BLIND EXTRACTION OF SPARSE SOURCES

Nasser Mourad and James P. Reilly

Department of Electrical & Computer Eng.
McMaster University, 1280 Main St. W., Hamilton, Ontario, Canada L8S 4K1
email: mouradna@mcmaster.ca, reillyj@mcmaster.ca

ABSTRACT

In this paper we propose a new algorithm for solving the blind source extraction (BSE) problem when the desired source signals are sparse. Previous approaches for solving this problem are based on the independent component analysis (ICA) technique, that extracts a source signal by finding a separating vector that maximizes the *non–Gaussianity* of the extracted source signal. These algorithms are general purpose algorithms and are not designed specifically for extracting sparse signals. In this paper we propose a new algorithm for extracting sparse source signals. The proposed algorithm is based on finding a separating vector that maximizes the *sparsity* of the extracted source signal. In the proposed algorithm, a non-convex objective function that measures the sparsity of the separated signal is locally replaced by a quadratic convex function. This results in an iterative algorithm in which a new estimate of the separating vector is obtained by solving an eigenvalue decomposition problem. A numerical example is presented to investigate the superiority of the proposed algorithm in comparison with one of the well known ICA algorithm for extracting sparse sources. .

Indexing Terms: *blind source separation, sparse component analysis, independent component analysis.* .

I. INTRODUCTION

The blind source separation (BSS) problem is defined as the problem of reconstructing n *unknown* source signals from m linear measurements when the mixing matrix is *unknown*. The relation between the measured signals and the original source signals can be expressed mathematically as

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S}, \tag{1}$$

where $\boldsymbol{X} \in \mathbb{R}^{m \times T}$ is a matrix of measured signals, $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is an *unknown* mixing matrix, $\boldsymbol{S} \in \mathbb{R}^{n \times T}$ is a matrix of it unknown source signals, m is the number of observations, n is the number of sources, and T is the number of samples. In this paper we consider the case $m = n$.

Over the last two decades, the BSS problem has been solved using the independent component analysis technique (ICA). One approach for solving the BSS problem via ICA is to estimate the hidden sources simultaneously. This is usually done by finding a separating matrix $\boldsymbol{B}$ such that the estimated sources $\boldsymbol{Y} = \boldsymbol{B}\boldsymbol{X}$ are mutually independent [1]. Another approach for solving the BSS problem via ICA is to sequentially extract the source signals one after the other, a technique known in the literature as blind signal extraction (BSE) [2]. The most popular algorithm that follows this strategy is called FastICA [3]. FastICA is based on finding a separating vector $\boldsymbol{b}$ by maximizing the *non Gaussianity* of the separated signal $\boldsymbol{y} = \boldsymbol{b}^T \boldsymbol{X}$.

As stated in [2], the BSE approach has the following advantages over the simultaneous BSS approach; 1) signals can be extracted in a specific order according to some features of the source signals, 2) the approach is very flexible in the sense that various criteria can be used in each stage depending on the features of the source to be extracted, 3) only interesting source signals need to be extracted, and 4) extensive computing time and resources can be saved. For these reasons, BSE has become one of the more promising techniques for solving the BSS problem, especially in cases when only few sources with specific stochastic properties or features need to be extracted.

In this paper we consider the case when the desired sources are *sparse*. The ICA based algorithms that can extract sparse sources, e.g., the FastICA algorithm [3], are based on modeling the sparse source signal as a super–Gaussian signal. Hence, a separating vector is estimated by *maximizing* the kurtosis of the extracted source signal. However, in this paper we propose an iterative algorithm that uses an objective function that explicitly measures the *diversity* (antisparsity) of the extracted source signal. As will be shown in Section IV, the proposed algorithm can extract sparse source signals with relatively smaller residual error than the FastICA algorithm. The primary computational operation in each iteration is an $m \times m$ eigendecomposition, so the method is not computationally demanding. Since the proposed method minimizes the sparsity of the extracted signals, it may be more appropriate in some cases than previous methods which maximize kurtosis.

