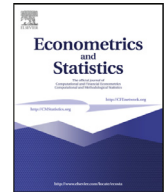Contents lists available at ScienceDirect

Econometrics and Statistics

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

Spatial-Temporal Analysis of Multi-Subject Functional Magnetic Resonance Imaging Data

Tingting Zhang [a,2,1,*], Minh Pham [b,1], Guofen Yan [c], Yaotian Wang [a], Sara Medina-DeVilliers [d], James A. Coan [d]

[a] Department of Statistics, University of Pittsburgh, United States
[b] Department of Decision Sciences, San Francisco State University, United States
[c] Department of Public Health Sciences, University of Virginia, United States
[d] Department of Psychology, University of Virginia, United States

ARTICLE INFO

ABSTRACT

Functional magnetic resonance imaging (fMRI) is one of the most popular neuroimaging technologies used in human brain studies. However, fMRI data analysis faces several challenges, including intensive computation due to the massive data size and large estimation errors due to a low signal-to-noise ratio of the data. A new statistical model and a computational algorithm are proposed to address these challenges. Specifically, a new multi-subject general linear model is built for stimulus-evoked fMRI data. The new model assumes that brain responses to stimuli at different brain regions of various subjects fall into a low-rank structure and can be represented by a few principal functions. Therefore, the new model enables combining data information across subjects and regions to evaluate subject-specific and region-specific brain activity. Two optimization functions and a new fast-to-compute algorithm are developed to analyze multi-subject stimulus-evoked fMRI data and address two research questions of a broad interest in psychology: evaluating every subject's brain responses to different stimuli and identifying brain regions responsive to the stimuli. Both simulation and real data analysis are conducted to show that the new method can outperform existing methods by providing more efficient estimates of brain activity.

© 2021 EcoSta Econometrics and Statistics. Published by Elsevier B.V. All rights reserved.

1. Introduction

Functional magnetic resonance imaging (fMRI) measures human brain activity non-invasively through blood-oxygen-level-dependent contrast (Ogawa et al., 1992) and records activity of the entire brain with a high spatial resolution. Therefore, fMRI is one of the most popular neuroimaging technologies used in human brain research. In many psychological and medical studies, researchers use different stimuli to evoke subjects' brain activities during fMRI recordings. The purposes of this paper are to analyze these multi-subject, stimulus-evoked fMRI data, evaluate each subject's brain responses to different stimuli, and identify voxels that have different responses to different stimuli.

---

* Corresponding author.
 E-mail addresses: tiz67@pitt.edu (T. Zhang), jcoan@virginia.edu (J.A. Coan).
[1] Equally contributing first authors.

ARTICLE IN PRESS

JID: ECOSTA [m3Gsc;March 25, 2021;7:38]

T. Zhang, M. Pham, G. Yan et al. Econometrics and Statistics xxx (xxxx) xxx

The analysis of stimulus-evoked fMRI data usually uses a general linear model (GLM) (Friston et al., 1995; Goutte et al., 2000; Worsley and Friston, 1995), in which the fMRI time series follows a linear regression of the stimulus sequence through hemodynamic response functions (HRF). Under the GLM, the HRF describes each subject's brain response to each stimulus type at each region and, therefore, is the estimation target of fMRI data analysis.

Several challenges are present when analyzing multi-subject, stimulus-evoked fMRI data. First, the data are massive in size. Typical fMRI data of one subject from one experimental session consist of 100,000 to 200,000 spatially indexed time series. Each time series contains measurements of brain activity at one brain voxel and at several hundred time points—with a time unit ranging from 1 second(s) to 2 s. A voxel is a small 3D cubic volume in the brain (2 cm×2 cm×2 cm for the data under study). Consequently, fMRI data analysis can be computationally intensive. Second, the HRF varies across regions, subjects, and stimulus types presented to the subject. Third, fMRI data have complex spatial and temporal properties. Fourth, each subject's fMRI data usually have a weak signal-to-noise ratio (SNR) (Lindquist, 2008), which brings an additional difficulty to accurate HRF estimation.

Most existing approaches to estimating HRFs attend to one or part of the challenges described above. For example, to address the first two challenges, researchers commonly use voxel-wise analysis (Friston et al., 1995; Goutte et al., 2000; Worsley and Friston, 1995), that is, analyze fMRI time series at each voxel independently. Although this approach is computationally fast and accommodates variations of brain activity across voxels, the approach ignores spatial information of fMRI data and association among voxels. Despite the development of joint models for brain activities at different locations Zhang et al. (2014a, 2016a); Bowman et al. (2008); Degras and Lindquist (2014a); Makni et al. (2005, 2008); Pedregosa et al. (2015), these approaches tend to use many parameters to accommodate the variation of brain responses across many voxels and subjects and require strong regularization or strong priors to reduce estimation errors. Moreover, existing methods are mainly focused on examining brain responses to each stimulus, which limits the ability to compare brain responses to different stimuli, as explained below.

Many psychological fMRI studies use the brain response to one stimulus as the reference and compare the brain responses to other stimuli with respect to the reference. This approach is to correct for the nuisance brain response caused by the fMRI data collection process, for example, the brain response due to anxiety and uncomfortableness experienced during fMRI scanning (Mazard et al., 2002; Szameitat et al., 2009). In this situation, the brain response to the stimulus of interest beyond the response to the reference stimulus should represent more accurately the brain activity due to the stimulus of interest in reality. Existing methods for comparing brain responses are still through a two-step approach: estimating responses to different stimuli first and then performing tests on the response estimates through voxel-wise analysis. However, this two-step approach leads to many false discoveries (Zhang et al., 2018).

We develop a new joint model for multi-subject, stimulus-evoked fMRI data to address limitations in existing fMRI data analysis methods. With an assumption of a low-rank structure for HRFs of different voxels and subjects, the new model, called multi-subject, low-rank general linear model (MSLRGLM), uses much fewer parameters than nonparametric models to accommodate the variation of brain responses across different subjects and voxels. In addition to model flexibility, the new model also enjoys a better estimation efficiency of HRFs than existing methods.

We formulate the MSLRGLM estimation as an optimization problem. Specifically, we propose two optimization functions to address two research questions: estimating subject-specific and voxel-specific HRFs and identifying voxels with different responses to different stimuli. We use regularization penalties in the optimization functions to incorporate spatial and temporal information of fMRI data. We include a sparsity-inducing penalty in the optimization function to identify voxels that have different responses to different stimuli. We use the sparsity to produce stable and interpretable brain response estimates, similar to many existing approaches (Baldassarre et al., 2012; Ge et al., 2016), because sparse neural responses have been reported in the literature (Olshausen and Field, 1996; 2004; Beloozerova et al., 2003; Wixted et al., 2014). To address the computational challenge in estimating HRFs of many subjects and voxels, we develop a new computational algorithm to minimize the optimization functions. We show that the proposed new model and computational algorithm provide more efficient HRF estimates with smaller variances than existing methods through both simulation and real data analysis. In summary, our new modeling and estimating approaches address the challenges in fMRI data analysis.

The rest of the article is organized as follows. Section 2 reviews existing HRF estimation methods under the GLM framework and introduces a new multi-subject GLM for fMRI data of different subjects and voxels. We present our estimation approach with new optimization functions in Section 3 and propose a new computational algorithm to minimize the functions in Section 4. The proposed HRF estimation method is compared with several existing ones through simulation studies in Section 5. Section 6 presents analysis results of multi-subject fMRI data collected in a psychological study on emotion. We evaluate each subject's brain responses to emotional stimuli and identify voxels responsive to the stimuli using the proposed method. Section 7 concludes.

2. Multi-Subject General Linear Model

2.1. The General Linear Model

Let $y^i_v(t)$ be the fMRI measurement of brain activity at voxel v of subject i at time t, $t = \kappa,\ 2\kappa, \ldots, T\kappa$, $i = 1, 2, \cdots, n$, and $v = 1, 2, \cdots, V$, where T is the total number of time points of fMRI time series, V is the total number of voxels under study, and κ is the time interval between consecutive 3D images. For the fMRI data under analysis, κ equals 2 s.

Suppose K different stimuli are presented in the fMRI experiment. Let $s_k^i(t)$ be the stimulus function that indicates the evoked times of the kth stimulus in the fMRI experiment session for subject i. That is, if the kth stimulus is evoked at time t, $s_k^i(t) = 1$, and 0 otherwise. The GLM assumes the fMRI time series of one subject i in voxel v follows a linear regression of K stimulus functions:

$$y_v^i(t) = C(t)\ d_v^i + \sum_{k=1}^{K} \int_0^m h_{v,k}^i(u) \cdot s_k^i(t-u)du + \varepsilon_v^i(t), \tag{1}$$

where $h_{v,k}^i(u)$ is the hemodynamic response function (HRF) of subject i and voxel v in response to the kth stimulus, $[0, m]$ is the given domain for the HRF, and $\varepsilon_v^i(t)$ is an error term. The HRF $h_{v,k}^i(t)$ describes subject i's brain response to the kth stimulus at voxel v.

We let $C(t) = (\sqrt{2/T}\cos(\pi\ \frac{t}{T}), \ldots, \sqrt{2/T}\cos(r\ \pi\ \frac{t}{T}))$ and $d_v^i \in \Re^r$ be an r-dimensional coefficient vector for $C(t)$. The values of d_v^i are estimated based on data. We use $C(t)\ d_v^i$ to characterize a low frequency drift in fMRI data caused by machine noise, subject's motion, respiration, and heartbeat (Friman et al., 2004; Friston et al., 2000). Here, we choose $r = 7$, equivalent to applying a high-pass, 128-second filtering to the data.

