Underdetermined Blind Source Separation based on Fuzzy C-Means and Semi-Nonnegative Matrix Factorization

Ossama. S. Alshabrawy[1,*], M. E. Ghoniem, W. A. Awad[2,*] and Aboul Ella Hassanien,[*]

[1]Math. Dept., Faculty of science, Mansoura University, New Dam., Egypt.
E-Mail: ossama_alshabrawy87@yahoo.com
[2]Math. & Comp. Sci. Dept., Faculty of Science, Port Said University, Egypt.
E-Mail: wealak@yahoo.com
[*]Scientific Research Group in Egypt (SRGE)
http://www.egyptscience.net

***Abstract*—**Conventional blind source separation is based on over-determined with more sensors than sources but the underdetermined is a challenging case and more convenient to actual situation. Non-negative Matrix Factorization (NMF) has been widely applied to Blind Source Separation (BSS) problems. However, the separation results are sensitive to the initialization of parameters of NMF. Avoiding the subjectivity of choosing parameters, we used the Fuzzy C-Means (FCM) clustering technique to estimate the mixing matrix and to reduce the requirement for sparsity. Also, decreasing the constraints is regarded in this paper by using Semi-NMF. In this paper we propose a new two-step algorithm in order to solve the underdetermined blind source separation. We show how to combine the FCM clustering technique with the gradient-based NMF with the multi-layer technique. The simulation results show that our proposed algorithm can separate the source signals with high signal-to-noise ratio and quite low cost time compared with some algorithms.

Keywords: Blind Source Separation;Underdetermined; Fuzzy C-Means Clustering; Nonnegative Matrix Factorization

I. INTRODUCTION

Blind Source Separation (or Blind Signal Separation, BSS) combining with artificial neural network, information theory and computer science, has received a great deal of attention in the fields of digital communication systems, speech processing, medical imaging, water marking, biomedical engineering and data mining [13-17] in recent years. Blindness means that very little information is known about the source signals or the mixing system. The source signals can be extracted only by the matrix of observations from a group of sensors.

The objective of BSS [1] is to extract the original source signals from their observation mixtures using only the information of the observed signals with no or very limited knowledge about the source signals or the mixing system. During the last few years, Researchers have developed many approaches to solve the problem of blind source separation. These approaches can be largely classified into two methodologies namely over-determined BSS and underdetermined BSS according to the number of source signals and observable mixed signals. The BSS of number of sensors or observable mixed signals less than source signals is called underdetermined BSS while the BSS of number of sources less than sensors is called over-determined BSS. However, the approaches for the blind source separation in the underdetermined case are rarely involved. In this latter case the approach of classic independent component analysis (ICA) fails to solve this problem.

Unlike the traditional BSS which is assumed that the mixing matrix is of full column rank, the mixing matrix in the underdetermined case is not of full column rank. Consequently, multiplying the observable data mixtures by the pseudo inverse of the mixing matrix cannot be used to recover the source signals. This makes recovering the source signals a very challenging task even if the mixing matrix is known [3]. In practical field, the over-determined mixture assumption does not always hold e.g. in radio- communications the probability of receiving more sources than sensors increases with the increase of reception bandwidth; thus it is necessary to solve the problem of underdetermined blind source separation (UBSS) [2]. Zhang and Zheng [4] divided the algorithms to solve underdetermined problems into three categories. The first category is the iteration algorithms that estimate the source signals and the mixing matrix simultaneously. The Second is the algorithms based on the statistical probability model of the source signals. The third is the “two-step” algorithms clustering-then-optimization.

In this paper we focus on the UBSS in the instantaneous mixture case utilizing the “two-step” algorithm. In the first step, the mixing matrix is estimated by fuzzy C-means clustering technique using only the matrix of observable mixtures. In the second step, we use the estimated mixing matrix as an input to the projected-gradient based NMF to estimate the source signals matrix.

