NeuroImage 173 (2018) 580–591

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

A low-rank multivariate general linear model for multi-subject fMRI data and a non-convex optimization algorithm for brain response comparison

Tingting Zhang [a,*,1], Minh Pham [a,1], Jianhui Sun [a], Guofen Yan [b], Huazhang Li [a], Yinge Sun [a], Marlen Z. Gonzalez [c], James A. Coan [c,**]

[a] Department of Statistics, University of Virginia, Charlottesville, VA 22904, USA
[b] Division of Biostatistics, University of Virginia, Charlottesville, VA 22904, USA
[c] Department of Psychology, University of Virginia, Charlottesville, VA 22904, USA

ARTICLE INFO

Keywords:
fMRI
Hemodynamic response function
Matrix factorization
Optimization
Regularization
Spline

ABSTRACT

The focus of this paper is on evaluating brain responses to different stimuli and identifying brain regions with different responses using multi-subject, stimulus-evoked functional magnetic resonance imaging (fMRI) data. To jointly model many brain voxels' responses to designed stimuli, we present a new low-rank multivariate general linear model (LRMGLM) for stimulus-evoked fMRI data. The new model not only is flexible to characterize variation in hemodynamic response functions (HRFs) across different regions and stimulus types, but also enables information "borrowing" across voxels and uses much fewer parameters than typical nonparametric models for HRFs. To estimate the proposed LRMGLM, we introduce a new penalized optimization function, which leads to temporally and spatially smooth HRF estimates. We develop an efficient optimization algorithm to minimize the optimization function and identify the voxels with different responses to stimuli. We show that the proposed method can outperform several existing voxel-wise methods by achieving both high sensitivity and specificity. We apply the proposed method to the fMRI data collected in an emotion study, and identify anterior dACC to have different responses to a designed threat and control stimuli.

Introduction

Functional magnetic resonance imaging (fMRI) is widely used in psychology experiments to measure brain responses to a designed sequence of stimuli, because it provides noninvasive measurements of human brain activity with a high spatial resolution of the entire brain. Typical fMRI data of each subject from one experimental session consist of hundreds of thousands of spatially indexed time series of identical length. Each time series, usually of several hundred time points with unit time (repetition time) ranging from 0.5 s to 2 s, indicates brain activity changes over time at one brain location corresponding to a small 3D cubic volume, called voxel, in the brain. In typical multi-subject fMRI experiments, dozens or hundreds of subjects' functional brain responses to designed stimulus sequences are measured. Since human brain activity varies by time, regions, subjects, and inputs to the brain, such multi-subject, stimulus-evoked fMRI data not only are massive in size, but also are highly complex and contain immense noise.

For fMRI experiments using a designed stimulus sequence to evoke human subject brain activity (in either event-related or block designs), a typical research focus is on evaluating the brain response to each stimulus type, usually within the framework of a generalized functional linear model, also called the general linear model (GLM) in neuroimaging literature (Friston et al., 1995; Goutte et al., 2000; Worsley and Friston, 1995). The GLM connects fMRI time series of each voxel to the stimulus sequence through an unknown hemodynamic response function (HRF). The focus of fMRI data analysis within the GLM is to estimate the HRF, as it summarizes each brain region's response to a stimulus. Many approaches have been developed in the literature to estimate HRFs, as explained in detail below.

Parametric methods for HRFs include the Poisson model (Friston et al., 1994b), the canonical HRF (Worsley and Friston, 1995) as a difference of two Gamma density functions (Worsley et al., 2002), a linear combination of the canonical HRF and its first order derivative (Calhoun et al., 2004; Friston et al., 1998; Glover, 1999; Liao et al., 2002), an

* Corresponding author. Halsey Hall 111, University of Virginia, Charlottesville, VA 22904, USA.
** Corresponding author. 102 Gilmer Hall, PO Box 400400, University of Virginia, Charlottesville, VA 22904, USA.
E-mail addresses: tz3b@virginia.edu (T. Zhang), jcoan@virginia.edu (J.A. Coan).
[1] Equally contributing authors.

https://doi.org/10.1016/j.neuroimage.2017.12.032
Received 12 May 2017; Accepted 12 December 2017
Available online 27 December 2017
1053-8119/© 2018 Elsevier Inc. All rights reserved.

inverse logit model (Lindquist and Wager, 2007), and many others. Nonparametric methods, usually representing the HRF by a linear combination of quite a few functional bases, are more flexible in characterizing the variation of HRFs across different brain areas, subjects, and stimuli. Popular nonparametric methods include the finite impulse response (FIR) method (Dale, 1999; Lange et al., 1999), smooth FIR (SFIR) (Goutte et al., 2000) and many other nonparametric estimates, which impose regularization on the roughness of HRF estimates (Casanova et al., 2008, 2009; Marrelec et al., 2003; Marrelec et al., 2001; Vakorin et al., 2007; Zhang et al., 2007). (Zhang et al., 2013, 2014b) built flexible semi-parametric models for subjects' HRFs to characterize both population-wide and subject-specific properties of the brain activity while accommodating variability in HRFs across different brain areas and stimuli.

The approaches mentioned above for HRF estimation within the GLM framework perform massive univariate analysis, that is, analyze one voxel's fMRI time series at a time. Since spatially adjacent voxels tend to have similar functions and similar fMRI data, it is more efficient to incorporate spatial information of voxels into the HRF modeling and estimation. Several approaches have been developed in this direction. Assuming that the HRFs of voxels within the same parcel share the same functional shape (Chaari et al., 2012; Makni et al., 2005, 2008; Vincent et al., 2010), proposed to use a spatially adaptive prior for HRF heights of voxels in the same parcel. Using a Poisson HRF (Buxton and Frank, 1997; Friston et al., 1994b) (Zhang et al., 2014a, 2016), developed Bayesian methods that can accommodate the complex spatial and temporal correlations among fMRI time series. The reference Degras and Lindquist, 2014 have recently developed a hierarchical model with a penalization approach to simultaneously detect activated brain regions and estimate HRFs in multi-subject fMRI studies but with the assumption that HRF shapes are the same across different stimulus types.

The existing approaches to HRF estimation within the GLM, as mentioned above, are mainly focused on estimating the HRF of each stimulus and mapping the brain's activated regions. In this paper, we focus on comparing HRFs. Specifically, in many psychological fMRI experiments—such as the one presented in Section 3, the researchers compare the brain responses to a stimulus of interest and to a control stimulus and aim to locate the brain regions with different responses to the two stimuli. A conventionally used approach is to extract characteristics, usually the height of HRF estimates, and compare extracted HRF characteristics through statistical tests. This approach, though intuitive and straightforward, has the following three limitations. First, subjects' HRF estimates tend to have large variability due to the low signal-to-noise ratio (SNR) of each subject's fMRI data. As a result, the characteristics of HRF estimates have large variability. Second, tests on one HRF characteristic may fail to identify differences in other HRF characteristics. For example, a *t*-test on comparing HRF heights usually fails to detect HRF shape differences. Third, tests on HRF characteristics are essentially voxel-wise analysis, ignoring the spatial property of fMRI data.

To address these limitations, we propose a new multivariate GLM (MGLM) to jointly model all voxels' responses to stimuli, and develop a new estimation method to compare HRFs of different stimuli. Specifically, with basis representation of all voxels' HRFs, we factorize the matrix of all voxels' basis coefficients into a product of two low-rank matrices, one corresponding to the temporal dimension of fMRI data and the other corresponding to the spatial dimension. We refer to these two matrices as the temporal and spatial parameter matrices and the ensuing MGLM as the low-rank MGLM (LRMGLM). This new model brings three concrete advantages. First, with the low-rank representation of many voxels' HRFs, the number of model parameters is significantly reduced. Second, the low-rank representation implies a common structure across different voxels' HRFs, enabling "borrowing" information across different voxels, thus increasing HRF estimation efficiency. Third, by employing the temporal and spatial parameter matrices (to summarize temporal and spatial properties of the brain activity), the proposed model enables separating the spatial and temporal information in the data, and thus, directly addresses the complexity in analyzing fMRI data.

We formulate estimation of the new LRMGLM as an optimization problem, and introduce a new penalization criterion with regularization on the temporal and spatial matrices. Specifically, we use penalty functions of the two matrices to ensure spatial smoothness of voxels' HRFs, temporal continuity of HRFs, and sparsity of identified brain regions. Moreover, we develop an efficient optimization algorithm to minimize the criterion, evaluate population-wide brain activity, and map the brain regions that respond differently to stimuli.

The rest of the article is organized as follows. In Section 2, we introduce the LRMGLM for stimulus-evoked fMRI data, formulate the ensuing model estimation into a non-convex optimization problem, and develop a fast optimization algorithm to estimate the model. A fast procedure to choose the penalty parameters in the optimization function is also recommended. In Section 3, we compare the proposed method with several standard approaches to HRF estimation via simulation studies. We then apply the proposed method to the data collected in an fMRI experiment involving emotion stimuli and identify brain voxels with different responses to two designed stimuli. Finally, section 4 has a discussion.

Materials and methods

A low-rank multivariate general linear model (LRMGLM)

Let $y_j^i(t),\ t=\delta,\ 2\delta,\ldots,T\delta,\ i=1,2,\cdots,n,$ and $j=1,2,\cdots,J,$ be the observed fMRI time series at brain region *j* for subject *i*, where δ, ranging from 0.5 to 4 s, is the experiment unit time for each 3D brain image to be captured. The *J* brain regions under study are either *J* voxels in a region of interest (ROI) or defined anatomically or functionally by some brain parcellation technique (Flandin et al., 2002) or clustering algorithm (Thirion and Faugeras, 2003).

The GLM connects the observed fMRI data to the brain response to a stimulus through a convolution between a stimulus function and a HRF. For an fMRI experiment with *K* different stimuli, the GLM is

$$y_j^i(t) = D^i(t)\, d_j^i + \sum_{k=1}^{K} \int_0^m h_{j,k}^i(u) \cdot v_k^i(t-u)du + \varepsilon_j^i(t), \tag{1}$$

where $D^i(t) \in \Re^r$ is a vector of given functions at time *t*, d_j^i is a column vector of *r* coefficients, $v_k^i(t-u)$, $k=1,\ldots,K$, is a known function, *m* is a fixed constant, and $\varepsilon_j^i(t)$ is an error term. The stimulus function, $v_k^i(t)$, describes the evoked time of the *k*th stimulus: $v_k^i(t)=1$ if the stimulus is evoked at time *t* in the experimental session for subject *i*; otherwise, it equals 0. The HRF $h_{j,k}^i(u)$, defined on the domain $[0,m]$, characterizes subject *i*'s brain response to the *k*th stimulus at brain voxel *j*. Here, the term $D^i(t)\, d_j^i$ characterizes a low frequency drift in the observed fMRI data due to subject's motion, respiration, heartbeat, and machine noise. Different approaches have been developed to correct for this drift effect, including applying a high-pass filter to the fMRI data and modeling the drift by a low-order polynomial function of time (Brosch et al., 2002; Luo and S Puthusserypady, 2008; Smith et al., 1999) or spline functions. Following the practice by (Friman et al., 2004; Friston et al., 2000), in this application, we let the covariates $D^i(t)$ be discrete cosine transform basis set functions $D(t) = \left(\sqrt{2/T}\cos\left(\pi\ \frac{t}{T}\right), \ldots,\ \sqrt{2/T}\cos\left(r\ \pi\ \frac{t}{T}\right)\right)$ (superscript *i* is omitted, because the covariates are identical across subjects). We use a high-pass 128-s cutoff, that is, $r = 2\ T\ RT/128 + 1 = 7$.

Different from standard approaches applying the GLM, (1), to one voxel's fMRI time series at one time and estimating each voxel's HRFs independently, we propose a joint model for all voxels' fMRI data within the GLM framework. We first represent $h_{j,k}^i(t)$ with fourth-order B-spline bases:

$$h^i_{j,k}(t) = \sum_{l=1}^{L} \omega^i_{jl,k} \cdot b_l(t),$$

where the basis functions $b_l(t)$ are chosen based on an equally-spaced partition with a unit-one inter-knot distance of the interval $[0, m]$. That is, the partition knots of $[0, m]$ are $(t_0 = 0, t_1 = 1, \ldots, t_m = m)$ and the number of basis functions L equals $4 + m - 1$, which is determined by the fourth-order and unit-one inter-knot distance accordingly. Moreover, we let $\omega^i_{j1,k} = \omega^i_{jL,k} = 0$ to satisfy the boundary condition that $h^i_{j,k}(0) = h^i_{j,k}(m) = 0$. The choice of basis functions is the same as that by Zhang et al., 2013. We use B-spline bases because they are effective for representing smooth functional curves (Boor, 2001; Eubank, 1999; Ruppert et al., 2003; Wahba, 1990). Existing studies have already shown that HRFs have a smooth curve, associated with the underlying smooth changes in blood oxygenation (Aguirre et al., 1998). Considering the above, (Zhang et al., 2007; Vakorin et al., 2007) also used B-splines bases to represent HRFs.