The HRF estimation under the GLM falls into two categories: parametric approaches that assume HRFs to have certain functional forms usually with only a few free parameters, and nonparametric approaches that use many functional bases to represent HRFs. The former includes the canonical HRF (Worsley and Friston, 1995; Worsley et al., 2002), Poisson model (Friston et al., 1994a), 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 inverse logit model (Lindquist and Wager, 2007), and many others. Nonparametric approaches include but are not limited to the finite impulse response (FIR) method (Dale, 1999; Lange et al., 1999), smooth FIR (Goutte et al., 2000), and many basis-representation methods that use penalties to regularize the variances of the estimates (Casanova et al., 2008; 2009; Marrelec et al., 2003; 2001; Vakorin et al., 2007; Zhang et al., 2007).

The above voxel-wise methods analyze the fMRI time series of one subject and one voxel at a time. However, fMRI time series at spatially close voxels are strongly correlated. It is more efficient to utilize spatial information of fMRI data to estimate HRFs. In this line of thought, Chaari et al. (2012); Makni et al. (2005, 2008); Vincent et al. (2010) used the brain parcellation, assumed voxels within the same parcel share the same HRF shapes, and assigned a spatially adaptive prior to HRF heights. Zhang et al. (2014b, 2016b) used a Poisson HRF (Buxton and Frank, 1997; Friston et al., 1994a) and characterized spatial and temporal associations between the GLM errors within a Bayesian framework. Degras and Lindquist (2014b) developed a hierarchical model for multi-subject fMRI data and imposed spatial and temporal regularization on HRF estimates.

Overall, existing nonparametric approaches to HRF estimation rely on enormous free parameters to accommodate variations of HRFs across many subjects and voxels. We propose a new nonparametric joint model for HRFs of all voxels and subjects to address this issue and assume these HRFs have a low-dimensional structure. Consequently, the number of parameters for characterizing many HRFs is significantly reduced compared to existing nonparametric or semiparametric methods. The details of the new model are given below.

We start with a nonparametric representation of $h_{jk}^i(t)$ with spline bases (de Boor, 2001; Eubank, 1999; Ruppert et al., 2003; Wahba, 1990): $h_{v,k}^i(t) = \sum_{l=1}^{L} \varpi_{lv,k}^i \cdot b_l(t)$, where $(b_1(t), \ldots, b_L(t))$ are fourth-order B-spline bases defined on an equally-spaced partition of the HRF domain $[0, m]$. The GLM (1) then becomes

$$y_v^i(t) = C(t)\ d_v^i + \sum_{k=1}^{K} \sum_{l=1}^{L} \varpi_{lv,k}^i \cdot \int_0^m b_l(u) \cdot s_k^i(t-u)du + \varepsilon_v^i(t).$$

Let $Y_v^i = (y_v^i(1), \ldots, y_v^i(T))'$, $\Lambda_{v,k}^i = (\varpi_{1v,k}^i, \ldots, \varpi_{Lv,k}^i)'$, $\mathbf{C}$ be a $T \times r$ matrix with the tth row equal to $C(t)$, $\mathbf{X}_k^i$ be a $T \times L$ matrix with the (t, l)th entry equal to $\int_0^m b_l(u) \cdot s_k^i(t-u)du$, and $\boldsymbol{\varepsilon}_v^i = (\varepsilon_v^i(1), \ldots, \varepsilon_v^i(T))'$. The above GLM has a matrix form:

$$Y_v^i = \mathbf{C}\ d_v^i + \sum_{k=1}^{K} \mathbf{X}_k^i\ \Lambda_{v,k}^i + \boldsymbol{\varepsilon}_v^i.$$

Let $\mathbf{Y}^i$ be a $T \times V$ matrix whose vth column equals Y_v^i. The GLM for $\mathbf{Y}^i$ can be written in a matrix form:

$$\mathbf{Y}^i = \mathbf{C}\ \mathbf{d}^i + \sum_{k=1}^{K}\ \mathbf{X}_k^i\ \Lambda_k^i + \mathbf{E}^i, \tag{2}$$

where Λ_k^i is an $L \times V$ matrix with the (l, v)th entry being $\varpi_{lv,k}^i$, $\mathbf{d}^i$ is an $r \times V$ matrix with the vth column equal to d_j^i, and $\mathbf{E}^i$ is a $T \times V$ matrix with the (t, v)th entry being $\varepsilon_v^i(t)$.

2.2. A Multi-Subject Low-Rank General Linear Model

The model (2) is the most flexible nonparametric model for HRFs of n ($i = 1, \ldots, n$) subjects and V ($v = 1, \ldots, V$) voxels in response to K ($k = 1, \ldots, K$) stimuli within the GLM framework. However, this model contains $n \cdot V \cdot K \cdot L$ basis coefficients

Table 1
Notations of Key Parameters

Parameter	Descriptions
$\mathbf{Y}^i$	Subject i's fMRI data (in a $T \times V$ matrix form) at V voxels and T time points.
$C(t)$	Functions $(\sqrt{2/T}\cos(\pi \frac{t}{T}), \ldots, \sqrt{2/T}\cos(r \pi \frac{t}{T}))$ for characterizing a low-frequency drift in fMRI data.
$\mathbf{C}$	A $T \times r$ matrix with the tth row equal to $C(t)$.
d_v^i	An $r \times 1$ vector of coefficients for $C(t)$ of subject i and voxel v.
$\mathbf{d}^i$	A $T \times V$ matrix with the vth column equal to d_v^i.
$\mathbf{X}_k^i$	A $T \times L$ matrix with the (t, l)th entry equal to $\int_0^m b_l(u) \cdot s_k^i(t - u)du$.
$\mathbf{W}_k$	An $L \times R$ matrix (in the model (3)) of coefficients of principal functions for representing HRFs of all subjects and voxels in response to the kth stimulus.
$W_{lr,k}$	The (l, r)th entry of the matrix $\mathbf{W}_k$.
$\mathbf{P}_k^i$	An $R \times V$ matrix (in the model (3)) of subject-specific and voxel-specific coefficients for representing subject i's HRFs in response to the kth stimulus.
$P_{rv,k}^i$	The (r, v)th entry of the matrix $\mathbf{P}_k^i$.
$\mathcal{P}_k$	$\{\mathbf{P}_k^i, i = 1, \ldots, n\}$.
Θ	All the parameters in the MSLRGLM $\{\mathbf{d}^i, \mathbf{W}_k, \mathbf{P}_k^i, i = 1, \ldots, n, \; k = 1, \ldots, K\}$.

$(\boldsymbol{\Lambda}_k^i)$ for representing many HRFs. In a typical fMRI study, n can be easily more than one hundred, V is in thousands, and L is in dozens. Therefore, Model (2) can contain enormous free parameters. To reduce the number of free parameters and also to enable combining data information across subjects and voxels to estimate HRFs, we assume

$$\boldsymbol{\Lambda}_k^i = \mathbf{W}_k \, \mathbf{P}_k^i,$$

where $\mathbf{W}_k$ and $\mathbf{P}_k^i$ are $L \times R$ and $R \times V$ matrices, respectively. The rank of $\boldsymbol{\Lambda}_k^i$, R, is given and much smaller than V. Under the low-rank assumption of $\boldsymbol{\Lambda}_k^i$, (2) becomes

$$\mathbf{Y}^i = \mathbf{C} \, \mathbf{d}^i + \sum_{k=1}^K \mathbf{X}_k^i \, \mathbf{W}_k \, \mathbf{P}_k^i + \mathbf{E}^i. \tag{3}$$

We refer to this model as *multi-subject, low-rank general linear model* (MSLRGLM). The key model parameters of the MSLRGLM are summarized in Table 1.

The MSLRGLM has a similar idea to matrix factorization (Agarwal et al., 2012; Koren et al., 2009; Feng et al., 2013; Samdin et al., 2017; Rendle, 2012) or factor models (Bai and Li, 2002; 2012), and the rank R of the MSLRGLM is similar to the dimension of factor matrices. Let $B(t) = (b_1(t), \ldots, b_L(t))$. Under the model (3), the HRFs, $h_{v,k}^i(t)$, of n subjects and V voxels are decomposed into a functional part $B(t)\mathbf{W}_k$ and a spatial part $\mathbf{P}_k^i$. The former is R principal functions for representing HRFs of different voxels and subjects, and the latter is subject-specific and voxel-specific coefficients for functional bases. As such, the MSLRGLM enables borrowing information across subjects and voxels to estimate HRFs while accommodating subject-specific and voxel-specific variation of HRFs. In addition, since the MSLRGLM uses much fewer parameters than the nonparametric method (with $L \times V$ parameters in each subject i's $\boldsymbol{\Lambda}_k^i$), the MSLRGLM leads to HRF estimates with smaller variances, as demonstrated in Section 5. We choose $R = 2$ in our study, because HRFs of different subjects and voxels all have a finite domain and a single-mode shape and differ mainly in height and time-to-peak (Lindquist, 2008). We also found that using R larger than 2 does not improve estimation or change analysis results much but rather adds computational costs.

3. Model Estimation

We use the model (3) to estimate HRFs of different subjects, voxels, and stimulus types and identify brain voxels with different responses to different stimuli. We propose two optimization functions with different penalties to achieve these two goals.

