33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

A New Spatially Constrained NMF with Application to fMRI

Saideh Ferdowsi, Vahid Abolghasemi, Bahador Makkiabadi, and Saeid Sanei

Abstract— **In this paper the problem of BOLD detection is addressed. The focus here is on non-negative matrix factorization (NMF), which is a data driven method and able to provide part-based representation of data. A new constrained optimization problem is proposed for the purpose of BOLD detection. The proposed constraint imposes some prior spatial information of active area inside the brain, on the decomposition process. The constraint is built up based on the type of stimulus and available physiological knowledge of the brain performance. The simulation results on both synthetic and real fMRI data show that applying the proposed constraint improves the BOLD detection performance.**

I. INTRODUCTION

Functional Magnetic Resonance Imaging (fMRI) is an imaging technique which provides useful information of brain functioning. The main goal of fMRI processing techniques is detecting the blood oxygen level dependent (BOLD) response as the result of brain activity [1]. Activation detection in brain functional imaging experiments helps to retrieve the neuronal activity as much as possible in response to cognitive or behavioral stimuli tasks. In general, fMRI analysis methods are divided into two groups: model-based techniques and data-based techniques. Model-based methods need prior knowledge about the activation task-waveform to work properly. In contrast, data-based methods which also called model-free do not rely on any assumed model or prior knowledge. However, exploiting some available *a priori* information can improve the performance of these methods.

The most common model-free methods which are applied to fMRI data are independent component analysis (ICA) and non-negative matrix factorization. Both ICA and NMF decompose a mixture of fMRI data into a set of time courses and their corresponding spatial sources [2][3]. Extracted time courses represent the brain temporal response to stimuli or artifacts. Each time course is related to a source known as an activated map and represent the activated area in the brain. Magnus et al. [4] state that a suitable fMRI analysis technique should provide sparse sources. That means having a small number of active (non-zero) voxels in each source. Sparsity is required because the brain networks of interest such as the motor or the visual cortex typically have sparse spatial structure [4]. ICA which is one of the fundamental methods in fMRI analysis, uses statistical moments to find maximally independent activated maps. ICA is not able to extract the sparse sources effectively; mainly due to a number of zero components which makes higher order averages hard to handle. However, it provides reliable results.

Non-negative matrix factorization has been applied to fMRI data during the past few years as an effective tool in order to detect the BOLD signal [5][6]. NMF decomposes a given nonnegative matrix into a product of two nonnegative matrices [7]. The nonnegativity constraint allows only additive combination, so, it can produce a part-based representation of data and consequently sparse results. Moreover, NMF is a flexible method since different constraints can be added. Prior information about an fMRI experiment can be used as a constraint on a model-free method to analyze the fMRI data. Temporal response of the brain to a specific task or stimulation is one of this prior information. Temporal brain response can be predicted by convolving the task-waveform and the hemodynamic response function (HDR) [8]. We use the predicted temporal response of brain as a constraint for the factorized time courses by NMF in our previous work [6]. Spatial information about the activated area inside the brain also can be predicted based on the type of stimulation and physiological knowledge about the performance of each lobe of the brain. The use of spatial information as a constraint for NMF is attractive because NMF has the ability to learn a part-base representation of all the sources. In this work, we explore the potential of NMF to take advantage of available spatial information about the activated areas and improve the procedure of BOLD detection.

In the next section, fundamentals of NMF are presented. Then, in section III, the details of the proposed method are discussed. In section IV, the results of the proposed method are given. Finally, section V concludes the paper.

