Compressed Dictionary Learning for Detecting Activations in fMRI using Double Sparsity

Shuangjiang Li, Hairong Qi

Department of Electrical Engineering and Computer Science
University of Tennessee, Knoxville

Dec. 4, 2014

sli22@vols.utk.edu

The 2nd IEEE Global Conf. on Signal and Info. Processing (GlobalSIP)
December 3-5, 2014. Atlanta, Georgia, USA

Outline

Outline

Background and Motivation

-fMRI

- A non-invasive technique for studying brain activity.
- During the course of an fMRI experiment, a series of brain images are acquired while the subject performs a set of tasks.

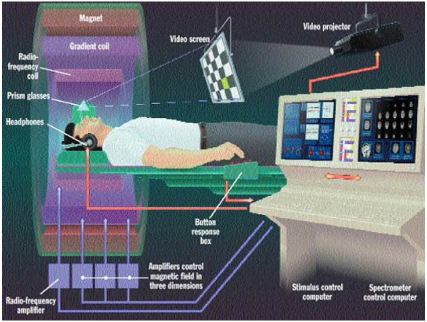


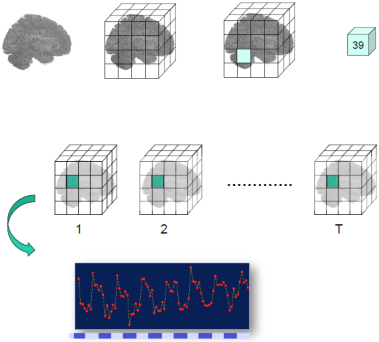
Figure 1: Example of a fMRI course[1].

[1]http://www.metinc.net/products/FMRI/products/img/fmri-sys.jpg

Background and Motivation

-fMRI Data

- Each image consists of 100,000 'voxels' (cubic volumes that span the 3D space of the brain). Each voxel corresponds to a spatial location and has a number associated with it that represents its intensity.
- During the course of an experiment several hundred images are acquired ($\approx$ one every 2s).



Background and Motivation

-GLM Analysis

- The General Linear Model (GLM) is a classical univariate approach toward the detection of task-related activations in the brain.

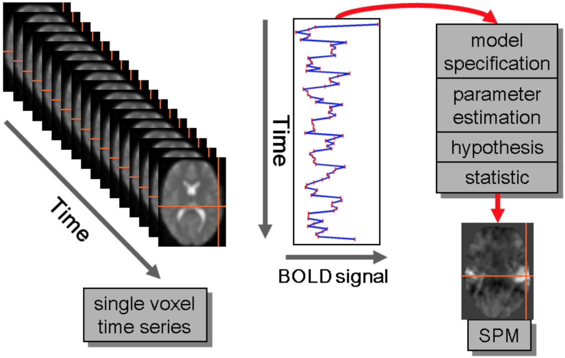


Figure 2: Computing activations based on GLM.

Background and Motivation

-Motivation

- A typical fMRI dataset is usually composed of time series, the blood-oxygenation-level-dependent (BOLD) signal, of tens of thousands voxels. Such high volume has become quite a burden for existing fMRI research. ⇒ Compressed Sensing
- Prof. Daubechies, et al. [2] 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. ⇒ Sparsity

[2] I. Daubechies, et. al "Independent component analysis for brain fMRI does not select for independence," PNAS, 2009

Outline

The General Linear Model (GLM)

- General Linear Model (GLM) 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.
- The GLM for the observed response variable y_j at voxel j, $j = 1, \cdots, N$, is given by:

$$\boxed{y_j = X\beta_j + e_j} \tag{1}$$

  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.
- Each column of the design matrix X is defined by the task/stimulus-related function convolved with a hemodynamic response function (HRF), typically either a gamma function or the difference between two gamma functions.

The General Linear Model (GLM)

- 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:

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

- In order to identify columns of interests that corresponding to the task-related design in the contribution of the BOLD signal, a contrast vector $c = [c_1, c_2, \cdots, c_L]$ 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).

Outline

A Motivating Example

We first give an illustration example on the result generated using CDL as compared with existing algorithms based on fixed design matrix X.

- 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. 3
- 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).

