voxel size of $3.2 \times 3.2 \times 3.5 \text{mm}^3$. All experiments described in this section are applied to preprocessed data belonging to subjects 1, 13, and 14 in a voxel-by-voxel basis. The preprocessing tasks include time

and motion correction as well as detrending; further details are described in [24].

Given the acceptance of the SPM software in practice [2], in this work we use the outcome yielded by SPM as a benchmark for performing comparisons in order to know whether the proposed method yields reliable results or not. In order to ensure that the assumptions of random field theory used by SPM hold [3], an additional preprocessing step is performed on the fMRI data. To be more precise, a spatially filtering operation with a 3D Gaussian kernel is implemented

TABLE 1: Signal-to-noise ratio of the synthetic dataset.

Group	1	2	3	4
SNR	0.1419	0.2838	0.4256	0.5675
SNR (dB)	-16.9604	-10.9398	-7.4200	-4.9207

as in [19]. During spatial smoothing SPM generates a binary 3D array m , which mask out the nonbrain voxels. That is, $m(i, j, k) = 1$ if the voxel located at position (i, j, k) is within the brain, otherwise $m(i, j, k) = 0$. The analysis is performed on the set of voxels inside the brain according to the SPM mask. Based on the structure of the dictionary two experiments are considered.

4.2.1. Experiment 1. In this experiment the dictionary is designed following the framework used by SPM [32]. That is, $\mathbf{X} \in \mathbb{R}^{500 \times 14}$ is constructed by combining the 13 convolved stimulus functions assembled in \mathcal{D}_c and the all-one column vector $\mathbf{1}$ of size 500. Thus, the designed dictionary models the whole-brain activity [3]. Furthermore, to remove confounds SPM implements as part of the analysis a high-pass filtering based on the DCT basis set was described in (16) to both data and design matrix; therefore, in order to have comparable experimental conditions, this high-pass filter is also applied to \mathbf{y}_j and \mathbf{X} before estimating the regression model with the ℓ_0 -LAD method.

4.2.2. Experiment 2. In the second experiment, predictors that model confounds are incorporated to the dictionary. Specifically, the 13 DCT basis functions defined by (16) for high-pass filtering purposes are incorporated into \mathbf{X} leading to the dictionary $\mathbf{X} = [\mathcal{D}_c \mid \mathcal{D}_2] \in \mathbb{R}^{500 \times 26}$. In this case, nontemporal smoothing is applied to the data nor to the designed dictionary. The model parameters are then estimated by using the iterative ℓ_0 -LAD regression algorithm and the activation maps are obtained by thresholding the statistical map according to the approach described above.

5. Results and Discussions

5.1. Synthetic Data. In order to identify the voxels activated by the stimulus \mathbf{s} , the synthetic datasets described above are analyzed with the proposed method. All results in this experiment are obtained by choosing $p = 0.975$ in (11), so that the percentage of probability of false alarms is 2.5% (equivalent to 2.7 voxels).

Given the knowledge of the ground truth, the proposed method is evaluated by comparing the set of voxels detected as activated against the set of true activation. From this information, it is determined the number of true detections, false alarms, and missed activated voxels. Figure 3(a) shows an illustrative example of the activation maps obtained for set 20. Activated voxels in all groups are shown overlaid on the ground truth, colored in gray. From the information shown, it is clear that the proposed method detects properly the set of activated voxels by the stimulus \mathbf{s} when $b \geq 2$ (i.e., $\text{SNR} \geq 0.2838$), keeping the number of false alarms and

missed activated voxels below the established error margin. In fact, similar results are obtained for the rest of sets as can be noted from Figure 3(b) that shows, for each group, the number of false alarms (top) and missed detections (bottom) as a function of the set number. It is important to point out that the average of false alarms and missed detections for groups 2 to 4 is close or below the percentage of error; see Table 2. Regarding group 1, the number of false alarms and missed activated voxels overcomes the expected rate. This behavior, typical of all sets of group 1 as can be seen in Figure 3(b), is caused by the high level of noise present in the signal (Table 1), that is, a consequence of the low energy level of activation of this group. That is, the activation level $b = 1$ maintains the signal below the noise level making the activation detection more difficult.

5.2. In Vivo fMRI Data. To illustrate the performance of the proposed method in detecting activation we choose the data of subject 14, run 1 and two representative tasks: Instructions and Faces. The selection of these particular tasks is based on the a priori knowledge of the anatomical and functional localization of auditory and visual cortex. With regard to the parameters α and K used by the ℓ_0 -LAD method in the stages of estimation and basis selection, they are selected following the recommendations given in [23]. According to this, in both experiments, α and K are set to 0.95 and 100, respectively. On the other hand, with the purpose that the total number of voxels classified as activated voxels by both the SPM software and the proposed method could be as close as possible, the activation maps generated by using SPM are obtained by adjusting the Family-wise error (FWE) rate at 0.05 [34], while for the proposed method the values of p in (11) are set to 0.975 for Instructions task (in all experiments) and to 0.998 and 0.999 for experiments 1 and 2, respectively, in case of Faces task.

5.2.1. Results of the First Experiment. Figures 4 and 5 show the activation maps for the Instructions and Faces tasks, respectively, obtained with (a) the proposed method with reduced dictionary (14 atoms) and (b) the SPM software.