3.1. Optimization Functions

Let $\Theta = \{\mathbf{d}^i, \mathbf{W}_k, \mathbf{P}_k^i, i = 1, \ldots, n, \; k = 1, \ldots, K\}$. One can estimate Θ by minimizing the sum of squared errors (SSE) of the model (3):

$$\text{SSE}(\Theta) = \frac{1}{n} \sum_{i=1}^n \| \mathbf{Y}^i - \mathbf{C} \, \mathbf{d}^i - \sum_{k=1}^K \mathbf{X}_k^i \, \mathbf{W}_k \, \mathbf{P}_k^i \|_F^2 .$$

To reduce the estimation variance of Θ and also to produce temporally (Buxton and Frank, 1997) and spatially smooth HRF estimates, we propose to minimize a penalized SSE (PSSE) with two penalties:

$$\Upsilon_1(\mathbf{W}_k) = \sum_{r=1}^R \int_0^m (\sum_{l=1}^L W_{lr,k} \cdot b_l^{(2)}(t))^2 dt \quad \text{and} \quad \Upsilon_2(\mathbf{P}_k^i) = \sum_{r=1}^R \sum_{v \sim \tilde{v}} (P_{rv,k}^i - P_{r\tilde{v},k}^i)^2,$$

where the notation $v \sim \tilde{v}$ stands for neighboring voxels v and $\tilde{v}$. The first penalty is to regularize the roughness of principal functions $B(t)\mathbf{W}_k$, and the second one is to make spatially close voxels have similar voxel-specific coefficients $P^i_{rv,k}$ for $B(t)\mathbf{W}_k$, and thus, similar HRFs. In summary, we propose to estimate HRFs by minimizing the PSSE:

$$\text{PSSE}(\mathbf{\Theta}) = \text{SSE}(\mathbf{\Theta}) + \lambda \sum_{k=1}^{K} \Upsilon_1(\mathbf{W}_k) + \tau \sum_{k=1}^{K} \sum_{i=1}^{n} \Upsilon_2(\mathbf{P}^i_k). \tag{4}$$

Another important research problem in fMRI data analysis is to compare brain responses to different stimuli. Without loss of generality, we compare HRFs of the first two stimuli and identify voxels with different population-wide responses to the two stimuli. Let $\mathcal{P}_k = \{\mathbf{P}^i_k, i = 1, \ldots, n\}$. Following many voxel selection approaches (Baldassarre et al., 2012; Ge et al., 2016), we take into account the sparsity that only a few voxels have different responses to the first two stimuli, and address the HRF comparison problem by minimizing the following PSSE:

$$\text{PSSE}_c(\mathbf{\Theta}) = \text{SSE}(\mathbf{\Theta}) + \lambda \sum_{k=1}^{K} \Upsilon_1(\mathbf{W}_k) + \tau \sum_{k=1}^{K} \sum_{i=1}^{n} \Upsilon_2(\mathbf{P}^i_k) + \mu \Upsilon_3(\mathcal{P}_1, \mathcal{P}_2), \tag{5}$$

subject to $\mathbf{W}_2 = \mathbf{W}_1$, where

$$\Upsilon_3(\mathcal{P}_1, \mathcal{P}_2) = \sum_{v=1}^{V} \sqrt{\sum_{i=1}^{n} \sum_{r=1}^{R} \left(P^i_{rv,1} - P^i_{rv,2} \right)^2}.$$

The penalty $\Upsilon_3(\mathcal{P}_1, \mathcal{P}_2)$, a generalization of the penalty in the group lasso (Yuan and Lin, 2006), is a sparsity-inducing penalty on the differences between voxel-specific parameters of the HRFs for the first two stimuli. As $\mathbf{W}_1$ and $\mathbf{W}_2$ are set equal, the differences between $\mathbf{P}^i_1$ and $\mathbf{P}^i_2$ capture the differences between subject i's brain responses to the two stimuli. Using the penalty $\Upsilon_3(\mathcal{P}_1, \mathcal{P}_2)$ on the differences between $\mathbf{P}^i_1$ and $\mathbf{P}^i_2$ enables identifying voxels with different responses to the first two stimuli.

4. Optimization Algorithms

A crucial challenge in analyzing fMRI data is computationally efficiently estimating massive HRFs of many subjects and voxels. For the presented real data analysis under study (Section 6), despite the low-rank structure for HRFs, the total number of model parameters needed for representing all the HRFs is more than 5.5 million. Consequently, we develop a computationally efficient algorithm to solve the above two optimization problems and estimating massive HRFs.

Let $\mathcal{W} = \{\mathbf{W}_k, k = 1, \ldots, K\}$, $\tilde{\mathbf{d}} = \{\mathbf{d}^i, i = 1, \ldots, n\}$, and $\tilde{\mathcal{P}} = \{\mathcal{P}_k, k = 1, \ldots, K\}$. Standard approaches (Zhang et al., 2018) to optimizing objective functions $\text{PSSE}(\mathbf{\Theta})$ and $\text{PSSE}_c(\mathbf{\Theta})$ are through an alternating search algorithm, which iterates between finding the minimizer $\mathcal{W}$ of PSSE given $(\tilde{\mathcal{P}}, \tilde{\mathbf{d}})$ and finding the minimizer $(\tilde{\mathcal{P}}, \tilde{\mathbf{d}})$ given $\mathcal{W}$. However, this algorithm is computationally intensive due to the large data dimension and many parameters of the objective functions. Instead, we minimize $\text{PSSE}(\mathbf{\Theta})$ in (4) using a gradient descent method (Nesterov, 2013; Gong et al., 2013), an iterative algorithm that generates a series of estimates of $\mathbf{\Theta}$ by moving a small step along the descent direction of the optimization function each time.

Direct application of the gradient descent method to minimizing $\text{PSSE}_c(\mathbf{\Theta})$ in (5) is not feasible, because $\text{PSSE}_c(\mathbf{\Theta})$ is non-differentiable for $\mathcal{P}_1$ and $\mathcal{P}_2$, and the gradient of $\text{PSSE}_c(\mathbf{\Theta})$ does not exist for $\mathcal{P}_1 = \mathcal{P}_2$. To address this, we develop a new gradient descent method, which is computationally efficient and has the same computational complexity as the conventional one. We explain the details below.

The objective function $\text{PSSE}_c(\mathbf{\Theta})$ in (5) is composed of a convex but non-differentiable component, $\mu \Upsilon_3(\mathcal{P}_1, \mathcal{P}_2)$, and a differentiable component, $\text{SSE}(\mathbf{\Theta}) + \lambda \sum_{k=1}^{K} \Upsilon_1(\mathbf{W}_k) + \tau \sum_{k=1}^{K} \sum_{i=1}^{n} \Upsilon_2(\mathbf{P}^i_k)$, i.e., $\text{PSSE}(\mathbf{\Theta})$.

We denote the parameter value at the tth step by $\mathbf{\Theta}^t$ for $t = 0, 1, \ldots$. We specify the initial value $\mathbf{\Theta}^0$, let a step-size parameter $\rho = 1$, and choose a step-size multiplier $\alpha > 1$ ($\alpha = 2$ in our analysis). The proposed gradient descent algorithm iterates between the following steps.

- At the tth iteration, calculate $\nabla \text{PSSE}(\mathbf{\Theta}^t)$, where ∇ denotes the gradient operator.
- Given the step-size parameter ρ, solve the following optimization problem:

$$\mathbf{\Theta}^* = \text{argmin}_{\mathbf{\Theta}} \langle \nabla \text{PSSE}(\mathbf{\Theta}^t), \mathbf{\Theta} \rangle + \mu \Upsilon_3(\mathcal{P}_1, \mathcal{P}_2) + \frac{\rho}{2} \| \mathbf{\Theta} - \mathbf{\Theta}^t \|_2^2, \tag{6}$$

  where $< \cdot, \cdot >$ denotes an inner product between two vectors, and $\|.\|_2$ denotes the Euclidean norm.
- If $\mathbf{\Theta}^*$ improves the objective function value at $\mathbf{\Theta}^t$, update $\mathbf{\Theta}^{t+1} = \mathbf{\Theta}^*$. Otherwise, increase the step-size parameter by $\rho = \alpha \rho$.
- Repeat until convergence.

Each iteration in the above algorithm involves solving a strongly convex subproblem (6). We show in the appendix that subproblem (6) has a closed form solution, and the computational cost of solving the subproblem is linear in terms of the number of parameters. The use of the parameter ρ is to control the step size in (6), ensuring that the new $\mathbf{\Theta}^{t+1}$ is not too far from $\mathbf{\Theta}^t$ in the previous step. We use a backtracking line search to tune the step-size parameter ρ. We find that this is more stable and faster than the commonly used constant step-size. Given penalty parameters, it took the proposed optimization algorithm only about 5 minutes to analyze 106 subjects' fMRI data at almost 7000 voxels.

4.1. Convergence

We show the convergence of the proposed algorithm with given penalty parameters λ, τ, and μ. We first make the following assumptions regarding the optimization function (5).

(1) PSSE($\boldsymbol{\Theta}$) is continuously differentiable with Lipschitz continuous gradient, that is, there exists a positive constant L such that for any two $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$:

$$\|\nabla\mathrm{PSSE}(\boldsymbol{\Theta}_1) - \nabla\mathrm{PSSE}(\boldsymbol{\Theta}_2)\|_2 \le L\|\boldsymbol{\Theta}_1 - \boldsymbol{\Theta}_2\|_2.$$

(2) The objective function in (5) is bounded from below.

Definition 4.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a proper convex function, i.e., $f(x) > -\infty \forall x \in \mathbb{R}^n$, and $f(x) < +\infty$ for at least one x. A vector $g \in \mathbb{R}^n$ such that $f(y) - f(x) \ge \langle g, y - x \rangle \ \forall x, y \in \mathbb{R}^n$ is called a subgradient of f at x. The set of all subgradients of f at x is called the subdifferential of f at x, denoted by $\partial f(x)$.