II. NON-NEGATIVE MATRIX FACTORIZATION

Suppose that a non-negative observation matrix $\mathbf{V} \in \mathbb{R}^{M\times N}$ and a positive number $K < min(M,N)$ are available. NMF seeks a decomposition of $\mathbf{V}$ into two non-negative matrices $\mathbf{W} \in \mathbb{R}^{M\times K}$ and $\mathbf{H} \in \mathbb{R}^{K\times N}$ such that:

$$\mathbf{V}_{ij} \simeq (\mathbf{WH})_{ij} = \sum_{k=1}^{K} \mathbf{W}_{ik}\mathbf{H}_{kj} \tag{1}$$

where $\mathbf{W}$ contains the basis vectors as its columns and $\mathbf{H}$ is the associated sources. In order to approximate $\mathbf{W}$ and $\mathbf{H}$, an optimization algorithm should be formulated by minimizing a cost function. One of the widely used cost functions is the following Euclidean distance:

$$D(\mathbf{W},\mathbf{H}) = \|\mathbf{V} - \mathbf{WH}\|_F^2, \tag{2}$$

This work is supported by The Leverhulme Trust.

The authors are with the Department of Computing, Faculty of Engineering and Physical Sciences, University of Surrey, GU2 7XH, United Kingdom. {s.ferdowsi, v.abolghasemi, b.makkiabadi, s.sanei}@surrey.ac.uk

978-1-4244-4122-8/11/$26.00 ©2011 IEEE

where the symbol $\|.\|_F$ represents the Frobenius norm. Lee and Seung proposed multiplicative update rules [9] which are driven using traditional gradient descent. The following equations show the derived update rules for Euclidean distance function:

$$\mathbf{W} \leftarrow \mathbf{W} \odot (\mathbf{V}\mathbf{H}^T) \oslash (\mathbf{W}\mathbf{H}\mathbf{H}^T)$$
$$\mathbf{H} \leftarrow \mathbf{H} \odot (\mathbf{W}^T\mathbf{V}) \oslash (\mathbf{W}^T\mathbf{W}\mathbf{H}). \tag{3}$$

Here, $\odot$ and $\oslash$ represent the element wise matrix multiplication and division, respectively. It has been proved that under these update rules the Euclidean distance is monotonically nonincreasing [9]. In this work the Euclidean distance is selected as the main cost function. The proposed constraints are added to the main cost function to incorporate the effect of prior knowledge about the spatial pattern of activated area.

III. Proposed Algorithm

The proposed optimization problem uses a decomposition model which separates the factorization result into two parts: task related and non-task related. Task-related part refers to the source and time course of BOLD and non task-related part refers to the remained sources and time courses composed of artifacts, transiently task-related and noise components. In this work, the factorization problem is defined as follows:

$$\mathbf{V} = \mathbf{W}\mathbf{H} + \mathbf{w}\mathbf{h}, \tag{4}$$

where $\mathbf{V}^{M\times N}$ contains the observed fMRI data, $\mathbf{wh}$ represents the task-related part and $\mathbf{WH}$ represents the non-task related part. $\mathbf{W}$ contains the extracted time courses as its column and $\mathbf{H}$ contains their corresponding sources as rows. Task-related part includes BOLD time course and the spatial pattern of BOLD. $\mathbf{w}^{M\times 1}$ is a column vector representing the BOLD time course and $\mathbf{h}^{1\times N}$ is a row vector representing the activated map of the BOLD.

The proposed constrains are made by the prior knowledge about the spatial pattern of BOLD which is applied to $\mathbf{h}$ using an appropriate penalty function. $D_1(\mathbf{h},\hat{\mathbf{h}})$ is the penalty term corresponding to BOLD spatial pattern $\mathbf{h}$:

$$D_1(\mathbf{h},\hat{\mathbf{h}}) = \lambda \left(\| \mathbf{h} \| \,.\, \| \hat{\mathbf{h}} \| - \mathbf{h}\hat{\mathbf{h}}^T \right) \tag{5}$$

where $\hat{\mathbf{h}}^{1\times N}$ is the predicted spatial pattern of the activated area. As it is seen, the proposed constraint attempts to maximize the correlation between the learnt $\mathbf{h}$ and predicted spatial pattern. λ is a small positive coefficient that makes a trade off between the factorization error and the constraint strength.