Let $Y^i_j = (y^i_j(1), \ldots, y^i_j(T))'$, $\mathbf{Y}^i$ be a $T \times J$ matrix with the (t, j)th element equalling $y^i_j(t)$, and $\mathbf{D}$ be a $T \times r$ matrix with the tth row equalling $D(t)$. Let $B_t = (b_1(t), \ldots, b_L(t))$, $\mathbf{X}^i_k$ be a $T \times L$ matrix with the (t, l)th element equalling $\int_0^m b_l(u) \cdot v^i_k(t-u)du$, $\Omega^i_{j,k} = (\omega^i_{j1,k}, \ldots, \omega^i_{jL,k})'$, $\mathbf{\Omega}^i_k$ be an $L \times J$ matrix with the jth column equalling $\Omega^i_{j,k}$, $\boldsymbol{d}^i$ be an $r \times J$ matrix with the jth column equalling d^i_j, $\boldsymbol{\varepsilon}^i_j = (\varepsilon^i_j(1), \ldots, \varepsilon^i_j(T))'$, and $\mathbf{E}^i$ be a $T \times J$ matrix with the jth column equalling $\boldsymbol{\varepsilon}^i_j$.

The GLMs (1) for all voxels' fMRI time series with the above basis representation of HRFs can be written in a matrix form

$$Y^i_j = \mathbf{D} d^i_j + \sum_{k=1}^{K} \mathbf{X}^i_k \Omega^i_{j,k} + \boldsymbol{\varepsilon}^i_j \quad \text{and } \mathbf{Y}^i = \mathbf{D}\boldsymbol{d}^i + \sum_{k=1}^{K} \mathbf{X}^i_k \mathbf{\Omega}^i_k + \mathbf{E}^i. \tag{2}$$

We call this model the multivariate GLM (MGLM). Although basis coefficients $\mathbf{\Omega}^i_k$ for $k = 1, \ldots, K$ can be estimated using ordinary least squares (OLS) estimation approach, the ensuing OLS estimates may have large variances and are unreliable to map activated brain voxels/regions, due to many parameters in the MGLM and a typically low SNR of fMRI data. To address this, we assume that $\mathbf{\Omega}^i_k$ falls into a parameter space of a much lower dimension, and has the form $\mathbf{\Omega}^i_k = \mathbf{U}^i_k \mathbf{V}^i_k$, where $\mathbf{U}^i_k$ and $\mathbf{V}^i_k$ are $L \times P$ and $P \times J$ matrices respectively with P much smaller than J. This form leads to a bilinear regression model of $\mathbf{Y}^i$ in $(\mathbf{U}^i_k, \mathbf{V}^i_k)$:

$$\mathbf{Y}^i = \mathbf{D}\, \boldsymbol{d}^i + \sum_{k=1}^{K} \mathbf{X}^i_k\, \mathbf{U}^i_k\, \mathbf{V}^i_k + \mathbf{E}^i. \tag{3}$$

In practice, we choose $P = 2$ to capture major differences in magnitude and latency of HRFs across subjects and voxels and also to keep the model simple.

The MGLM with a low-rank representation in (3), called the low-rank MGLM (LRMGLM), decomposes the fMRI data into temporal and spatial two parts. The matrix $\mathbf{U}^i_k$—whose elements $U^i_{lp,k}$, $l = 1, \ldots, L$ and $p = 1, \ldots, P$, are associated with L functional bases—captures principal functional shapes shared by J voxels' HRFs in response to the kth stimulus; the matrix $\mathbf{V}^i_k$—whose elements $V^i_{pj,k}$, $p = 1, \ldots, P$ and $j = 1, \ldots, J$, are associated with J voxels—characterizes each voxel's unique property, i.e., voxel-specific characteristic, in response to the kth stimulus. We refer to $\mathbf{U}^i_k$ and $\mathbf{V}^i_k$ as ith subject's temporal and spatial matrices, respectively. Independent component analysis (ICA) (Arbabshirani et al., 2013; Beckmann and Smith, 2005; Calhoun and Adali, 2006; Calhoun et al., 2001; Guo, 2008; Guo and Tang, 2013; Shi and Guo, 2014) uses a similar idea of separating spatial and temporal information, but is usually applied to resting-state fMRI, i.e., not considering stimulus effects. The proposed LRMGLM analyzed simultaneously different stimulus effects on the brain activity.

In comparison with the MGLM (2), the LRMGLM has three significant advantages. First, the LRMGLM captures the common temporal properties of HRFs shared by J voxels and translates such common properties to a rank constraint on the vector of estimates. As such, the LRMGLM combines information across voxels to estimate their HRFs. Second, the LRMGLM uses much fewer parameters than existing nonparametric methods to represent HRFs, and thus the ensuing estimates have much smaller variances. Third, by allowing different HRF shapes across stimulus types and brain regions, the LRMGLM provides adequate model flexibility for accommodating the variability of brain activity.

Outside the imaging literature while inside the statistics literature, regression models in a formulation similar to (3) have been proposed by (Gabriel, 1998; Potthoff and Roy, 1964; Srivastava et al., 2009). In these regressions, the expectation of the response is a product of three matrices: $\mathbb{E}Y = X\Theta Z$, where X and Z are known and Θ is the parameter matrix. The reference Hoff, 2015 proposed a similar multilinear tensor regression model in which $\mathbb{E}Y = \Theta_1 X \Theta_2'$, where Θ_1 and Θ_2 are unknown parameter matrices.

Note that the proposed LRMGLM is different from typical tensor regression models (Basu et al., 2012; Li et al., 2013; Shi et al., 2014; Zhou et al., 2013). The tensor regression, also called scalar-on-image regression in the imaging application, uses ultra-high dimensional neuroimaging data as predictors and low-dimensional scalar variables as the response. In contrast, in the proposed LRMGLM, the predictor matrix is of low dimension, and the response and coefficient matrices are of similarly high dimensions. As such, the proposed model is more related to matrix factorization (Agarwal et al., 2012; Koren et al., 2009; Feng et al., 2013; Balqis Samdin et al., 2017; Rendle, 2012) and factor models (Bai and Li, 2002, 2012). Other related factor-analytic models for brain data (fMRI and EEG), such as those by (Allen et al., 2014; Balqis Samdin et al., 2017; Ting et al., 2017; Wang et al., 2016), typically conduct factor analysis or matrix decomposition of the observed data in matrix form. In contrast, the LRMGLM decomposes the coefficient matrix after regressing for the stimulus effects on the brain activity.

Population-wide LRMGLM. The model (3) is for within-subject fMRI data analysis. To evaluate population-wide brain activity, we propose a new population-wide LRMGLM

$$\mathbf{Y}^i = \mathbf{D}\, \boldsymbol{d}^i + \sum_{k=1}^{K} \mathbf{X}^i_k\, \overline{\mathbf{U}}_k\, \overline{\mathbf{V}}_k + \tilde{\mathbf{E}}^i. \tag{4}$$

Here, $\overline{\mathbf{U}}_k$ with elements $\overline{U}_{lp,k}$, $l = 1, \ldots, L$ and $p = 1, \ldots, P$, and $\overline{\mathbf{V}}_k$ with elements $\overline{V}_{pj,k}$, $p = 1, \ldots, P$, and $j = 1, \ldots, J$, are population-average temporal and spatial parameter matrices, respectively. Intuitively, $\overline{\mathbf{U}}_k$ and $\overline{\mathbf{V}}_k$ are the means of the subject-specific parameters $\mathbf{U}^i_k$ and $\mathbf{V}^i_k$ across subjects. In comparison with the model (3), the model (4) puts the differences between subject-specific and population-average parameters into the error term $\tilde{\mathbf{E}}^i$.

Unlike typical group analysis (Guo, 2008; Guo and Tang, 2013; Xu et al., 2009; Zhang et al., 2016) of multi-subject fMRI data, we do not introduce subject-specific parameters for HRFs in the model (4). Since the focus of this paper is on studying the population-wide brain activity, we use (4) to reduce the number of free parameters so that we can evaluate population-wide brain activity with high efficiency and computationally fast. One may modify (4) and introduce subject-specific parameters to accommodate variation in brain activity across subjects. The key model parameters in (4) are summarized in Table 1.

Optimization function for parameter estimation

Detecting Regions with Different Responses. We estimate the model (4) and detect brain regions with different responses to a pair of stimuli by minimizing a penalized sum of squared errors (PSSE). The PSSE uses three penalties to mitigate three practical complications. First, the HRF is a continuous function, thus a roughness penalty is imposed on the matrix $\overline{\mathbf{U}}_k$ to induce temporal smoothness of HRFs, i.e.,

Table 1
Notations of key parameters.

Parameter	Description
$\mathbf{Y}^i$	The fMRI data in a $T \times J$ matrix form for J voxels of the ith subject.
$\mathbf{X}_k^i$	The $T \times L$ design matrix of the ith subject associated with the kth stimulus.
$D(t)$	A vector of values of r discrete cosine transform basis set functions at time t.
$\mathbf{D}$	A $T \times r$ matrix with the tth row equalling $D(t)$.
d_j^i	Drift coefficients for voxel j of the ith subject.
$\boldsymbol{d}^i$	A $T \times J$ matrix of drift coefficients with the jth column equalling d_j^i.
$\mathbf{V}_k^i$	The ith subject's $P \times J$ spatial matrix introduced in Model (3) for characterizing voxel-specific responses of J voxels to the kth stimulus.
$\overline{\mathbf{V}}_k$	The $P \times J$ population-average spatial parameter matrix introduced in Model (4) corresponding to the kth stimulus.
$\overline{V}_{pj,k}$	The element in the pth row and jth column of the matrix $\overline{\mathbf{V}}_k$.
$\mathbf{U}_k^i$	The ith subject's $L \times P$ temporal matrix introduced in Model (3) for characterizing common HRF shapes shared across J voxels in response to the kth stimulus.
$\overline{\mathbf{U}}_k$	The $L \times P$ population-average temporal parameter matrix introduced in Model (4) corresponding to the kth stimulus
$\overline{U}_{lp,k}$	The element in the lth row and pth column of the matrix $\overline{\mathbf{U}}_k$.

$$\mathscr{P}(\overline{\mathbf{U}}_k) = \sum_{p=1}^{P} \int_0^m \left(\sum_{l=1}^{L} \overline{U}_{lp,k} \cdot b_l^{(2)}(t) \right)^2 dt.$$

Second, since spatially close voxels tend to have similar fMRI time series and thus similar HRFs, a regularization that utilizes voxels' spatial information is imposed on $\overline{\mathbf{V}}_k$ to induce spatial smoothness of HRFs, i.e.,

$$\mathscr{R}(\overline{\mathbf{V}}_k) = \sum_{p=1}^{P} \sum_{j \sim \tilde{j}} \left(\overline{V}_{pj,k} - \overline{V}_{p\tilde{j},k} \right)^2,$$

where $j \sim \tilde{j}$ denotes spatially adjacent voxels j and $\tilde{j}$. The penalty $\mathscr{R}(\overline{\mathbf{V}}_k)$ ensures that adjacent voxels have similar voxel-specific parameters $\overline{V}_{pj,k}$ and thus similar HRFs. We here use a one-neighbor dependence to avoid oversmoothing, especially for boundary voxels. Third, we impose a sparsity-inducing penalty on spatial parameters $\overline{\mathbf{V}}_k$ to locate regions with different responses to stimuli. Without loss of generality, we compare brain responses to the first two stimuli and impose the following sparsity penalty on the difference of spatial parameter matrices,

$$\mathscr{S}(\overline{\mathbf{V}}_1, \overline{\mathbf{V}}_2) = \sum_{j=1}^{J} \sqrt{\sum_{p=1}^{P} \left(\overline{V}_{pj,1} - \overline{V}_{pj,2} \right)^2}.$$

Let $\boldsymbol{\Theta} = \{\boldsymbol{d}^i, \overline{\mathbf{U}}_k, \overline{\mathbf{V}}_k, i = 1, \ldots, n, k = 1, \ldots, K\}$. Overall, we estimate the LRMGLM and identify brain voxels/regions with different responses to stimuli 1 and 2 through minimizing the following PSSE:

$$\mathrm{PSSE}(\boldsymbol{\Theta}) = \mathrm{SSE}(\boldsymbol{\Theta}) + \lambda \cdot \sum_{k=1}^{K} \mathscr{P}(\overline{\mathbf{U}}_k) + \tau \cdot \sum_{k=1}^{K} \mathscr{R}(\overline{\mathbf{V}}_k) + \mu \cdot \mathscr{S}(\overline{\mathbf{V}}_1, \overline{\mathbf{V}}_2), \tag{5}$$

where $\mathrm{SSE}(\boldsymbol{\Theta}) = \frac{1}{n}\sum_{i=1}^{n} \| \mathbf{Y}^i - \mathbf{D}\, \boldsymbol{d}^i - \sum_{k=1}^{K} \mathbf{X}_k^i\, \overline{\mathbf{U}}_k\, \overline{\mathbf{V}}_k \|_F^2$.