It is worth mentioning that, there is a significant difference between the proposed algorithm and the sparse component analysis (SCA) technique. The SCA technique solves the BSS problem by finding an estimate of the mixing matrix via clustering the columns of the measured matrix $\boldsymbol{X}$ [4], [5]. Therefore, the SCA technique solves the *simultaneous* BSS problem, and is restricted to the case where *all* the sources are sparse. However, in the proposed algorithm only the desired sources are assumed

This work was supported in part by the Egyptian Ministry of High Education and the Natural Sciences and Engineering Research Council of Canada (NSERC).

to be sparse, and these sources are extracted sequentially. The proposed algorithm is referred to as *blind sparse signal extraction* (BSSE).

This paper is organized as follows. In Section II we derive an analytic solution for extracting a sparse source signal. The proposed algorithm is proposed in Section III. Section IV presents a numerical example for assessing the performance of the proposed algorithm. Finally, conclusions are given in Section V.

II. ANALYTIC SOLUTION

Consider the case where the set of indices of the zero entries of the desired source, say the ith source, is known. Let $\mathcal{I}_i$ refer to these indices, i.e., $\boldsymbol{s}^i(\mathcal{I}_i) = 0$. Due to the assumption that the supports of the original source signals are independent, then, with high probability, there are some indices for which $\boldsymbol{s}^j(\mathcal{I}_i) \neq 0, \forall j \neq i$, i.e., $||\boldsymbol{s}^j(\mathcal{I}_i)||_{\ell_2} > 0$. Therefore, if $\boldsymbol{B}$ is a separating matrix, and neglecting the permutation indeterminacy, the entries of the ith row of the matrix

$$\boldsymbol{Y}_{\mathcal{I}_i} = \boldsymbol{B}\boldsymbol{X}_{\mathcal{I}_i}. \tag{2}$$

are all zeros, where $\boldsymbol{X}_{\mathcal{I}_i}$ is a sub–matrix of $\boldsymbol{X}$ with columns corresponding to the indices in $\mathcal{I}_i$. Accordingly, the ith row of the separating matrix $\boldsymbol{B}$ can be obtained by solving the following optimization problem

$$\tilde{\boldsymbol{b}}^i = \arg\min_{\bar{\boldsymbol{b}}} \mathcal{J}_{\bar{b}} \triangleq ||\bar{\boldsymbol{b}}\boldsymbol{X}_{\mathcal{I}_i}||^2_{\ell_2} \quad \text{subject to} \quad ||\bar{\boldsymbol{b}}||_{\ell_2} = 1. \tag{3}$$

where the constraint $||\bar{\boldsymbol{b}}||_{\ell_2} = 1$ is utilized to prevent the trivial solution $\bar{\boldsymbol{b}} = 0$.

Following the standard method of Lagrangian multipliers (see, e.g. [6]), the optimization problem (3) can be readily solved with a solution vector satisfying

$$\boldsymbol{X}_{\mathcal{I}_i}\boldsymbol{X}^T_{\mathcal{I}_i}\bar{\boldsymbol{b}}^T = \lambda\bar{\boldsymbol{b}}^T. \tag{4}$$

where λ is the Lagrangian multiplier. Equation (4) shows that λ and $\bar{\boldsymbol{b}}^T$ are eigenvalue and eigenvector of the correlation matrix $\boldsymbol{X}_{\mathcal{I}_i}\boldsymbol{X}^T_{\mathcal{I}_i}$, respectively. Moreover, substituting (4) into the expression of the objective function in (3) we get