The following objective function is achieved by introducing the constraints:

$$D = \frac{1}{2}\|\mathbf{V} - \mathbf{W}\mathbf{H} - \mathbf{w}\mathbf{h}\|_F^2 + \lambda \left(\| \mathbf{h} \| \,.\, \| \hat{\mathbf{h}} \| - \mathbf{h}\hat{\mathbf{h}}^T \right) \tag{6}$$

Update rules for $\mathbf{W},\mathbf{H},\mathbf{w}$ and $\mathbf{h}$ are derived by taking the gradient of objective function with respect to each one and using the Karush Kuhn Tucker (KKT) [10] conditions. The gradients of objective function (6) with respect to all factors are as follows:

$$\frac{\partial D}{\partial \mathbf{W}} = -\mathbf{V}\mathbf{H}^T + \mathbf{W}\mathbf{H}\mathbf{H}^T + \mathbf{w}\mathbf{h}\mathbf{H}^T \tag{7}$$

$$\frac{\partial D}{\partial \mathbf{w}} = -\mathbf{V}\mathbf{h}^T + \mathbf{W}\mathbf{H}\mathbf{h}^T + \mathbf{w}\mathbf{h}\mathbf{h}^T \tag{8}$$

$$\frac{\partial D}{\partial \mathbf{H}} = -\mathbf{W}^T\mathbf{V} + \mathbf{W}^T\mathbf{W}\mathbf{H} + \mathbf{W}^T\mathbf{w}\mathbf{h} \tag{9}$$

$$\frac{\partial D}{\partial \mathbf{h}} = -\mathbf{w}^T\mathbf{V} + \mathbf{w}^T\mathbf{W}\mathbf{H} + \mathbf{w}^T\mathbf{w}\mathbf{h} + \lambda \left(\frac{\mathbf{h}}{\| \mathbf{h} \|} \,.\, \| \hat{\mathbf{h}} \| - \hat{\mathbf{h}} \right) \tag{10}$$

Based on the KKT conditions the multiplicative update rules are obtained as:

$$\mathbf{W} \leftarrow \mathbf{W} \odot \frac{\mathbf{V}\mathbf{H}^T}{\mathbf{W}\mathbf{H}\mathbf{H}^T + \mathbf{w}\mathbf{h}\mathbf{H}^T} \tag{11}$$

$$\mathbf{w} \leftarrow \mathbf{w} \odot \frac{\mathbf{V}\mathbf{h}^T}{\mathbf{W}\mathbf{H}\mathbf{h}^T + \mathbf{w}\mathbf{h}\mathbf{h}^T} \tag{12}$$

$$\mathbf{H} \leftarrow \mathbf{H} \odot \frac{\mathbf{W}^T\mathbf{V}}{\mathbf{W}^T\mathbf{W}\mathbf{H} + \mathbf{W}^T\mathbf{w}\mathbf{h}} \tag{13}$$

$$\mathbf{h} \leftarrow \mathbf{h} \odot \frac{\mathbf{w}^T\mathbf{V} + \lambda\hat{\mathbf{h}}}{\mathbf{w}^T\mathbf{W}\mathbf{H} + \mathbf{w}^T\mathbf{w}\mathbf{h} + \lambda\frac{\mathbf{h}}{\|\mathbf{h}\|}\|\hat{\mathbf{h}}\|}. \tag{14}$$

All matrices and vectors are calculated by iteratively updating the above rules until achieving an acceptable error. The initial value is an important issue for non-negative matrix factorization as it influences the convergence of algorithm. Using the proper initial value for all factors helps to avoid being trapped in any local minima. In this work, we applied the multi-initialization method to find the best initial value.