Under these experimental conditions, for the Instructions task, the proposed method activates 676 voxels contained in 26 slices (2 to 26 and 28) and SPM activates 708 voxels contained in 22 slices (2 to 23). Although the activated slices are not exactly the same, the number of matching activated slices is high, to be more precise 22 (slices 2 to 23) for a matching percentage of 84.61% at slices level. At voxels level, however, the percentage of matching is 57.99%. It can be seen in Figure 4 that the activated regions yielded by both algorithms are similar in each one of the matching activated slices, although the proposed method tends to exhibit more

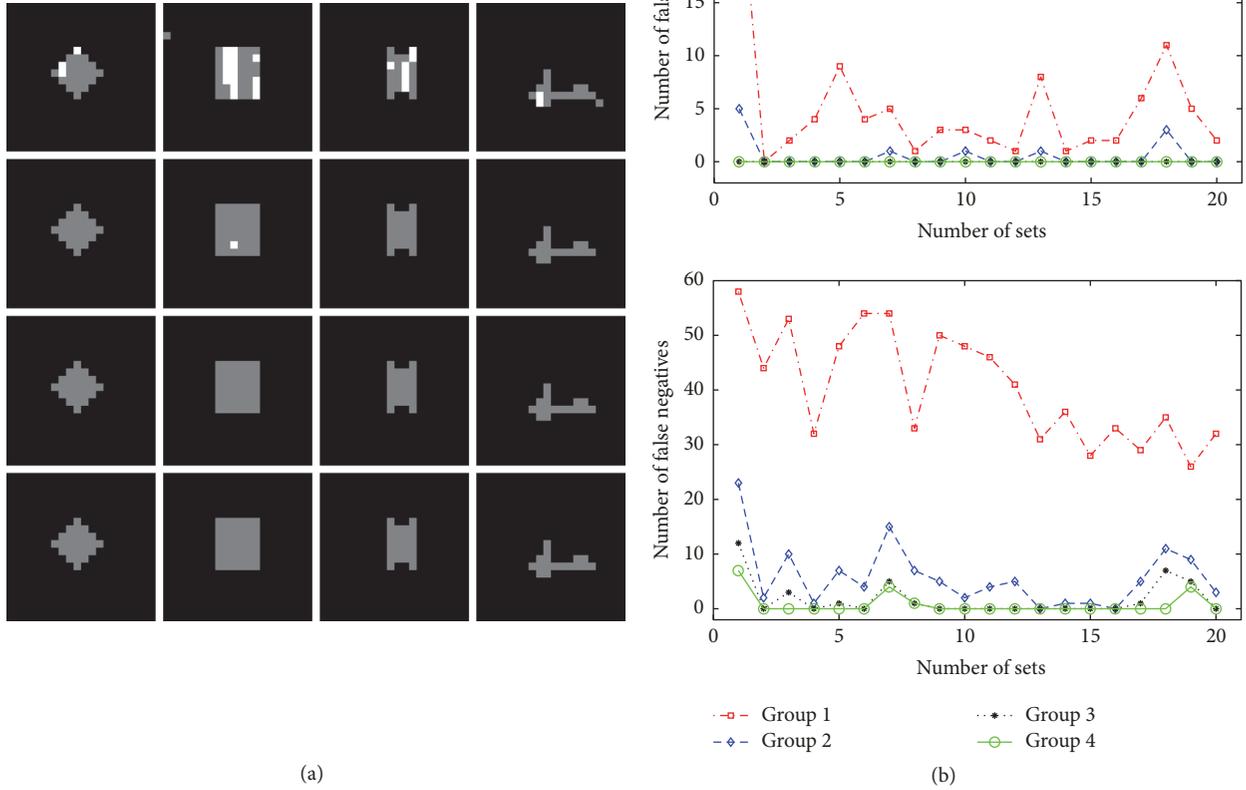


FIGURE 3: (a) Activation maps (left to right: slice 1 to slice 4; top to bottom: group 1 to group 4) for stimulus s , set 20. (b) Number of false positives (top) and false negatives (bottom) for sets 1 to 20 and groups 1 to 4.

TABLE 2: Average of activated voxels, true detections, false alarms, and missed activations for synthetic data base.

Group	Activated voxels	True detections	False alarms	Missed activated voxels
g_1	72.40	67.45	4.95	40.55
g_2	102.80	102.25	0.55	5.75
g_3	106.25	106.25	0	1.75
g_4	107.20	107.20	0	0.8

isolated voxels than SPM which seems to promote clustering. Despite these differences, it is clear that the activated regions yielded by the proposed method are consistent with those achieved by SPM. Furthermore, as expected for this kind of stimulus, the activated areas detected by both methods appear to be in the auditory cortex. For the Faces task (Figure 5), SPM activates 91 voxels belonging to 11 slices (9 to 19), whereas the proposed method activates 95 voxels contained in 17 slices (4, 6, 9–19, 21, 22, 24, and 28), so the percentage of coincidence at slice level is 64.71%. Although only 50 of the voxels classified as activated by both methods are localized at the same position, both activated regions and activation patterns are quite similar. This is because,

except for a very small number of voxels, the locations of the mismatched voxels classified as activated by the proposed method are located at positions sufficiently close to voxels activated by SPM. More importantly, areas detected as activated by both methods appear to be on the visual cortex as expected.