$$\mathcal{J}_{\bar{b}} = ||\bar{\boldsymbol{b}}\boldsymbol{X}_{\mathcal{I}_i}||^2_{\ell_2} = \bar{\boldsymbol{b}}\boldsymbol{X}_{\mathcal{I}_i}\boldsymbol{X}^T_{\mathcal{I}_i}\bar{\boldsymbol{b}}^T = \lambda\bar{\boldsymbol{b}}\bar{\boldsymbol{b}}^T = \lambda. \tag{5}$$

where in the last equality we used the fact that $||\bar{\boldsymbol{b}}||_{\ell_2} = 1$. Therefore, for minimizing the cost function $\mathcal{J}_{\bar{b}}$, λ must be minimum. Accordingly, the solution of the optimization problem (3) is given by

$$\left(\tilde{\boldsymbol{b}}^i\right)^T = \text{eig}_{min}(\boldsymbol{X}_{\mathcal{I}_i}\boldsymbol{X}^T_{\mathcal{I}_i}). \tag{6}$$

where $\text{eig}_{min}(\boldsymbol{R})$ is the eigenvector of the correlation matrix $\boldsymbol{R}$ corresponding to the minimum eigenvalue.

Equation (6) states that the separating vector of the ith source can be readily estimated from the measured signals once the indices at which the ith source equals zero are known. However, this is of little direct value, since these indices are generally unknown. We now propose an iterative algorithm that converges in its limit to (6).

III. PROPOSED ITERATIVE ALGORITHM

As described in Section I, the ICA-based algorithms that solve the BSE problem are based on finding a separating vector $\boldsymbol{b}$, which maximizes the *non Gaussianity* of the separated signal $\boldsymbol{y} = \boldsymbol{b}^T\boldsymbol{X}$. In this section we propose an iterative algorithm that solves the BSE when the desired source signals are sparse. The proposed algorithm is based on estimating a separating vector that minimizes the diversity (the number of nonzero samples) of the separated source signal. Before deriving the proposed algorithm we provide a justification to minimizing the diversity of the separated signal as an objective function.

III-A. The Diversity of the separated signal as an objective function

In this subsection we provide a justification to minimizing the diversity of the separated source signal as an objective function for extracting sparse source signals. Consider the case of estimating the ith row of the separating matrix $\boldsymbol{B}$ such that the corresponding row $\boldsymbol{y}^i \in \mathbb{R}^T$ of $\boldsymbol{Y}$, corresponding to the ith separated source, is sparse. The ith row of $\boldsymbol{Y}$ is expressed as

$$\boldsymbol{y}^i = \boldsymbol{b}^i\boldsymbol{X} = \boldsymbol{b}^i\boldsymbol{A}\boldsymbol{S} = \boldsymbol{c}^i\boldsymbol{S} = \sum_{j=1}^{n} c^i[j]\boldsymbol{s}^j. \tag{7}$$

where $\boldsymbol{b}^i$ is the ith row of $\boldsymbol{B}$, $\boldsymbol{c}^i = \boldsymbol{b}^i\boldsymbol{A}$, and $\boldsymbol{s}^j$ is the jth source signal, i.e., the jth row of $\boldsymbol{S}$. Since the order of the estimated source is irrelevant, the superscript ^{i} will be dropped from the sequel equations, and we use the notation $\bar{(\,)}$ to refer to a row vector. Accordingly (7) can be rewritten as

$$\bar{\boldsymbol{y}} = \bar{\boldsymbol{b}}\boldsymbol{X} = \bar{\boldsymbol{c}}\boldsymbol{S} = \sum_{j=1}^{n} c[j]\boldsymbol{s}^j, \tag{8}$$

which shows that the estimated source signal is a linear combination of the original source signals.