Iterative algorithm for minimizing the PSSE

Let $\mathscr{U} = \{\overline{\mathbf{U}}_k, k = 1, \ldots, K\}$, $\mathscr{V} = \{\overline{\mathbf{V}}_k, k = 1, \ldots, K\}$, and $\tilde{\mathbf{d}} = \{\boldsymbol{d}^i, i = 1, \ldots, n\}$. We propose the following alternating convex search algorithm to minimize $\mathrm{PSSE}(\boldsymbol{\Theta})$, which iterates between two major convex optimization steps until convergence:

1. Given $\mathscr{U}$ and $\tilde{\mathbf{d}}$, find minimizer $\mathscr{V}$ of $\mathrm{PSSE}(\mathscr{V} \big| \tilde{\mathbf{d}}, \mathscr{U}) = \mathrm{SSE}(\boldsymbol{\Theta}) + \tau \sum_{k=1}^{K} \mathscr{R}(\overline{\mathbf{V}}_k) + \mu \cdot \mathscr{S}(\overline{\mathbf{V}}_1, \overline{\mathbf{V}}_2)$.
2. Given $\mathscr{V}$, find minimizer $\tilde{\mathbf{d}}$ and $\mathscr{U}$ of $\mathrm{PSSE}(\tilde{\mathbf{d}}, \mathscr{U} \big| \mathscr{V}) = \mathrm{SSE}(\boldsymbol{\Theta}) + \lambda \cdot \sum_{k=1}^{K} \mathscr{P}(\overline{\mathbf{U}}_k)$.

Step 2 is straightforward, because given $\mathscr{V}$, $\mathrm{PSSE}(\tilde{\mathbf{d}}, \mathscr{U} \big| \mathscr{V})$ is a quadratic function of $\tilde{\mathbf{d}}$ and $\mathscr{U}$, and the analytic formula of the minimizer exists. See Appendix A for the derivation of the minimizer. Step 2 can be efficiently solved by using a computational algorithm developed in (Robinson, 1991; Ruppert et al., 2003). For Step 1, since the objective function $\mathrm{PSSE}(\mathscr{V} \big| \tilde{\mathbf{d}}, \mathscr{U})$ is convex and non-differentiable, we use an iterative method called Alternating Linearization Proximal Gradient [Kiwiel et al., 2006, Lin et al., 2014, ALPG] to find its minimizer. The main idea of the ALPG is to iteratively solve two subproblems, each involving minimization of a linear approximation of the objective function, as explained in detail below.

We can view the objective function $\mathrm{PSSE}(\mathscr{V} \big| \tilde{\mathbf{d}}, \mathscr{U})$ as a sum of two component functions: the differentiable part $\mathrm{DPSSE}(\mathscr{V} \big| \tilde{\mathbf{d}}, \mathscr{U}) = \mathrm{SSE}(\boldsymbol{\Theta}) + \tau \sum_{k=1}^{K} \mathscr{R}(\overline{\mathbf{V}}_k)$ and the non-differentiable part $\mu \cdot \mathscr{S}(\overline{\mathbf{V}}_1, \overline{\mathbf{V}}_2)$. Each linear approximation consists of one component function and a linearization of the other component function. Specifically, at each iteration, $\mathrm{DPSSE}(\mathscr{V} \big| \tilde{\mathbf{d}}, \mathscr{U})$ and $\mu \cdot \mathscr{S}(\overline{\mathbf{V}}_1, \overline{\mathbf{V}}_2)$ are linearized at points $\mathscr{Z}_1$ and $\mathscr{Z}_2$ in subproblems 1 and 2, respectively, where $\mathscr{Z}_1$ and $\mathscr{Z}_2$, of the same dimension as $\mathscr{V}$, are the output from solving the previous subproblem. Let $\mathscr{V}^*$ be the estimate of $\mathscr{V}$ from the previous iteration, i.e., the output from solving the subproblems in the last iteration. The ALPG algorithm for minimizing $\mathrm{PSSE}(\mathscr{V} \big| \tilde{\mathbf{d}}, \mathscr{U})$ in Step 1 is outlined in the following.

1.I Initialization. Set $\mathscr{V}^*$, $\mathscr{Z}_1$, and $\mathscr{Z}_2$, all in the same dimension as $\mathscr{V}$, to be 0 at the starting point. Let $\mathscr{G}_1$ be the gradient of the $\mathrm{DPSSE}(\mathscr{V} \big| \tilde{\mathbf{d}}, \mathscr{U})$ at $\mathscr{Z}_1$.

1.II.a In the subproblem 1, solve the optimization

$$\mathscr{Z}_2 = \arg\min_{\mathscr{V}} \left\langle \mathscr{G}_1, \mathscr{V} \right\rangle + \mu \cdot \mathscr{S}(\overline{\mathbf{V}}_1, \overline{\mathbf{V}}_2) + \frac{\rho}{2} \| \mathscr{V} - \mathscr{V}^* \|_2^2, \tag{6}$$

- where $\langle \cdot, \cdot \rangle$ stands for inner product, $\langle \mathscr{G}_1, \mathscr{V} \rangle$ is a linear approximation of $\mathrm{DPSSE}(\mathscr{V} \big| \tilde{\mathbf{d}}, \mathscr{U})$ at $\mathscr{Z}_1$, and $\rho > 0$ is a positive constant. The term $\frac{\rho}{2} \| \mathscr{V} - \mathscr{V}^* \|_2^2$ is used to penalize the distance between the candidate $\mathscr{Z}_2$ and the estimate $\mathscr{V}^*$ so that the solution to the subproblem is in the proximity of $\mathscr{V}^*$, controlled by ρ. We choose $\rho = 1$ in our implementation. Through our studies, we found that the magnitude of ρ does not influence the efficiency of the whole algorithm much.

1.II.b Calculate a sub-gradient $\mathscr{G}_2$ of the non-differentiable part at $\mathscr{Z}_2$ according to:

$$\mathscr{G}_2 = -\mathscr{G}_1 - \rho(\mathscr{Z}_2 - \mathscr{V}^*). \tag{7}$$

Then $\langle \mathscr{G}_2, \mathscr{V} - \mathscr{Z}_2 \rangle$ is a linear lower-approximation of $\mu \cdot \mathscr{S}(\overline{\mathbf{V}}_1, \overline{\mathbf{V}}_2)$ since $\mathscr{G}_2$ is a sub-gradient of $\mu \cdot \mathscr{S}(\overline{\mathbf{V}}_1, \overline{\mathbf{V}}_2)$ at $\mathscr{Z}_2$. If $\mathscr{Z}_2$ improves the objective function $\mathrm{PSSE}(\mathscr{V} \big| \tilde{\mathbf{d}}, \mathscr{U})$, update $\mathscr{V}^* = \mathscr{Z}_2$.

1.III.a In the subproblem 2, solve

$$\mathscr{Z}_1 = \arg\min_{\mathscr{V}} \mathrm{SSE}(\boldsymbol{\Theta}) + \tau \sum_{k=1}^{K} \mathscr{R}(\overline{\mathbf{V}}_k) + \left\langle \mathscr{G}_2, \mathscr{V} - \mathscr{Z}_2 \right\rangle + \frac{\rho}{2} \| \mathscr{V} - \mathscr{V}^* \|_2^2. \tag{8}$$

1.III.b The gradient of the differentiable part at $\mathscr{Z}_1$ can be calculated by the following:

$$\mathscr{G}_1 = -\mathscr{G}_2 - \rho(\mathscr{Z}_1 - \mathscr{V}^*). \tag{9}$$

If $\mathscr{Z}_1$ improves the objective function value of PSSE($\mathscr{V} \big| \tilde{\mathbf{d}}, \mathscr{U}$), update $\mathscr{V}^* = \mathscr{Z}_1$.

1.IV Repeat from 1.II.a until convergence.

The efficiency of the ALPG algorithm lies in the efficiency of solving the above two subproblems. The optimization function in (8) is quadratic of $\mathscr{V}$, so the analytic formulation for the minimizer $\mathscr{Z}_1$ exists. The optimization function in (6) is not differentiable of $\mathscr{V}$ and it is computationally challenging to solve (6) due to the high dimension and complex structure of the function $\mu \cdot \mathscr{S}(\overline{\mathbf{V}}_1, \overline{\mathbf{V}}_2)$. Despite these issues, the function is separable in terms of the columns of $\overline{\mathbf{V}}_1 - \overline{\mathbf{V}}_2$, therefore, it can be decomposed into J smaller problems. Each of these smaller problems, although still non-differentiable, has a simple dual form and can be solved efficiently with a closed-form solution. In summary, the special form of the objective function PSSE($\mathscr{V} \big| \tilde{\mathbf{d}}, \mathscr{U}$) enables us to minimize it efficiently by using the ALPG. The detailed description of solving the two subproblems is given in Appendix B.

In summary, the alternating convex search algorithm converges monotonically with improvement made after every iteration. The algorithm is terminated when the improvement made is less than a threshhold of 10^{-4}. Under mild conditions such as continuity of the objective function, the compactness of the solution set, and uniqueness of the subproblems' solutions, which are satisfied in our model, the sequence of points generated by the algorithm converge to a partial optimum of the objective function Gorski et al., 2007.

Strategy for selecting penalty parameter

Three penalty parameters, λ, τ and μ, control the temporal smoothness of the HRFs, spatial smoothness of estimated HRFs of spatially distributed voxels, and the extent of sparsity in the identified brain voxels, respectively. Ordinary cross-validation (OCV) and generalized cross-validation (GCV) (Wahba, 1990) are conventional methods for choosing the penalty parameters in penalized optimization problems. In the context of functional data analysis (Reiss and Ogden, 2007, 2009), proposed GCV-based selection procedures, and (Wood, 2011) investigated a restricted maximum likelihood (REML) method. To select penalty values, because these approaches are either time consuming (i.e., OCV) or not directly applicable (i.e., REML), we propose a fast computing method.

We start with a large number of candidate values, ranging from e^{-1} to e^{10}, for each penalty parameter. The boundary values (e^{-1} and e^{10}) are chosen such that the penalty parameters beyond this range result in either almost no or all voxels selected. Applying the proposed alternating convex search algorithm to fMRI data with different combinations of the three penalty parameter values, we identify the combinations that lead to the result satisfying two criteria: (1) Between 5% and 50% of candidate voxels are selected; and (2) the selected voxels in the largest cluster account for more than 80% of all selected voxels. The criterion (1) ensures sparsity while excluding the penalty parameters that result in too few voxels selected. The criterion (ii) ensures that the selected voxels are spatially clustered together. Based on simulation studies, we found that the threshold varying from 50% to 80% in the criterion (ii) leads to the same result. We here choose 80%, because it screens away most parameters (> 80%) and the ensuring computational time is least for penalty parameter selection. For each combination of the rest penalty parameters, we calculate the correlations between each selected voxel with its neighboring unselected voxels. (If all the neighboring voxels of one selected voxel are selected, we exclude this voxel from the calculation.) Then, we take an average of all these correlations, and choose the penalty parameters that give the smallest average correlation.

The intuition behind the above selection approach is two fold. First, if too few or too many voxels are selected, many selected voxels at the boundary of selected regions have fMRI data similar to those of the neighboring unselected voxels. Then, the average correlation tends to be high. Second, the average correlation is a measure of similarity between the two groups of selected and unselected voxels at the boundary of selected brain regions. The penalty parameter selection based on the smallest average correlation is essentially to minimize the similarity between the two groups.

The above correlation-based procedure selects parameters that can balance well between selecting clusters of voxels with different activities from unselected voxels and "borrowing" information from neighboring voxels based on one-neighbor dependency. Moreover, the proposed penalty parameter selection procedure is parallelizable. Through our simulation studies below, we show that the proposed penalty parameter selection procedure is computationally fast and leads to voxel/region selection with a high true positive rate (TPR) and a low false positive rate (FPR).

Results

Simulations

Data Generation. We simulated fMRI data using the experimental design that is same as the real fMRI experiment under study (Coan, 2010, 2011; Coan et al., 2006; Zhang et al., 2013). Each subject's fMRI data contain 205 scans (after removing the first a few fMRI scans) with repetition time (TR) 2s. The experiment used 4 different stimuli. We generated different HRFs associated with the 4 stimuli, and used the proposed LRMGLM to identify voxels with different responses to the first two stimuli, consistent with the objective of the real psychology experiment under study, to be explained in Section Illustrative example.

