

Compressed Dictionary Learning for Detecting Activations in fMRI using Double Sparsity

Shuangjiang Li, Hairong Qi
Electrical Engineering and Computer Science
University of Tennessee, Knoxville, TN 37996
{shuangjiang, hqi}@utk.edu

Abstract—This paper focuses on detecting activated voxels in fMRI data by exploiting the sparsity of the BOLD signal. Due to the large volume of the data, we propose to learn a dictionary from the compressed measurements of the BOLD signal. The solution to the inverse problem induced by the General Linear Model is then sought through sparse coding using the double sparsity model, where sparsity is imposed on both the learnt dictionary and the generated coefficients. The estimated sparse coefficients are then used to decide whether or not a stimulus is presented in the observed BOLD signal. Experimental results on real fMRI data demonstrate that the proposed method leads to similar activated regions as compared to those activated by the Statistical Parametric Mapping (SPM) software but with much less samples needed.

Index Terms—Compressed sensing, fMRI activation detection, dictionary learning, double sparsity

I. INTRODUCTION

In functional Magnetic Resonance Imaging (fMRI), an fMRI scanner measures the blood-oxygenation-level-dependent (BOLD) signal at all points in a 3-D grid. It has been widely accepted as a noninvasive technique for localizing brain voxels activation. A typical fMRI dataset is usually composed of time series (BOLD signals) of tens of thousands voxels [1]. Such high volume has become quite a burden for existing fMRI research where statistical analysis is often conducted to detect which voxels are activated by a task or stimulus.

The General Linear Model (GLM) is a classical univariate approach toward the detection of task-related activations in the brain. Under this approach, a linear model is fit to the fMRI time series of each voxel resulting in a set of voxel-specific parameters, which are then used to form statistical parametric maps (SPMs) [2].

Recently, there has been a growing interest in the fMRI data analysis based on sparse representation of BOLD signal [3]–[5]. Specially, Daubechies, et al. [3] showed that the most influential factor for the ICA algorithm is the sparsity of the components rather than independence, and suggested to develop decomposition methods based on the GLM where the BOLD signal may be regarded as a linear combination of a sparse set of brain activity patterns. As a result, Lee et al. [5] proposed a data-driven sparse GLM for fMRI data analysis, Guillen et al. [4] utilized the ℓ_0 -regularized Least Absolute Deviation (ℓ_0 -LAD) to solve for sparse representation coefficient under the GLM setting.

Compressed Sensing (CS) is a new methodology to capture signals at lower rate than the Nyquist sampling rate when the signals are sparse or sparse in some domain [6], [7]. In [8], the sparsity of MRI signals has been exploited that resulted in the reduction of the scan time and the improvement of the resolution of MR imagery.

In this paper, we propose a compressed dictionary learning (CDL) approach, where the problem of detecting activated voxels is addressed using sparse representation of BOLD signals with a learnt dictionary. The high volume problem of the signal is handled using CS, where the dictionary is learnt only from the compressed measurements of the BOLD signal. The solution to the inverse problem induced by GLM is then sought through sparse coding using the double sparsity model, where sparsity is imposed on both the learnt dictionary and the BOLD signal sparse representation coefficients. Our contribution in this work is two-fold. First, we use the compressed measurements to learn the dictionary. Second, we adopt sparse coding based on double sparsity model. Specifically, our method requires much less samples to effectively detect the voxel activations. In addition, no prior assumptions of the task-related stimulus (i.e., design matrix) are needed since it is automatically incorporated in the learnt dictionary. The avoidance of the task-related stimulus is the key enabler for practical adoption of the proposed approach. For example, the patient that goes through the fMRI experiment for medical treatment could be in coma, the task-related stimulus would not function in this case.

The rest of this paper is organized as follows. In Section II, we introduce the GLM approach and how to detect activations based on the GLM output. In Section III, we discuss the proposed compressed dictionary learning (CDL) approach for activation detection. Performance evaluation on using SPM analysis and CDL are shown in Section IV. Section V concludes the paper.