As mentioned before, λ should balance the trade off between the constraint and factorization error. In this work we proposed a simple rule to update the value of λ as follows:

$$\lambda \leftarrow \lambda + \mu . \left(1 - \frac{\mathbf{h}\hat{\mathbf{h}}^T}{\| \mathbf{h} \| \,.\, \| \hat{\mathbf{h}} \|} \right) \tag{15}$$

where μ is a very small positive value selected as 0.05. In order to obtain the best regularization parameter (λ), the above rule is updated until reaching an acceptable normalized correlation between the source of interest $\mathbf{h}$ and spatial constraint $\hat{\mathbf{h}}$. The normalized correlation which has been used as a threshold was set to 0.5.

IV. Results

In this section the performance of the proposed algorithm is evaluated. The proposed algorithm is examined using two different datasets consisting of synthetic and real fMRI.

A. Experimental Results with Synthetic Data

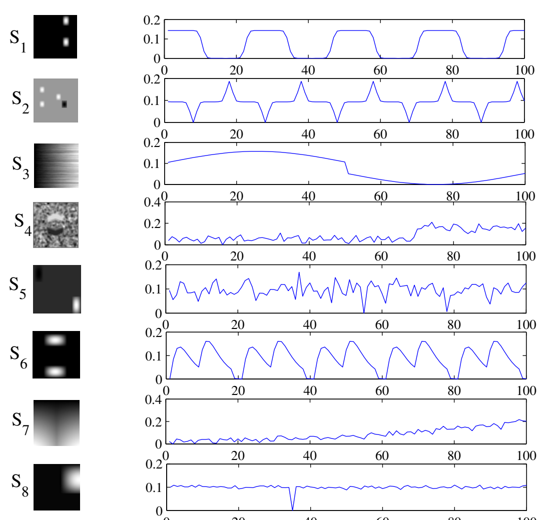
The simulated data which are used in this work has been provided by Machine Learning for Signal Processing laboratory [11]. The statistical properties of the fMRI sources are used as the basic knowledge for producing this dataset. This dataset consists of two matrices, one containing 8 simulated fMRI sources and the other containing their corresponding time courses (Figure 1). The mixture matrix is obtained by multiplying these two matrices. Generally, there are two main groups of fMRI sources: those that are related to performing a task by the subject, and the sources which are related to artifacts. S_1 shows the task-related simulated source, S_2 and S_6 are transient task-related, and the rest are the artifact related simulated sources.

In the first part of the experiments, we applied the proposed method to the synthetic mixture. Figure 2 shows the spatial constraint, which is vectorized to $\hat{\mathbf{h}}$. The proposed algorithm was applied to the simulated fMRI data for several times and with different noise levels added to the mixture. Based on the experiments, the optimum value of λ obtained for the case of no noise, $SNR = 5dB$ and $SNR = 15dB$ are 0.11, 0.15 and 0.15, respectively. Table I shows the computed SIR value for the reconstructed source of interest (s_1) and different levels of noise. The Signal-to-Interference-Ratio (SIR) and Signal-to-Noise-Ratio (SNR) were computed based on the following definitions:

$$SIR_h = 20\log \frac{\|h\|_2}{\|h-\hat{h}\|_2} \quad (16)$$

$$SNR = 20\log \frac{\|\mathbf{V}\|_2}{\|\mathbf{N}\|_2}, \quad (17)$$

where $\mathbf{N}^{M\times N}$ is a matrix representing the random Gaussian noise which is added to the mixture. As can be seen from the table, the SIR value in the proposed method is higher (almost double) compared to the unconstrained case. This improvement is robust against different noise levels as it is reflected in the table. The results of unconstrained and constrained source separation approaches for $SNR = 5dB$ are shown in Figures 3 and 4. Comparing the extracted simulated BOLD in these figures, reveals the ability of the proposed constraint to effectively take advantages of sparse part-based representation feature of NMF. Correspondingly, the extracted time-course as a result of the proposed method is more correlated with the actual square wave given in Figure 1.



(a) (b)

Fig. 1. Simulated fMRI data; (a) source image and (b) their corresponding time courses.

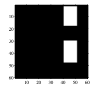


Fig. 2. Spatial constraint for the simulated data.

