Image Space Reconstruction Based Overcomplete ICA Algorithm

Shangming Yang, Min Wan and Hua Yan
Computational Intelligence Laboratory
School of Computer Science and Engineering
University of Electronic Science and Technology of China, Chengdu, 610054, China
minn003@163.com, wan_min@126.com

***Abstract*— In this paper, based on image space reconstruction algorithm and regression ICA, a new algorithm for overcomplete ICA is developed. In this algorithm, it first reconstructs part of the observations and then predicates the missing observations from their probability distributions and the available observations. After these steps, the standard ICA algorithm is employed to estimate all the independent components. In the paper, the convergence of the proposed algorithm is also analyzed. Experiment demonstrates that the algorithm can obtain good fidelity for blind signal separation. In the simulation, three speech signals are successfully separated while only two mixtures of the three signals are given.**

I. Introduction

Up to now, independent component analysis (ICA) and blind source separation (BSS) have been applied in many areas such as speech recognition systems, telecommunications, and medical signal processing. ICA is one of the most important tools for recovering independent sources with sensor observations that are unknown linear mixtures of the unobserved independent source signals [1-6]. A number of publications have proposed various methods for the standard ICA noise model and the ICA noise free model which require at least as many sensors as sources. For those models, however, if the number of original sources is often unknown, or the number of observed signals from sensors is less or greater than the original ones, it is very hard to estimate their original sources. To solve this sort of problems, researchers developed the overdetermine and overcomplete ICA algorithms. In this paper we only interested in the solutions of overcomplete ICA model, in which the number of the observations is less than that of the original signals. In overcomplete ICA, Lewicki and Sejnowski in [7] derived a gradient-based method called learning overcomplete representations of the data that allowed for more basis vectors than dimensions in the inputs, and this algorithm has a requirement for the assumption of a low level of noise. In [8], Girolami presented an expectation-maximization algorithm for learning sparse and overcomplete data representations. The proposed algorithm exploited a variational approximation to a range of heavy-tailed distributions whose limit was the Laplacian. Yuanqing Li et al. [9] presented a sparse decomposition approach of observed data matrix which is used in blind source separation with less sensors than sources. Lee et al. [10] proposed an overcomplete ICA technique which assumes a linear mixing model with additive noise and involve two steps, the first is to learn an overcomplete representation and the second is to infer sources given a sparse prior on the coefficients. Shi et al. [11] introduced a blind separation technique which also includes two steps. The first step was to learn the mixing matrix for the observed data using the sparse mixture model, and the second was to infer the sources by solving a linear programming problem after the mixing matrix is estimated. This algorithm further assumes if the sources were sparse, the mixing matrix could be estimated by a method called the generalized exponential mixture model. After estimating the mixing matrix, the sources could be obtained by a posteriori approach. If we consider noisy sources to be variables in the ICA, its observations will be less than original sources and can be considered as one of the overcomplete models. In noisy ICA, Employing bias removal, Cichocki et al. [12] proposed modified algorithm, which can reduce the noise to very low level. Jianting Cao et al. [13] proposed an approach to high level noisy ICA which can separate the mixtures of sub-Gaussian and super-Gaussian source components. Mathis [14] introduced a three-step blind signal separation algorithm for noisy environment data in which three different approaches are used to mitigate the effects of additive noise in the transfer medium. Motivated by these methods, we present a new algorithm for the mixture model that is able to expand overcomplete ICA and noisy ICA to the standard ICA. After the n-m missing observations are estimated by regression ICA, the sources are estimated by using a standard ICA algorithm. For the proposed algorithm, experiments with speech signals demonstrate good separation results.

II. ICA Regressions

For overcomplete ICA model, the number of the observation variables is less than that of the original signals. In standard ICA, the numbers of observations and original sources are equal, and the observations are noise free. A random vector $\mathbf{x}(t)$ for this model is defined as

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t), \tag{1}$$

where $\boldsymbol{A}$ is an $(m \times n)$ mixing matrix $(m \geq n)$ and $\mathbf{s}(t) = [s_1(t)...s_n(t)]^T$ is a source vector of stochastically independent signals. ICA is obtained by estimating the mixing matrix $\mathbf{A}$. As the estimation of a separating or de-mixing matrix $\mathbf{W}$ and/or a mixing matrix $\mathbf{A}$ for noisy and overcompltete ICA is rather difficult, the majority of past research efforts were devoted to the noiseless case.