II. THE GENERAL LINEAR MODEL APPROACH

A widely used statistical method for analyzing fMRI time series is the general linear model. It models the time series as a linear combination of several different signal components and tests whether activity in a brain region is systematically related to any of these known input functions for each voxel in an fMRI imaging system. More precisely, the GLM for the observed response variable y_j at voxel j , $j = 1, \dots, N$, is

given by:

$$y_j = X\beta_j + e_j \quad (1)$$

where, $y_j \in \mathbb{R}^M$ with M being the number of scans, $X \in \mathbb{R}^{M \times L}$ denotes the design matrix, $\beta_j \in \mathbb{R}^L$ represents the signal strength at the j -th voxel, and $e_j \in \mathbb{R}^M$ is the noise. Typically, each column of the design matrix X is defined by the task/stimulus-related function convolved with a hemodynamic response function (HRF), i.e., the predicted task related BOLD response [2]. Basically, the design matrix is constructed with the expected task-related BOLD response as described above. Sometimes we also add extra columns to the design matrix with nuisance components that model the confounding effects. The stimulus is assumed to be equivalent to the experimental paradigm, while the HRF is modeled using a canonical HRF, typically either a gamma function or the difference between two gamma functions [2].

Under GLM, various methods for estimating β may be used. The Ordinary Least Squares (OLS) has been traditionally adopted where no prior information is applied:

$$\beta_j = (X^T X)^{-1} X^T y_j \quad (2)$$

In order to identify columns of interests that corresponding to the task-related design in the contribution of the BOLD signal, a contrast vector $c = [c_1, c_2, \dots, c_L]$ is applied on the estimated coefficient $\hat{\beta}_j$ by $c^T \hat{\beta}_j$. This hypothesis testing is then performed on a voxel-by-voxel basis using either a t -test or F-test. The resulting test statistic will then be calculated and formatted in an image termed statistical parametric map (SPM).

III. THE CDL APPROACH

In this section, we first give an illustration example on the result generated using CDL as compared with existing algorithms based on fixed design matrix. We then present the CDL problem formulation as well as methods on how to solve it by transforming it into the standard LASSO problem [9].

A. A Motivating Example

In order to gain some insight on the performance of the OLS and ℓ_0 -LS (i.e., sparse decomposition but still using fixed design matrix) and the proposed CDL approach when the parameter vector is sparse, we generate some synthetic BOLD signal, as shown in Fig. 1. We model the fMRI time series z of a particular voxel as a sparse linear combination of various stimuli and additive noise. That is, $z = X\alpha + \epsilon$, where α is a sparse vector of length $L = 13$ and a support 3 (i.e., only 3 entries in α are non-zero). Thus, the voxel related to the observed time series z is activated by three stimuli. Given a design matrix X which corresponds to an experimental design, we are interested in identifying the stimuli in the design matrix that activate the corresponding voxel as well as the contributions of those stimuli (i.e., coefficients). We follow the same procedures as in [4] and use all 13 convolved task functions obtained from Pittsburgh Brain Activity Interpretation Competition 2007 (PBAIC 2007) [10] to generate

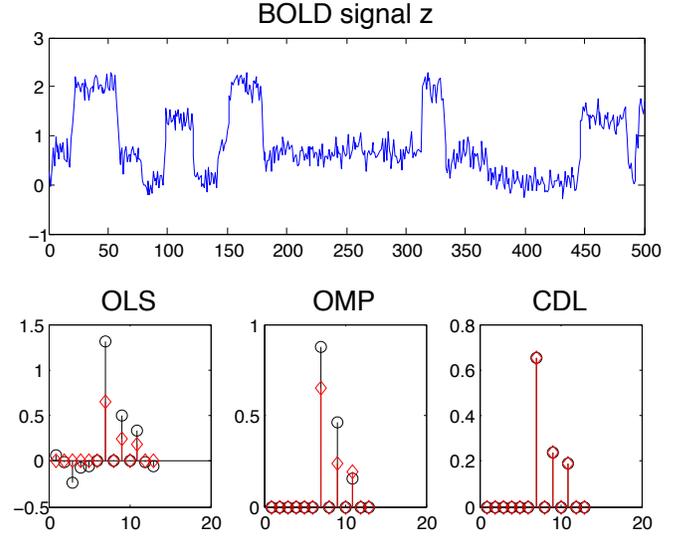


Fig. 1: Solution of the inverse problem: $z = X\alpha + \epsilon$. Top: observed time series z . Bottom: solutions obtained by OLS, OMP, and CDL; here \diamond denotes the original parameter vector and \circ denotes the estimated solution. CDL uses 250 projected samples from z , the sparse solution is truncated to show only the first 13 entries in this case.

the design matrix $X \in \mathbb{R}^{500 \times 13}$. We then add a 9dB noise to the signal. From Fig. 1, we see that OLS generates many entries that are not sparse. ℓ_0 -LS, implemented using OMP [11], successfully detects the right support of the sparse signal, but it fails to estimate the contribution of the stimuli, α . For the proposed CDL method, a compressed measurement matrix $\Phi \in \mathbb{R}^{250 \times 500}$ is randomly generated as Gaussian random matrix, and a dictionary $D \in \mathbb{R}^{500 \times 500}$ is obtained from the design matrix X . CDL then generates a sparse estimation of α with 500 entries. We observe that CDL does correctly identify the sparse support as well as contributions of the stimuli.