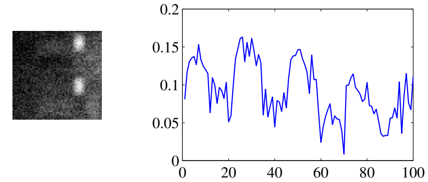


Fig. 3. Extracted active map and its time course using unconstrained NMF for $SNR = 5dB$.

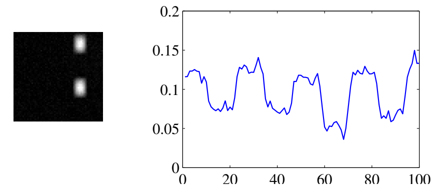


Fig. 4. Extracted active map and its time course using proposed constrained method for $SNR = 5dB$.

TABLE I

AVERAGE SIR VALUES FOR DIFFERENT EXPERIMENTS.

Noise level	SIR(dB)	
	unconstrained	proposed method
noiseless	1.6564	3.4091
SNR=15dB	1.4343	2.6020
SNR=5dB	1.0226	2.1247

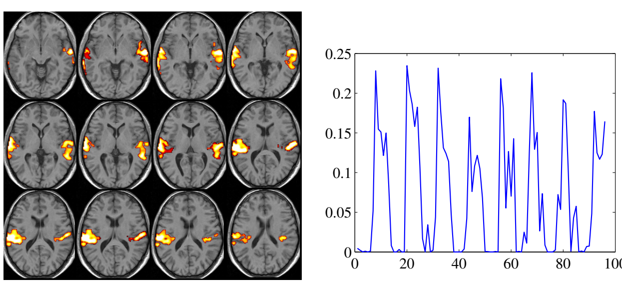
B. Experimental Results on Real fMRI Data

The real fMRI dataset used to evaluate the proposed algorithm was taken from the SPM website [12]. This dataset has obtained in an auditory fMRI experiment from a single subject. It comprises the whole brain and was acquired by a 2T Siemens MAGNETOM Vision scanner and the scan to scan repeat time (TR) is 7seconds. The Auditory stimulus is bi-syllabic words presented binaurally at a rate of 60 per minute. The dataset contains 96 scans and each scan consists of 64 contiguous slices $(64\times 64\times 64, 3\times 3mm \times 3mm^3 voxels)$.

As it was mentioned before the spatial constraint is designed based on the knowledge about the type of stimulation. We constructed the spatial constraint using the WFU-PickAtlas toolbox [13] . As the stimulation does not involve any high level of auditory processing, the “primary auditory cortex" and "superior temporal gyrus" are considered as the regions activated due to hearing the words.

In order to evaluate the performance of the proposed method and finding the voxels contributing to the activation process, the estimated task-related source (**h**) using the proposed method, was scored to z-score [14]. After a proper thresholding, the activated region is obtained. Figure 5 represents the activated area and its corresponding time course obtained using the proposed method. To determine the value of parameter λ for real data and examine the convergence of the proposed method, we ran the algorithm 70 times. The results of experiments show that the parameter λ converges to a value in the range of $[0.11\ 0.12]$ according to equation (15). The number of sources existing in this dataset was estimated by minimum description length (MDL) method [15] as 35 sources. However, we set $K = 34$ as the total number of non task-related sources, and one remaining source indicating the source of interest (**h**). The normalized correlation between the estimated BOLD time course ($\mathbf{w}$) and the expected BOLD time course was calculated to compare the performance of the algorithm. The expected BOLD time course is obtained by convolving the HDR and task-wave form of the fMRI experiment [8]. Our experiments show that the normalized correlation between the estimated BOLD time course ($\mathbf{w}$) and the expected BOLD time course has the maximum value of 0.9123. This correlation value is 0.8584 for the case of unconstrained separation, which is smaller.

Finally, we compared the results of the proposed method to the results of SPM which is a benchmark in this field. The SPM results are shown in Figure 6. It is found from the figures 5 and 6 the detected active region using the proposed method matches the results of SPM, which confirms the correctness of the result of the proposed method.