Our work in this paper introduces three contributions in the study of underdetermined blind source separation. First, we speed the convergence and solve the problem of initialization for NMF which is sensitive to the initial values of mixing matrix, source signals matrix and the other parameters by using

FCM to estimate the mixing matrix and then use this estimation as an input to the Semi-NMF algorithm to estimate the source signals. Second, we decrease the complexity and increase the performance of the separation. The final contribution is using the projected gradient based NMF which is rather computationally soft and decreases the constraints using semi-NMF.

The rest of the paper is organized as follows. In Section II, we introduce an overview and basic concepts of NMF and FCM clustering. In Section III, we present the detailed proposed system. In section IV, we show the analysis of the typical experiments and the results of different BSS methods and the simulation results show the effectiveness and high performance of the proposed algorithm. Finally, a short conclusion and future work are drawn in Section V.

I. Background

A. Fuzzy C-Means Clustering

The fuzzy set theory developed by zadeh 1965 provides a powerful analytical tool for soft clustering method. The fuzzy c-means (FCM) algorithm (developed by Dunn in 1973 and improved by Bezdek in 1981) is the best known method for fuzzy clustering, based on optimizing objective function, it has been used most widely in many applications as a conventional tool in clustering, and has the most perfect algorithm theory. The FCM clustering algorithm is a variation and an extension for the traditional k-means clustering algorithm, in which for each data point a degree of membership or membership function $\xi_{ij} \in [0,1]$ of clusters is calculated. However, almost no data points have a crisp membership $\xi_{ij} \in \{0,1\}$.That is, FCM always produces fuzzy memberships ξ_{ij} in the open interval (0, 1) [6]. The centroids of the clusters are computed based on the degree of memberships as well as sample data points.

According to unsupervised learning clustering literatures, FCM algorithm is a fuzzy clustering method for finding cluster centers with the objective function:

$$Z_{FCM}(\xi, b) = \sum_{i=1}^{C}\sum_{j=1}^{T}\xi_{ij}^{\beta} d(y_j, b_i) = \sum_{i=1}^{C}\sum_{j=1}^{T}\xi_{ij}^{\beta} ||Y_j - b_i|| \quad (1)$$

where, C is the number of desired clusters, T is the number of samples, the weighting exponent $\beta > 1$ is the degree of fuzziness,$\xi = \xi_i(y_j)$ is the fuzzy c-partition matrix, $b = \{b_1, b_2, \dots \dots, b_C\}$ is the cluster centers, $d_{ij} = ||Y_j - b_i||$ is the distance between the data sample point and the cluster center. The necessary conditions for minimizing ξ, b are the following update rules equations:

$$\xi_{ij}^{(n)} = \frac{||Y_j - b_i||^{-2/(\beta-1)}}{\sum_{k=1}^{C} ||Y_j - b_k||^{-2/(\beta-1)}} \quad (2)$$

and,

$$b_i = \frac{\sum_{j=1}^{T}\xi_{ij}^{\beta} y_j}{\sum_{j=1}^{T}\xi_{ij}^{\beta}} \quad (3)$$

where, $i = 1, 2, \dots \dots, C$, $j = 1,2, \dots \dots, T$.

B. Nonnegative Matrix Factorization

Many researchers have proposed the nonnegative matrix factorization (NMF). Since it was first proposed in 1999, nonnegative matrix factorization (NMF) has attracted more and more attention. Recently, it has been widely used in many areas and one of these areas is BSS [5,18,19]. However, the solution is not unique since, NMF is a non-convex programming, and in most algorithms it frequently results in local optima.

The basic NMF decomposition model for BSS is as follows:

$$Y = AX + E \quad (4)$$

Where, $Y \in R_{+}^{I\times T}$ ($with\, Y \geq 0$) is the observable mixtures, $A \in R_{+}^{I\times J}$ is the mixing matrix, $X \in R_{+}^{J\times T}$ is the source signals matrix with $A, X \geq 0$ and E is the additive noise. For BSS I is the number of mixtures or sensors, T is the number of sample time points, and J is the number of sources. With only the data observable mixtures Y is the only known, the mixing matrix A and the source signals X are estimated by formula (4).