TABLE I: List of variable notations.

$Y_{M \times N}$	BOLD signal of N voxels
$D_{M \times p}$	The dictionary of p atoms
$A_{p \times N}$	Set of N coefficient vectors
$Q_{K \times N}$	Set of N projected measurements
$\Phi_{K \times M}$	The measurement matrix
$\Psi_{M \times M}$	The basis for the dictionary D
$\Theta_{M \times p}$	Set of p sparse coefficient vectors

B. CDL - Problem Formulation

Table I outlines the notations that are used in this paper. Given a BOLD signal matrix $Y \in \mathbb{R}^{M \times N}$ (N is the total number of voxels), contrary to the design matrix X in the GLM approach, the dictionary learning approach tries to learn a dictionary $D \in \mathbb{R}^{M \times p}$ and its corresponding coefficient matrix $A \in \mathbb{R}^{p \times N}$ as follows:

$$\min_{D, A} \left\{ \frac{1}{2} \|Y - DA\|_2^2 + \lambda_A \|A\|_1 \right\} \quad (3)$$

This can be efficiently solved by recursively updating the sparse coefficients A and the dictionary D [12]. First, given the BOLD signal Y , an intermediate sparse approximation with respect to the dictionary $D^{(t-1)}$ from step $t-1$ is computed by solving the following LASSO problem:

$$\min_{A^{(t)}} \left\{ \frac{1}{2} \|Y - D^{(t-1)} A^{(t)}\|_2^2 + \lambda_A \|A^{(t)}\|_1 \right\} \quad (4)$$

The dictionary is subsequently updated to minimize the representation error while $A^{(t)}$ is fixed:

$$D^{(t)} = \arg \min_{D^{(t)}} \left\{ \frac{1}{2} \|Y - D^{(t)} A^{(t)}\|_2^2 \right\} \quad (5)$$

Since the BOLD signal Y is of high volume, in this work, we are interested in the case where only a linear projection of Y onto a measurement matrix Φ is available. Under this compressed measurement setting, $A^{(t)}$ is computed using compressed samples as will be explained later. Then the dictionary update step in Eq. (5) becomes the following under-determined problem:

$$\min_{D^{(t)}} \left\{ \frac{1}{2} \|Q - \Phi D^{(t)} A^{(t)}\|_2^2 \right\}, \text{ s.t. } Q = \Phi Y \quad (6)$$

which does not have unique solution for $D^{(t)}$ for a CS measurement matrix $\Phi \in \mathbb{R}^{K \times M}$ which has less rows than columns. In what follows, we will discuss how to add additional sparse structure constraint on the dictionary D to help us solve Eq. (6).

C. Sparse Dictionary Model

The sparse dictionary model suggests that each atom of the dictionary has itself a sparse representation over some prespecified base dictionary Ψ [13]. The dictionary is therefore expressed as:

$$D = \Psi \Theta \quad (7)$$

where $\Psi \in \mathbb{R}^{M \times M}$ is the basis and Θ is the atom representation matrix, assumed to be sparse. Since D mimics the design matrix, Ψ is prespecified and served as the basis of D which obviously affects the success of the entire model, and thus we prefer to initialize Ψ with some prior knowledge about the data (i.e., information from the experimental paradigm). In comparison to the learnt dictionary from Eq. (5). The dictionary model in Eq. (7) provides adaptability via the sparse matrix Θ , which can be viewed as an extension to the existing dictionaries, adding a new layer of adaptivity.

By substituting the $D = \Psi \Theta$ with a sparse Θ , Eq. (3) now becomes:

$$\min_{D, \Theta} \left\{ \frac{1}{2} \|Y - \Psi \Theta A\|_2^2 + \lambda_A \|A\|_1 + \lambda_\Theta \|\Theta\|_1 \right\} \quad (8)$$

D. The proposed CDL approach

Similar to the dictionary learning approach in sec. III-B, there are two steps in the CDL algorithm. In the sparse coding step, the dictionary $D^{(t-1)}$ is fixed and obtained from the

previous iteration. The sparse coefficient $A^{(t)}$ can be obtained by minimizing the following problem:

$$\min_{A^{(t)}} \left\{ \frac{1}{2} \|Q - \Phi D^{(t-1)} A^{(t)}\|_2^2 + \lambda_A \|A^{(t)}\|_1 \right\} \quad (9)$$

Optimizing over $A^{(t)}$ is straightforward LASSO problem. While in the dictionary update step, the optimization problem becomes:

$$\min_{\Theta^{(t)}} \left\{ \frac{1}{2} \|Q - \Phi \Psi \Theta^{(t)} A^{(t)}\|_2^2 + \lambda_\Theta \|\Theta^{(t)}\|_1 \right\} \quad (10)$$

Here, optimizing over $\Theta^{(t)}$ is not directly LASSO which requires the following Lemma to reformulate into the standard LASSO problem.