The simulated data distributed over a $15 \times 15 \times 15$ lattice. Each voxel's 4 HRFs are simulated from a semi-parametric model by Zhang et al., 2013

$$h_{j,k}^i(t) = A_{j,k}^i \cdot \phi_{j,k}\left(t + D_{j,k}^i\right), \tag{10}$$

where the function $\phi_{j,k}(t)$ characterizes the mean HRF shape of voxel j in response to the kth stimulus across subjects, and $A_{j,k}^i$ and $D_{j,k}^i$ are subject i's brain response magnitude and latency, respectively, at voxel j in response to the kth stimulus. We use the semi-parametric model to generate HRFs, because it can well describe the variation of HRFs in magnitude and latency across subjects, voxels, and stimulus types.

Following (Worsley et al., 2002), we used a difference of two gamma density functions to represent $\phi_{j,k}(t)$:

$$\phi_{j,k}(t) = b_{1,jk}^{a_{1,jk}} \frac{t^{a_{1,jk}-1} \exp\left(-b_{1,jk}\, t\right)}{\Gamma\left(a_{1,jk}\right)} - c \cdot b_{2,jk}^{a_{2,jk}} \frac{t^{a_{2,jk}-1} \exp\left(-b_{2,jk}\, t\right)}{\Gamma\left(a_{2,jk}\right)}. \tag{11}$$

We let HRF parameters $a_{1,jk}$, $a_{2,jk}$, $b_{1,jk}$, $b_{2,jk}$ and c fixed at values of 6, 16, 1, 1, and 1/6, respectively, leading to the canonical HRF shape (Worsley and Friston, 1995), which is widely used in the neuroimaging literature for describing brain activities in motor and visual cortices (Handwerker et al., 2004; Lindquist, 2008).

We let $\boldsymbol{A}_k^i = \{A_{j,k}^i, j = 1 \ldots, J\}$ and $\boldsymbol{D}_k^i = \{D_{j,k}^i, j = 1 \ldots, J\}$ follow multivariate normal distributions MVN($\boldsymbol{\mu}_{\boldsymbol{A}_k}$, $\boldsymbol{\Sigma}_{\boldsymbol{A}_k}$) and MVN($\boldsymbol{\mu}_{\boldsymbol{D}_k}$, $\boldsymbol{\Sigma}_{\boldsymbol{D}_k}$), respectively. As such, $\boldsymbol{\mu}_{\boldsymbol{A}_k}$ can be viewed as the population-mean brain

Table 2
Parameters of simulated HRFs $h^i_{j,k}$. Note that the HRF parameters associated with the second stimulus are for the voxels on the center $7 \times 7 \times 7$ lattice grids while the HRF parameters of the rest voxels are identical to those under the first stimulus.

	k	μ_{A_k}	σ_{A_k}	ρ_{A_k}	μ_{D_k}	σ_{D_k}	ρ_{D_k}	$a_{1,i}$	$a_{2,i}$	$b_{1,i}$	$b_{2,i}$	c
	1	300	50	0.9	−1	0.5	0.9	6	16	1	1	1/6
Example 1	2	350	50	0.9	−4	0.5	0.9	6	16	1	1	1/6
	3	300	50	0.9	0	0.5	0.9	6	16	1	1	1/6
	4	600	100	0.5	0	0.3	0.5	6	16	1	1	1/6

response magnitudes, and $\boldsymbol{\mu}_{D_k}$ is the population-mean latency at J voxels using the canonical HRF as the baseline.

Multivariate normal distributions can be conveniently used to characterize the spatial correlations among HRFs of different voxels. Specifically, we let $\boldsymbol{\Sigma}_\theta(j_1, j_2) = \sigma^2_\theta \cdot \rho_\theta^{\|j_1 - j_2\|_2^2}$, where ρ_θ is a positive constant between 0 and 1, $\|j_1 - j_2\|_2$ denotes the Euclidean distance between voxels j_1 and j_2, and σ^2_θ is the variance of the multivariate normal distribution.

Echoing the real problem of comparing brain responses to the first two stimuli, we let the HRFs of the first two stimuli be different at the center $7 \times 7 \times 7$ voxels. That is, around 10.0% of voxels have different responses to the first two stimuli. The parameter values of the multivariate normal distributions are given in Table 2. The HRFs of the first two stimuli differ both in magnitudes and latency. We let the mean $\boldsymbol{\mu}_{D_k}$ be different from 0 to mimic the real problem where the activity of emotion-related regions is different from of motor and visual cortices. Moreover, we choose large values for ρ_θ to impose a strong spatial dependence among HRF parameters. Fig. 1(a) shows the simulated HRFs of two subjects at one voxel, which has different HRFs to the first two stimuli.

We generated GLM error terms $\tilde{\mathbf{E}}^i$ from an autoregressive model of order 4 (AR(4)) (Casanova et al., 2008; Zhang et al., 2013) with lag-1 correlation of 0.45:

$$\varepsilon^i_j(t) = 0.37\, \varepsilon^i_j(t-1) + 0.14\, \varepsilon^i_j(t-2) + 0.05\, \varepsilon^i_j(t-3) + 0.02\, \varepsilon^i_j(t-4) + e^i_j(t),$$

where $\boldsymbol{e}^i_t = (e^i_1(t), \ldots, e^i_J(t))' \overset{i.i.d}{\sim} \mathrm{MNV}(0, \Phi)$, $\Phi(j_1, j_2) = \tau^2 \cdot \rho_e^{\|j_1 - j_2\|_2^2}$, $\rho_e = 0.99$ and $\tau = 200$. We let the spatial dependence among errors be stronger than that of HRFs considering strong smoothing of fMRI data due to the pre-processing steps.

We simulated $n = 106$ subjects' fMRI data—the same as the real fMRI data, each having 205 scans separated by 2s (TR). We first used parameters in Table 2 to generate subjects' HRFs for voxels distributed on the $15 \times 15 \times 15$ lattice, calculated the voxels' drift terms and AR(4) errors, and obtained fMRI time series based on the GLM (1).

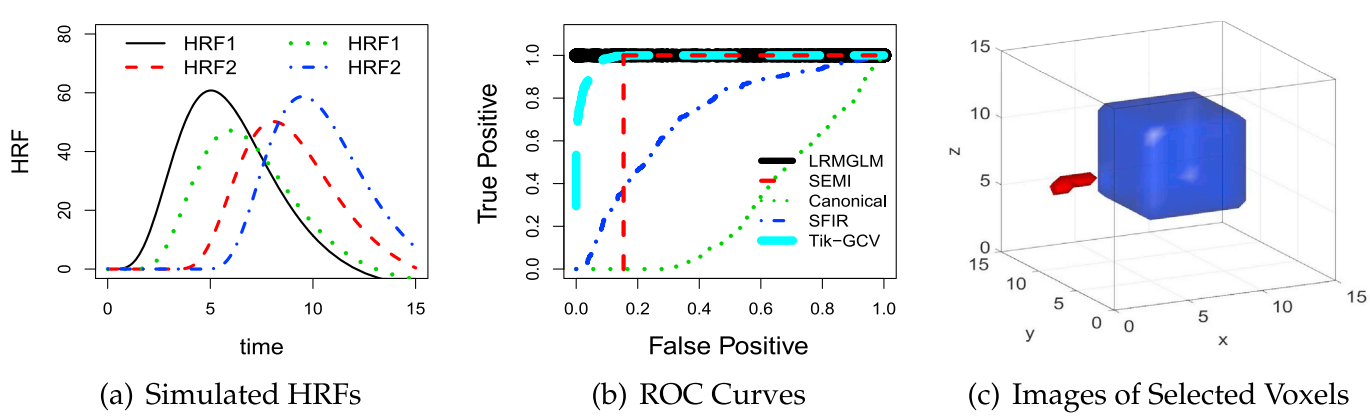
Note that the simulated data do not exactly follow the proposed LRMGLM (3). Even with this, we show in the following that the LRMGLM can outperform the voxel-wise methods, which actually follow the data generating model.

Statistical Analysis and Discussion. We applied the proposed LRMGLM to the simulated data. We compared our result to those by the nonparametric method using Tikhonov regularization and generalized cross validation (Tik-GCV) (Casanova et al., 2008, 2009), the smooth finite impulse response (SFIR) method (Glover, 1999; Goutte et al., 2000), the method using the canonical HRF and its first order derivative as functional bases (Liao et al., 2002) (referred to as the canonical method), and the semi-parametric (SEMI) method (Zhang et al., 2013) developed based on the HRF generating model (10). We first used these methods to estimate HRFs, then extracted their heights, and used t-tests on height estimates to compare subjects' HRFs in response to the first two stimuli. By using different thresholds for the P-values from the t-tests, we created the ROC curve, which shows pairs of TPRs and FPRs for different thresholds. For comparison, we also presented the ROC curve of the LRMGLM by using different combinations of penalty parameters.

Fig. 1(b) shows the ROC curves of the four competing methods and the LRMGLM for simulated data. Among the four voxel-wise methods that have been analyzed, Tik-GCV has the best ROC curve, achieving 100% TPR at the lowest FPR (around 8%). The other three methods have much larger FPRs. For example, SEMI can select all the true voxels, but with a FPR larger than 15%. In contrast, the LRMGLM, with an appropriate combination of three penalty values, can achieve 100% TPR and 0% FPR.

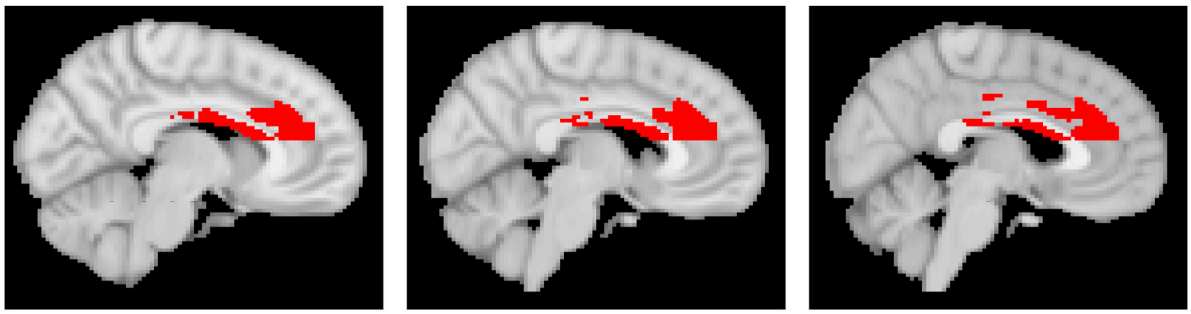
Fig. 1(c) shows the 3D image of voxels identified to have different responses to the first two stimuli by using the LRMGLM with the selected penalty parameters. The voxels in blue correspond to true positives and those in red correspond to false positives. The LRMGLM achieved 0.3% FPR and 100% TPR. This simulation indicates that the proposed penalty parameter selection procedure can identify appropriate parameters values that leads to voxel selection with both high sensitivity and specificity.

We conducted another simulation example where the HRF shape $\phi_{j,k}(t)$ is given based on the real data and two clusters of voxels have



(a) Simulated HRFs (b) ROC Curves (c) Images of Selected Voxels

Fig. 1. (a) Two subjects' HRFs associated with the first two stimuli. (b) ROC curves of five methods. (c) 3D Image of voxels identified to have different responses to the first two stimuli using the penalty parameters selected by the approach in Section 2.4. Blue cells correspond to true positives and red cells correspond to false positives.



(a) Slice X=42 (b) Slice X=43 (c) Slice X=44

Fig. 2. (a)–(c) Images of dACC voxels identified by the LRMGLM to have different responses to threat and safety cues.

different responses to the first two stimuli. This simulation is to mimic real data analysis results. The LRMGLM still outperformed all the other methods. The details are provided in the supplementary file.

Illustrative example

Subjects. The fMRI data under study were collected in a psychology study, which examined brain responses to negative emotional stimuli (Coan, 2010, 2011; Coan et al., 2006; Coan et al., 2013). The study recruited 106 mentally and physically healthy subjects from a larger community sample (Allen et al., 2007).

Experimental Design. The fMRI experiment used mild electric shocks to evoke subjects' emotional responses. The electric shocks, at 4 mA with 1 s duration, were delivered by using an isolated physiological stimulator (Coulbourn Instruments, Allentown, PA). Overall, the experiment used a block design consisting of 4 different stimuli: threat cues, indicating 20% chance of receiving a mild electric shock; safety cues indicating no shock; resting periods between the cues; and 2 real mild electric shocks. These four stimuli occurred at frequencies of 27.9%, 21.0%, 50.8%, and 0.2%.