Assume that the *regions of support* of the original source signals are independent, where the support of the signal refers to the indices of the nonzero samples. Then, with high probability, the *diversity* of the summation of two source signals is greater than the diversity of any one of the two sources. Accordingly, the sparsest estimated signal $\bar{\boldsymbol{y}}$ can be obtained if $\bar{\boldsymbol{c}}$ has only one nonzero entry. Since $\bar{\boldsymbol{c}} = \bar{\boldsymbol{b}}\boldsymbol{A}$, this implies that $\bar{\boldsymbol{b}}$ corresponds to a scaled version of one row of $\boldsymbol{A}^{-1}$, the inverse of the mixing matrix. Accordingly, one sparse source signal can be obtained by finding a separating vector $\bar{\boldsymbol{b}}$ that minimizes the diversity of the estimated source vector $\bar{\boldsymbol{y}} = \bar{\boldsymbol{b}}\boldsymbol{X}$.

III-B. Extracting a sparse source signal

Since the prior information available about the desired source signal is its sparsity, the separating vector $\bar{\boldsymbol{b}}$ can be estimated by minimizing the following objective function

$$\tilde{\bar{\boldsymbol{b}}} = \arg\min_{\bar{\boldsymbol{b}}} g\left(\bar{\boldsymbol{y}}\right) = g(\bar{\boldsymbol{b}}\boldsymbol{X}) \quad \text{subject to} \quad ||\bar{\boldsymbol{b}}||_{\ell_2} = 1. \tag{9}$$

where $g(\bar{\boldsymbol{y}})$ is a function that measures the diversity (antisparsity) of the separated vector $\bar{\boldsymbol{y}}$. In this paper we consider a class of objective functions of the form $g(\bar{\boldsymbol{y}}) = \sum_{t=1}^{T} g_c(y[t])$, where $g_c(\cdot)$ is a symmetric and monotonically increasing concave function on the nonnegative orthant $\mathcal{O}_1$ [7]. In [7] we proposed replacing $g(\bar{\boldsymbol{y}})$ by the quadratic convex function $f(\bar{\boldsymbol{y}})$ defined as

$$\begin{aligned} f(\bar{\boldsymbol{y}}) &= g(\bar{\boldsymbol{y}}_0) + \nabla g(\bar{\boldsymbol{y}}_0)(\bar{\boldsymbol{y}} - \bar{\boldsymbol{y}}_0)^T \\ &- 0.5(\bar{\boldsymbol{y}} - \bar{\boldsymbol{y}}_0)\nabla^2 g(\bar{\boldsymbol{y}}_0)(\bar{\boldsymbol{y}} - \bar{\boldsymbol{y}}_0)^T \end{aligned} \quad (10)$$

where $\bar{\boldsymbol{y}} \in \mathbb{R}^T$, $g(\bar{\boldsymbol{y}}) = \sum_i g_c(y[i])$ and $\nabla g(\bar{\boldsymbol{y}}_0)$, and $\nabla^2 g(\bar{\boldsymbol{y}}_0)$ are the gradient and Hessian of $g(\bar{\boldsymbol{y}})$ at $\bar{\boldsymbol{y}} = \bar{\boldsymbol{y}}_0$, respectively. A more general expression for $f(\bar{\boldsymbol{y}})$ is presented in [8].

For the case of $g_c(y[i]) = \log(|y[i]|)$, it is straightforward to show that $f(\bar{\boldsymbol{y}})$ in this case (at the kth iteration) is reduced to

$$f(\bar{\boldsymbol{y}}) = ||\bar{\boldsymbol{y}}\boldsymbol{W}_k||_{\ell_2}^2 + C, \quad (11)$$

where C is a constant that does not depend on $\bar{\boldsymbol{y}}$, and $\boldsymbol{W}_k$ is a diagonal matrix, whose tth diagonal element is given by

$$W_k[t,t] = \frac{1}{|y_{k-1}[t]|}, \quad t = 1, \dots, T. \quad (12)$$

where $y_{k-1}[t]$ is the estimate of $y[t]$ at the previous iteration. The proposed algorithm is based on minimizing the objective function $g(\bar{\boldsymbol{y}})$ by iteratively minimizing $f(\bar{\boldsymbol{y}})$. Therefore, given an initial estimate of the separating vector $\bar{\boldsymbol{b}}_k$, a new estimate can be obtained by solving the following optimization problem

$$\bar{\boldsymbol{b}}_{k+1} = \arg\min_{\bar{\boldsymbol{b}}} ||\bar{\boldsymbol{b}}\boldsymbol{X}\boldsymbol{W}_k||_{\ell_2}^2 \quad \text{subject to} \quad ||\bar{\boldsymbol{b}}||_{\ell_2} = 1. \quad (13)$$