5.2.2. *Results of the Second Experiment.* Figures 6 and 7 show activation maps for the Instructions and Faces tasks, respectively. Under the Instructions experimental condition, the proposed approach activates 717 voxels contained in 22 slices (2, 3, 4, 7 to 23, 26, 27) whereas the SPM software activates the slices 2 to 23, coinciding with 17 slices activated

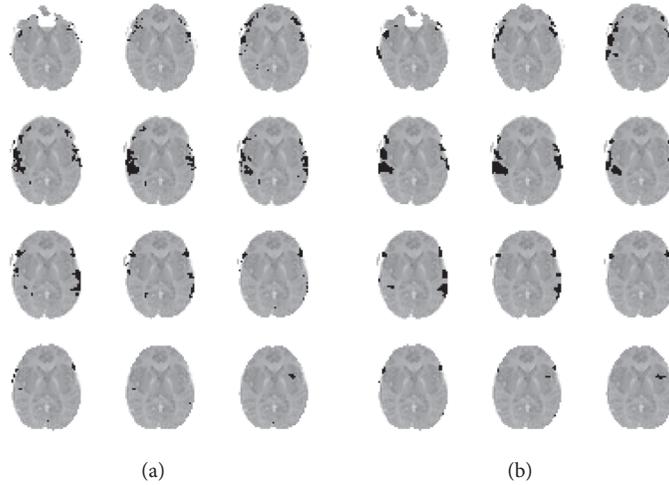


FIGURE 4: Activation maps (left to right, top to bottom: slices 11 to 22) for the Instructions task obtained with (a) proposed method with reduced dictionary (14 atoms) and (b) SPM software with FWE rate.

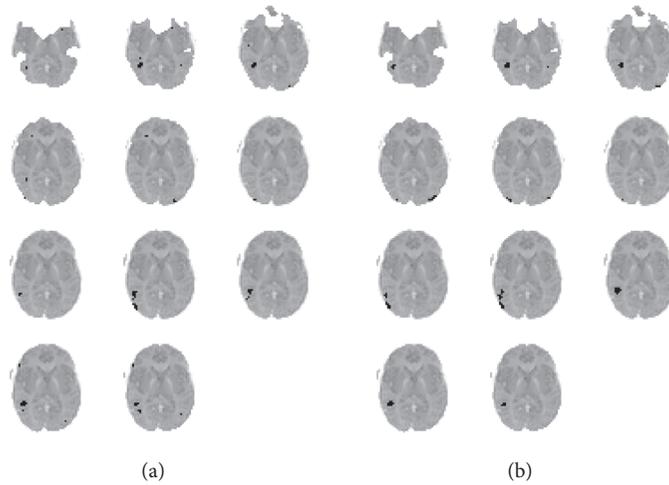


FIGURE 5: Activation maps (left to right, top to bottom: slices 9 to 19) for the Faces task obtained with (a) proposed method with reduced dictionary (14 atoms) and (b) SPM software with FWE rate.

simultaneously by both methods for a matching percentage at slice level of 77.27%. In this case, there are 453 activated voxels whose spatial localization agree with those of SPM maps for a matching percentage at voxel level of 63.18%. For the Faces task, the proposed method activates 120 voxels contained in 17 slices (8 to 24), while SPM activates 91 voxels contained in 11 slices (9 to 19). Therefore, the percentage of matching slices for this condition is 64.71%, whereas the number of voxels activated by both methods at the same spatial localization is 57 for a matching percentage at voxel level of 47.5%. An interesting issue to evaluate in this experiment is the effect of incorporating the DCT basis set to the dictionary instead of performing a previous temporal filtering of the data. Although the dictionary and the design matrix are no longer the same, from Figures 6 and 7, it is clear that the activated regions and the activation patterns look quite similar for

both methods and, as expected, seem to be localized at the auditory and visual cortex, respectively. More interesting, using the extended dictionary produces an effect of “filling” in the activation maps that is caused by an increment in the number of activated voxels in relation to the amount of activated voxels using the reduced dictionary. For example, in case of the Instructions task (Table 3) 41 additional voxels are declared as activated while the number of activated slices decreases to 22. This new group of activated voxels tends to fill some empty spaces near to the activated regions in the activation maps obtained with the reduced dictionary. A similar behavior is exhibited in case of the Faces task. This effect is highlighted in Figures 8 and 9 by comparing the activation maps obtained with the proposed method by using the reduced dictionary (a) and the extended dictionary (b) for the Instructions and Faces tasks, respectively.

TABLE 3: Results of experiments 1 and 2 indicating number of activated voxels (av), number of activated slices (as), and the matching percentages at voxel level (mpvl) and at slice level (mpsl) with respect to SPM software.

Stimulus	SPM		Reduced dictionary				Extended dictionary			
	av	as	av	as	mpvl	mpsl	av	as	mpvl	mpsl
Instructions	708	22	676	26	57.99%	84.62%	717	22	63.18%	77.27%
Faces	91	11	95	17	54.96%	64.71%	120	17	47.50%	64.71%

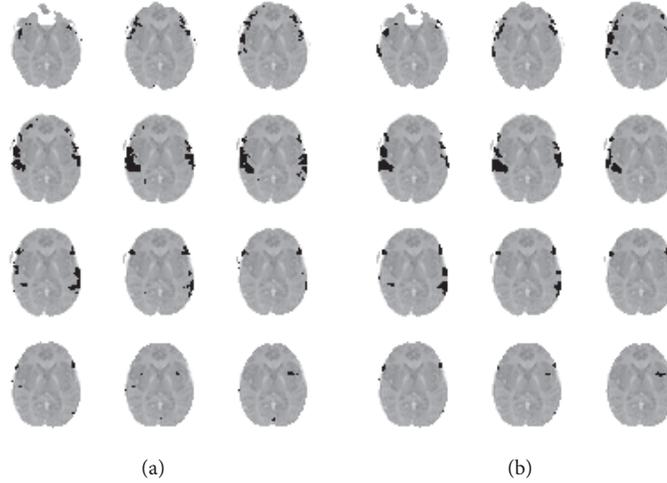


FIGURE 6: Activation maps (left to right, top to bottom: slices 11 to 22) for the Instructions task generated by (a) the proposed method with extended dictionary (26 atoms), and (b) SPM software with FWE rate.