(a) (b)

Fig. 5. Results of the proposed method: (a) Active area in auditory region, and (b) its corresponding time course.

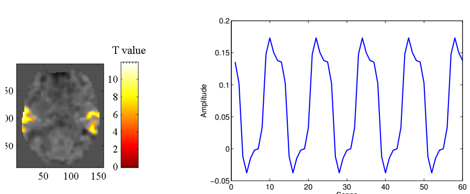


Fig. 6. SPM result for the same real data set.

V. Conclusions

In this paper a constrained NMF technique was proposed to detect the task-related component in fMRI. The aim of applying the proposed constraint is to use the additional available physiologic information related to the stimuli of interest. Using such a constraint on the sources within NMF factorization algorithm allows us to benefit more from the part-based representation property of NMF. The extensive simulation results confirm the ability of the proposed method to correctly detect the activated area. However, the results of the proposed method is strongly affected by the accuracy of prior information or constraint. This drawback restrict the performance of the algorithm specially in cases which there is no information about the time course of BOLD to evaluate the results. Improving the performance of the spatially constrained NMF in terms of developing some methods to provide more accurate spatial constraint is the most important aim in our future work.

References

[1] S. Ogawa, T. Lee, A. Nayak, and P. Glynn, "Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields.," vol. 14, pp. 68–78, 1990.

[2] V. D. Calhoun, T. Adali, L. K. Hansen, J. Larsen, and J. J. Pekar, "ICA of functional mri data: An overview," *in Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, pp. 281–288, 2003.

[3] X. Wang, J. Tian, X. Li, J. Dai, and L. Ai, "Detecting brain activations by constrained non-negative matrix factorization from task-related BOLD fMRI," *Medical Images*, vol. 5369, pp. 675–682, 2004.

[4] M. O. Ulfarsson and V. Solo, "Sparse variable principal component analysis with application to fMRI," *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, pp. 460 –463, 2007.

[5] G. Lohmann, K. G. Volz, and M. Ullsperger, "Using non-negative matrix factorization for single-trial analysis of fMRI data," *NeuroImage*, vol. 37, no. 4, pp. 1148 – 1160, 2007.

[6] S. Ferdowsi, V. Abolghasemi, and S. Sanei, "A constrained nmf algorithm for bold detection in fMRI," *Machine Learning for Signal Processing (MLSP), 2010 IEEE International Workshop on*, pp. 77 –82, 2010.

[7] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, October 1999.

[8] K. J. Friston, P. Fletcher, O. Josephs, A. Holmes, M. D. Rugg, and R. Turner, "Event-related fmri: Characterizing differential responses," *NeuroImage*, vol. 7, no. 1, pp. 30 – 40, 1998.

[9] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *In NIPS*, pp. 556–562, 2000.

[10] N. Gillis and F. Glineur, "Nonnegative factorization and the maximum edge biclique problem," Tech. Rep., Universit catholique de Louvain, Center for Operations Research and Econometrics (CORE), Nov. 2008.

[11] Machine Learning for Signal Processing Laboratory, "http://mlsp.umbc.edu," .

[12] Statitstical Parameter Mapping (SPM) homepage, "http://www.fil.ion.ucl.ac.uk/spm/," .

[13] WFU PickAtlas Toolbox , "http://fmri.wfubmc.edu/software/pickatlas," .

[14] K. J. Friston, K. J. Worsley, R. S. J. Frackowiak, J. C. Mazziotta, and A. C. Evans, "Assessing the significance of focal activations using their spatial extent.," *Human Brain Mapping*, , no. 1, pp. 214–220, 1994.

[15] E. Fishler, M. Grosmann, and H. Messer, "Detection of signals by information theoretic criteria: general asymptotic performance analysis," *Signal Processing, IEEE Transactions on*, vol. 50, no. 5, pp. 1027 –1036, May 2002.