Definition 4.2. Parameter $\boldsymbol{\Theta}^*$ is a critical point of the problem in (5) if the following holds:

$$\nabla\mathrm{PSSE}(\boldsymbol{\Theta}^*) \in \partial(-\mu\Upsilon_3(\mathcal{P}_1^*, \mathcal{P}_2^*)).$$

Theoretical Result 4.1. Assume that the above assumptions (1) and (2) hold. All the limiting points of the sequence of estimates $\{\boldsymbol{\Theta}^t, t = 0, 1, \ldots\}$ generated by the proposed gradient descent method are critical points of (5).

The above theorem indicates that the proposed gradient descent method converges to a point that has a zero sugradient. The proof is in the Appendix.

4.2. Penalty Parameter Selection

We develop new cross-validation-based methods to select penalty parameters for PSSE($\boldsymbol{\Theta}$) and PSSE$_c$($\boldsymbol{\Theta}$). The fMRI data are randomly split into a training set consisting of 80% of the data and a testing set consisting of the rest 20% of the data. Let $\Im$ be the set of subjects in the training set, $\aleph$ be the set of subjects in the testing set, and $\hbar$ be the number of subjects in $\aleph$.

We select penalty parameters for PSSE($\boldsymbol{\Theta}$) first. For a given combination of λ and τ, we use the training data to calculate $\mathbf{W}_k$, $k = 1, \ldots, K$, through minimizing PSSE($\boldsymbol{\Theta}$). For data $\mathbf{Y}^i$ of each subject i in the testing set, we solve

$$(\hat{\mathbf{P}}_k^i, \hat{\mathbf{d}}^i) = \mathrm{argmin}_{\mathbf{P}_k^i, \mathbf{d}^i} \| \mathbf{Y}^i - \mathbf{C}\,\mathbf{d}^i - \sum_{k=1}^{K} \mathbf{X}_k^i\, \mathbf{W}_k\, \mathbf{P}_k^i \|_F^2 + \tau \sum_{k=1}^{K} \Upsilon_2(\mathbf{P}_k^i),$$

and calculate $\mathcal{T}_i^1 = \| \mathbf{Y}^i - \mathbf{C}\,\hat{\mathbf{d}}^i - \sum_{k=1}^{K} \mathbf{X}_k^i\, \mathbf{W}_k\, \hat{\mathbf{P}}_k^i \|_F^2$ as the testing error for subject i. We choose the combination of (λ, τ) that yields the smallest average testing error $\sum_{i \in \aleph} \mathcal{T}_i^1 / \hbar$.

Standard cross-validation-based methods are time consuming for selecting penalty parameters for PSSE$_c$($\boldsymbol{\Theta}$), as it involves three different penalty parameters. To reduce the computational burden, we propose to select λ and (τ, μ) in two separate steps. The intuition behind this approach is that $\lambda \sum_{k=1}^{K} \Upsilon_1(\mathbf{W}_k)$ regularizes the roughness of principal functions $B(t)\mathbf{W}_k$ and does not involve $\tilde{\mathcal{P}}$. The details are given below.

1. Selection of λ.
    1.a Given a candidate value of λ, we use the training set to calculate $\mathbf{W}_k$, $k = 1, \ldots, K$, through minimizing

    $$\min_{\boldsymbol{\Theta}} \mathrm{SSE}(\boldsymbol{\Theta}) + \lambda \sum_{k=1}^{K} \Upsilon_1(\mathbf{W}_k).$$

    1.b Given $\mathbf{W}_k$, $k = 1, \ldots, K$, from the previous step, calculate the testing error for each $\mathbf{Y}^i$ in the testing set as

    $$\mathcal{T}_i^2 = \min_{\mathbf{P}_k^i, \mathbf{d}^i} \| \mathbf{Y}^i - \mathbf{C}\,\mathbf{d}^i - \sum_{k=1}^{K} \mathbf{X}_k^i\, \mathbf{W}_k\, \mathbf{P}_k^i \|_F^2 .$$
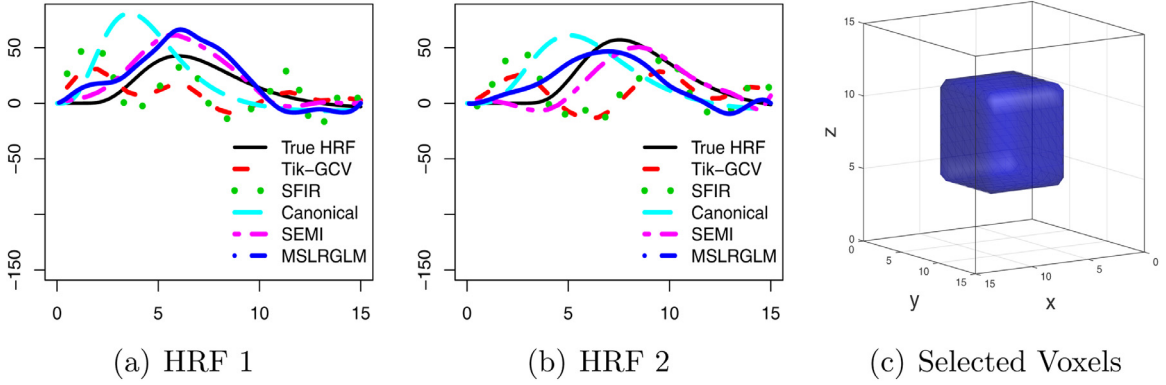
    1.c Select λ^* that leads to the smallest average testing error $\sum_{i \in \aleph} \mathcal{T}_i^2 / \hbar$.
2. Given λ^*, we select two parameters τ and μ.
    2.a Given a combination (τ, μ), we use the training data to obtain $\mathbf{W}_k$, $k = 1, \ldots, K$, through solving the problem

    $$\min_{\boldsymbol{\Theta}} \mathrm{PSSE}_c(\boldsymbol{\Theta}) = \mathrm{SSE}(\boldsymbol{\Theta}) + \lambda^* \sum_{k=1}^{K} \Upsilon_1(\mathbf{W}_k) + \tau \sum_{k=1}^{K} \sum_{i \in \Im} \Upsilon_2(\mathbf{P}_k^i)$$
    $$+ \mu\Upsilon_3(\mathcal{P}_1, \mathcal{P}_2),$$

    where $\Upsilon_3(\mathcal{P}_1, \mathcal{P}_2)$ is a function of $\mathbf{P}_1^i$ and $\mathbf{P}_2^i$ for i in the training set $\Im$ only.

Fig. 1. HRF Estimates of Simulated Data. (a) Estimated HRFs of one subject and one voxel in response to the first stimulus. (b) Estimated HRFs of the same subject and voxel as (a) in response to the second stimulus. (c) Selected true voxels (in blue) and selected null voxels (in red if any).

Table 2
The AREs of estimated HRFs by the proposed low-rank methods and competing methods.

HRF k	MSLRGLM-h	MSLRGLM-c	SEMI	Tik-GCV	SFIR	Canonical
1	0.30	0.30	0.52	0.92	1.10	0.80
2	0.42	0.44	0.52	0.90	1.10	0.85

2.b For each $\mathbf{Y}^i$ in the testing set $\aleph$, we solve

$$(\hat{\mathbf{P}}^i_k, \hat{\mathbf{d}}^i) = \mathrm{argmin}_{\mathbf{P}^i_k, \mathbf{d}^i} \parallel \mathbf{Y}^i - \mathbf{C}\,\mathbf{d}^i - \sum_{k=1}^{K} \mathbf{X}^i_k\,\mathbf{W}_k\,\mathbf{P}^i_k \parallel^2_F .$$

2.c Calculate the testing error, $\mathcal{T}^3_i = \parallel \mathbf{Y}^i - \mathbf{C}\,\hat{\mathbf{d}}^i - \sum_{k=1}^{K} \mathbf{X}^i_k\,\mathbf{W}_k\,\hat{\mathbf{P}}^i_k \parallel^2_F$, for $\mathbf{Y}^i$.
2.d We choose the combination of (τ, μ) that minimizes the average testing error $\sum_{i \in \aleph} \mathcal{T}^3_i / \hbar$.

5. Simulation Study

We generate fMRI data from the GLM (1) using the same experimental design as that of the real data. The simulated data contain $n = 106$ subjects' fMRI time series at $V = 15^3$ voxels, which are distributed over a $15 \times 15 \times 15$ lattice grid. The data have $K = 4$ stimuli. We let voxels in the center $7 \times 7 \times 7$ grid have different HRFs associated with the first two stimuli.

We simulate HRFs from a semi-parametric model developed by Zhang et al. (2013). This model characterizes subject-specific and voxel-specific brain responses:

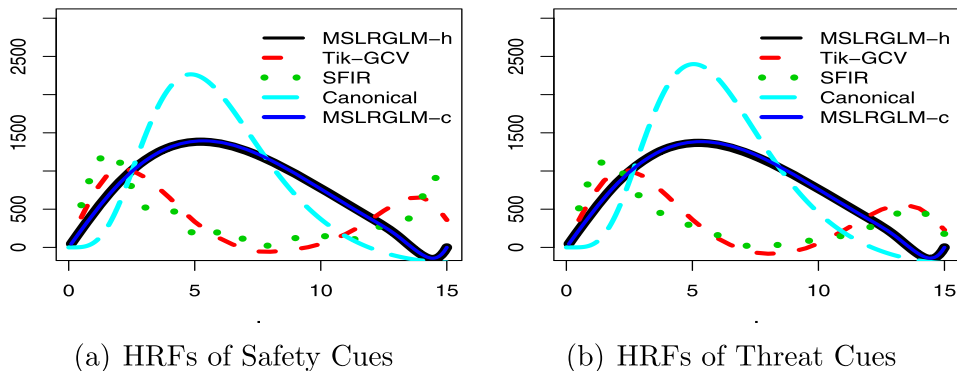$$h^i_{v,k}(t) = A^i_{v,k} \cdot f_{v,k}(t + O^i_{v,k}), \tag{7}$$

where $f_{v,k}(t)$ is the functional shape shared in common across n subjects' HRFs at voxel v in response to the kth stimulus, and $A^i_{v,k}$ and $O^i_{v,k}$ are subject-specific and voxel-specific response magnitude and latency, respectively. For easy comparison, we let HRF parameters and fMRI data be the same as those in the paper Zhang et al. (2018).