Lemma 1. *Let $Q \in \mathbb{R}^{K \times N}$ and $\Phi \in \mathbb{R}^{K \times M}$ be two matrices, and $u \in \mathbb{R}^M$ and $v \in \mathbb{R}^N$ be two vectors. Also assume that $v^T v = 1$. Then the following holds:*

$$\|Q - \Phi u v^T\|_2^2 = \|Qv - \Phi u\|_2^2 + f(Q, v). \quad (11)$$

The detailed derivation can be found in [13]. Based on Lemma 1, each column of $\Theta^{(t)}$, denoted as $\theta_j^{(t)}$, in Eq. (10) can be solved by the following LASSO-like problem:

$$\theta_j^{(t)} = \arg \min_{\theta_j^{(t)}} \left\{ \frac{1}{2} \|E_{\theta_j^{(t)}} a_j^{(t)T} - \Phi \Psi \theta_j^{(t)}\|_2^2 + \lambda_\Theta \|\theta_j^{(t)}\|_1 \right\} \quad (12)$$

where $E_{\theta_j^{(t)}}$ is the projected estimation error associated with the dictionary atom θ_j and $a_j^{(t)}$ is the j -th column of matrix $A^{(t)}$ as follows:

$$E_{\theta_j^{(t)}} := Q - \sum_{i=1, i \neq j}^p \Phi \theta_i^{(t-1)} a_j^{(t)} \quad (13)$$

Now Eq. (10) can be efficiently solved using Eq. (12) and Eq. (13). Also Lemma 1 requires that the column of the dictionary $D^{(t)}$ should be in unit norm, which is done by re-scaling the learnt dictionary in each iteration.

IV. EXPERIMENTS

In this section, we demonstrate the result comparison on activation detection using the GLM with a design matrix and the CDL with a learnt dictionary. We first present the data used as well as how the activation maps are generated using SPM software then we give the experimental results with some discussions.

A. Data

We use the dataset from Pittsburgh Brain Activity Interpretation Competition 2007 (PBAIC 2007) [10]. The brain images were collected from three subjects using a Siemens 3T Allegra scanner. The BOLD signal was acquired by using EpiBOLD sequence, with the imaging parameters TR and TE being set to 1.75s and 25ms. Each subject's data consists of three runs lasting for approximately 20 minutes with 704 volumes in each segment. Each volume contains $64 \times 64 \times 34$ voxels with a voxel size of $3.2 \times 3.2 \times 3.5$ mm³. In this experiment, we use the preprocessed data where slice time correction,

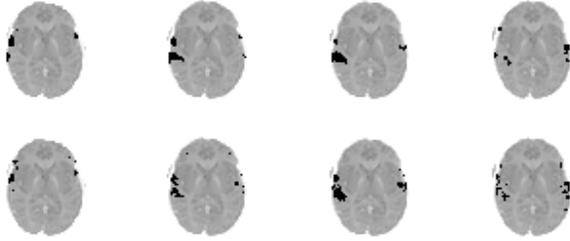


Fig. 2: Activation maps for the *Instructions* task. Top: results generated using SPM with design matrix $X \in \mathbb{R}^{500 \times 14}$, Bottom: results generated using CDL method, with Gaussian measurement matrix $\Phi \in \mathbb{R}^{250 \times 500}$ and a learnt dictionary $D \in \mathbb{R}^{500 \times 500}$. Slice number from left to right are 13, 14, 15, and 16 in both rows.

motion correction and detrending have been performed on the functional and structural data using NeuroImage software (AFNI). A fixed period is extracted from the preprocessed dataset leading to a total of 500 volumes in each run.

For the design matrix X , the first 13 columns of X are constructed by considering the thirteen convolved stimuli/task function that are part of the features set provided by PBAIC 2007¹, done by the SPM software package². A temporal smoothing is applied to the data in order to remove the low frequency variations (i.e., confounds) due to artifacts such as aliased biorhythms as well as other drift terms. The highpass filter is obtained from DCT basis with a cut-off frequency set as 1/128 Hz. We also add one column of all ones that models the whole brain activity [14]. The design matrix $X \in \mathbb{R}^{500 \times 14}$ is then used in SPM to generate the activation maps for comparison purpose.