There are many algorithms for NMF Such as multiplicative algorithms but these methods depend on choosing an auxiliary function and are rather computation intensive. On the other hand, the projected gradient algorithms are faster and have low complexity. The projected gradient based update rules take the following general form of iterative updates as follows [5]:

$$X^{(n+1)} = [X^{(n)} - \alpha_X P_X]_{+} \quad (5)$$

$$A^{(n+1)} = [A^{(n)} - \alpha_A P_A]_{+} \quad (6)$$

where, $[\,]_{+}$ means that the entries are forced to be nonnegative by replacing the negative values by zero in the theoretical field but for practical purposes it is replaced by a small positive value ε to avoid numerical instabilities, thus giving component-wise $X = \max\{\varepsilon, X\}$,and P_A and P_X are the descent directions of A and X respectively and α_X, α_A are the learning rates.

One of the projected gradient based algorithms is Projected Sequential Subspace Optimization (PSESOP) method [5, 12] that performs a projected minimization of a smooth objective function over a subspace spanned by several directions. These include the current gradient and the gradient from previous iterations, together with the Nemirovski directions. Nemirovski showed that convex smooth unconstrained optimization is optimal if the optimization in the n-th iterative step is performed over a subspace which includes the current gradient $Q^{(n)}$, the directions

$$P^{(n)} = X_t - X_0 \quad (7)$$

and the linear combination of the previous gradients

$$Lp^{(n)} = \sum_{l=0}^{n-1} W_l Q^{(l)} \quad (8)$$

With the coefficients $W_0 = 0$ and,

$$W_l = 0.5 + \sqrt{\left(0.25 + W_{(n-1)}^2\right)}\,;\, l = 1, 2, \dots, n-1 \quad (9)$$

These directions should be orthogonal to the current gradient. And the line search vector can be given in a closed form as follows:

$$\delta^{(n)} = -\left(\left(P_X^{(n)}\right)^T A^T A P_X^{(n)} + \lambda I\right)^{-1} \left(P_X^{(n)}\right)^T \nabla_{x_t} D_F(y_t||x_t)$$

$$, t = 1, 2, 3 \dots T . \quad (10)$$

Where, $D_F(y_t||Ax_t) = \frac{1}{2}||y_t - Ax_t||_2^2$

C. Semi Nonegative Matrix Factorization

Ding et al [9] has considered variations of NMF where the elements of one factor but not the other, are constrained to be nonnegative, and so allowing the data matrix Y to have mixed signs [7].In some applications the observed input data are unsigned (unconstrained) as indicated by $Y_\pm \in R^{I\times T}$ which allows us to relax the constraints regarding nonnegativity of one factor. This leads to approximated semi-NMF which can take the following form:

$$Y_\pm = A_\pm X_+ + E \quad (11)$$

Where, the sub index X_+ indicates that a matrix is forced to be nonnegative.

II. THE PROPOSED SYSTEM

In this section we present Fuzzy C-Means Semi-NMF as a new method for underdetermined blind source separation. It's the first work that combines Fuzzy C-Means Clustering algorithm with Semi-Nonnegative Matrix Factorization which allows relaxation to the nonnegativity constraints to the input data mixtures matrix and the mixing matrix. This implies that a fast convergence in the system we proposed. We use PSESOP algorithm that has additive update rules with modified format which we use it in the Semi-NMF methodology to estimate the sources only exploiting the high performance and low complexity time.