Intense noise and enclosed space inside the fMRI scanner (Mazard et al., 2002; Szameitat et al., 2009) can cause anxiety and uncomfortableness to subjects. To correct for their brains' emotional activity due to the uncomfortable fMRI scanning process, we treat the brain response to safety cues as the baseline and identify brain voxels with different responses to threat and safety cues.

Data Acquisition and Preprocessing. Data collection was completed at Fontaine Research Park in the University of Virginia with a Siemen's 3.0 T MAGNETOM Trio high-speed magnetic imaging device with a circularly polarized transmit/receive head coil and integrated mirror. Before collection of functional images, 176 high-resolution T1-magnetization-prepared rapid-acquisition gradient echo images were acquired to determine the localization of function (1-mm slices, TR = 1900 ms, echo time (TE) = 2.53 ms, flip angle = 9°, field of view (FOV) = 250 mm, voxel size = 1 × 1 × 1 mm). Two-hundred and sixteen functional T2*- weighted echo planar images (EPIs) sensitive to blood-oxygen-level-dependent contrasts were collected in volumes of 28 3.5-mm transversal echo-planar slices (1-mm slice gap) covering the whole brain (1-mm slice gap, TR = 2000 ms, TE = 40 ms, flip angle = 90°, FOV = 192 mm, matrix = 64 × 64, voxel size = 3 × 3 × 3.5 mm).

Data were preprocessed and analyzed using FMRIB's Software Library (FSL) software [Friston et al., 1994a, Version5.98; www.fmrib.ox.ac.uk/fsl]. Motion correction was completed using FMRIB's Linear Image Registration Tool, and intra-modal correction algorithm tool [Jenkinson et al., 2002, MCFLIRT], with slice scan time correction and a high-pass filtering cutoff point of 100 s. Brain extraction was completed using BET (Smith, 2002) and smoothing completed with a 5-mm full width at half minimum Gaussian kernel. Images were registered to Montreal Neurological Institute (MNI) space using FLIRT (Jenkinson et al., 2002). Trials in which participants received shocks were excluded due to movement artifacts.

We focus on the fMRI data in dorsal anterior cingulate cortex (dACC), which is an emotion-related brain region of interest (Knutson et al., 2000; Milad et al., 2007) and implicated in mental health (Dedovic et al., 2016). We used Harvard subcortical brain atlas to determine dACC in the fMRI data.

Statistical Analysis and Discussion. The brain region dACC consists of functionally distinct subregions (Schulz et al., 2011), which potentially process emotional information differently (Bush et al., 2002). Applying the LRMGLM to the fMRI data in dACC, we aim to identify the dACC subregion that has the most pronounced response to a negative emotional stimulus in comparison to a control stimulus.

In total, 33% voxels have been selected. Fig. 2 shows the dACC voxels identified to have different responses to the threat and safety cues. The voxels identified are mainly in the anterior portion of dACC (adACC). This result indicates the existence of distinction among subregions in dACC, which is in line with the existing studies regarding heterogeneity in dACC (Bush et al., 2002). According to (Etkin et al., 2011; Morris et al., 1998), adACC plays an important role in appraising and expressing negative emotion. Moreover, adACC is involved in regulating amygdala's affective response to fear (Hong et al., 2014). Our analysis confirms different extents of responses to negative emotion in dACC subregions and identifies the subregion with the strongest response.

We also detected responses to negative emotional stimuli in the white matter corpus callosum, in line with a growing body of evidence on fMRI activation in white matter (Fraser et al., 2012; Yarkoni et al., 2009; Zhang et al., 2013), especially in corpus callosum (D'arcy et al., 2006; Fabri et al., 2011; Gawryluk et al., 2011; Mazerolle et al., 2008). The reference (Gawryluk et al., 2014) provides potential reasons for detecting white matter activity through fMRI. The LRMGLM is able to characterize the white matter activity. We can use the proposed method to further investigate white matter and gray matter brain functions.

We also applied SFIR, Tik-GCV, Canonical, and SEMI methods to the dACC data. We found that p-values obtained were extremely small for all the voxels (below 10^{-7}). For comparison, we selected the top 33% voxels with the smallest P-values and show them in Fig. 3. In general, we found that the voxels identified by the nonparametric methods, SFIR and Tik-GCV, tend to be scattered, possibly due to large variances of nonparametric estimates. The SEMI method also identified voxels in adACC to have different responses. Overall, the voxels identified by the SEMI method are more clustered than those by SFIR and Tik-GCV, but more spread out than those by the LRMGLM. The canonical method identified two separate clusters of voxels, one at the boundary of dACC, and the other in the corpus callosum. Though the canonical method selected closely clustered voxels, it is difficult to provide meaningful scientific interpretation of the ensuing results.

Overall, voxel-wise methods tend to have large FPRs for two potential reasons. First, fMRI data have a low SNR, while second, the errors in fMRI data at spatially close regions are strongly correlated. Consequently, the null voxels spatially close to true voxels tend to have significant P-values,

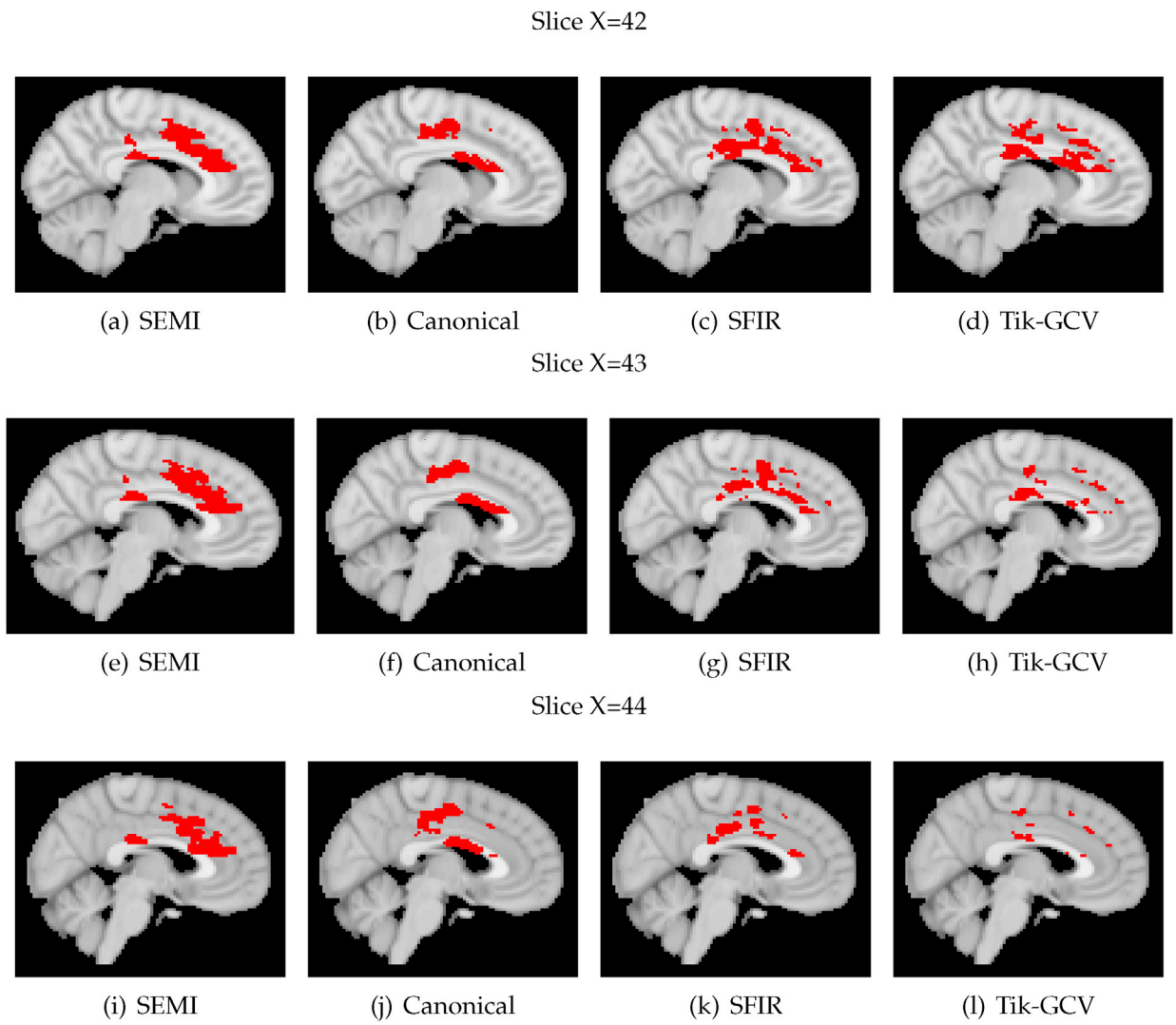


Fig. 3. Images of dACC voxels identified by four voxel-wise methods—SEMI, Canonical, SFIR, and Tik-GCV—to have different responses to threat and safety cues.

and voxel-wise methods without accounting for voxels' spatial information tend to select many null voxels. In general, the more neighboring voxels of true voxels and the stronger spatial dependence of fMRI data, the more false discoveries. In contrast, the proposed LRMGLM method takes into account the spatial smoothness of the HRFs of selected voxels, and controls the size of selected voxels based on the similarity between selected and unselected voxels at the boundary. Thus, voxel selection by the LRMGLM has a much smaller FPR.

Discussion

In the simulated example, HRFs of the first two stimuli differ both in magnitude and latency. The canonical method performed poorly and selected the voxels mainly located at the boundary of the space of candidate voxels. Similarly, the dACC voxels with the most significant P-values by the canonical method are also located at the boundary of dACC. The canonical method is able to select voxels in a close cluster, possibly because the HRF of each subject and voxel has only two free parameters, the ensuing estimates have much less variability and the P-values of spatially close voxels are strongly correlated due to the strong correlations of the voxels' fMRI data.

The proposed method does not compare magnitude estimates of HRFs but rather the entire HRFs. However, the proposed LRMGLM can also be used to detect magnitude differences, especially if there is a prior belief in this. Simulation studies (see supplementary files) show that the proposed LRMGLM outperformed voxel-wise methods for comparing magnitude estimates by having better sensitivity and specificity, even if the underlying HRFs differ only in magnitude, because the proposed method utilizes all voxels' information. Nevertheless, one drawback of the proposed method is that it cannot tell a direction of difference, positive or negative. Thus, further investigation of HRF estimates is needed to address this question.

We did not use cross validation to select penalty parameters, because we found that the ensuing voxel selection can have a high FPR. This is possibly because though the observed fMRI data have a linear relationship with HRFs in the GLM, the proposed LRMGLM is bilinear in spatial and temporal matrices, that is, linear in $\overline{\mathbf{U}}_k$ given $\overline{\mathbf{V}}_k$ and linear in $\overline{\mathbf{V}}_k$ given $\overline{\mathbf{U}}_k$, but nonlinear in $(\overline{\mathbf{U}}_k, \overline{\mathbf{V}}_k)$. As such, the optimization function used is non-convex, and the penalty parameters leading to small estimated prediction errors do not necessarily lead to small voxel selection errors.

It is possible to incorporate the spatial and temporal dependence structure of the error matrix $\tilde{\mathbf{E}}^i$ into the optimization function PSSE($\boldsymbol{\Theta}$) by replacing SSE($\boldsymbol{\Theta}$) with

$$\frac{1}{n}\sum_{i=1}^{n}\mathrm{tr}\left(\left(\mathbf{Y}^i - \mathbf{D}\,\boldsymbol{d}^i - \sum_{k=1}^{K}\mathbf{X}_k^i\,\overline{\mathbf{U}}_k\,\overline{\mathbf{V}}_k\right)'\boldsymbol{\Gamma}\left(\mathbf{Y}^i - \mathbf{D}^i\,\boldsymbol{d}^i - \sum_{k=1}^{K}\mathbf{X}_k^i\,\overline{\mathbf{U}}_k\,\overline{\mathbf{V}}_k\right)\right),$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix, and the matrix $\mathbf{\Gamma}$ characterizes the temporal and spatial correlations among the error terms. The unknown $\mathbf{\Gamma}$ can be estimated using the model fitting errors from the voxel-wise analysis of the fMRI data under the standard GLM (1). In addition, we can impose a structure over $\mathbf{\Gamma}$ (Aubry et al., 2012) to improve its estimation efficiency. This approach will be investigated in the future research.