Based on the probability distributions of the original sources, regression ICA [15] can be used to estimate the missing observations for the overcomplete ICA. In general, the observations in vector $\mathbf{x}$ are separated into two parts, the observed variables and the missing variables. Since in overcomplete ICA, the number of observations m is less than the number of the original signals n, that is, we only have the data of m observations, and from which we expect to find other n-m missing observation signals so that we can use standard ICA algorithm to solve overcomplete ICA problems. We consider this m observations to be the observed variables, and the other n-m unknown variables to be the predicting variables in regression ICA. For simplicity, the m first variables form the vector of the observed variables $\mathbf{x}_o = (x_1, ..., x_m)^T$, and the other n-m variables form the vector of the missing variables $x_m = (x_{m+1}, ..., x_n)^T$. The regression ICA model can be written as

$$\begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_m \end{bmatrix} = \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{A}_m \end{bmatrix} \mathbf{s}. \tag{2}$$

Therefore, for a given observation of $\mathbf{x}_o$, if we can obtain $\mathbf{x}_m$ for the overcomplete ICA, then it will be transformed to standard ICA. Now the problem is how to predict $\mathbf{x}_m$. To predict $\mathbf{x}_m$ correctly, the joint probability distribution of $\mathbf{x}$ must be used. To be more precisely, we must have some previous observations of $\mathbf{x}_m$ to be able to estimate the joint probability distribution, which means we need to measure how the predicted variables depend on the predicting variables. The regression $\mathbf{x}_m$ is conventionally defined as the conditional expectation:

$$\hat{\mathbf{x}}_m = E\{\mathbf{x}_m|\mathbf{x}_o\} = E\{\mathbf{A}_m\mathbf{s}|\mathbf{x}_o\} = \mathbf{A}_m \int_{\mathbf{A}_o s = \mathbf{x}_o} \mathbf{s} p(\mathbf{s}) d\mathbf{s}. \tag{3}$$

Denote the probability density of each s_i by p_i, and by $g_i(u) = p_i'(u)/p_i(u) + cu$ a function that equals the negative score function p_i'/p_i of the probability density of s_i plus an arbitrary linear term. In practical application, the negative score function is usually considered the same for all i. For example, we can choose the $tanh$ function for the score function of a mildly super-Gaussian distribution. Denote further by g the multi-dimensional function that consists of applying g_i on the ith component of its argument, for every i. After the above preprocessing and assumptions we have the following result [15]:

$$E\{\mathbf{x}_m|\mathbf{x}_o\} \approx \mathbf{A}_m g(\mathbf{A}_o^T \mathbf{x}_o), \tag{4}$$

where we can take $g_i(u) = -tanh(u) + u$ for all i. In this case, the vector $\mathbf{A}_o^T\mathbf{x}_o$ can be interpreted as an initial linear estimate of $\mathbf{s}$. In a standard ICA $\mathbf{x} = \mathbf{A}\mathbf{s}$, after whitening, we can have an orthogonal matrix $\bar{\mathbf{A}}$ to replace A, and therefore $\mathbf{A}_o^T$ can be consider the pseudoinverse of $\mathbf{A}_o$. Obtaining initial estimation $\hat{\mathbf{s}} = g(\mathbf{A}_o^T\mathbf{x}_o)$ is the key step in this algorithm. We will discuss this in the following section.