The method starts with producing initial base cluster centers, apply the FCM algorithm, and then use PSESOP. Finally, apply the multi-layer algorithm. The whole procedure is summarized as:

(1) Apply the FCM clustering algorithm to estimate the mixing matrix with normal distribution random numbers initial cluster center

(2) Extract the sources using PSESOP in a semi-NMF fashion

(3) Apply the multi-layer algorithm [8] to the preceding steps and get the final source signals

A. Mixing matrix estimation

Using the FCM clustering algorithm we can solve the problem of initialization for NMF and eliminate the selection of parameters, and get a very good result. Here, in this paper unlike [4] we use FCM method directly to the source signals without any transformation although signals are not always sparse in time domain since, the NMF then the multi-layer technique will treat this problem. Given only the observable mixtures matrix Y as follows:

$Y = [Y_1, Y_2, \ldots\ldots, Y_t]$, $t = 1,2,3,\ldots\ldots T$

Where T is the number of sampling time points we can get the J columns of the mixing matrix *A*. We have chosen the initial base cluster centers as a random numbers with normal distribution. The final cluster centers which are the columns of the mixing matrix are computed according to the following algorithm.

FCM clustering algorithm

Input: the observable mixtures $Y = [Y_1, Y_2, \ldots\ldots, Y_T]$, T the number of sample time points, the dimension *I* of each data point of *Y*, the number of desired clusters *C*, Maximum number of iterations *N*, degree of fuzziness β, initial centroids for clusters, tolerance ε for convergence.

Output: the final cluster centers which are the columns of A

Initialize the cluster centers.

While (neither convergence nor reach the maximum number of iterations)

Repeat for i = 1 to C and for j = 1 to T

Compute $d_{ij} = ||Y_j - b_i||$

Repeat for i = 1 to C and for j = 1 to T

If $d_{kj} = 0$ for some $k = k_0$ set $\xi_{k_0 j} = 1$ and $\xi_{ij} = 0$ for all $i \neq k_0$, otherwise compute the membership as in (2)

Repeat for I = 1 to C

Update the cluster centers, using equation (3)

End while

B. Source signals Recovery with Semi-NMF PSESOP

After the columns of the mixing matrix are estimated i.e. we get the mixing matrix, we use the Semi-NMF in order to estimate the source signals matrix. In Semi-NMF the observable mixtures matrix Y and the mixing matrix A are unconstrained (i.e. unsigned) with only the sources matrix is constrained to be nonnegative. PSESOP is then performed in a modified fashion where the matrices Y and A is unconstrained i.e. Semi-NMF. The source signals can be obtained by Semi-NMF PSESOP algorithm. In this work we regard the source signals which is forced to be nonnegative obtained by PSESOP and the mixing matrix obtained by FCM as a preprocessing step for the multi-layer technique which be stated below in detail. The algorithm of Semi-NMF PSESOP is stated below

Semi-NMF PSESOP Algorithm

Input: the observable mixtures $Y = [Y_1, Y_2, \ldots\ldots, Y_t]$, the mixing matrix A obtained from FCM algorithm, number of components J, Maximum number of iterations N

Output: the source signals X

Initialize the matrix X randomly.

Repeat for t=1 to T

Begin

$x_t = x_t^{(0)}, Lp_t = 0, W_t = 0$

Repeat

$P_t = x_t - x_t^{(0)}$

$Q_t = A^T A x_t - A_t y_t$

Compute W_t as in equation (7)

$Lp_t = Lp_t + w_t Q_t$

$P_X = [P_t, Lp_t, Q_t]$

Compute δ as in equation (10)

$x_t = [x_t + P_X\,\delta]_+$ Or,

// $xt = \max\{\varepsilon, x_t + P_X\,\delta\}$

Until stopping criteria is met

End

End for

C. *Multi-layer FCM and Semi-NMF PSESOP algorithm*

The multi-layer technique proposed by A. Cichocki and R. Zdunek 2006 [8,11], also known hierarchical multi-layer technique. The mixing matrix A is replaced in multi-layer NMF by a set of cascaded matrices as follows:

$Y = A^{(1)}A^{(2)}A^{(3)} \dots \dots A^{L}X + E$, where L is the number of layers [5]. Since the model is linear, all the matrices can be merged into a single matrix A if no special constraints are imposed upon the individual matrices A^{l} $(l = 1, 2, 3, \dots, L)$. The multi-layer technique is used to improve the performance of the NMF or Semi-NMF as in this paper. In the first step of the multi-layer algorithm the basic approximate decomposition $Y \cong A^{(1)}X^{(1)} \in R^{I\times T}$ can be performed. Then the result obtained from the first step can be used to build up a new assignment, so$X^{(1)} \cong A^{(2)}X^{(1)} \in R^{J\times T}$. The learning update rules are performed hierarchically or sequentially i.e. layer by layer. Repeating these update rules according to the number of layers to get the final form of the multi-layer model as follows:

$$Y \cong A^{(1)}A^{(2)}A^{(3)} \dots \dots A^{(L)}X^{L} \tag{12}$$

Where, $A = A^{(1)}A^{(2)}A^{(3)} \dots \dots A^{(L)}$ *and* $X = X^{L}$. The multi-layer FCM and Semi-NMF PSESOP (proposed algorithm) is demonstrated as follows:

Multi-layer FCM and Semi-NMF PSESOP algorithm

Input: The observations matrix Y and the number of layers L

Output: The source signals X and signal-to-noise ratio

Repeat for l=1 to L

If l >1 then

Set $Y = XH$; where XH is the estimated source signals

Otherwise, keep Y as it is

Compute A^{l} from FCM clustering algorithm

Compute XH from Semi-NMF PSESOP algorithm

Compute the final result for A as:

$$A = A^{(1)}A^{(2)}A^{(3)} \dots \dots A^{(L)}$$

End for

Get the final result for XH which is the final estimation of the source signals, and then compute the signal-to-noise ratio as in the simulation results.

III. Simulation Results

In this section, we present a good simulation to demonstrate the effectiveness of our proposed algorithm. The experimental analysis is conducted on six speech signals coming from http://sourceforge.net/projects/emu/files/examples/aetobi/ and the number of sensors (mixtures) is set $I = 2$. The mixture signals that we perform our experiments on are obtained by multiplying a matrix of random numbers with normal distribution by the source signals. The source signals are not always sparse in the time domain. The analysis aims at comparing mainly the reconstruction index Signal-to-Noise Ratio (SNR) to evaluate the performance of the proposed method. SNR [10] is defined as:

$$\mathrm{SNR} = -\log\left(\frac{\|\hat{X}-X\|^2}{\|X\|^2}\right) \tag{13}$$

Where, $\hat{X}$ is the estimation of the source signals X. When $SNR \geq 25\,dB$, the efficiency of the separation results is good. The proposed algorithm is implemented by Matlab 7.10.0 R2010a.

Without loss of generality, the experiments are performed by multi-layer FCM and Semi-NMF PSESOP (proposed algorithm) stated above with which the number of layers is only 5 layers and the observation matrix as an input without the constraint of nonnegativity. The sources signals and mixture signals are shown in fig. 1 (a) and (b) respectively.