A Motivating Example

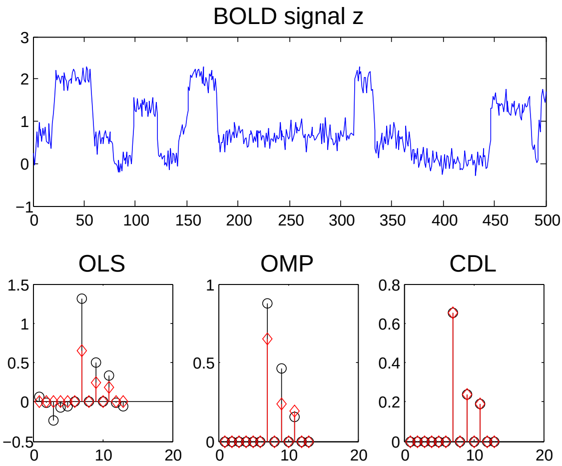


Figure 3: Solution of the inverse problem: $z = X\alpha + \epsilon$. Top: observed time series z. Bottom: solutions obtained by OLS, OMP, and CDL; here $\diamond$ denotes the original parameter vector and $\circ$ denotes the estimated solution. CDL uses 250 projected samples from z, the sparse solution is truncated to show only the first

A Motivating Example

- OLS generates many entries that are not sparse.
- ℓ_0-LS, implemented using OMP, 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.

CDL - Problem Formulation

-List of notations

Table 1: 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

CDL - Problem Formulation

- 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:

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

- This can be efficiently solved by recursively updating the sparse coefficients A and the dictionary D [3].
- 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:

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

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

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

[3] J. Mairal et al., "Online dictionary learning for sparse coding," in Proc. of the 26th Annual Intl. Conf. on Machine Learning. 2009

CDL - Problem Formulation

- 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.
- Then the dictionary update step in Eq. (5) becomes the following under-determined problem:

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

  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).

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 [4]. The dictionary is therefore expressed as:

$$\boxed{D = \Psi\Theta} \tag{7}$$

  where $\Psi \in \mathbb{R}^{M\times M}$ is the basis and Θ is the atom representation matrix, assumed to be sparse.
- 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.

[4] R. Rubinstein et al., "Double sparsity: Learning sparse dictionaries for sparse signal approximation," IEEE Trans. on Signal Processing, 2010

Sparse Dictionary Model

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

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

The proposed CDL approach

- 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:

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

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

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

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

The proposed CDL approach

Lemma

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 [12]:

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

- 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:

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

- 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:

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

Outline

Experimental Results

-Experiments Settings

- We demonstrate the result comparison on activation detection using the GLM with a design matrix and the CDL with a learnt dictionary.
- We use the dataset from Pittsburgh Brain Activity Interpretation Competition 2007 (PBAIC 2007) [5]
- In this experiment, we use the preprocessed data where slice time correction, 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.

[5] `http://pbc.lrdc.pitt.edu/`

Experimental Results

-Experiments Settings (cont'd)

- 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 2007[6], done by the SPM software package[7]. We also add one column of all ones that models the whole brain activity.
- The design matrix $X \in \mathbb{R}^{500\times 14}$ is then used in SPM to generate the activation maps for comparison purpose.

[6] http://www.lrdc.pitt.edu/ebc/2007/docs/CompetitionGuideBook2007v7.pdf

[7] http://www.fil.ion.ucl.ac.uk/spm/

Experimental Results

-Experiments Settings (cont'd)

- The measurement matrix is randomly generated using the Gaussian i.i.d measurement matrix with the CS measurement ratio set as 0.5.
- 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 software package [8] for solving the LASSO. Fig. 4 shows the activation maps from both methods, while the detailed comparisons are listed in Table 2.

[8] http://spams-devel.gforge.inria.fr/

Experimental Results

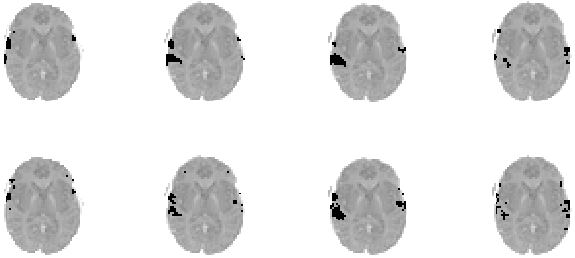


Figure 4: Activation maps for the *Instructions* task. Top: results generated using SPM with design matrix $X \in \mathbb{R}^{500\times 14}$, Bottom: results generated using CDL method, with Gaussian measurement matrix $\Phi \in \mathbb{R}^{250\times 500}$ and a learnt dictionary $D \in \mathbb{R}^{500\times 500}$. Slice number from left to right are 13, 14, 15, and 16 in both rows.

Experimental Results

	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 2: Detected activations comparison of SPM and CDL.

Outline

Conclusion

- 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.

Thank you!
Any questions?