6. Conclusions and Future Works

In this paper, a new method for detecting activation in fMRI data that exploits the sparseness of the BOLD signal in both time and spatial domains is presented. By adopting a sparse approach in both domains two objectives are achieved. First, in time domain, the ℓ_0 pseudonorm promotes the concentration of energy (related to stimuli) of a fMRI signal in a reduced set of model coefficients. Thus, during the parameter estimation a preselection of candidate voxels to be activated is performed by identifying those whose relationship with a selected stimulus is significant in the sense that if a voxel is activated by a particular stimulus its regression coefficient is significantly high. Second, in spatial domain, by adopting a Laplacian model it is possible to determine the threshold parameter which allows deciding how “significant” is the magnitude of the spatial parameter to consider the voxel as activated. The proposed approach is validated in two ways, first, using synthetic data where the ground truth is known in advance. Secondly, the method is evaluated through the comparison of activation maps with those generated by SPM software using real fMRI data. For synthetic data, results demonstrate that our approach is able to identify most activated voxels without any false activation. In case of real data, the activated regions detected by our approach are similar to those yielded by SPM. The results show the efficacy of our method and suggest that incorporating prior knowledge about spatial sparsity into spatial classification can reduce significantly the burdensome posterior spatial analysis

needed when sparsity is ignored. Although the proposed method does not promote clustering, the activated regions exhibit mostly connected patterns. However, given that fMRI data are spatially correlated an approach that jointly exploits sparsity and could induce clustering is under consideration as a future work.

Appendix

Statistical Analysis of Spatial Parameters

As mentioned in Section 3.2, selecting a Laplacian model for the estimated spatial parameters is supported by a set of statistical tests on the parameters outputted by the ℓ_0 -LAD based regression approach which, in turn, are associated with 13 stimuli belonging to the PBAIC 2007 database. The statistical tests consisted of a visual analysis through histograms, fitting distributions and QQ-plots. Furthermore, a quantitative analysis based on the AIC is performed for each stimulus.

Data analysis through histograms strongly suggests as a plausible model the Laplacian distribution mainly due to the concentration of mass centered around zero and the heaviness of the tails of the empirical distribution. Nevertheless, two probability distribution models are fitted to data: specifically, the Gaussian model which has been the standard model assumed for spatial analysis of fMRI data and the Laplacian model. For both models, maximum likelihood

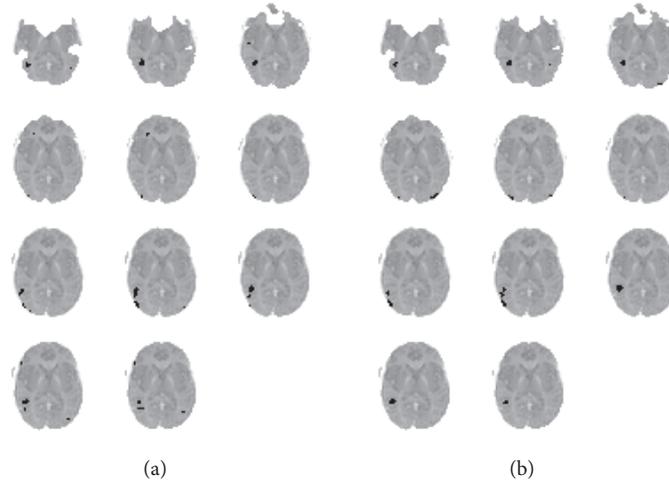


FIGURE 7: Activation maps (left to right, top to bottom: slices 9 to 19) for the Faces task generated by (a) the proposed method with extended dictionary and (b) SPM software with FWE rate.

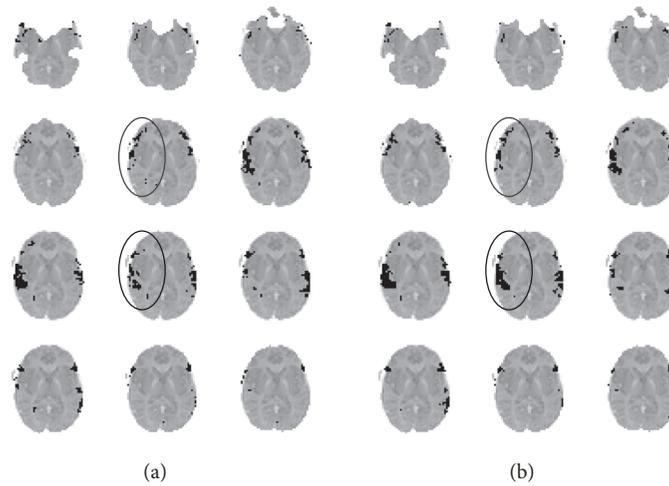


FIGURE 8: Activation maps for the Instructions task generated with the proposed method by using the reduced dictionary (a) and extended dictionary (b). The activated regions contained in the ellipses show the effect of “filling” observed when the parameter estimation is performed using the extended dictionary.