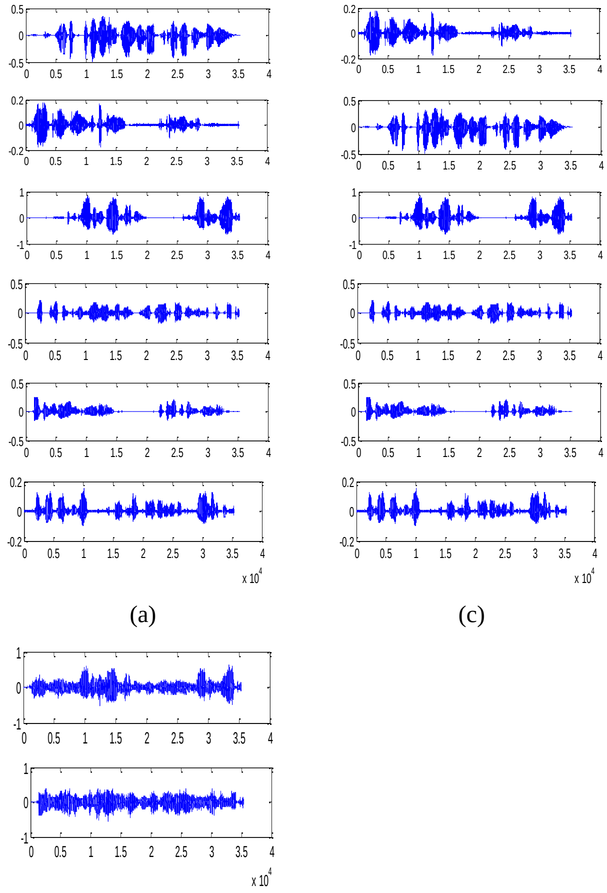

Figure 1. (a) six source signals; (b) two mixture signals; (c) six separated signals

The parameter inputs of FCM clustering algorithm step in the proposed algorithm are 35376, 2, 6, 20, 3, 0.005 for the number of sample points, the dimension of the sample time points, number of clusters, maximum number of iterations, degree of fuzziness, and tolerance respectively. We consider the dimension of the sample points equals to the number of sensors (observable mixtures). The final cluster centers got from this step are the columns of the mixing matrix A.

The parameter inputs of Semi-NMF PSESOP algorithm set in our proposed algorithm are the observable mixtures matrix Y, the mixing matrix A obtained from the previous step instead of initialize it as in [5], the number of components (sources) equals 6, the maximum number of iteration is only 5 iterations.

In order to analyze performance of algorithm proposed, we calculate reconstruction index (SNR) for each separated signal using as (13) respectively. A comparison between our proposed method and other three methods is shown in fig. 2. The three methods are PSESOP, multi-layer PSESOP [5, 11], fuzzy c-means PSESOP that we also propose. The separated signals by our proposed algorithm are shown in fig. 1 (c).

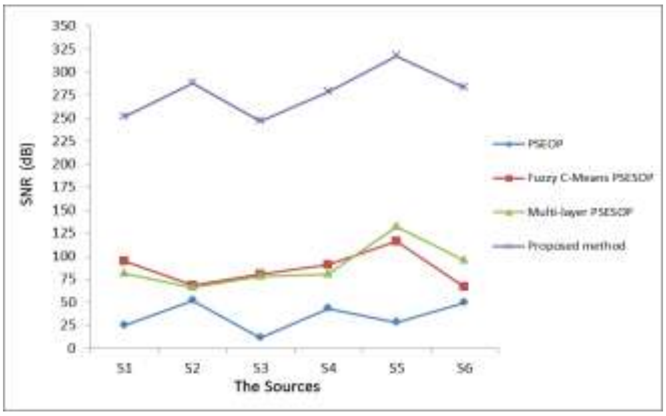

Figure 2. A comparison between different methods

Another evaluation for our proposed method compared with the method in [4] and PSESOP method using the dataset of four speech signals in [4], is recorded in Table 1. From the results in Table 1 and fig. 2, we can see that separation performance of our algorithm is very high and has faster convergence with respect to the other methods.

TABLE I. PERFORMANCE COMPARISON (SNR) USING THE DATESET IN [4]

SNR	The source signals			
	S1	***S2***	***S3***	***S4***
Method in [4]	9.7533	8.0677	5.948	6.422
PSESOP method	24.9054	49.6603	11.4642	43.3386
Proposed method	245.7718	264.1283	191.6499	229.6291

IV. CONCLUSION AND FUTURE WORK

In this paper, we address the problem of underdetermined blind source separation. We proposed a new two-step method of estimating original source signals in the underdetermined case based on Multi-layer technique using the FCM to estimate the mixing matrix in the first step and then using the Semi-NMF PSESOP to estimate the sources in the second step repeating these steps for a number of layers. The proposed algorithm partially relaxes the nonnegativity and the sparsity conditions. Also, this new algorithm can effectively improve the mixing matrix estimation, and significantly reduce the computation costs of source signals extraction. The simulation experiments illustrate the validity, some advantages, and the superior performance of the proposed algorithm.

Our future work will make a reduction to the sample time points, enhance the mixing matrix estimation, and separate non-sparse signals. We would also like to reduce the running time.

