

# A COMPARATIVE STUDY OF $\alpha$ -DIVERGENCE BASED NMF TECHNIQUES FOR fMRI ANALYSIS

*Saideh Ferdowsi, Vahid Abolghasemi, and Saeid Sanei*

NICE Research Group, Faculty of Engineering and Physical Sciences, University of Surrey, UK  
Emails: {s.ferdowsi, v.abolghasemi, s.sanei}@surrey.ac.uk

## ABSTRACT

The main objective of fMRI analysis methods is to detect the Blood Oxygenation Level Dependent (BOLD) from fMRI sequences. Algorithms which are discussed here are known as data-driven methods. The main advantage of these types of algorithms over data-based methods is that there is no need for prior information. Here, we focus on one of the powerful matrix factorization algorithms which has been recently applied to fMRI called Non-negative Matrix Factorization (NMF) [1]. There exist many different NMF techniques in the literature and no comprehensive assessment of their performances on fMRI data has been reported. So, in this work the performance in terms of BOLD detection, using  $\alpha$ -Divergence based methods are investigated and then compared with the commonly used Euclidean distance based method. The aim is to highlight the advantages that such techniques can have in practice. We explored the performance of these techniques for two types of real fMRI and also synthetic data. We observed that the  $\alpha$ -Divergence based methods are also applicable to fMRI data and reveal acceptable performance.

## 1. INTRODUCTION

Nonnegative Matrix Factorization (NMF) [2] has been known as a powerful method for nonnegative data analysis such as rank reduction and separating hidden sources. NMF has found many applications in different areas, for example, machine learning, clustering, pattern recognition, and data mining. The aim of NMF is to find an approximate factorization for a nonnegative matrix  $\mathbf{V}$  into two nonnegative matrices  $\mathbf{W}$  and  $\mathbf{H}$ . The columns of  $\mathbf{W}$  are called basis functions, while the rows of  $\mathbf{H}$  represent the hidden nonnegative sources which correspond to each basis function. In contrast to other methods such as Principal Component Analysis (PCA) or Independent Component Analysis (ICA), the non-negativity constraint in NMF produces sparse results which is advantageous [3, 4].

There are various cost functions to evaluate the error of factorization. Squared Euclidean distance and generalized Kullback Leibler divergence are the best known and the most frequently used cost functions in NMF [5]. Some other cost functions are Csiszar's divergence [6], Bregman divergence [7], generalized divergence measure [8] and  $\alpha$

or  $\beta$  divergences [9]. A suitable cost function can be determined based on the assumption about noise distribution. When  $\mathbf{E} = \mathbf{V} - \mathbf{WH}$  is modeled as normal distribution, the Squared Euclidean distance is an optimal choice. In some applications, such as pattern recognition, image processing and statistical learning, noise is not necessarily Gaussian and the cost functions based on information divergence are often used.

In this paper, we investigate the application of NMF algorithms based on  $\alpha$ -Divergence which are recently derived by Cickocki's group, for functional magnetic resonance imaging (fMRI) data [10]. The main goal of this work is to analyze the results of these algorithms for fMRI data and compare their performances with other NMF algorithms.

fMRI is a useful neuroimaging technique to study both functional and anatomical behaviors of brain. fMRI measures the hemodynamic response (HDR) resulted from neural activity. When brain neurons are activated, their consumption of oxygen increases and as a result the blood flow in the activation zone increases. This change in the level of oxygenation and blood flow affects the intensity of the recorded MRI images. The aim of fMRI techniques is to detect the resulting signal change called BOLD signal. For more detailed description about fMRI principles one can refer to [11].

The rest of the paper is organized as follows. In the next section, we mathematically represent the NMF techniques which are used in this paper. Then, in section 3, the experimental results of applying different methods are given and compared. Finally, the conclusion is drawn in section 4.

## 2. NON-NEGATIVE MATRIX FACTORIZATION FOR BOLD DETECTION

The following problem is considered in NMF formulation. Given a nonnegative matrix  $\mathbf{V} \in \mathbb{R}^{m \times n}$  containing the input data, we must find two nonnegative matrices  $\mathbf{W} \in \mathbb{R}^{m \times k}$  and  $\mathbf{H} \in \mathbb{R}^{k \times n}$  such that:

$$\mathbf{V}_{ij} \simeq (\mathbf{WH})_{ij} = \sum_{a=1}^k \mathbf{W}_{ia} \mathbf{H}_{aj}, \quad (1)$$

where  $k$  presents the rank of factorization and is assumed to be known or estimated by information-theoretic criteria. In order to complete the factorization process an appropriate cost function should be defined and then minimized. Equation (2) shows the most widely used NMF cost function

The authors acknowledge the support by The Leverhulme Trust.

which employs Euclidean distance between the data matrix  $\mathbf{V}$  and multiplication of the factors:

$$D = \|\mathbf{V} - \mathbf{WH}\|^2 = \sum_{ij} [\mathbf{V}_{ij} - [\mathbf{WH}]_{ij}]^2 \quad (2)$$

Multiplicative update rules for the above optimization problem have been derived by Lee and Seung [5]. They proved that the referred cost functions would converge to a local minima under these update rules.