Comparing (13) with (3) we readily find that

$$\bar{\boldsymbol{b}}_{k+1}^T = \text{eig}_{min}\left(\boldsymbol{X}_k\boldsymbol{X}_k^T\right). \quad (14)$$

where $\boldsymbol{X}_k = \boldsymbol{X}\boldsymbol{W}_k$. Note that the t-th column of $\boldsymbol{X}_k$ is a scaled version of the t-th column of $\boldsymbol{X}$, with a scaling factor given by $W_k[t,t]$ defined in (12). Since $W_k[t,t]$ is inversely proportional to $|y_k[t]|$, multiplying the data matrix $\boldsymbol{X}$ by the diagonal matrix $\boldsymbol{W}_k$ has the effect of selecting the columns of the data matrix with indices corresponding to the indices of the small entries in the estimated source vector $\bar{\boldsymbol{y}}_k$, in a manner similar to the iterative reweighted least squares (IRLS) algorithms. As the algorithm converges to the desired sparse signal, the diagonal matrix becomes more selective and $\boldsymbol{X}_k$ converges to $\boldsymbol{X}_{\mathcal{I}_i}$, defined in (2). The proposed algorithm is summarized in Table I

III-C. Removing the extracted source from the mixture.

After estimating the source signal $\bar{\boldsymbol{y}}$ using the BSSE algorithm presented in Table I, the estimated source must be removed from the mixture before extracting a new source. This can be done by finding a vector $\boldsymbol{h} \in \mathbb{R}^n$ that minimizes the following objective function [2]

$$\mathcal{J}(\boldsymbol{h}) = \sum_{t=1}^{T} \boldsymbol{x}_{jt}^T \boldsymbol{x}_{jt}. \quad (15)$$

where $\boldsymbol{x}_{jt}$ is the t-th column of the matrix $\boldsymbol{X}^j = \boldsymbol{X}^{j-1} - \boldsymbol{h}\bar{\boldsymbol{y}}$, where j is an index corresponding to the number of the previously extracted source signals. Minimizing (15) we have

$$\boldsymbol{h} = \frac{\boldsymbol{X}^{j-1}\bar{\boldsymbol{y}}^T}{||\bar{\boldsymbol{y}}||_{\ell_2}^2} = \frac{\boldsymbol{X}^{j-1}\left(\boldsymbol{X}^{j-1}\right)^T\bar{\boldsymbol{b}}^T}{||\bar{\boldsymbol{y}}||_{\ell_2}^2} \quad (16)$$

where $\bar{\boldsymbol{b}}$ is the separating vector obtained from the BSSE algorithm in Table I. Note that, $\boldsymbol{h}$ is an estimate of the mixing column associated with the extracted source $\bar{\boldsymbol{y}}$. The new matrix $\boldsymbol{X}^j$ can then be used as an input to the BSSE algorithm for extracting the $(j+1)$–st source signal, which in turn has to be removed from $\boldsymbol{X}^j$ before extracting a new source signal. This process can be repeated until all the desired source signals are extracted. Note that, after removing $j \geq 1$ source signals, and for estimating the separating vector of the $(j+1)$–st source signal, the second step in Table I must be changed into

$$\bar{\boldsymbol{b}}_{k+1}^T = \text{eig}_{j+1}\left(\boldsymbol{X}^j\boldsymbol{W}_k^2\left(\boldsymbol{X}^j\right)^T\right).$$

where $\text{eig}_{j+1}(\boldsymbol{R})$ is the eigenvector corresponding to the $(j+1)$–st smallest eigenvalue of the correlation matrix $\boldsymbol{R}$.