VI. REFERENCES

[1] Wang Rongjie, Zhan Yiju, Zhou Haifeng, “A method of underdetermined blind source separation with an unknown number of sources,” Engineering Applications of Artificial Intelligence, vol. , pp. , June 2011

[2] Fengbo Lu, Zhitao Huang,Wenli Jiang, “Underdetermined blind separation of non-disjoint signals in time–frequency domain based on matrix diagonalization,” Signal Processing 91,pp. 1568–1577, Jan. 2011

[3] Dezhong Peng, Yong Xiang, “Underdetermined blind separation of non-sparse sources using spatial time-frequency distributions,” Digital Signal Processing 20, pp. 581-596, August 2009

[4] Chaozhu Zhang, Cui Zheng, “Underdetermined Blind Source Separation Based on Fuzzy C-Means Clustering and Sparse Representation,”, 2004

[5] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, Shun-ichi Amari, “Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation”, John Wiley, 2009

[6] Kuo-Lung Wu,“Analysis of parameter selections for fuzzy c-means,” Pattern Recognition 45, pp. 407-415, July 2011

[7] Vladimir Nikulin, Tian-Hsiang Huangb, Shu-Kay Ngc, Suren I. Rathnayake, Geoffrey J. McLachlan, “A very fast algorithm for matrix factorization,” Statistics and Probability Letters 81, pp. 773-782, Feb. 2011

[8] A. Cichocki and R. Zdunek. “Multilayer nonnegative matrix factorization. Electronics Letters,” 42(16),pp. 947–948, 2006.

[9] Ding, C., Li, T., Jordan, M.I, “Convex and semi-nonnegative matrix factorizations”. IEEE Transactions on Pattern Analysis and Machine Intelligence 32, pp.45–55, 2010.

[10] Pau Bofill, Michael Zibulevsky, “Underdetermined blind source separation using sparse representations,” Signal Processing 81, pp. 2353–2362, 2001

[11] A. Cichocki and R. Zdunek. “Multilayer nonnegative matrix factorization using projected gradient approaches”, In The 13th International Conference on Neural Information Processing (ICONIP06), Hong Kong, October 3–6 2006. http://iconip2006.cse.cuhk.edu.hk/.

[12] G. Narkiss and M. Zibulevsky. Sequential subspace optimization method for large-scale unconstrained problems. Optimization Online, page 26, October 2005.

[13] Yadong, Liu,Zongtan,Zhou,Dewen,Hu,2005. Anovel method for spatio-temporal pattern analysis of brain fMRI data. Science in China Series F:Information Sciences 48(2),pp. 151–160.

[14] Araki S.,Makino,S.,Blin,A.,2004. Underdetermined blind separation for speechin real environment with sparseness and ICA. In:ProceedingsoftheICASSP’04, Montreal, Canada, pp.881–884.

[15] Ohnishi, Naoya,Imiya,Atsushi,2008. Independent component analysis of optical flow for robot navigation.Neurocomputing71(10–12),pp. 2140–2163.

[16] Tonazzini, Anna,Bedini,Luigi,Salerno,Emanuele, A Markov model for blind image separation by a mean-field EM algorithm. IEEE Transactions on Image Processing 15(2), pp.473–482,2005.

[17] Er-Wei, Bai,QingYu,Li,Zhiyong,Zhang. Blind source separation/ channel equalization of nonlinear channels with binary inputs. IEEE Transactions on Signal Processing53(7), pp.2315–2323, 2005.

[18] Tuomas Virtanen, "Monaural Sound Source Separation by Nonnegative Matrix Factorization with Temporal Contiuity and Sparse Criteria", IEEE Transactions on Audio, Speech, and Language Processing, Vol. 15, No. 3, March 2007

[19] A. Cichocki, R. Zdunek, and S. Amari. New algorithms for non-negative matrix factorization in applications to blind source separation. In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP2006, volume 5, pages 621–624, Toulouse, France, May 14–19 2006.