Recently, a family of new algorithms based on the  $\alpha$ -Divergence has been proposed by Cichocki et al. [10]. The following equation shows the basic  $\alpha$ -Divergence between  $\mathbf{V}$  and  $\mathbf{WH}$ :

$$D = \frac{1}{\alpha(\alpha-1)} \sum_{ij} [\mathbf{V}_{ij}^\alpha (\mathbf{WH})_{ij}^{1-\alpha} - \alpha \mathbf{V}_{ij} + (\alpha-1)(\mathbf{WH})_{ij}] \quad (3)$$

The above equation is normally applied for  $\alpha \in [0, 2]$ . The four following distances are those that are commonly used for different values of  $\alpha$ :

- Kulback-Leibler (KL) I-Divergence ( $\alpha \rightarrow 1$ ):

$$D = \sum_{ij} [\mathbf{V}_{ij} \ln \frac{\mathbf{V}_{ij}}{(\mathbf{WH})_{ij}} - \mathbf{V}_{ij} + (\mathbf{WH})_{ij}] \quad (4)$$

- Dual KL I-Divergence ( $\alpha \rightarrow 0$ ):

$$D = \sum_{ij} [(\mathbf{WH})_{ij} \ln \frac{(\mathbf{WH})_{ij}}{\mathbf{V}_{ij}} + \mathbf{V}_{ij} - (\mathbf{WH})_{ij}] \quad (5)$$

- Squared Hellinger Divergence ( $\alpha = 0.5$ ):

$$D = \sum_{ij} (\sqrt{(\mathbf{WH})_{ij}} - \sqrt{\mathbf{V}_{ij}})^2 \quad (6)$$

- Pearson Divergence ( $\alpha = 2$ ):

$$D = \sum_{ij} \frac{(\mathbf{V}_{ij} - (\mathbf{WH})_{ij})^2}{(\mathbf{WH})_{ij}} \quad (7)$$

The choice of optimal  $\alpha$  depends on the application and the data. Equations (8) and (9) present the main learning rules for  $\alpha$ -Divergence algorithms. These update rules are suitable for large scale NMF [9] and can be expressed as:

$$\mathbf{W}_{ij} \leftarrow (\mathbf{W}_{ik} \sum_j \mathbf{H}_{kj} (\frac{\mathbf{V}_{ij}}{(\mathbf{WH})_{ij}})^\alpha)^{(\omega/\alpha)} (1 + \alpha_{sw}) \quad (8)$$

$$\mathbf{H}_{ij} \leftarrow (\mathbf{H}_{kj} \sum_i \mathbf{W}_{ij} (\frac{\mathbf{V}_{kj}}{(\mathbf{WH})_{kj}})^\alpha)^{(\omega/\alpha)} (1 + \alpha_{sh}), \quad (9)$$

where  $\omega$  is the over relaxation parameter and is typically selected between 0.5 and 2. Over relaxation parameter accelerates the convergence and stabilizes the algorithm.  $\alpha_{sw}$  and  $\alpha_{sh}$  are small positive parameters which are used to enforce the sparsity constraint on the algorithm.

In the next section, we apply these update rules for detecting the BOLD area in the brain. The aim is to illustrate

the performance of these NMF techniques and see whether they can be extended as a main tool for this application. We also compare these results to the results of Euclidean distance based method which is the common NMF technique already applied to this problem [12]. It is also worth noting that although data driven techniques can be applied directly and with no prior information on fMRI data, incorporating some available prior knowledge may be possible in the form of additional constraints, which has been reported recently in [1] but is not dealt here.

### 3. EXPERIMENTAL RESULTS

We used the  $\alpha$ -Divergence based NMF algorithms to detect brain activation in a set of synthetic and real fMRI data. In order to find optimal value of  $\alpha$ , source separation procedure was repeated with different  $\alpha$  values. Moreover, the performance of the groups of  $\alpha$  algorithms were compared with more common NMF algorithms such as Euclidean distance based methods.