After generating the coefficients β from the GLM, for each voxel, the contrast vector c is applied to estimate the BOLD response to a single stimulus. For example, to evaluate the *instructions* task in the PBAIC 2007 dataset, all the entries of the contrast vector c is set to zero except for the sixth that is set to one. Finally, by calculating the $c^T \beta$, an activation map will be generated corresponding to the specific task.

B. Results

In order to generate the sparse coefficients using the CDL approach, the measurement matrix is randomly generated using the Gaussian i.i.d measurement matrix with the CS measurement ratio set as 0.5 (i.e., $K/M = 0.5$ in Φ). The basis Ψ for the dictionary is randomly generated using DCT coefficients, with the first 13 columns from the design matrix X used in SPM, and $p = 500$. We set $\lambda_A = \lambda_\Theta = 0.1$, and use the SPAMS [12] software package for solving the LASSO. Fig. 2 shows the activation maps from both methods, while the detailed comparisons are listed in Table II. This result confirms that CDL enjoys a very similar detected activation maps as compared with SPM while using much less data samples.

¹<http://www.lrdc.pitt.edu/ebc/2007/docs/CompetitionGuideBook2007v7.pdf>

²<http://www.fil.ion.ucl.ac.uk/spm/>

	Activated slice indices (totally 34 slices)	Avg. slice-wise matches (%)	Avg. voxel-wise matches (%)
SPM	5-22	83.33%	50.14%
CDL	3, 4, 8-22		

TABLE II: Detected activations comparison of SPM and CDL.

V. CONCLUSIONS

In this paper, we presented CDL, a compressed dictionary learning approach for detecting activations in fMRI data. The double sparsity model was applied in solving the inverse problem induced by the general linear model in the analysis, where sparsity was imposed on both the learnt dictionary and the sparse representation of the BOLD signal. Compressed sensing measurements were used for learning the dictionary instead of the entire BOLD signal and thus reducing the data volume to be processed. Experimental results on real fMRI data demonstrated that CDL could successfully detect the activated voxels similar to the results generated by the SPM software but with much less data samples used.

ACKNOWLEDGEMENT

The authors would like to thank Dr. Blanca Guillen at the Department of Mathematics, Bioengineering Group, UNET, San Cristóbal, Venezuela for the intriguing discussions and the help on the experiment.

REFERENCES

- [1] Y. Li, P. Namburi, Z. Yu, C. Guan, J. Feng, and Z. Gu, "Voxel selection in fMRI data analysis based on sparse representation," *Biomedical Engineering, IEEE Trans. on*, vol. 56, no. 10, pp. 2439–2451, 2009.
- [2] M. A. Lindquist *et al.*, "The statistical analysis of fMRI data," *Statistical Science*, vol. 23, no. 4, pp. 439–464, 2008.
- [3] I. Daubechies, E. Roussos, S. Takerkart, M. Benharrosh, C. Golden, K. D'ardenne, W. Richter, J. Cohen, and J. Haxby, "Independent component analysis for brain fMRI does not select for independence," *Proceedings of the National Academy of Sciences*, vol. 106, no. 26, pp. 10415–10422, 2009.
- [4] B. Guillen, J. L. Paredes, and R. Medina, "A sparse based approach for detecting activations in fMRI," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 7816–7819.
- [5] K. Lee, S. Tak, and J. C. Ye, "A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion," *Medical Imaging, IEEE Trans. on*, vol. 30, no. 5, pp. 1076–1089, 2011.
- [6] D. Donoho, "Compressed sensing," *Information Theory, IEEE Trans. on*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [7] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Trans. on*, vol. 52, no. 2, pp. 489–509, 2006.
- [8] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic resonance in medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [9] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani *et al.*, "Least angle regression," *The Annals of statistics*, vol. 32, no. 2, pp. 407–499, 2004.
- [10] Pittsburgh-EBC-Group, "PBAIC homepage," <http://pbc.lrdc.pitt.edu/>.
- [11] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Information Theory, IEEE Trans. on*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [12] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009, pp. 689–696.
- [13] R. Rubinstein, M. Zibulevsky, and M. Elad, "Double sparsity: Learning sparse dictionaries for sparse signal approximation," *Signal Processing, IEEE Trans. on*, vol. 58, no. 3, pp. 1553–1564, 2010.
- [14] R. A. Poldrack, J. A. Mumford, and T. E. Nichols, *Handbook of functional MRI data analysis*. Cambridge University Press, 2011.