We apply the MSLRGLM to the simulated data by optimizing two functions $\mathrm{PSSE}(\mathbf{\Theta})$ and $\mathrm{PSSE}_c(\mathbf{\Theta})$. We denote the two optimizations by MSLRGLM-h and MSLRGLM-c, respectively. The HRF estimates by MSLRGLM-h and MSLRGLM-c are almost identical. We also estimate HRFs by competing methods, including the semi-parametric (SEMI) method (Zhang et al., 2013), a nonparametric method with 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), and a method by representing the HRF with the canonical HRF and its first order derivative (Liao et al., 2002) (referred to as the canonical method).

Fig. 1 shows estimated HRFs by six methods in comparison to the true HRFs. We use the average relative error (ARE) as the criterion to evaluate HRF estimation errors:

$$\mathrm{ARE}_k = \frac{1}{nV} \sum_{i=1}^{n} \sum_{v=1}^{V} \frac{\parallel h^i_{v,k} - \hat{h}^i_{v,k} \parallel_2}{\parallel h^i_{v,k} \parallel_2},$$

where $\hat{h}^i_{v,k}$ denotes an estimate of the HRF, $h^i_{v,k}$. Table 2 summarizes the AREs of different HRF estimation methods. MSLRGLM-h and MSLRGLM-c outperform other four methods by having much smaller estimation errors. Note that the proposed MSLRGLM is different from the data-generating model (7). SEMI is directly based on the HRF-generating model (7),

ARTICLE IN PRESS

JID: ECOSTA [m3Gsc;March 25, 2021;7:38]

T. Zhang, M. Pham, G. Yan et al. Econometrics and Statistics xxx (xxxx) xxx

(a) HRFs of Safety Cues   (b) HRFs of Threat Cues

Fig. 2. Estimated HRFs for one selected voxel in dACC. (a) Estimated HRFs in response to the safety cue. (b) Estimated HRFs in response to the threat cue.

and the nonparametric methods, Tik-GCV and SFIR, are consistent with (7). However, these voxel-wise methods are not as efficient as the MSLRGLM for estimating HRFs.

We use MSLRGLM-c to identify voxels that have different HRFs of the first two stimuli. Fig. 1(c) shows voxels selected by MSLRGLM-c. All the selected voxels are true ones (in blue). MSLRGLM-c achieves 0% false discovery rate (FPR) and 100% true positive rate (TPR). The voxel selection method by Zhang et al. (2018) has a slightly higher FPR, 0.3%, with 100% TPR. For comparison, we also perform voxel-wise t-tests to compare HRF estimates produced by the other four competing methods mentioned above. The FPR of SEMI is 15.4% for 100% TPR, the FPR of Tik-GCV is 16.5% for 100% TPR, and the FPR of the other two methods, SFIR and Canonical, is 100% for 100% TPR. Overall, the existing voxel-wise methods without using the spatial information of the data have much larger FPRs than MSLRGLM-c in identifying voxels that have different responses to different stimuli.

In summary, the proposed MSLRGLM-c can simultaneously estimate HRFs of different subjects and voxels efficiently and select active voxels with high accuracy. In contrast, existing methods cannot achieve accurate HRF estimation and voxel selection simultaneously.

6. Real Data Analysis

The fMRI data under analysis were collected from a psychology study on the human brain's emotion function (Coan, 2010; 2011; Coan et al., 2006; 2013). The fMRI experiment used threats of mild electric shocks as negative emotional stimuli to evoke $n = 106$ subjects' brain activities. The experiment protocol consisted of four stimuli: threat cues with 20% chances of a mild electric shock, safety cues with no electric shock, resting periods between cues, and end of trial cues with sometimes simultaneous real electric shocks. We preprocessed the fMRI data using FMRIB's Software Library software (Friston et al., 1994b, Version 5.98; www.fmrib.ox.ac.uk/fsl) and used FLIRT (Jenkinson et al., 2002) to register images to Montreal Neurological Institute (MNI) space.

The purpose of this study is to evaluate subjects' brain responses to negative emotional stimuli and identify voxels that are responsive to the stimuli. To correct for subjects' emotional brain activity due to the fMRI scanning (Mazard et al., 2002; Szameitat et al., 2009), we used subjects' brain responses to safety cues as the reference and compared subjects' HRFs due to threat and safety cues. We performed a region-of-interest analysis of the fMRI data in dorsal anterior cingulate cortex (dACC). Fig. 3(a) shows the dACC brain area. The dACC is reported to be involved in affective processing in the literature (Knutson et al., 2000) and be associated with subjects' mental health (Dedovic et al., 2016). We used the Harvard subcortical brain atlas to extract subjects' fMRI data in dACC.

We show the estimated HRFs in response to safety and threat cues in Fig. 2. The HRF estimates by MSLRGLM-h and MSLRGLM-c are almost identical. For comparison, we also show the same subject's estimated HRFs by other methods. The HRF estimates by voxel-wise nonparametric methods, SFIR and Tik-GCV, have large variances and two modes, against the common belief of a single mode in HRFs. The HRF estimates by SEMI have too large variances to be shown in the figure.

We note that dACC voxels have a different functional shape from the canonical HRF, which most often describes responses of visual and motor cortices to visual-motor tasks (Lindquist, 2008). The HRFs of emotional stimuli tend to have more gradual changes over time in contrast to the abrupt increase and decrease of the canonical HRF. The difference between the HRF in visual and motor areas compared with the HRFs of emotional stimuli is consistent with the prior evidence of heterogeneous HRF waveforms in different brain regions (David et al., 2008). The gradual hemodynamic response to emotional stimuli is likely due to a need for top-down control of attention and emotion regulation, which is in line with the literature on early positive event-related potentials in the temporo-occipital brain region (MacNamara et al., 2009) related to the dACC. Additionally, prior studies suggest that this gradual change may also be explained by the strong decrease in the autoregulation constant of cerebral blood flow (David et al., 2008). In contrast, abrupt changes of the canonical HRFs related

(a) dACC Voxels (b) $Z = 43$ (c) $X = 49$ (d) $Y = 69$
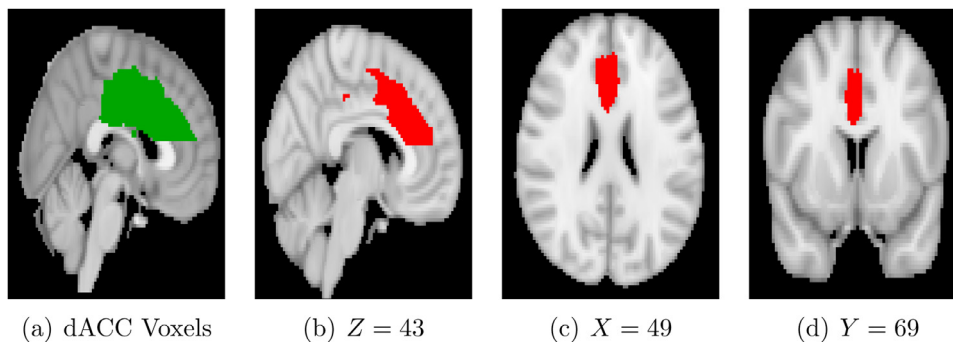
Fig. 3. (a) All the dACC voxels (in green). (b)-(d) Selected dACC voxels (in red) shown in axial, coronal, and sagittal views.

to visual and motor functions of the brain occur earlier and are likely due to bottom-up processing. This result provides evidence of the necessity to use a flexible model for HRFs of different voxels, subjects, and stimulus types. In summary, the proposed MSLRGLM outperforms all the existing methods by providing more stable and scientifically more interpretable HRF estimates.

We use MSLRGLM-c to select voxels that have different responses to safety and threat cues. The differences between HRFs of threat and safety cues are mainly in magnitude but very small within each subject. Nevertheless, we identified around 27% voxels with different responses to the two cues. Figs. 3(b)-3(d) show the selected dACC voxels in axial, coronal, and sagittal views. We found that the voxels with the most significant responses to negative emotional stimuli are clustered in a region that spans both the medial part of the dACC and anterior of the midcingulate cortex (aMCC). It is likely that this particular region is distinct from other regions of the dACC and may indeed be a separate functional unit (Vogt, 2016). The dACC has a well-established role in emotion and autonomic regulation, while the MCC has been largely implicated in decision making and skeletomotor control (Bush et al., 2000; Vogt, 2009). This particular area may be an important site for the interaction between negatively salient stimuli and motor signals in the brain and may be involved in defensive responses like freezing (Pereira et al., 2010; Vogt, 2008). This result indicates a distinction between sub-regions in the dACC/aMCC that is consistent with prior literature that supports functional heterogeneity within this region (Bush et al., 2002). Furthermore, this area of the dACC/aMCC is associated with pain and negative affect (Vogt, 2008). Our analysis confirms different extents of responses to emotionally salient stimuli in dACC/aMCC subregions and identifies the subregion with the strongest response.