Though the proposed method can outperform voxel-wise methods in detecting brain voxels with different population-wide responses to stimuli, the former may not be as efficient as the latter for estimating each subject's HRFs for two potential reasons. First, variation of brain activity across subjects is large. The proposed LRMGLM (4) is developed mainly for evaluating population-average brain activity rather than for estimating subject-specific HRFs. We use this model to achieve high efficiency in voxel selection by significantly reducing the number of free parameters that accommodate the variation of brain activity across subjects. Second, with the temporal and spatial matrix parameters $(\overline{\mathbf{U}}_k, \overline{\mathbf{V}}_k)$ having a nonlinear formulation, a high efficiency in voxel selection does not directly translate to a small error in HRF estimation. After all, variable selection is *not* the same problem as parameter estimation. Extending the LRMGLM to accommodate subject-specific properties of brain activity will be the focus of the future research.

It took the proposed alternating convex search algorithm no more than 2 min to finish analyzing 106 subjects' fMRI data at almost 7000 voxels in dACC on a personal laptop using one i7 core. Roughly speaking, for a similar experimental design with four different stimulus types, the computation complexity of the proposed method is $O(n{\cdot}J^2)$. The alternating convex search algorithm is easily parallelizable and thus, can scale up to much larger data by using parallel computing.

Acknowledgments

T. Zhang's research is funded by the Enhancement Grant by the NSF CHARGE program at the University of Virginia and the Quantitative Collaborative seed grant program at the UVa. J. Coan's research was partially funded by a grant issued by the National Institute of Mental Health. The project described was supported by Award Number R01MH080725 to Coan. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIMH, the National Institutes of Health or SAMSI. We thank the editor and two reviewers for comments. Support from Casey Brown, Karen Hasselmo, Alexander Tatum, and Zoe Englander is also acknowledged.

Appendix A. Details of optimization step 1

In the first step of the proposed optimization procedure, we find the minimizer for $\mathscr{U}$ and $\tilde{\mathbf{d}}$ given fixed $\mathscr{V}$:

$$\min_{\mathscr{U},\tilde{\mathbf{d}}} \sum_{i=1}^{n} \| \mathbf{Y}^i - \mathbf{D}\,\boldsymbol{d}^i - \sum_{k=1}^{K} \mathbf{X}_k^i\, \overline{\mathbf{U}}_k\, \overline{\mathbf{V}}_k \|_F^2 + \lambda{\cdot} \sum_{k=1}^{K} \mathscr{P}(\overline{\mathbf{U}}_k). \tag{12}$$

The roughness penalty $\mathscr{P}(\overline{\mathbf{U}}_k)$ is given by

$$\mathscr{P}(\overline{\mathbf{U}}_k) = \sum_{p=1}^{P} \int_0^m \left(\sum_{l=1}^{L} \overline{U}_{k,lp}{\cdot} b_l^{(2)}(t) \right)^2 dt = \overline{U}_k^T R \overline{U}_k,$$

where R is an $L \times L$ matrix with element $R_{l1,l2} = \int_0^m b_{l1}^{(2)}(t) b_{l2}^{(2)}(t) dt$ for $l_1, l_2 = 1, \ldots, L$.

We have $\mathrm{vec}(\mathbf{X}_k^i \overline{\mathbf{U}}_k \overline{\mathbf{V}}_k) = \overline{\mathbf{V}}_k^T \otimes \mathbf{X}_k^i\ \mathrm{vec}(\overline{\mathbf{U}}_k)$, $\mathrm{vec}(\mathbf{D}\boldsymbol{d}^i) = \mathbb{I}_J \otimes \mathbf{D}\ \mathrm{vec}(\boldsymbol{d}^i)$, where $\mathrm{vec}(\cdot)$ stands for vectorization, $\otimes$ denotes the Kronecker product, and $\mathbb{I}_J$ is the identity matrix of size $J \times J$. Thus,

$$\sum_{k=1}^{K} \mathbf{X}_k^i \overline{\mathbf{U}}_k \overline{\mathbf{V}_k} = \begin{bmatrix} \overline{\mathbf{V}}_1^T \otimes \mathbf{X}_1^i & \cdots & \overline{\mathbf{V}}_K^T \otimes \mathbf{X}_K^i \end{bmatrix} \begin{bmatrix} \mathrm{vec}(\overline{\mathbf{U}}_1) \\ \vdots \\ \mathrm{vec}(\overline{\mathbf{U}}_K) \end{bmatrix}$$

Let $\mathscr{X}^i = \begin{bmatrix} \overline{\mathbf{V}}_1^T \otimes \mathbf{X}_1^i & \cdots & \overline{\mathbf{V}}_K^T \otimes \mathbf{X}_K^i \end{bmatrix}$, $\mathbb{R} = \begin{bmatrix} \mathbb{I}_P \otimes R & 0 & \ldots & 0 \\ 0 & \mathbb{I}_P \otimes R & 0 & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \mathbb{I}_P \otimes R \end{bmatrix}$, $\mathrm{vec}(\mathscr{U}) = \begin{bmatrix} \mathrm{vec}(\overline{\mathbf{U}}_1) \\ \vdots \\ \mathrm{vec}(\overline{\mathbf{U}}_K) \end{bmatrix}$, and $\mathscr{D} = \mathbb{I}_J \otimes D$. Thus, $\mathscr{X}^i \in \Re^{JT \times KPL}$, $\mathbb{R} \in \Re^{KPL \times KPL}$, and $\mathscr{D} \in \Re^{JT \times Jr}$, and (12) can be written in the following formulation

$$\min_{\mathscr{U},\tilde{\mathbf{d}}^i} \sum_{i=1}^{n} \left\| \mathbf{Y}^i - \mathbf{D}\,\boldsymbol{d}^i - \mathscr{X}^i\ \mathrm{vec}(\mathscr{U}) \right\|_F^2 + \lambda \mathrm{vec}(\mathscr{U})'\ \mathbb{R}\ \mathrm{vec}(\mathscr{U}). \tag{13}$$

The Hessian matrix of the above quadratic function is

$$Q = \begin{bmatrix} \sum_{i=1}^{n} (\mathscr{X}^i)^T \mathscr{X}^i + \lambda \mathbb{R} & (\mathscr{X}^1)^T \mathscr{D} & (\mathscr{X}^2)^T \mathscr{D} & \cdots & (\mathscr{X}^n)^T \mathscr{D} \\ \mathscr{D}^T(\mathscr{X}^1) & \mathscr{D}^T \mathscr{D} & 0 & \cdots & 0 \\ \mathscr{D}^T(\mathscr{X}^2) & 0 & \mathscr{D}^T \mathscr{D} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathscr{D}^T(\mathscr{X}^n) & 0 & 0 & \cdots & \mathscr{D}^T \mathscr{D} \end{bmatrix}.$$

Optimizing (13) is equivalent to solving the linear system

$$\begin{bmatrix} \sum_{i=1}^{n} (\mathscr{X}^i)^T \mathscr{X}^i + \lambda\mathbb{R} & (\mathscr{X}^1)^T \mathscr{D} & (\mathscr{X}^2)^T \mathscr{D} & \cdots & (\mathscr{X}^n)^T \mathscr{D} \\ \mathscr{D}^T(\mathscr{X}^1) & \mathscr{D}^T\mathscr{D} & 0 & \cdots & 0 \\ \mathscr{D}^T(\mathscr{X}^2) & 0 & \mathscr{D}^T\mathscr{D} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathscr{D}^T(\mathscr{X}^n) & 0 & 0 & \cdots & \mathscr{D}^T\mathscr{D} \end{bmatrix} \begin{bmatrix} \mathrm{vec}(\mathscr{U}) \\ \mathrm{vec}(\boldsymbol{d}^1) \\ \vdots \\ \mathrm{vec}(\boldsymbol{d}^n) \end{bmatrix} = \begin{bmatrix} f^{\mathscr{U}} \\ f^1 \\ \vdots \\ f^n \end{bmatrix},$$

where $f^{\mathscr{U}} = \sum_{i=1}^{n} (\mathscr{X}^i)^T vec(\mathbf{Y}^i)$ and $f^i = \mathscr{D}^T vec(\mathbf{Y}^i)$ for $i = 1, \ldots, n$. The system has special diagonal block structure, which can be solved by substitution. From the last n blocks of the linear system, we can solve for $\mathrm{vec}(\boldsymbol{d}^i)$ in terms of $\mathrm{vec}(\mathscr{U})$ and substitute them back to the first block. The resulting system only has $\mathrm{vec}(\mathscr{U})$ as unknown variables and can be solved efficiently using any linear system solver. In our implementation, we use conjugate gradient method in MATLAB.

Appendix B. Details of optimization step 2

Solution to Subproblem 1. We present the solution to the first subproblem of the ALPG algorithm. To ease notations and complexity, we remove the subscript and superscript of $\mathscr{G}_1$ and $\mathscr{G}_2$ to $\mathscr{G}$, and of $\mathscr{Z}_1$ and $\mathscr{Z}_2$ to $\mathscr{Z}$. Notice that $\overline{\mathbf{V}}_k, \overline{\mathbf{Z}}_k, \overline{\mathbf{G}}_k, \overline{\mathbf{V}}_k^* \in \Re^{P\times J}, k = 1, \ldots, K$; $\mathscr{Z}$, $\mathscr{V}$, $\mathscr{G}$, and $\mathscr{V}^*$ have the same dimension, i.e., $\mathscr{V} = \{\overline{\mathbf{V}}_k, k = 1, \ldots, K\}$, $\mathscr{Z} = \{\overline{\mathbf{Z}}_k, k = 1, \ldots, K\}$, $\mathscr{G} = \{\overline{\mathbf{G}}_k, k = 1, \ldots, K\}$, and $\mathscr{V}^* = \{\overline{\mathbf{V}}_k^*, k = 1, \ldots, K\}$. Denote $\overline{\mathbf{V}}_{j,k}$ to be the jth column of $\overline{\mathbf{V}}_k$, and $\overline{\mathbf{V}}_{pj,k}$ to be the (p,j)th element of $\overline{\mathbf{V}}_k$.

The first subproblem is to solve

$$\mathscr{Z} = \arg\min_{\mathscr{V}} \left\langle \mathscr{G}, \mathscr{V} \right\rangle + \mu \sum_{j=1}^{J} \sqrt{\sum_{p=1}^{P} \left(\overline{\mathbf{V}}_{pj,1} - \overline{\mathbf{V}}_{pj,2}\right)^2} + \frac{\rho}{2} \| \mathscr{V} - \mathscr{V}\|_2^{*2} \tag{14}$$

The non-differentiable term involves $\overline{\mathbf{V}}_1$ and $\overline{\mathbf{V}}_2$ only, parts of $\mathscr{V}$ corresponding to stimuli 1 and 2, so the solution of the problem for $k \neq \{1, 2\}$ is

$$\overline{\mathbf{Z}}_k = \frac{\rho\overline{\mathbf{V}}_k^* - \overline{\mathbf{G}}_k}{\rho}.$$

The optimization problem in (14) for $k = 1, 2$ is divided into J smaller problems, each corresponding to one voxel j in the following form

$$\min_{\overline{\mathbf{V}}_{j,1}, \overline{\mathbf{V}}_{j,2}} \left\langle \overline{\mathbf{G}}_{j,1}, \overline{\mathbf{V}}_{j,1} \right\rangle + \left\langle \overline{\mathbf{G}}_{j,2}, \overline{\mathbf{V}}_{j,2} \right\rangle + \mu \left\| \overline{\mathbf{V}}_{j,1} - \overline{\mathbf{V}}_{j,2} \right\|_2 + \frac{\rho}{2} \left\| \overline{\mathbf{V}}_{j,1} - \overline{\mathbf{V}}_{j,1}^* \right\|_2^2 + \frac{\rho}{2} \left\| \overline{\mathbf{V}}_{j,2} - \overline{\mathbf{V}}_{j,2}^* \right\|_2^2,$$

where $\| \cdot \|_2$ denotes the Euclidean norm. We introduce a slack vector $S = \overline{\mathbf{V}}_{j,1} - \overline{\mathbf{V}}_{j,2} \in \Re^{P\times 1}$, and the problem above is equivalent to

$$\min_{\overline{\mathbf{V}}_{j,1}, \overline{\mathbf{V}}_{j,2}} \left\langle \overline{\mathbf{G}}_{j,1}, \overline{\mathbf{V}}_{j,1} \right\rangle + \left\langle \overline{\mathbf{G}}_{j,2}, \overline{\mathbf{V}}_{j,2} \right\rangle + \mu \|S\|_2 + \frac{\rho}{2} \left\| \overline{\mathbf{V}}_{j,1} - \overline{\mathbf{V}}_{j,1}^* \right\|_2^2 + \frac{\rho}{2} \left\| \overline{\mathbf{V}}_{j,2} - \overline{\mathbf{V}}_{j,2}^* \right\|_2^2,$$

$$\text{such that } S = \overline{\mathbf{V}}_{j,1} - \overline{\mathbf{V}}_{j,2}.$$