Table I. Extraction of a sparse source

$[\bar{\boldsymbol{y}}, \bar{\boldsymbol{b}}]$ = BSSE($\boldsymbol{X}$, $\bar{\boldsymbol{b}}_0$)

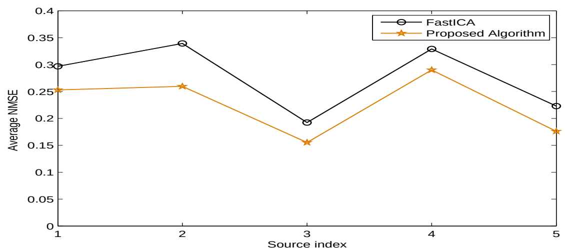
Select a small threshold parameter ϵ, and calculate $\bar{\boldsymbol{y}}_0 = \bar{\boldsymbol{b}}_0\boldsymbol{X}$.

1) For $k = 0, 1, \dots$, repeat until convergence:
   - Calculate $W_k[t,t] = \frac{1}{|y_k[t]|}, \quad t = 1, \dots, T.$
   - Update the separating vector: $\bar{\boldsymbol{b}}_{k+1}^T = \text{eig}_{min}\left(\boldsymbol{X}\boldsymbol{W}_k^2\boldsymbol{X}^T\right)$.
   - Get a new estimate of the separated signal: $\bar{\boldsymbol{y}}_{k+1} = \bar{\boldsymbol{b}}_{k+1}\boldsymbol{X}$.
   - if $||\bar{\boldsymbol{y}}_{k+1} - \bar{\boldsymbol{y}}_k||_{\ell_2}/||\bar{\boldsymbol{y}}_{k+1}||_{\ell_2} < \epsilon$, stop.
2) Output $\bar{\boldsymbol{y}}_{k+1}$ and $\bar{\boldsymbol{b}}_{k+1}$ as the solutions.

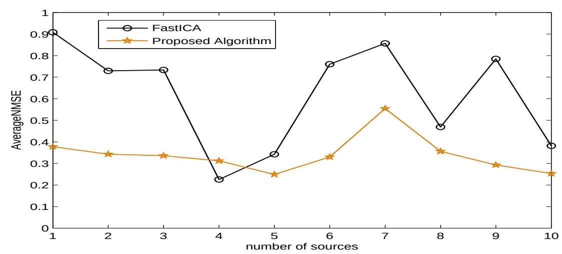
End

IV. SIMULATION RESULTS

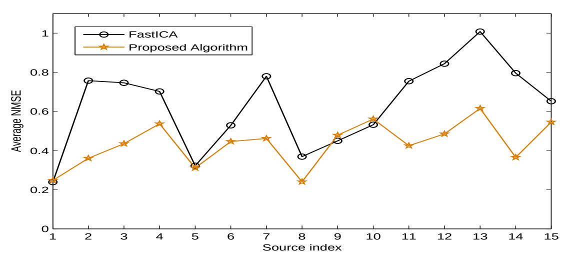
In this example we provide a comparison between the BSSE algorithm, summarized in Table I, and the FastICA algorithm in solving the BSS problem when the hidden sources are sparse. The measured signals are generated by multiplying a (5×50) randomly generated sparse source matrix by a square mixing matrix $\boldsymbol{A} \in \mathbb{R}^{5\times5}$ randomly generated from an $i.i.d.$ normal distribution with zero mean and unit variance. Each row of the sparse source matrix $\boldsymbol{S}$ is constructed such that it has exactly 8 nonzero entries. The indices of these 8 entries are randomly selected, and their amplitudes are chosen from a uniform distribution between ±1. Figure 1 provides a comparison between the proposed BSSE algorithm and the FastICA algorithm in