(ML) estimators are used to compute the respective distribution’s parameters (i.e., the location and dispersion parameters) using the spatial parameters $\mathbf{z}^l = [\hat{\beta}_{l1}, \hat{\beta}_{l2}, \dots, \hat{\beta}_{lN}]$ associated with the l th stimulus, $l = 1, \dots, 13$. Figure 10 shows the logarithm of the probability distribution along with the Laplacian and the Gaussian distribution for the stimuli: Dog, Faces, Instructions, and WeaponsTools. Furthermore, we focus our attention in examining the heaviness of distribution tails since the selection of threshold parameter θ depends mainly on the behavior of those tails for each selected distribution model. As can be seen in Figure 10, in general, the estimated distribution exhibits heavier-than-Gaussian tails. Furthermore, comparing the tails of the empirical distribution with respect to those of the Laplacian distribution, it seems to have good match for all stimuli.

These findings are consistent with those obtained through QQ-plots. In particular, in case of the 4 stimuli considered above, all QQ-plots of the Gaussian distribution versus the estimated distribution, see Figure 11, show that on the right side the plot is above the straight line and on the left side it is below of the straight line. Departures from this straight line indicate departures from normality. Furthermore, the shape of the normal plot indicates that the estimated distribution has heavier tails than the Gaussian model.

Since graphical tools lack rigor, a quantitative measure of the distance between the estimated distribution and the proposed models based on the AIC is used. The AIC is defined by $AIC_m = -2 \ln(\mathfrak{L}_m(\hat{\pi} \mid z_1, z_2, \dots, z_N)) + 2K$, where K is the number of parameters in the model, \mathfrak{L}_m is the likelihood function of the model with $m \in \{\mathcal{L}, \mathcal{N}\}$, and $\hat{\pi}$

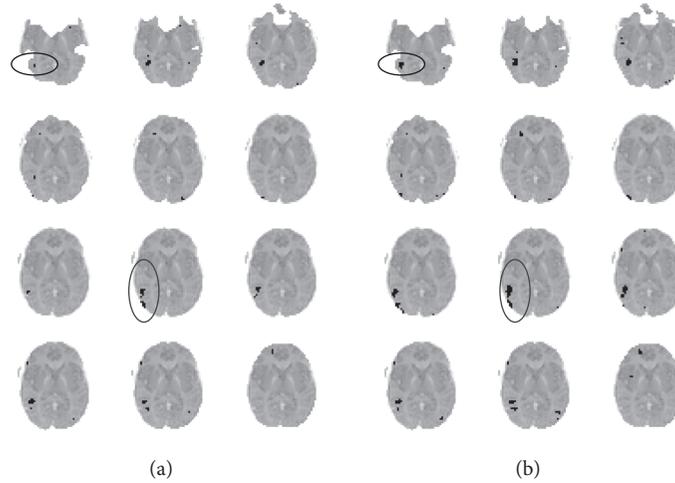


FIGURE 9: Activation maps for the Faces task generated with the proposed method by using the reduced dictionary (a) and extended dictionary (b). The activated regions contained in the ellipses show the effect of “filling” observed when the parameter estimation is performed by using the extended dictionary.

is the ML estimated of the parameters corresponding to the model [35]. According to this, the Laplacian model better fits the data if $AIC_{\mathcal{L}} - AIC_{\mathcal{N}} < 0$. From Figure 12, it is clear that the Laplacian model fits the data for all stimuli except 7 and 12, for which, according to the AIC, the nearest model is the Gaussian one.

Previous analysis shows that the tails of the estimated distribution are generally heavier than those of a Gaussian distribution, but not as heavy as those of a Laplacian distribution. This finding suggests the feasibility of adopting a more general statistical model: the Generalized Gaussian (GG) distribution, which contains the Gaussian and Laplacian distributions as special cases. The GG density has the following form [29]:

$$\mathcal{GG}(x; \mu, \sigma, \rho) = \frac{1}{2\sigma\rho^{1/\rho}\Gamma(1 + 1/\rho)} e^{-(|x-\mu|^\rho/\rho\sigma^\rho)}, \quad (\text{A.1})$$

where $\Gamma(\cdot)$ defines the gamma function; μ , σ , and $\rho > 0$ denote, respectively, the location, scale, and shape parameters. This last parameter determines the distribution shape and it is linked to the thickness of the tails [29]. An interesting aspect of the GG model is that the shape parameter ρ can be determined from the data; this fact allows us to estimate the value of ρ that gives the highest probability of producing the observed data. In order to estimate the value of ρ the *normalp* package [29] is used to analyze the estimated spatial parameters of both synthetic and *in vivo* data. For both databases the parameters σ and ρ are determined by using the function *estimatep* assuming the median as the location parameter. After that, the threshold θ is estimated by using the function *qnormp* with probability $pr = 0.025$ as in our experiments.