Let $\tilde{\alpha} \in \Re^{P\times 1}$ be the vector of dual variables corresponding to the constraint $S = \overline{\mathbf{V}}_{j,1} - \overline{\mathbf{V}}_{j,2}$. The Lagrangian function of the above optimization problem is

$$\mathscr{L}(\overline{\mathbf{V}}_{j,1}, \overline{\mathbf{V}}_{j,2}, S, \tilde{\alpha}) = \left\langle \overline{\mathbf{G}}_{j,1}, \overline{\mathbf{V}}_{j,1} \right\rangle + \left\langle \overline{\mathbf{G}}_{j,2}, \overline{\mathbf{V}}_{j,2} \right\rangle + \mu \|S\|_2 + \frac{\rho}{2} \left\| \overline{\mathbf{V}}_{j,1} - \overline{\mathbf{V}}_{j,1}^* \right\|_2^2 \tag{15}$$

$$+ \frac{\rho}{2} \left\| \overline{\mathbf{V}}_{j,2} - \overline{\mathbf{V}}_{j,2}^* \right\|_2^2 - \tilde{\alpha}^T \left(S - \left(\overline{\mathbf{V}}_{j,1} - \overline{\mathbf{V}}_{j,2}\right)\right).$$

We see that the minimum of the Lagrangian with respect to S is finite if and only if $\|\tilde{\alpha}\|_2 \leq \mu$ (Lin et al., 2014). Then the minimum value of the S-term is 0 and we can eliminate them to get reduced form of Lagrangian:

$$\widehat{\mathscr{L}}(\overline{\mathbf{V}}_{j,1}, \overline{\mathbf{V}}_{j,2}, \tilde{\alpha}) = \left\langle \overline{\mathbf{G}}_{j,1}, \overline{\mathbf{V}}_{j,1} \right\rangle + \left\langle \overline{\mathbf{G}}_{j,2}, \overline{\mathbf{V}}_{j,2} \right\rangle + \frac{\rho}{2} \left\| \overline{\mathbf{V}}_{j,1} - \overline{\mathbf{V}}_{j,1}^* \right\|_2^2 + \frac{\rho}{2} \left\| \overline{\mathbf{V}}_{j,2} - \overline{\mathbf{V}}_{j,2}^* \right\|_2^2 + \tilde{\alpha}^T \left(\overline{\mathbf{V}}_{j,1} - \overline{\mathbf{V}}_{j,2}\right).$$

To find the dual function, we minimize $\widehat{\mathscr{L}}(\overline{\mathbf{V}}_{j,1}, \overline{\mathbf{V}}_{j,2}, \tilde{\alpha})$ over $\overline{\mathbf{V}}_{j,1}, \overline{\mathbf{V}}_{j,2}$:

$$\overline{\mathbf{V}}_{j,1} = \frac{-\tilde{\alpha} - \overline{\mathbf{G}}_{j,1} + \rho\overline{\mathbf{V}}_{j,1}^*}{\rho} \quad \text{and} \quad \overline{\mathbf{V}}_{j,2} = \frac{\tilde{\alpha} - \overline{\mathbf{G}}_{j,2} + \rho\overline{\mathbf{V}}_{j,2}^*}{\rho}. \tag{16}$$

Substitute these back to (15), we obtain the dual problem

$$\min_{\tilde{\alpha}} \frac{1}{2}\left\langle \tilde{\alpha}, \tilde{\alpha} \right\rangle + \left\langle \tilde{\alpha}, \frac{\left(\overline{\mathbf{G}}_{j,1} - \overline{\mathbf{G}}_{j,2}\right) - \rho\left(\overline{\mathbf{V}}^*_{j,1} - \overline{\mathbf{V}}^*_{j,2}\right)}{2} \right\rangle, \text{ such that } \|\tilde{\alpha}\|_2 \le \mu.$$

Let $\tilde{h} = ((\overline{\mathbf{G}}_{j,1} - \overline{\mathbf{G}}_{j,2}) - \rho(\overline{\mathbf{V}}^*_{j,1} - \overline{\mathbf{V}}^*_{j,2}))/2$. The above problem has closed form solution: if $\left\|\tilde{h}\right\|_2 < \mu$ then $\tilde{\alpha}^* = -\tilde{h}$, otherwise $\tilde{\alpha}^* = -\mu \cdot \frac{\tilde{h}}{\|\tilde{h}\|_2}$. Solution of $\overline{\mathbf{V}}_{j,1}$ and $\overline{\mathbf{V}}_{j,2}$ can be calculated by plugging $\tilde{\alpha}^*$ into (16).

Solution to Subproblem 2. The second subproblem involves solving

$$\min_{\mathscr{V}} \sum_{i=1}^{n} \| \mathbf{Y}^i - \mathbf{D}\, \boldsymbol{d}^i - \sum_{k=1}^{K} \mathbf{X}^i_k\, \overline{\mathbf{U}}_k\, \overline{\mathbf{V}}_k \|_F^2 + \tau \sum_{k=1}^{K} \mathscr{R}(\overline{\mathbf{V}}_k) + \left\langle \mathscr{G}_2, \mathscr{V} \right\rangle + \frac{\rho}{2} \| \mathscr{V} - \mathscr{V}^* \|_2^2. \tag{17}$$

To ease notations, we let $\mathbf{Y}^i - \mathbf{D}\, \boldsymbol{d}^i \rightarrow \mathbf{Y}^i$ and $\mathbf{X}^i_k\, \overline{\mathbf{U}}_k \rightarrow \mathbf{X}^i_k$. The spatial penalty can be written in matrix form, $\mathscr{R}(\overline{\mathbf{V}}_k) = \left\|\overline{\mathbf{V}}_k C\right\|_F^2$.The problem becomes

$$\min_{\mathscr{V}} \sum_{i=1}^{n} \| \mathbf{Y}^i - \sum_{k=1}^{K} \mathbf{X}^i_k\, \overline{\mathbf{V}}_k \|_F^2 + \tau \sum_{k=1}^{K} \| \overline{\mathbf{V}}_k C \|_F^2 + \left\langle \mathscr{G}_2, \mathscr{V} \right\rangle + \frac{\rho}{2} \| \mathscr{V} - \mathscr{V}^* \|_2^2. \tag{18}$$

This problem is a quadratic programming with positive definite Hessian. To solve this problem efficiently, we use the conjugate gradient method implemented in MATLAB.

Appendix C. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.neuroimage.2017.12.032.

References

Agarwal, Alekh, Negahban, Sahand, Wainwright, Martin J., 2012. Noisy matrix decomposition via convex relaxation: optimal rates in high dimensions. Ann. Stat. 40 (2), 1171–1197. https://doi.org/10.1214/12-AOS1000, 04.

Aguirre, Geoffrey Karl, Zarahn, E., D'esposito, M., 1998. The variability of human, bold hemodynamic responses. Neuroimage 8 (4), 360–369.

Allen, Joseph P., Porter, Maryfrances, McFarland, Christy, McElhaney, Kathleen Boykin, Marsh, Penny, 2007. The relation of attachment security to adolescents paternal and peer relationships, depression, and externalizing behavior. Child Dev. 78 (4), 1222–1239.

Allen, Genevera I., Grosenick, Logan, Taylor, Jonathan, 2014. A generalized least-square matrix decomposition. J. Am. Stat. Assoc. 109 (505), 145–159.

Arbabshirani, M.R., Havlicek, M., Kiehl, K.A., Pearlson, G.D., Calhoun, V.D., 2013. Functional network connectivity during rest and task conditions: a comparative study. Hum. Brain Mapp. 34 (11), 2959–2971.

Aubry, Augusto, De Maio, Antonio, Pallotta, Luca, Farina, Alfonso, 2012. Maximum likelihood estimation of a structured covariance matrix with a condition number constraint. IEEE Trans. Signal Process. 60 (6), 3004–3021.

Bai, J., Li, K., 2002. Forecasting using principal components from a large number of predictors. J. Am. Stat. Assoc. 97 (460), 1167–1179.

Bai, J., Li, K., 2012. Statistical analysis of factor models of high dimension. Ann. Stat. 40 (1), 436–465.

Balqis Samdin, S., Ting, Chee-Ming, Ombao, Hernando, Salleh, Sh-Hussain, 2017. A unified estimation framework for state-related changes in effective brain connectivity. IEEE Trans. Biomed. Eng. 64 (4), 844–858.

Basu, S., Dunagan, J., Duh, K., Muniswamy-Reddy, K.K., 2012. Blr-d: applying bilinear logistic regression to factored diagnosis problems. ACM SIGOPS - Oper. Syst. Rev. 45, 31–38.

Beckmann, Christian F., Smith, Stephen M., 2005. Tensorial extensions of independent component analysis for multisubject fmri analysis. Neuroimage 25 (1), 294–311.

Boor, Carl de, 2001. A Practical Guide to Splines (Applied Mathematical Sciences, vol. 27.

Brosch, J., Talavage, T., Ulmer, J., Nyenhuis, J., 2002. Simulation of human respiration in fMRI with a mechanical model. IEEE Trans. Biomed. Eng. 49, 700–707.

Bush, George, Vogt, Brent A., Holmes, Jennifer, Dale, Anders M., Greve, Douglas, Jenike, Michael A., Rosen, Bruce R., 2002. Dorsal anterior cingulate cortex: a role in reward-based decision making. Proc. Natl. Acad. Sci. Unit. States Am. 99 (1), 523–528.

Buxton, Richard B., Frank, Lawrence R., 1997. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. J. Cerebr. Blood Flow Metabol. 17 (1), 64–72.

Calhoun, V.D., Adali, T., 2006. Unmixing fmri with independent component analysis. IEEE Eng. Med. Biol. Mag. 25 (2), 79–90.

Calhoun, Vince D., Adali, Tulay, Pearlson, Godfrey D., Pekar, J.J., 2001. A method for making group inferences from functional mri data using independent component analysis. Hum. Brain Mapp. 14 (3), 140–151.

Calhoun, V.D., Stevens, M.C., Pearlson, G.D., Kiehl, K.A., 2004. Fmri analysis with the general linear model: removal of latency-induced amplitude bias by incorporation of hemodynamic derivative terms. Neuroimage 22, 252–257.

Casanova, R., Ryali, S., Serences, J., Yang, L., Kraft, R., Laurienti, P.J., Maldjian, J.A., 2008. The impact of temporal regularization on estimates of the BOLD hemodynamic response function: a comparative analysis. Neuroimage 40 (4), 1606–1618.

Casanova, R., Yang, L., Hairston, W.D., Laurienti, P.J., Maldjian, J.A., 2009. Evaluating the impact of spatio-temporal smoothness constraints on the BOLD hemodynamic response function estimation: an analysis based on Tikhonov regularization. Physiol. Meas. 30 (5), 37–51.

Chaari, L., Forbes, F., Vincent, T., Ciuciu, P., 2012. Adaptive hemodynamic-informed parcellation of fmri data in a variational joint detection estimation framework. In: 15th Proceedings MICCAI, LNCS, 7512, pp. 180–188.

Coan, J.A., 2010. Adult attachment and the brain. J. Soc. Pers. Relat. 27, 210–217.

Coan, J.A., 2011. The social regulation of emotion. In: Oxford Handbook of Social Neuroscience. Oxford University Press, New York, pp. 614–623.

Coan, J.A., Schaefer, H.S., Davidson, R.J., 2006. Lending a hand: social regulation of the neural response to threat. Psychol. Sci. 17, 1032–1039.

Coan, J.A., Beckes, L., Allen, J.P., 2013. Childhood maternal support and neighborhood quality moderate the social regulation of neural threat responding in adulthood. Int. J. Psychophysiol. 88, 224–231.

Dale, A., 1999. Optimal experimental design for event-related fMRI. Hum. Brain Mapp. 8, 109–114.

Dedovic, Katarina, Slavich, George M., Muscatell, Keely A., Irwin, Michael R., Eisenberger, Naomi I., 2016. Dorsal anterior cingulate cortex responses to repeated social evaluative feedback in young women with and without a history of depression. Front. Behav. Neurosci. 10.

Degras, D., Lindquist, M.A., 2014. A hierarchical model for simultaneous detection and estimation in multi-subject fmri studies. Neuroimage 98, 61–72.

D'arcy, Ryan CN, Hamilton, Andrew, Jarmasz, Mark, Sullivan, Sara, Stroink, Gerhard, 2006. Exploratory data analysis reveals visuovisual interhemispheric transfer in functional magnetic resonance imaging. Magn. Reson. Med. 55 (4), 952–958,.

Etkin, Amit, Egner, Tobias, Kalisch, Raffael, 2011. Emotional processing in anterior cingulate and medial prefrontal cortex. Trends Cognit. Sci. 15 (2), 85–93.

Eubank, Randall L., 1999. Nonparametric Regression and Spline Smoothing. CRC press.