7. Discussions

The two optimization functions PSSE($\boldsymbol{\Theta}$) and PSSE$_c$($\boldsymbol{\Theta}$) produce similar HRF estimates, possibly because the differences between subject-specific coefficients $\mathcal{P}_1$ and $\mathcal{P}_2$ are small in comparison to the data error. With an additional sparsity penalty, PSSE$_c$($\boldsymbol{\Theta}$) enables identifying voxels with different responses to different stimuli. This voxel selection has slightly better accuracy than using a population-mean GLM for fMRI data (Zhang et al., 2018). We suppose that the larger variation of HRFs across subjects, the better performance of the MSLRGLM-c.

In general, the differences in height and time-to-peak of HRFs across different voxels and subjects are the easiest to estimate, so we let the proposed model have rank 2. The literature has reported the initial dip in HRFs (Bandettini, 2012). However, we found that increasing the rank of the proposed model from 2 to 3 does not enable detecting the initial dip, possibly due to the low temporal resolution of the data under analysis. The initial dip lasts briefly for only 2 to 3 s (Cheng et al., 2001). The fMRI data under analysis has a time unit of 2 s, making the detection of the initial dip difficult. It is possible that for fMRI data with a much higher temporal resolution, the proposed model with a larger rank can detect the initial dip in HRFs.

In our optimization approach, we incorporate the spatial information of the data by imposing regularization only on the differences between neighboring voxels' HRF estimates. It is possible to account for more complex distances between voxels in the HRF estimation, for example, structural connectivity between voxels. The structural connectivity is usually measured using diffusion tensor imaging (DTI) (Ethofer et al., 2013) and can be incorporated as weights for the differences between voxels' HRFs in the optimization function. There are also many other methods for combining fMRI and DTI data study the human brain functions (Zhu et al., 2014).

Although the proposed method can estimate subject-specific HRFs and select voxels simultaneously, the estimation of the MSLRGLM requires significantly more computational time than the population-mean GLM. This is because the former has n times more parameters than the latter. The scalability of the proposed model and computational algorithm to much larger datasets will be the focus of future research.

The proposed MSLRGLM can be used to build scalar-on-image regression, where subjects' covariates are the response, and their fMRI data are predictors. Since the MSLRGLM characterizes subject-specific brain activities with much fewer parameters

ARTICLE IN PRESS

JID: ECOSTA [m3Gsc;March 25, 2021;7:38]

T. Zhang, M. Pham, G. Yan et al. Econometrics and Statistics xxx (xxxx) xxx

than most existing models for fMRI data, we expect that the ensuing scalar-on-image regression will have a better prediction power than exiting methods. The scalar-on-image regression based on the MSLRGLM will be another focus of future research.

Appendix

Analytical Solution to Subproblem (6).

The subproblem in the proposed gradient descent method is

$$\boldsymbol{\Theta}^* = \text{argmin}_{\boldsymbol{\Theta}} \langle \nabla \text{PSSE}(\boldsymbol{\Theta}^t), \boldsymbol{\Theta} \rangle + \mu \Upsilon_3(\mathcal{P}_1, \mathcal{P}_2) + \frac{\rho}{2} \|\boldsymbol{\Theta} - \boldsymbol{\Theta}^t\|_2^2.$$

Let $\boldsymbol{\Theta}^* = \{\mathbf{d}^{i*}, \mathbf{W}_k^*, \mathbf{P}_k^{i*}, i = 1, \ldots, n, \ k = 1, \ldots, K\}$, $\mathcal{P}_k^* = \{\mathbf{P}_k^{i*}, i = 1, \ldots, n\}$, $\boldsymbol{\Theta}^t = \{\mathbf{d}^{i,t}, \mathbf{W}_k^t, \mathbf{P}_k^{i,t}, i = 1, \ldots, n, \ k = 1, \ldots, K\}$, and $\mathcal{P}_k^t = \{\mathbf{P}_k^{i,t}, i = 1, \ldots, n\}$.

Let $\boldsymbol{\Phi} = \{\mathbf{d}^i, \mathbf{W}_k, i = 1, \ldots, n, \ k = 1, \ldots, K\} \cup \{\mathbf{P}_k^i, i = 1, \ldots, n, \ k = 3, \ldots, K\}$. The optimization (6) is separable into two optimization problems: the optimization for $\boldsymbol{\Phi}$

$$\boldsymbol{\Phi}^* = \text{argmin}_{\boldsymbol{\Phi}} \left\langle \frac{\partial \text{PSSE}(\boldsymbol{\Theta}^t)}{\partial \boldsymbol{\Phi}}, \boldsymbol{\Phi} \right\rangle + \frac{\rho}{2} \|\boldsymbol{\Phi} - \boldsymbol{\Phi}^t\|_2^2;$$

and the other one for $(\mathcal{P}_1, \mathcal{P}_2)$

$$(\mathcal{P}_1^*, \mathcal{P}_2^*) = \text{argmin}_{(\mathcal{P}_1, \mathcal{P}_2)} \left\langle \frac{\partial \text{PSSE}(\boldsymbol{\Theta}^t)}{\partial (\mathcal{P}_1, \mathcal{P}_2)}, (\mathcal{P}_1, \mathcal{P}_2) \right\rangle + \frac{\rho}{2} \|(\mathcal{P}_1, \mathcal{P}_2) - (\mathcal{P}_1^t, \mathcal{P}_2^t)\|_2^2$$

$$+ \mu \sum_{v=1}^{V} \sqrt{\sum_{i=1}^{n} \sum_{r=1}^{R} (P_{rv,1}^i - P_{rv,2}^i)^2}. \tag{A.1}$$

We here explain in detail the solution to (A.1).

Let $\mathbf{P}_{v,k}^i = (P_{1v,k}^i, \ldots, P_{Rv,k}^i)^T$. We let $\mathbf{G}_{v,1}^i = \frac{\partial \text{PSSE}(\boldsymbol{\Theta}^t)}{\partial (\mathbf{P}_{v,1}^i)}$ and $\mathbf{G}_{v,2}^i = \frac{\partial \text{PSSE}(\boldsymbol{\Theta}^t)}{\partial (\mathbf{P}_{v,2}^i)}$, $i = 1, \ldots, n$, and $v = 1, \ldots, V$. For simple notations and without loss of generality, we present the solution for $n = 1$. The optimization problem in (A.1) is divided into V smaller problems, each corresponding to one voxel v in the following form:

$$\min_{\mathbf{P}_{v,1}^i, \mathbf{P}_{v,2}^i} \langle \mathbf{G}_{v,1}^i, \mathbf{P}_{v,1}^i \rangle + \langle \mathbf{G}_{v,2}^i, \mathbf{P}_{v,2}^i \rangle + \mu \|\mathbf{P}_{v,1}^i - \mathbf{P}_{v,2}^i\|_2^2 + \frac{\rho}{2} \|\mathbf{P}_{v,1}^i - \mathbf{P}_{v,1}^{i,t}\|_2^2 + \frac{\rho}{2} \|\mathbf{P}_{v,2}^i - \mathbf{P}_{v,2}^{i,t}\|_2^2.$$

We introduce a slack vector $\mathbf{S} = \mathbf{P}_{v,1}^i - \mathbf{P}_{v,2}^i \in \Re^{R \times 1}$. The problem above is equivalent to

$$\min_{\mathbf{P}_{v,1}^i, \mathbf{P}_{v,2}^i} \langle \mathbf{G}_{v,1}^i, \mathbf{P}_{v,1}^i \rangle + \langle \mathbf{G}_{v,2}^i, \mathbf{P}_{v,2}^i \rangle + \mu \|\mathbf{S}\|_2 + \frac{\rho}{2} \|\mathbf{P}_{v,1}^i - \mathbf{P}_{v,1}^{i,t}\|_2^2 + \frac{\rho}{2} \|\mathbf{P}_{v,2}^i - \mathbf{P}_{v,2}^{i,t}\|_2^2,$$

$$\text{such that } \mathbf{S} = \mathbf{P}_{v,1}^i - \mathbf{P}_{v,2}^i.$$

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

$$\mathcal{L}(\mathbf{P}_{v,1}^i, \mathbf{P}_{v,2}^i, \mathbf{S}, \tilde{\alpha}) = \langle \mathbf{G}_{v,1}^i, \mathbf{P}_{v,1}^i \rangle + \langle \mathbf{G}_{v,2}^i, \mathbf{P}_{v,2}^i \rangle + \mu \|\mathbf{S}\|_2 + \frac{\rho}{2} \|\mathbf{P}_{v,1}^i - \mathbf{P}_{v,1}^{i,t}\|_2^2$$

$$+ \frac{\rho}{2} \|\mathbf{P}_{v,2}^i - \mathbf{P}_{v,2}^{i,t}\|_2^2 - \tilde{\alpha}^T (\mathbf{S} - (\mathbf{P}_{v,1}^i - \mathbf{P}_{v,2}^i)). \tag{A.2}$$

Since the minimum of the Lagrangian with respect to $\mathbf{S}$ is finite if and only if $\|\tilde{\alpha}\|_2 \leq \mu$ (Lin et al., 2014), the minimum value of the $\mathbf{S}$ term is 0. As such, we can eliminate them in (A.2) and a get a reduced form of the Lagrangian function:

$$\hat{\mathcal{L}}(\mathbf{P}_{v,1}^i, \mathbf{P}_{v,2}^i, \mathbf{S}, \tilde{\alpha}) = \langle \mathbf{G}_{v,1}^i, \mathbf{P}_{v,1}^i \rangle + \langle \mathbf{G}_{v,2}^i, \mathbf{P}_{v,2}^i \rangle + \frac{\rho}{2} \|\mathbf{P}_{v,1}^i - \mathbf{P}_{v,1}^{i,t}\|_2^2$$

$$+ \frac{\rho}{2} \|\mathbf{P}_{v,2}^i - \mathbf{P}_{v,2}^{i,t}\|_2^2 + \tilde{\alpha}^T (\mathbf{P}_{v,1}^i - \mathbf{P}_{v,2}^i).$$