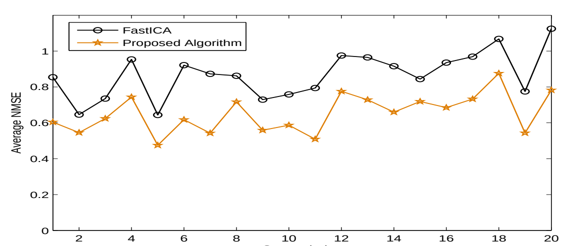
(a)



(b)



(c)



(d)

Fig. 1. The average NMSE between the estimated sources and the original ones. The x axis represents the indices of the original sources, and the y axis represents the average NMSE over 200 different runs corresponding to 200 different selections of the square mixing matrix. The figures (a)–(d) correspond to the cases n = 5, 10, 15, and 20, respectively.

terms of the average NMSE and the size of the BSS problem, where the NMSE between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined as

$$NMSE = \left|\left| \frac{\boldsymbol{x}}{||\boldsymbol{x}||_2} - \frac{\boldsymbol{y}}{||\boldsymbol{y}||_2} \right|\right|_2. \tag{17}$$

In Figure 1 we consider the following four different values of the number of sources, $n = 5, 10, 15,$ and 20, respectively. For each value of n, a sparse $(n \times 100)$ source matrix is generated, and 200 different square mixing matrices are generated. For each one of the mixing matrices, the BSS problem is solved using the proposed BSSE algorithm and the FastICA algorithm. The NMSE between each source and the corresponding estimate is calculated, and the average over the 200 different trials is plotted in Figure 1. As shown in this figure, and for each source in the four different cases, the average NMSE for the proposed algorithm is smaller than that of the FastICA algorithm. Accordingly, the proposed algorithm can estimate sparse sources more accurately than the FastICA algorithm.

V. CONCLUSION

In this paper we proposed a new algorithm for solving the BSE problem when the desired source signals are sparse. The proposed algorithm is based on finding a separating vector that maximizes the *sparsity* of the extracted source signal. The nonconvex objective function that measures the sparsity of the separated signal was locally replaced by a quadratic convex function. This resulted in an iterative algorithm in which a new estimate of the separating vector was obtained by solving an eigenvalue decomposition problem. A numerical example demonstrates the superior performance of the proposed algorithm in comparison with one of the well known ICA algorithms for extracting sparse sources.

VI. REFERENCES

[1] A.J. Bell, and T.J. Sejnowski, " An information-maximization approach to blind source separation and blind deconvolution," Neural Comp., vol. 7, no. 6, pp. 1129-1159, 1995.

[2] A. Cichocki and S. Amari, "Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications," John Wiley & Sons, Ltd.

[3] A. Hyvrinen, "Fast and robust fixed-point algorithms for independent component analyis," IEEE Trans. Neural Net., vol. 10, no. 3, pp. 626-634, 1999.

[4] M. B. Zadeh, C. Jutten, A. Mansour, "Sparse ICA via cluster-wise PCA," Neurocomputing, vol. 69, pp. 1458–1466, 2006.

[5] Z. He, A. Cichocki, Y. Li, S. Xie, and S. Sanei, "K-hyperline clustering learning for sparse component analysis," Signal Processing, vol. 89, no. 6, pp. 1011-1022, 2009.

[6] E. de Klerk, C. Roos, and T. Terlaky, "Nonlinear Optimization", Waterloo, August 26, 2004.

[7] N. Mourad and J. Reilly, "ℓ_p Minimization for Sparse Vector Reconstruction", *ICASSP2009*, Taipei, Taiwan, Apr. 19-24, 2009.

[8] Nasser Mourad and J. Reilly, "Minimizing Nonconvex Functions for Sparse Vector Reconstruction", In preparation.