Fabri, Mara, Polonara, Gabriele, Mascioli, Giulia, Salvolini, Ugo, Manzoni, Tullio, 2011. Topographical organization of human corpus callosum: an fmri mapping study. Brain Res. 1370, 99–111.

Feng, J., Xu, H., Yan, S., 2013. Online robust pca via stochastic optimization. NIPS 404–412.

Flandin, G., Kherif, F., Pennec, X., Riviere, D., Ayache, N., Poline, J.-B., 2002. Parcellation of brain images with anatomical and functional constraints for fmri data analysis. In: Biomedical Imaging, 2002. Proceedings. 2002 IEEE International Symposium on, pp. 907–910. https://doi.org/10.1109/ISBI.2002.1029408.

Fraser, L.M., Stevens, M.T., Steven, D.B., D'Arch, R.C.N., 2012. White versus gray matter: fmri hemodynamic responses show similar characteristics, but differ in peak amplitude. BMC Neurosci. 13, 91.

Friman, O., Borga, M., Lundberg, P., Knutssonb, H., 2004. Detection and detrending in fmri data analysis. Neuroimage 22, 645–655.

Friston, Karl J., Holmes, Andrew P., Worsley, Keith J., Poline, J.-P., Frith, Chris D., Frackowiak, Richard SJ., 1994a. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2 (4), 189–210.

Friston, K.J., Jezzard, P., Turner, R., 1994b. Analysis of functional MRI time-series. Hum. Brain Mapp. 1, 153–171.

Friston, K.J., Holmes, A., Worsley, K.J., Poline, J.-P., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical parametric mapping: a general linear approach. Hum. Brain Mapp. 2, 189–210.
Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., Turner, R., 1998. Event-related fMRI: characterizing differential responses. Neuroimage 7, 30–40.
Friston, K.J., Josephs, O., Zarahn, E., Holmes, A.P., Rouquette, S., Poline, J.B., 2000. To smooth or not to smooth? Neuroimage 12, 196–208.
Gabriel, K.R., 1998. Generalised bilinear regression. Biometrika 85, 689–700.
Gawryluk, Jodie R., D'arcy, Ryan CN., Mazerolle, Erin L., Brewer, Kimberly D., Beyea, Steven D., 2011. Functional mapping in the corpus callosum: a 4t fmri study of white matter. Neuroimage 54 (1), 10–15.
Gawryluk, Jodie R., Mazerolle, Erin L., D'Arcy, Ryan CN., 2014. Does functional mri detect activation in white matter? a review of emerging evidence, issues, and future directions. Front. Neurosci. 8.
Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage 9, 416–429.
Gorski, K., Pfeuffer, F., Klamroth, K., 2007. Biconvex sets and optimization with biconvex functions - a survey and extensions. Math. Meth. Oper. Res. 66 (3), 373–407.
Goutte, C., Nielsen, F.A., Hansen, L.K., 2000. Modeling the hemodynamic response in fMRI using smooth FIR filters. IEEE Trans. Med. Imag. 19, 1188–1201.
Guo, Ying, 2008. Group independent component analysis of multi-subject fmri data: connections and distinctions between two methods. In: BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on, vol. 2, pp. 748–752.
Guo, Ying, Tang, Li, 2013. A hierarchical model for probabilistic independent component analysis of multi-subject fmri studies. Biometrics 69 (4), 970–981.
Handwerker, D.A., Ollinger, J.M., D'Esposito, M., 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. Neuroimage 21, 1639–1651.
Hoff, P.D., 2015. Multilinear tensor regression for longitudinal relational data. Ann. Appl. Stat. 9, 1169–1193.
Hong, David S., Bray, Signe, Haas, Brian W., Hoeft, Fumiko, Reiss, Allan L., 2014. Aberrant neurocognitive processing of fear in young girls with turner syndrome. Soc. Cognit. Affect Neurosci. 9 (3), 255–264.
Jenkinson, Mark, Bannister, Peter, Brady, Michael, Smith, Stephen, 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17 (2), 825–841.
Kiwiel, K., Rosa, C., Ruszczyński, A., 2006. Proximal decomposition via alternating linearization. SIAM J. Optim. 9 (3), 668–689.
Knutson, Brian, Westdorp, Andrew, Kaiser, Erica, Hommer, Daniel, 2000. Fmri visualization of brain activity during a monetary incentive delay task. Neuroimage 12 (1), 20–27.
Koren, Y., Bell, R., Volinsky, C., 2009. Matrix factorization techniques for recommender systems. Computer 8, 30–37.
Lange, N., Strother, S.C., Anderson, J.R., Nielsen, F.A., Holmes, A.P., Kolenda, T., Savoy, R., Hansen, L.K., 1999. Plurality and resemblance in fMRI data analysis. Neuroimage 10, 282–303.
Li, X., Zhou, H., Li, L., 2013. Tucker tensor regression and neuroimaging analysis arXiv: 1404.4104.
Liao, C.H., Worsley, K.J., Poline, J.B., Aston, J.A.D., Duncan, G.H., Evans, A.C., 2002. Estimating the delay of the fmri response. Neuroimage 16, 593–606.
Lin, X., Pham, M., Ruszczyński, A., 2014. Alternating linearization for structured regularization problems. J. Mach. Learn. Res. 15 (1), 3447–3481.
Lindquist, M.A., 2008. The statistical analysis of fMRI data. Stat. Sci. 23, 439–464.
Lindquist, M.A., Wager, T.D., 2007. Validity and power in hemodynamic response modelling: a comparison study and a new approach. Hum. Brain Mapp. 28, 764–784.
Luo and S Puthusserypady, 2008. Analysis of fMRI datawith drift: modified general linear model and Bayesian estimator. IEEE Trans. Biomed. Eng. 55, 1504–1511.
Makni, S., Ciuciu, P., Idier, J., Poline, J.B., 2005. Joint detection-estimation of brain activity in functional mri: a multichannel deconvolution solution. IEEE Trans. Signal Process. 53 (9), 3488–3502.
Makni, S., Idier, J., Vincent, T., Thirion, B., Dehaene-Lambertz, G., Ciuciu, P., 2008. A fully bayesian approach to the parcel-based detection-estimation of brain activity in fmri. Neuroimage 41 (3), 941–969.
Marrelec, G., Benali, H., Ciuciu, P., Poline, J.B., 2001. Bayesian estimation of the hemodynamic of the hemodynamic response function in functional MRI. In: AIP Conference Proceedings, 617, pp. 229–247.
Marrelec, G., Benali, H., Ciuciu, P., Pelegrini-Issac, M., Poline, J.B., 2003. Robust estimation of the hemodynamic response function in event-related BOLD fMRI using basic physiological information. Hum. Brain Mapp. 19, 1–17.
Mazard, A., Mazoyer, B., Etard, O., Tzourio-Mazoyer, N., Kosslyn, S.M., Mellet, Emmanuel, 2002. Impact of fmri acoustic noise on the functional anatomy of visual mental imagery. J. Cognit. Neurosci. 14 (2), 172–186.
Mazerolle, Erin L., D'Arcy, Ryan CN., Beyea, Steven D., 2008. Detecting functional magnetic resonance imaging activation in white matter: interhemispheric transfer across the corpus callosum. BMC Neurosci. 9 (1), 84.
Milad, Mohammed R., Quirk, Gregory J., Pitman, Roger K., Orr, Scott P., Fischl, Bruce, Rauch, Scott L., 2007. A role for the human dorsal anterior cingulate cortex in fear expression. Biol. Psychiatr. 62 (10), 1191–1194.
Morris, John S., Friston, Karl J., Büchel, C., Frith, Christopher D., Young, Andrew W., Calder, Andrew J., Dolan, Raymond J., 1998. A neuromodulatory role for the human amygdala in processing emotional facial expressions. Brain 121 (1), 47–57.
Potthoff, R.F., Roy, S.N., 1964. A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika 51, 313–326.
Reiss, P.T., Ogden, R.T., 2007. Functional principal component regression and functional partial least squares. J. Am. Stat. Assoc. 102, 984–996.
Reiss, P.T., Ogden, R.T., 2009. Smoothing parameter selection for a class of semiparametric linear models. J. Roy. Stat. Soc. B 71, 505–523.
Rendle, S., 2012. Factorization machines with libfm. ACM Transactions on Intelligent Systems and Technology (TIST) 3, 3–57.
Robinson, George K., 1991. That blup is a good thing: the estimation of random effects. Stat. Sci. 15–32.
Ruppert, David, Wand, Matt P., Carroll, Raymond J., 2003. Semiparametric Regression. Cambridge university press. Number 12.
Shi, Ran, Guo, Ying, 2014. Modeling covariate effects in group independent component analysis with applications to functional magnetic resonance imaging. arXiv preprint arXiv:1402.4239.
Shi, J.V., Xu, Y., Baraniuk, R.G., 2014. Sparse Bilinear Logistic Regression. Preprint.
Smith, Stephen M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17 (3), 143–155.
Smith, A., Lewis, B., Ruttinmann, U., et al., 1999. Investigation of low frequency drift in fMRI signal. Neuroimage 9, 526–533.
Srivastava, M.S., von Rosen, T., von Rosen, D., 2009. Estimation and testing in general multivariate linear models with kronecker product covariance structure. Sankhya 71, 137–163.
Schulz, Kurt P., Bédard, Anne-Claude V., Czarnecki, Rosa, Fan, Jin, 2011. Preparatory activity and connectivity in dorsal anterior cingulate cortex for cognitive control. Neuroimage 57 (1), 242–250.
Szameitat, André J., Shen, Shan, Sterr, Annette, 2009. The functional magnetic resonance imaging (fmri) procedure as experienced by healthy participants and stroke patients–a pilot study. BMC Med. Imag. 9 (1), 14.
Thirion, B., Faugeras, O., 2003. Feature detection in fmri data: the information bottleneck approach. Med. Image Anal. 8.
Ting, Ombao, Hernando, Salleh, Sh-Hussain, 2017. Multi-scale factor analysis of high-dimensional brain signals. arXiv preprint arXiv:1705.06502.
Vakorin, V.A., Borowsky, R., Sarty, G.E., 2007. Characterizing the functional MRI response using Tikhonov regularization. Stat. Med. 26 (21), 3830–3844.
Vincent, T., Risser, L., Ciuciu, P., 2010. Spatially adaptive mixture modeling for analysis of fmri time series. IEEE Trans. Med. Imag. 29 (4), 1059–1074.
Wahba, G., 1990. Spline Models for Observational Data. SIAM, Philadelphia.
Wang, Yuxiao, Ting, Chee-Ming, Ombao, Hernando, 2016. Modeling effective connectivity in high-dimensional cortical source signals. IEEE J. Sel. Top. Signal Process. 10 (7), 1315–1325.
Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. J. Roy. Stat. Soc. B 73, 3–36.
Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited again. Neuroimage 2, 173–181.
Worsley, K.J., Liao, C.H., Aston, J., Petre, V., Duncan, G.H., Morales, F., Evans, A., 2002. A general statistical analysis for fMRI data. Neuroimage 15, 1–15.
Xu, Lei, Johnson, Timothy D., Nichols, Thomas E., Nee, Derek E., 2009. Modeling inter-subject variability in fmri activation location: a bayesian hierarchical spatial model. Biometrics 65 (4), 1041–1051.
Yarkoni, T., Barch, D.M., Gray, J.R., Conturo, T.E., Braver, T.S., 2009. Bold correlates of trial-by-trial reaction time variability in gray and white matter: a multistudy fmri analysis. PLos One 4, e4527.
Zhang, C.M., Jiang, Y., Yu, T., 2007. A comparative study of one-level and two-level semiparametric estimation of hemodynamic response function for fMRI data. Stat. Med. 26, 3845–3861.
Zhang, T., Li, F., Beckes, L., Coan, J.A., 2013. A semi-parametric model of the hemodynamic response for multi-subject fmri data. Neuroimage 75, 136–145.
Zhang, Linlin, Guindani, Michele, Versace, Francesco, Vannucci, Marina, 2014a. A spatio-temporal nonparametric bayesian variable selection model of fmri data for clustering correlated time courses. Neuroimage 95, 162–175.
Zhang, T., Li, F., Gonzalez, M.Z., Maresh, E.L., Coan, J.A., 2014b. A semi-parametric nonlinear model for event-related fmri. Neuroimage 97, 178–187.
Zhang, Linlin, Guindani, Michele, Versace, Francesco, Engelmann, Jeffrey M., Vannucci, Marina, et al., 2016. A spatiotemporal nonparametric bayesian model of multi-subject fmri data. Ann. Appl. Stat. 10 (2), 638–666.
Zhou, H., Li, L., Zhu, H., 2013. Tensor regression with applications in neuroimaging data analysis. J. Am. Stat. Assoc. 108 (502), 540–552.