In case of synthetic database, the shape parameter estimated by the function *estimatep* is $\rho = 1$ for each spatial parameters set associated with the synthetic stimulus s . According to these results the distribution model more

suitable for each set of synthetic spatial parameters is the Laplacian model. These results are consistent with those inferred from the histograms, estimated distribution, and the AIC for synthetic database. Certainly, by using $\rho = 1$ on the *qnormp* function the value estimated of θ coincides with those calculated by using the closed form expression (11). With regard to *in vivo* data, the results of analyzing the estimated spatial parameters of the 13 stimuli described above by using the *normalp* package are presented in Table 4. As can be seen in this table, in most cases, the shape parameter is equal to one, confirming, once again, the appropriateness of the Laplacian model. Note that for stimuli 6 and 12, the shape parameters are, respectively, 1.3637 and 1.7865, which by the closeness of the distribution it may be thought of as Laplacian and Gaussian, respectively. These findings are consistent with the previous analysis, even though the stimulus 7 is classified as Gaussian by the AIC criterion. This seemingly discrepancy in the classification of stimulus 7 can be explained from the estimated values of $AIC_{\mathcal{L}}$ and $AIC_{\mathcal{G}}$, which are close together.

Disclosure

A partial version of this manuscript was presented as a plenary at the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society in Boston, in 2011 [27].

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors would like to acknowledge ECOS-NORD-FONACIT under Grant PI-20100000299, the Council of Scientific, Humanistic, Technological and Artistic Development

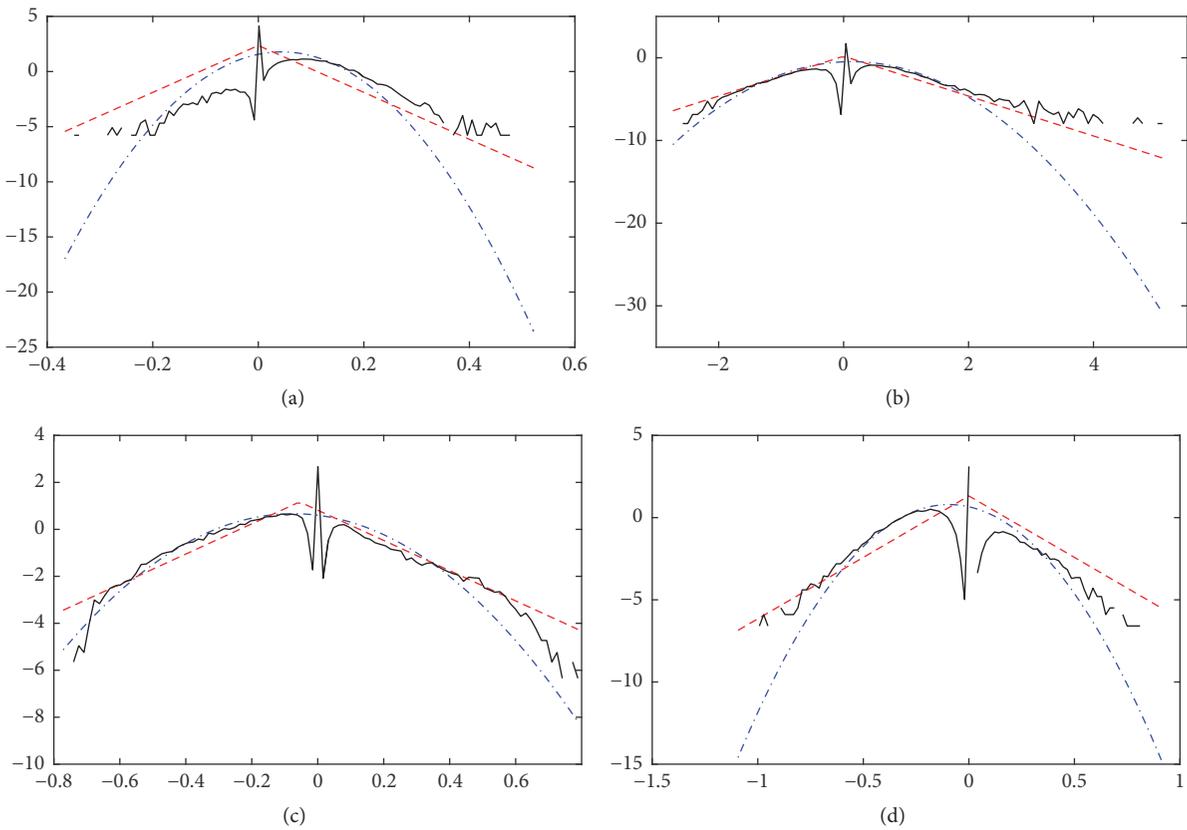


FIGURE 10: Statistical behavior of spatial parameters set $\{\hat{\beta}_{ij}, j = 1, \dots, N\}$ for the stimuli: (a) Dog, (b) Faces, (c) Instructions y , and (d) WeaponsTools. Solid line is the probability density estimated from the data, dashed line corresponds to the Laplacian density, and the dashed-dotted line corresponds to Gaussian density.