III. Proposed Overcomplete ICA Algorithm

A. Prediction of the Observations

In ICA, the only data we have are the observations. To obtain the regression $\mathbf{x}_m$ in (3), we need to find a way to reconstruct $\mathbf{x}_0$ so that we have a separation of the observations. Finding the mixing matrix $\mathbf{A}_0$ in (4) is the first step of solving the overcomplete ICA problem here. For the overcomplete ICA, it has the following model:

$$\mathbf{x}_0 = \mathbf{A}_0\mathbf{s} =$$

$$\begin{bmatrix} a_{11} & a_{12} & ... & a_{1m} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2m} & ... & a_{2n} \\ ... \\ a_{m1} & a_{m2} & ... & a_{mm} & ... & a_{mn} \end{bmatrix} (s_1, s_2, ...s_n)^T \tag{5}$$

where it assumes that the number of the independent components is larger than the number of observed variables($m<n$). For the overcomplete ICA model in (5), we denote the m observation components vector to be $\mathbf{x}_0$, because if we can find a way to estimate the another n-m components vector $\mathbf{x}_m$ of the observations, then it will be expanded to standard ICA, which has the following structure:

$$\begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_m \end{bmatrix} = \begin{bmatrix} \mathbf{A}_0 \\ \mathbf{A}_m \end{bmatrix} \mathbf{s} =$$

$$\begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ ... \\ a_{m1} & a_{m2} & ... & a_{mn} \\ a_{m+11} & a_{m+22} & ... & a_{m+2n} \\ ... \\ a_{n1} & a_{n2} & ... & a_{nn} \end{bmatrix} (s_1, s_2, ...s_n)^T. \tag{6}$$

To obtain the overcomplete ICA in (6), we introduce an algorithm to estimate the missing observations, which is the square Euclidean distance expressed by the Frobenius norm [17]:

$$D_F(\mathbf{A}, \mathbf{S}) = \frac{1}{2}||\mathbf{X} - \mathbf{A}\mathbf{S}||_F^2 = \frac{1}{2}\sum_{i=1}^{m}\sum_{k=1}^{N} |x_{ik} - [\mathbf{A}\mathbf{S}]_{ik}|^2$$

$$s.\ t. \quad a_{ij} \geq 0,\ x_j(k) = x_{jk} \geq 0,\ \forall\ i, j, k. \tag{7}$$

where $\mathbf{X}$ and $\mathbf{S}$ are all the observations and their corresponding original sources. From this loss function, Lee and Seung proposed the following multiplicative algorithm[18]:

$$a_{ij} \leftarrow a_{ij}\frac{[\mathbf{X}\mathbf{S}^T]_{ij}}{[\mathbf{A}\mathbf{S}\mathbf{S}^T]_{ij}}, \quad s_{jk} \leftarrow s_{jk}\frac{[\mathbf{A}^T\mathbf{S}]_{jk}}{[\mathbf{A}^T\mathbf{A}\mathbf{S}]_{jk}} \tag{8}$$

The algorithm in (7) and (8) is an extension of the image space reconstruction algorithm which is simply called ISRA [19]. From the update results of (8), we have a reconstruction of the observation $\mathbf{X}_0$: $\mathbf{X}_0 = \mathbf{A}_0\mathbf{S}_0$.

Now, using the above result, we can indicate initial components in vector $\hat{\mathbf{s}} = (s_1, s_2, ..., s_n)$ as follows:

$$\hat{\mathbf{s}} = \mathbf{A}_0^T \mathbf{x}_0. \tag{9}$$

which can be used for the expectation computing in (4). The rest part of the overcomplete ICA is to predict the other n-m observations by using (3) and (4). After the this, we achieved n observations for the n original sources ICA. The overcomplete ICA now is transformed to standard ICA. On the other words, using equations in (3) and (4), we can construct a standard ICA expression with $(m + l) = n$ observations and $(m + l) = n$ original sources. The standard ICA algorithms can be used now.

Denote $\mathbf{y}$ to be the estimation of original sources $\mathbf{s}$, most ICA learning algorithms are derived from heuristic considerations of a performance function. ICA and maximization of likelihood lead to the loss function as following:

$$L(\mathbf{y}, \mathbf{W}) = -log|det(\mathbf{W})| - \sum_{i=1}^{n} logp_i(y_i). \tag{10}$$

To compute the gradient of the loss function L, the total differential dL of L is derived when $\mathbf{W}$ is changed from $\mathbf{W}$ to $\mathbf{W} + d\mathbf{W}$, which is

$$dL = L(\mathbf{y}, \mathbf{W} + d\mathbf{W}) - L(\mathbf{y}, \mathbf{W}) = \sum_{i,j} \frac{\partial L}{\partial w_{ij}} dw_{ij}. \tag{11}$$