#### 3.1 Data sets

##### 3.1.1 Synthetic Data

The simulated data was provided by Machine Learning for Signal Processing laboratory [13]. The statistical properties of the fMRI sources are used as the basic knowledge for producing this data set. In general, fMRI sources are divided into two main groups: signals of interests and artifacts [14]. The signals of interests include task-related, function-related and transiently task-related signals and assumed to have a super-gaussian distribution. The artifacts include physiology-related, motion-related and scanner-related signals and assumed to have a sub-gaussian distribution. The simulated fMRI data contains both types of sources and their corresponding time courses (Figure 1). The aim of fMRI analysis methods is to separate task-related source or BOLD from the image sequence. The top row in Figure 1 shows the simulated BOLD which has to be detected.

In order to generate the mixtures, the matrix of time courses is multiplied by the matrix of sources. Further, in order to evaluate the performance of the applied NMF algorithms, we used the SIR value of separated task-related source. The SIR value is defined as follows:

$$SIR = 20 \log \frac{\|\mathbf{h}\|_2}{\|\mathbf{h} - \hat{\mathbf{h}}\|_2} \quad (10)$$

where  $\mathbf{h}$  is the extracted task-related source and  $\hat{\mathbf{h}}$  is the actual task-related source or BOLD.

We evaluated Amari's approach based on  $\alpha$ -Divergence [15, 16], Dual KL I-Divergence, KL I-Divergence, Hellinger Divergence and Pearson Divergence for this data set for 100 times. Our experiments show that the best convergence of these algorithms is obtained when the over-relaxation parameter is  $\omega = 1.9$  and sparsity regularization parameters are

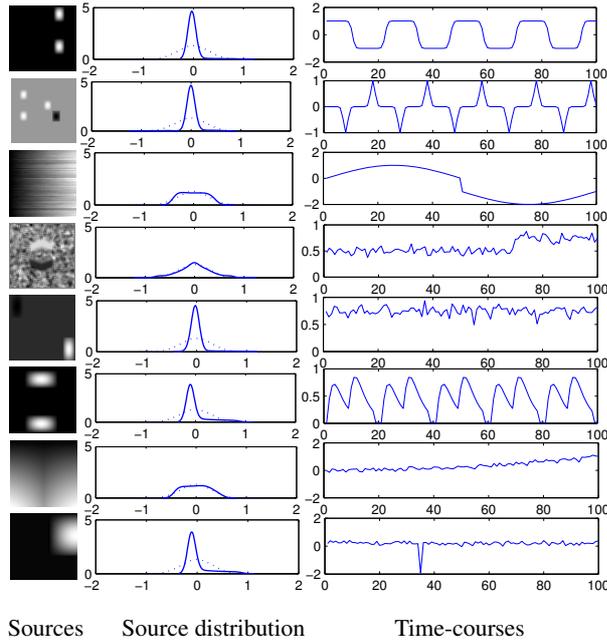


Figure 1: Simulated fMRI

$\alpha_{sW} = \alpha_{sH} = 0.001$ . The average SIR was computed using equation (10). Figure 2 shows the computed average SIR value for the results of different  $\alpha$ -Divergence based NMF algorithms. It is seen from the figure that the best SIR for Amari's approach is obtained when  $\alpha = 0.5$ . Moreover the KL I-Divergence and Square Hellinger Divergence show higher SIR compared to Dual KL I-Divergence and Pearson Divergence.

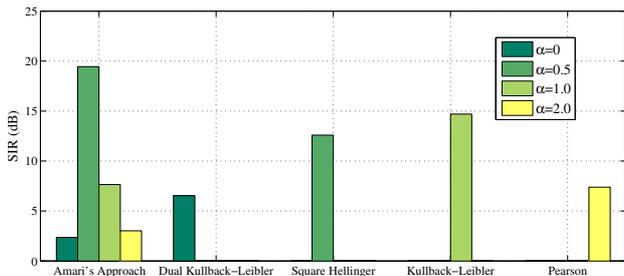


Figure 2: Computed SIR of source of interest for different methods.

In addition to the previous methods, we applied Euclidean distance to factorize the simulated mixtures into sources and time courses. The computed average SIR for this experiment was 29.33 dB, which is higher compared to the results of  $\alpha$ -Divergence based methods. Here, the sparsity constraint was used to obtain more accurate results. In addition, the regularization sparsity parameter was selected as 0.1 for this part of the experiment.