We minimize $\hat{\mathcal{L}}(\mathbf{P}_{v,1}^i, \mathbf{P}_{v,2}^i, \mathbf{S}, \tilde{\alpha})$ over $(\mathbf{P}_{v,1}^i, \mathbf{P}_{v,2}^i)$ and obtain

$$\mathbf{P}_{v,1}^i = \frac{-\tilde{\alpha} - \mathbf{G}_{v,1}^i + \rho \mathbf{P}_{v,1}^{i,t}}{\rho} \quad \text{and} \quad \mathbf{P}_{v,2}^i = \frac{\tilde{\alpha} - \mathbf{G}_{v,2}^i + \rho \mathbf{P}_{v,2}^{i,t}}{\rho}. \tag{A.3}$$

Substitute the above two equations back to (A.2), we obtain the dual problem

$$\min_{\tilde{\alpha}} \frac{1}{2} \langle \tilde{\alpha}, \tilde{\alpha} \rangle + \langle \tilde{\alpha}, \frac{(\mathbf{G}_{v,1}^i - \mathbf{G}_{v,2}^i) - \rho(\mathbf{P}_{v,1}^{i,t} - \mathbf{P}_{v,2}^{i,t})}{2} \rangle, \text{ such that } \|\tilde{\alpha}\|_2 \leq \mu.$$

Let $\tilde{h} = \frac{1}{2} \left[(\mathbf{G}_{v,1}^i - \mathbf{G}_{v,2}^i) - \rho(\mathbf{P}_{v,1}^{i,t} - \mathbf{P}_{v,2}^{i,t}) \right]$. The above problem has a closed form solution: if $\|\tilde{h}\|_2 < \mu$ then $\tilde{\alpha}^* = -\tilde{h}$, otherwise $\tilde{\alpha}^* = -\mu \frac{\tilde{h}}{\|\tilde{h}\|_2}$. Plugging $\tilde{\alpha}^*$ into (A.3) leads to the solution of $\mathbf{P}_{v,1}^i$ and $\mathbf{P}_{v,2}^i$.

Proof of Theorem 4.1.

At the tth iteration, $\boldsymbol{\Theta}^{t+1}$ is the minimizer of the problem (6). Therefore,

$$\langle \nabla \text{PSSE}(\boldsymbol{\Theta}^t), \boldsymbol{\Theta}^{t+1} - \boldsymbol{\Theta}^t \rangle + \mu \Upsilon_3(\mathcal{P}_1^{t+1}, \mathcal{P}_2^{t+1}) + \frac{\rho}{2} \|\boldsymbol{\Theta}^{t+1} - \boldsymbol{\Theta}^t\|_2^2 \leq \mu \Upsilon_3(\mathcal{P}_1^t, \mathcal{P}_2^t).$$

From Assumption (1), we have

$$\text{PSSE}(\boldsymbol{\Theta}^{t+1}) \leq \text{PSSE}(\boldsymbol{\Theta}^t) + \langle \nabla \text{PSSE}(\boldsymbol{\Theta}^t), \boldsymbol{\Theta}^{t+1} - \boldsymbol{\Theta}^t \rangle + \frac{L}{2} \|\boldsymbol{\Theta}^{t+1} - \boldsymbol{\Theta}^t\|^2.$$

Combining the above two equations, we have

$$\text{PSSE}(\boldsymbol{\Theta}^{t+1}) + \mu \Upsilon_3(\mathcal{P}_1^{t+1}, \mathcal{P}_2^{t+1}) \leq \text{PSSE}(\boldsymbol{\Theta}^t) + \mu \Upsilon_3(\mathcal{P}_1^t, \mathcal{P}_2^t) - \frac{\rho - L}{2} \|\boldsymbol{\Theta}^{t+1} - \boldsymbol{\Theta}^t\|^2. \qquad (A.4)$$

This indicates that there exists a step length parameter $\rho > L$ that will lead to a decrease in the objective function value. In practice, we use backtracking to find this step length. This also implies that the sequence $\{\boldsymbol{\Theta}^t\}$ generated by the algorithm satisfies:

$$\text{PSSE}(\boldsymbol{\Theta}_{t+1}) + \mu \Upsilon_3(\mathcal{P}_1^{t+1}, \mathcal{P}_2^{t+1}) \leq \text{PSSE}(\boldsymbol{\Theta}_t) + \mu \Upsilon_3(\mathcal{P}_1^t, \mathcal{P}_2^t).$$

Let $\boldsymbol{\Theta}^a$ be a limit point of the sequence $\{\boldsymbol{\Theta}^t\}$. That is, there exists a subsequence $\mathcal{K}$ such that:

$$\lim_{k \in \mathcal{K}} \boldsymbol{\Theta}^k = \boldsymbol{\Theta}^a.$$

The sequence of objective function values monotonically decreases and is bounded below, so the sequence of objective function values converges:

$$\lim_{k \to \infty} \text{PSSE}(\boldsymbol{\Theta}^k) + \mu \Upsilon_3(\mathcal{P}_1^k, \mathcal{P}_2^k) = \lim_{k \in \mathcal{K}} \text{PSSE}(\boldsymbol{\Theta}^k) + \mu \Upsilon_3(\mathcal{P}_1^k, \mathcal{P}_2^k)$$
$$= \text{PSSE}(\boldsymbol{\Theta}^a) + \mu \Upsilon_3(\mathcal{P}_1^a, \mathcal{P}_2^a).$$

Passing the limit to (A.4), we obtain

$$\lim_{k \in \mathcal{K} \to \infty} \|\boldsymbol{\Theta}^{k+1} - \boldsymbol{\Theta}^k\|^2 = 0.$$

At the tth iteration, $\boldsymbol{\Theta}^{t+1}$ is the minimizer of the problem (6), which satisfies the optimality condition

$$0 \in \nabla \text{PSSE}(\boldsymbol{\Theta}^t) + \rho(\boldsymbol{\Theta}^{t+1} - \boldsymbol{\Theta}^t) + \partial \mu \Upsilon_3(\mathcal{P}_1^{t+1}, \mathcal{P}_2^{t+1}).$$

Passing the limit on t, we have

$$0 \in \nabla \text{PSSE}(\boldsymbol{\Theta}^a) + \partial \mu \Upsilon_3(\mathcal{P}_1^a, \mathcal{P}_2^a).$$

The limit point $\boldsymbol{\Theta}^a$ is a critical point of (5).

References

Agarwal, A., Negahban, S., Wainwright, M.J., 2012. Noisy matrix decomposition via convex relaxation: Optimal rates in high dimensions. The Annals of Statistics 40 (2), 1171–1197.

Bai, J., Li, K., 2002. Forecasting using princpal components from a large number of predcitors. Journal of the American Statistical Association 97 (460), 1167–1179.

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

Baldassarre, L., Mourao-Miranda, J., Pontil, M., 2012. Structured sparsity models for brain decoding from fmri data. In: 2012 Second International Workshop on Pattern Recognition in NeuroImaging. IEEE, pp. 5–8.

Bandettini, P.A., 2012. Twenty years of functional mri: the science and the stories. Neuroimage 62 (2), 575–588.

Beloozerova, I.N., Sirota, M.G., Swadlow, H.A., 2003. Activity of different classes of neurons of the motor cortex during locomotion. Journal of Neuroscience 23 (3), 1087–1097.

de Boor, C., 2001. A practical guide to splines (applied mathematical sciences vol. 27).

Bowman, F.D., Caffo, B., Bassett, S.S., Kilts, C., 2008. A bayesian hierarchical framework for spatial modeling of fmri data. NeuroImage 39 (1), 146–156.

Bush, G., Luu, P., Posner, M.I., 2000. Cognitive and emotional influences in anterior cingulate cortex. Trends in cognitive sciences 4 (6), 215–222.

Bush, G., Vogt, B.A., Holmes, J., Dale, A.M., Greve, D., Jenike, M.A., Rosen, B.R., 2002. Dorsal anterior cingulate cortex: a role in reward-based decision making. Proceedings of the National Academy of Sciences 99 (1), 523–528.

Buxton, R.B., Frank, L.R., 1997. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. Journal of Cerebral Blood Flow & Metabolism 17 (1), 64–72.

Calhoun, V., Stevens, M., Pearlson, G., Kiehl, K., 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., Maldjian, J., 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., Laurienti, P., Maldjian, J., 2009. Evaluating the impact of spatio-temporal smoothness constraints on the BOLD hemodynamic response function estimation: an analysis based on Tikhonov regularization. Physiological Measurement 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, 180–188.

Cheng, K., Waggoner, R.A., Tanaka, K., 2001. Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging. Neuron 32 (2), 359–374.

Coan, J., 2010. Adult attachment and the brain. Journal of Social and Personal Relationships 27, 210–217.

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

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

Coan, J., Schaefer, H., Davidson, R., 2006. Lending a hand: Social regulation of the neural response to threat. Psychological Science 17, 1032–1039.

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

David, A., Guillemain, I., Saillet, S., Reyt, S., Deransart, C., Segebarth, C., Depaulis, A., 2008. Identifying neural drivers with functional mri: an electrophysio-logical validation. PLoS biology 6 (12), e315.