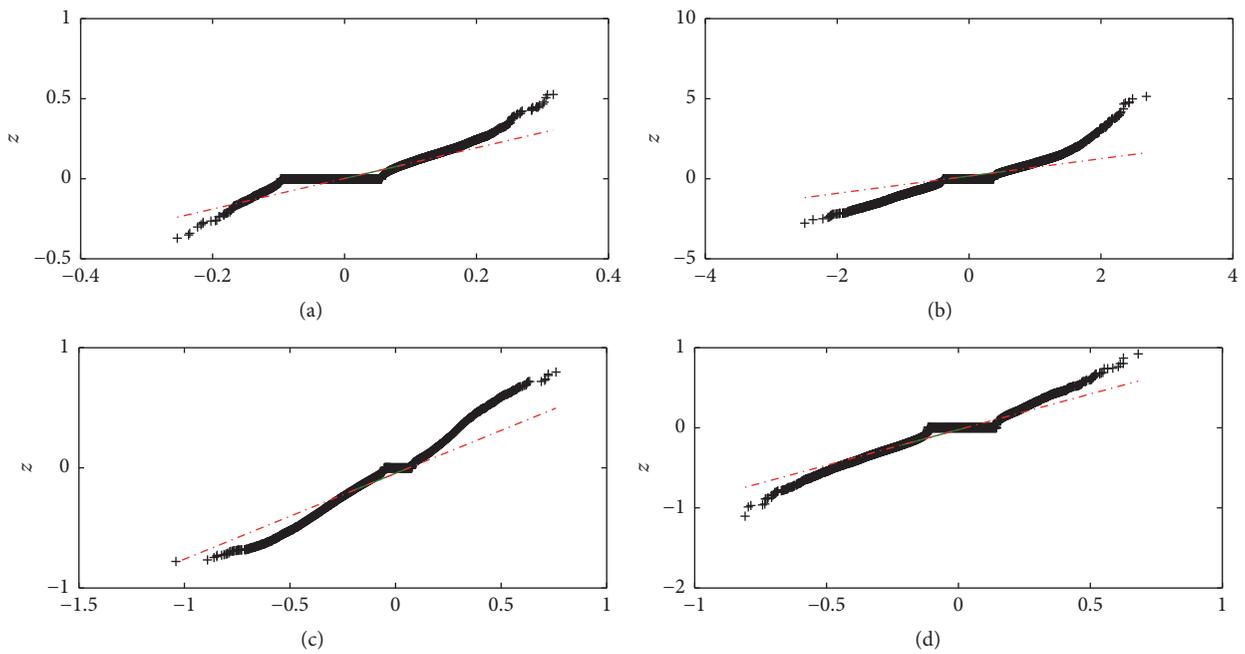


FIGURE 11: QQ-plots for Gaussian and estimated distributions of stimuli: (a) Dog, (b) Faces, (c) Instructions, and (d) WeaponsTools.

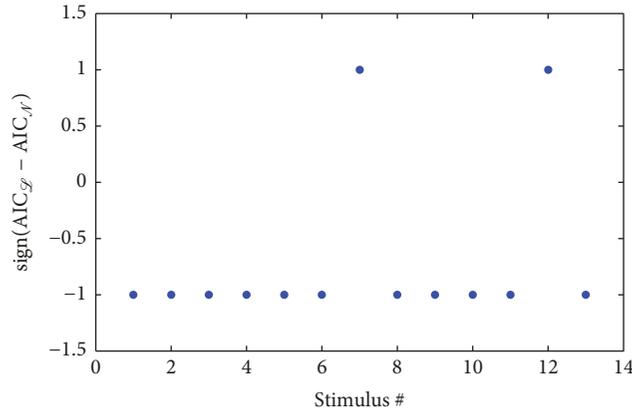


FIGURE 12: Sign of the difference $AIC_{\mathcal{G}} - AIC_{\mathcal{N}}$ for 13 stimuli of PBAIC 2007 data.

TABLE 4: Results of analyzing in vivo data with the normalp package indicating the shape parameter ρ and the threshold θ estimated with (i) the normalp package and (ii) the closed form expression (11).

Stimuli	ρ_{normalp}	θ_{normalp}	θ_{theorexp}
Arousal	1	0.00979738	0.00979711
Dog	1	0.1483275	0.14082884
Faces	1	1.24216047	1.24212600
FruitsVegetables	1	0.1328233	0.13281962
Hits	1	0.39250583	0.39249493
Instructions	1.3637	0.35744665	0.41064940
InteriorExterior	1	0.13646107	0.13645728
SearchFruit	1	0.04498026	0.04497901
SearchPeople	1	0.01944126	0.01944072
SearchWeapons	1	0.03387324	0.03387230
Valence	1	0.04011155	0.04011044
Velocity	1.7865	0.1221627	0.18050036
WeaponsTools	1	0.40075143	0.40074030

of the Universidad de los Andes (CDCHTA-ULA) under Project I-1336-12-02-B, and the Universidad Nacional Experimental del Táchira (UNET) for their financial support.