### 3.1.2 Real Data

The real data sets which are used in this work are taken from the SPM website [17]. The first data set is Auditory fMRI data from single subject experiment. This data set, taken from the whole brain, was acquired by a 2T Siemens MAGNETOM Vision scanner with the scan to scan repeat time (TR) of 7 seconds. The Auditory stimulus is bi-syllabic words presented binaurally at a rate of 60 per minute. The data set contains 96 scans and each scan consists of 64 contiguous slices ( $64 \times 64 \times 64, 3 \times 3 \text{mm}^3$  voxels). The 96 scans include 8 blocks of size 12, each of which containing 6 scans under rest and 6 scans under auditory stimulations.

The second data set is a Visual fMRI data and was collected using a 2T Siemens MAGNETOM vision system. There are 360 scans during the four runs in this data set. Each run consists of four conditions which are *fixation*, *attention*, *no attention* and *stationary*. In the *attention* condition the subject should detect changes and during the *no attention* condition the subject was instructed to just keep eye open. During the *attention* and *no attention* conditions subjects fixated centrally, while white dots emerged from the fixation point to the edge of the screen.

In our experiments on real fMRI data, we applied Amari's approach with  $\alpha = 0.5$  and KL I-Divergence to both data sets. Visual comparison of the results for different methods is difficult. However Figures 3 and 4 present the results obtained by KL I-Divergence. Figure 3 shows the extracted task-related source (BOLD) and its corresponding time course of the first data set. As it is seen, the activated region is correctly detected for different brain slices.

Figure 4 presents the results of applying KL I-Divergence to the visual fMRI data set. We can clearly see the detected BOLD in the occipital lobe which is responsible for visual processing tasks. The extracted time course also verifies the temporal behavior of activation.

In order to compare the performance of  $\alpha$ -Divergence based methods and Euclidean distance which is more popular for fMRI analysis, the normalized correlation between the extracted time-course and the predicted temporal response of brain has been calculated. Temporal response of brain to a specific task can be modeled by convolving the task-waveform and the hemodynamic response function (HDR) [1, 18]. The results show that the normalized correlation between the extracted time-course and the predicted temporal response of brain for the results of Euclidean distance has higher value than those for the two other  $\alpha$ -Divergence based method. The numerical results of this comparison for both data set are given in Table 1.

## 4. CONCLUSION

The problem of BOLD detection using a source separation method was presented. Five different variations of NMF technique were applied to fMRI data. These methods include Amari's approach, Dual KL I-Divergence, KL I-Divergence,

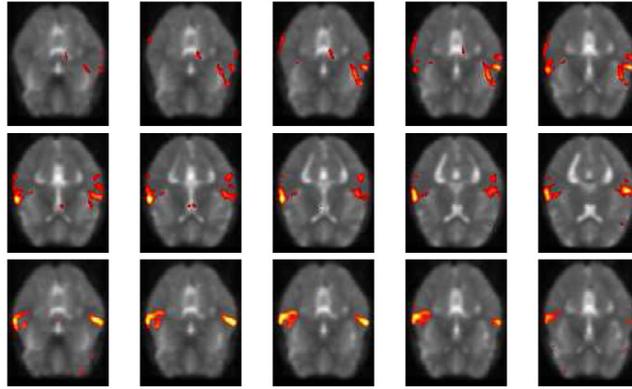
Table 1: Normalized correlation between the extracted BOLD time course and predicted temporal response of brain.

	Auditory data set	Visual data set
KL I-Divergence	0.6950	0.8736
Amari's approach $\alpha = 0.5$	0.6371	0.8152
Euclidean distance	0.8689	0.9102

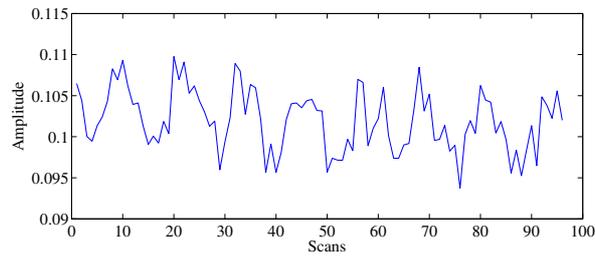
Hellinger Divergence, and Pearson Divergence. In contrast to Euclidean distance based method, the  $\alpha$ -Divergence based methods consider non-Gaussian distribution for noise in the factorization process. We then presented the results of these methods together with Euclidean distance based one. We observed that the  $\alpha$ -Divergence methods are very sensitive to selection of over-relaxation parameter which is a disadvantage. However, the separation results of these techniques were correct and comparable to those of obtained by applying Euclidean distance based method. We can further investigate the flexibility of  $\alpha$ -Divergence based algorithms in adding constraints, which has to be explored in the future.