Dedovic, K., Slavich, G.M., Muscatell, K.A., Irwin, M.R., Eisenberger, N.I., 2016. Dorsal anterior cingulate cortex responses to repeated social evaluative feed-back in young women with and without a history of depression. Frontiers in behavioral neuroscience 10.

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

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

Ethofer, T., Bretscher, J., Wiethoff, S., Bisch, J., Schlipf, S., Wildgruber, D., Kreifelts, B., 2013. Functional responses and structural connections of cortical areas for processing faces and voices in the superior temporal sulcus. Neuroimage 76, 45–56.

Eubank, R.L., 1999. Nonparametric regression and spline smoothing. CRC press.

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

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

Friston, K., Fletcher, P., Josephs, O., Holmes, A., Rugg, M., Turner, R., 1998. Event-related fMRI: characterizing differential responses. NeuroImage 7, 30–40.

Friston, K., Holmes, A., Worsley, K., Poline, P., Frith, C., Frackowiak, R., 1995. Statistical parametric maps in functional imaging: A general linear approach. Human Brain Mapping 2, 189–210.

Friston, K., Jezzard, P., Turner, R., 1994. Analysis of functional MRI time-series. Human Brain Mapping 1, 153–171.

Friston, K., Josephs, O., Zarahn, E., Holmes, A., Rouquette, S., Poline, J., 2000. To smooth or not to smooth? NeuroImage 12, 196–208.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P., Frith, C.D., Frackowiak, R.S., 1994. Statistical parametric maps in functional imaging: a general linear approach. Human brain mapping 2 (4), 189–210.

Ge, R., Wang, Y., Zhang, J., Yao, L., Zhang, H., Long, Z., 2016. Improved fastica algorithm in fmri data analysis using the sparsity property of the sources. Journal of neuroscience methods 263, 103–114.

Glover, G., 1999. Deconvolution of impulse response in event-related BOLD fMRI. NeuroImage 9, 416–429.

Gong, P., Zhang, C., Lu, Z., Huang, J., Ye, J., 2013. A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems.. ICML'13 Proceedings of the 30th International Conference on International Conference on Machine Learning 28, 37–45.

Goutte, C., Nielsen, F., Hansen, L., 2000. Modeling the hemodynamic response in fmri using smooth fir filters. IEEE Transactions on Medical Imaging 19, 1188–1201.

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17 (2), 825–841.

Knutson, B., Westdorp, A., Kaiser, E., Hommer, D., 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., Anderson, J., Nielsen, F., Holmes, A., Kolenda, T., Savoy, R., Hansen, L.K., 1999. Plurality and resemblance in fMRI data analysis. NeuroImage 10, 282–303.

Liao, C., Worsley, K., Poline, J., Aston, J., Duncan, G., Evans, A., 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. Journal of Machine Learning Research 15 (1), 3447–3481.

Lindquist, M., 2008. The statistical analysis of fmri data. Statistical Science 23, 439–464.

Lindquist, M., Wager, T., 2007. Validity and power in hemodynamic response modelling: a comparison study and a new approach. Human Brain Mapping 28, 764–784.

MacNamara, A., Foti, D., Hajcak, G., 2009. Tell me about it: neural activity elicited by emotional pictures and preceding descriptions.. Emotion 9 (4), 531.

Makni, S., Beckmann, C., Smith, S., Woolrich, M., 2008. Bayesian deconvolution of fmri data using bilinear dynamical systems. NeuroImage 42 (4), 1381–1396.

Makni, S., Ciuciu, P., Idier, J., Poline, J.-B., 2005. Joint detection-estimation of brain activity in functional mri: a multichannel deconvolution solution. IEEE Transactions on Signal Processing 53 (9), 3488–3502.

Marrelec, G., Benali, H., Ciuciu, P., Pelegrini-Issac, M., Poline, J., 2003. Robust estimation of the hemodynamic response function in event-related BOLD fMRI using basic physiological information. Human Brain Mapping 19, 1–17.

Marrelec, G., Benali, H., Ciuciu, P., Poline, J., 2001. Bayesian estimation of the hemodynamic of the hemodynamic response function in functional MRI. AIP Conference Proceedings 617, 229–247.

Mazard, A., Mazoyer, B., Etard, O., Tzourio-Mazoyer, N., Kosslyn, S., Mellet, E., 2002. Impact of fmri acoustic noise on the functional anatomy of visual mental imagery. Journal of Cognitive Neuroscience 14 (2), 172–186.

Nesterov, Y., 2013. Gradient methods for minimizing composite objective function.. Mathematical Programming 140 (1), 125–161.

Ogawa, S., Tank, D.W., Menon, R., Ellermann, J.M., Kim, S.G., Merkle, H., Ugurbil, K., 1992. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proceedings of the National Academy of Sciences 89 (13), 5951–5955.

Olshausen, B.A., Field, D.J., 1996. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381 (6583), 607–609.

Olshausen, B.A., Field, D.J., 2004. Sparse coding of sensory inputs. Current opinion in neurobiology 14 (4), 481–487.

Pedregosa, F., Eickenberg, M., Ciuciu, P., Thirion, B., Gramfort, A., 2015. Data-driven hrf estimation for encoding and decoding models. NeuroImage 104, 209–220.

Pereira, M.G., de Oliveira, L., Erthal, F.S., Joffily, M., Mocaiber, I.F., Volchan, E., Pessoa, L., 2010. Emotion affects action: midcingulate cortex as a pivotal node of interaction between negative emotion and motor signals. Cognitive, Affective, & Behavioral Neuroscience 10 (1), 94–106.

Rendle, S., 2012. Factorization machines with libfm. ACM Transactions on Intelligent Systems and Technology (TIST) 3, 3–57.

Ruppert, D., Wand, M.P., Carroll, R.J., 2003. Semiparametric regression. Cambridge university press.

Samdin, S.B., Ting, C.-M., Ombao, H., Salleh, S.-H., 2017. A unified estimation framework for state-related changes in effective brain connectivity. IEEE Transactions on Biomedical Engineering 64 (4), 844–858.

Szameitat, A.J., Shen, S., Sterr, A., 2009. The functional magnetic resonance imaging (fmri) procedure as experienced by healthy participants and stroke patients–a pilot study. BMC medical imaging 9 (1), 14.

Vakorin, V., Borowsky, R., Sarty, G., 2007. Characterizing the functional MRI response using Tikhonov regularization. Statistics in Medicine 26 (21), 3830–3844.

Vincent, T., Risser, L., Ciuciu, P., 2010. Spatially adaptive mixture modeling for analysis of fmri time series.. IEEE Transactions on Medical Imaging 29 (4), 1059–1074.

Vogt, B., 2008. Architecture, cytology and comparative organization of primate cingulate cortex. Cingulate neurobiology and disease (Vogt B, ed). Oxford: Oxford UP.

Vogt, B.A., 2009. Regions and subregions of the cingulate cortex. Cingulate neurobiology and disease 1.

Vogt, B.A., 2016. Midcingulate cortex: structure, connections, homologies, functions and diseases. Journal of chemical neuroanatomy 74, 28–46.

Wahba, G., 1990. Spline Models for Observational Data. SIAM, Philadelphia.

Wixted, J.T., Squire, L.R., Jang, Y., Papesh, M.H., Goldinger, S.D., Kuhn, J.R., Smith, K.A., Treiman, D.M., Steinmetz, P.N., 2014. Sparse and distributed coding of episodic memory in neurons of the human hippocampus. Proceedings of the National Academy of Sciences 111 (26), 9621–9626.

Worsley, K., Friston, K., 1995. Analysis of fMRI time-series revisited again. NeuroImage 2, 173–181.

Worsley, K., Liao, C., Aston, J., Petre, V., Duncan, G., Morales, F., Evans, A., 2002. A general statistical analysis for fMRI data. NeuroImage 15, 1–15.

Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68 (1), 49–67.

Zhang, C., Jiang, Y., Yu, T., 2007. A comparative study of one-level and two-level semiparametric estimation of hemodynamic response function for fMRI data. Statistics in Medicine 26, 3845–3861.

Zhang, L., Guindani, M., Versace, F., Engelmann, J.M., Vannucci, M., et al., 2016. A spatiotemporal nonparametric bayesian model of multi-subject fmri data. The Annals of Applied Statistics 10 (2), 638–666.

Zhang, L., Guindani, M., Versace, F., Engelmann, J.M., Vannucci, M., et al., 2016. A spatiotemporal nonparametric bayesian model of multi-subject fmri data. The Annals of Applied Statistics 10 (2), 638–666.

Zhang, L., Guindani, M., Versace, F., Vannucci, M., 2014. A spatio-temporal nonparametric bayesian variable selection model of fmri data for clustering correlated time courses. NeuroImage 95, 162–175.

Zhang, L., Guindani, M., Versace, F., Vannucci, M., 2014. A spatio-temporal nonparametric bayesian variable selection model of fmri data for clustering correlated time courses. NeuroImage 95, 162–175.

Zhang, T., Li, F., Beckes, L., Coan, J., 2013. A semi-parametric model of the hemodynamic response for multi-subject fmri data. NeuroImage 75, 136–145.

Zhang, T., Pham, M., Sun, J., Yan, G., Li, H., Sun, Y., Gonzalez, M.Z., Coan, J.A., 2018. A low-rank multivariate general linear model for multi-subject fmri data and a non-convex optimization algorithm for brain response comparison. NeuroImage 173, 580–591.

Zhu, D., Zhang, T., Jiang, X., Hu, X., Chen, H., Yang, N., Lv, J., Han, J., Guo, L., Liu, T., 2014. Fusing dti and fmri data: a survey of methods and applications. NeuroImage 102, 184–191.