References

- [1] M. A. Lindquist, "The Statistical Analysis of fMRI Data," *Statistical Science*, vol. 23, no. 4, pp. 439–464, 2008.
- [2] G. E. Sarty, "Computing brain activity maps from fMRI time-series images," *Computing Brain Activity Maps from fMRI Time-Series Images*, pp. 1–190, 2006.
- [3] W. D. Penny and K. J. Friston, "Advanced image processing in magnetic resonance imaging," in *Classical and Bayesian Inference in fMRI*, vol. 27, pp. 541–563, Taylor and Francis Group, 2005.
- [4] E. G. Larsson and Y. Seln, "Linear regression with a sparse parameter vector," *IEEE Transactions on Signal Processing*, vol. 55, no. 2, pp. 451–460, 2007.
- [5] S. J. Hanson and B. M. Bly, "The distribution of BOLD susceptibility effects in the brain is non-Gaussian," *NeuroReport*, vol. 12, no. 9, pp. 1971–1977, 2001.
- [6] C.-C. Chen, C. W. Tyler, and H. A. Baseler, "Statistical properties of BOLD magnetic resonance activity in the human brain," *NeuroImage*, vol. 20, no. 2, pp. 1096–1109, 2003.
- [7] W.-L. Luo and T. E. Nichols, "Diagnosis and exploration of massively univariate neuroimaging models," *NeuroImage*, vol. 19, no. 3, pp. 1014–1032, 2003.
- [8] I. Daubechies, E. Roussos, S. Takerkart et al., "Independent component analysis for brain fMRI does not select for independence," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 26, pp. 10415–10422, 2009.
- [9] B. A. Olshausen and D. J. Field, "Sparse coding of sensory inputs," *Current Opinion in Neurobiology*, vol. 14, no. 4, pp. 481–487, 2004.
- [10] J. Wolfe, A. R. Houweling, and M. Brecht, "Sparse and powerful cortical spikes," *Current Opinion in Neurobiology*, vol. 20, no. 3, pp. 306–312, 2010.
- [11] R. Quian Quiroga and G. Kreiman, "Measuring Sparseness in the Brain: Comment on Bowers (2009)," *Psychological Review*, vol. 117, no. 1, pp. 291–297, 2010.
- [12] J. M. Fadili and E. Bullmore, "Penalized partially linear models using sparse representations with an application to fmri time series," *IEEE Transactions on Signal Processing*, vol. 53, no. 9, pp. 3436–3448, 2005.
- [13] I. Khalidov, J. Fadili, F. Lazeyras, D. Van De Ville, and M. Unser, "Activelets: wavelets for sparse representation of hemodynamic responses," *Signal Processing*, vol. 91, no. 12, pp. 2810–2821, 2011.

- [14] K. Lee, S. Tak, and J. C. Ye, "A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1076–1089, 2011.
- [15] V. Abolghasemi, S. Ferdowsi, and S. Sanei, "Fast and incoherent dictionary learning algorithms with application to fMRI," *Signal, Image and Video Processing*, vol. 9, no. 1, pp. 147–158, 2015.
- [16] J. Lv, X. Jiang, X. Li et al., "Sparse representation of whole-brain fMRI signals for identification of functional networks," *Medical Image Analysis*, vol. 20, no. 1, pp. 112–134, 2015.
- [17] M. K. Carroll, G. A. Cecchi, I. Rish, R. Garg, and A. R. Rao, "Prediction and interpretation of distributed neural activity with sparse models," *NeuroImage*, vol. 44, no. 1, pp. 112–122, 2009.
- [18] B. Ng, A. Vahdat, G. Hamarneh, and R. Abugharbieh, "Generalized sparse classifiers for decoding cognitive states in fMRI," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 6357, pp. 108–115, 2010.
- [19] Y. Li, Z. Yu, Z. Gu, P. Namburi, C. Guan, and J. Feng, "Voxel Selection in fMRI Data Analysis Based on Sparse Representation," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 10, pp. 2439–2451, 2009.
- [20] Y. Li, J. Long, L. He, H. Lu, Z. Gu, and P. Sun, "A Sparse Representation-Based Algorithm for Pattern Localization in Brain Imaging Data Analysis," *PLoS ONE*, vol. 7, no. 12, Article ID e50332, 2012.
- [21] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale ℓ_1 -regularized least squares," *IEEE J. Sel. Top. Signal Process*, vol. 1, no. 4, pp. 606–617, 2007.
- [22] A. J. den Dekker and J. Sijbers, "Implications of the Rician distribution for fMRI generalized likelihood ratio tests," *Magnetic Resonance Imaging*, vol. 23, no. 9, pp. 953–959, 2005.
- [23] J. L. Paredes and G. R. Arce, "Compressive sensing signal reconstruction by weighted median regression estimates," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2585–2601, 2011.
- [24] Pittsburg-EBC-Group, "PBAIC Homepage," Accessed: March 2013.
- [25] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 58, no. 1, pp. 267–288, 1996.
- [26] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.
- [27] B. Guillen, J. L. Paredes, and R. Medina, "A sparse based approach for detecting activations in fMRI," in *Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS 2011*, pp. 7816–7819, USA, September 2011.
- [28] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, 2010.
- [29] A. Mineo and M. Ruggieri, "A software tool for the exponential power distribution: The normalp package," *Journal of Statistical Software*, vol. 12, no. 4, pp. 1–24, 2005.
- [30] P. Puig and M. A. Stephens, "Tests of fit for the laplace distribution, with applications," *Technometrics*, vol. 42, no. 4, pp. 417–424, 2000.
- [31] F. G. Meyer and X. Shen, "Classification of fMRI time series in a low-dimensional subspace with a spatial prior," *IEEE Transactions on Medical Imaging*, vol. 27, no. 1, pp. 87–98, 2008.
- [32] W. T. C. for Neuroimaging, "Statistical Parametric Map. <http://www.fil.ion.ucl.ac.uk/spm/>, Accessed: March 2015.
- [33] Pittsburg-EBC-Group, "2007 Guide Book," Accessed: March 2013.
- [34] V. Litvak, J. Mattout, S. Kiebel et al., "EEG and MEG data analysis in SPM8," *Computational Intelligence and Neuroscience*, vol. 2011, Article ID 852961, 2011.
- [35] H. Bozdogan, "Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions," *Psychometrika*, vol. 52, no. 3, pp. 345–370, 1987.



Hindawi

Submit your manuscripts at
www.hindawi.com