## REFERENCES

- [1] 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*, pages 77–82, 2010.
- [2] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.
- [3] P. O. Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, 5:1457–1469, 2004.
- [4] J. Eggert and E. Korner. Sparse coding and NMF. *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, 4:2529 – 2533 vol.4, jul. 2004.
- [5] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems 13*, pages 556–562. MIT Press, April 2001.
- [6] A. Cichocki, R. Zdunek, and S.-I. Amari. Csiszrs divergences for non-negative matrix factorization: Family of new algorithms. In *LNCS*, pages 32–39. Springer, 2006.
- [7] I. S. Dhillon and S. Sra. Generalized nonnegative matrix approximations with bregman divergences. In *In: Neural Information Proc. Systems*, pages 283–290, 2005.
- [8] R. Kompass. A generalized divergence measure for nonnegative matrix factorization. *Neural Comput*, 19:780–791, 2007.
- [9] A. Cichocki, S.-I. Amari, R. Zdunek, R. Kompass, G. Hori, and Z. He. Extended smart algorithms for non-negative matrix factorization. In *Artificial Intelligence and Soft Computing ICAISC 2006*, volume 4029 of *Lecture Notes in Computer Science*, pages 548–562. 2006.
- [10] A. Cichocki, H. Lee, Y.-D. Kim, and S. Choi. Non-negative matrix factorization with  $\alpha$ -divergence. *Pattern Recognition Letters*, 29(9):1433 – 1440, 2008.
- [11] K. J. Friston, C. D. Frith, R. Turner, and R. S. J. Frackowiak. Characterizing evoked hemodynamics with fmri. *NeuroImage*, 2(2, Part 1):157 – 165, 1995.
- [12] X. Wang, J. Tian, L. Yang, and J. Hu. Clustered cnmf for fMRI data analysis. *Medical Images*, 5746:631–638, 2005.
- [13] Machine Learning for Signal Processing Laboratory. <http://mlsp.umbc.edu>.
- [14] V. D. Calhoun, T. Adali, L. K. Hansen, J. Larsen, and J. J. Pekar. ICA of functional MRI data: An overview. In *in Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, pages 281–288, 2003.
- [15] S. Amari. Differential geometrical methods in statistics. page Springer, 1985.
- [16] S. I. Amari. Integration of Stochastic Models by Minimizing  $\alpha$ -Divergence. *Neural Comput.*, 19(10):2780–2796, October 2007.
- [17] Statistical Parameter Mapping (SPM) homepage. <http://www.fil.ion.ucl.ac.uk/spm/>.
- [18] K. J. Friston, P. Fletcher, O. Josephs, A. Holmes, M. D. Rugg, and R. Turner. Event-related fmri: Characterizing differential responses. *NeuroImage*, 7(1):30 – 40, 1998.

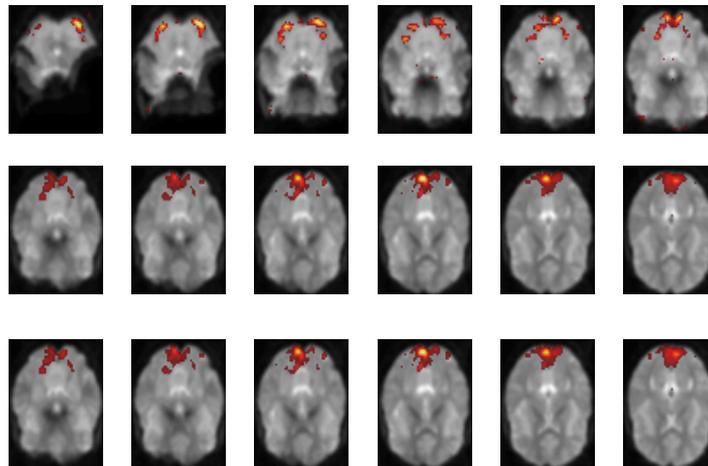


(a) The spatial map obtained by KL I-Divergence method

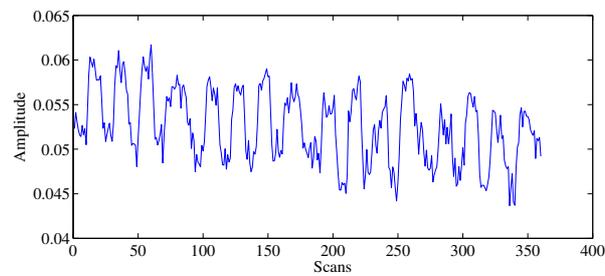


(b) The corresponding time-course

Figure 3: Auditory data analysis results.



(a) The spatial map obtained by KL I-Divergence method



(b) The corresponding time-course

Figure 4: Visual data analysis results.