This leads to the following stochastic gradient learning algorithm for ICA [8]:

$$\mathbf{W}(t+1) = \mathbf{W}(t) + \eta(t)[\mathbf{I} - \varphi(\mathbf{y}(t))\mathbf{y}^T(t)]\mathbf{W}(t). \tag{12}$$

where $\mathbf{W}$ is the demixing matrix and the estimation result $\mathbf{y} = \mathbf{W}\mathbf{x}$ includes all the real original sources. In our algorithm, if we know the number of the noisy sources, we can include the noisy sources to be its components. The eventual $\mathbf{W}$ will be used to recover all the original sources. The equilibrium point of the update equation (12) satisfies

$$E[\mathbf{I} - \varphi[\mathbf{y}(t)]\mathbf{y}^T(t)] = 0. \tag{13}$$

B. Convergence Analysis for the Algorithm

After we obtain the missing observations for the overcomplete ICA, we can use (12) to estimate their original sources. To guarantee the local convergence of the update equation (12), the continuous time version of the algorithm is considered. Using the expected version of the learning equation, it has

$$\dot{\mathbf{W}}(t) = \eta(t)E[\mathbf{I} - \varphi[\mathbf{y}(t)]\mathbf{y}^T(t)]\mathbf{W}(t). \tag{14}$$

To obtain the equilibrium point, Amari in [8] suggested the following terms for the stability conditions:

$$\sigma_i^2 = E[y_i^2], \tag{15}$$
$$k_i = E[\dot{\varphi}_i(y_i)], \tag{16}$$
$$m_i = E[y_i^2 \dot{\varphi}_i(y_i)]. \tag{17}$$

In equation (5), As we mentioned before, the noise components can be considered as part of the source signals, and the ICA estimation of this model will obtain all the original sources. Thus the noisy ICA can be solved as overcomplete ICA. The stability conditions (15)-(17) for this model will also include the real sources and noise. In most cases, the independent sources may have different types of distributions. Applying these terms to the estimated l variables, we have the following new terms:

$$\sigma_l^2 = E[n_l^2], \tag{18}$$
$$k_l = E[\dot{\varphi}_l(n_l)], \tag{19}$$
$$m_l = E[n_l^2 \dot{\varphi}_l(n_l)]. \tag{20}$$

The stable equilibrium of the learning algorithm for the separation solution needs to add the other l sources. For each of the pairwise sources $i, j (i \neq j)$, the conditions of the stability are developed as

$$E[y_i^2 \dot{\varphi}_i(y_i)] + 1 > 0, \tag{21}$$
$$E[\dot{\varphi}_i(y_i)] > 0, \tag{22}$$
$$E[y_i^2]E[y_j^2]E[\dot{\varphi}_i(y_i)]E[\dot{\varphi}_j(y_j)] - 1 > 0, \tag{23}$$

and for each of the pairwise sources $l, m (l \neq m)$ and the pairwise sources l, i, it has

$$E[n_l^2 \dot{\varphi}_l(n_l)] + 1 > 0, \tag{24}$$
$$E[\dot{\varphi}_l(n_l)] > 0, \tag{25}$$
$$E[n_l^2]E[n_m^2]E[\dot{\varphi}_l(n_l)]E[\dot{\varphi}_m(n_m)] - 1 > 0, \tag{26}$$
$$E[n_l^2]E[y_i^2]E[\dot{\varphi}_l(n_l)]E[\dot{\varphi}_i(y_i)] - 1 > 0. \tag{27}$$

The generalized Gaussian distribution is given by

$$p(y_i) = \frac{\alpha}{2\beta\Gamma(\frac{1}{\alpha})} e^{-(|y_i|/\beta)^\alpha}, \tag{28}$$

where $0 < \alpha < 2$, and $\alpha < 2$, the model is for supper-Gaussian distribution, $\alpha = 2$ for Gaussian and $\alpha > 2$ for sub-Gaussian. The derivation of $p(y_i)$ is

$$\dot{p}(y_i) = \alpha \frac{sign(y_i)}{\beta} \left(\frac{|y_i|}{\beta}\right)^{\alpha-1} \frac{\alpha}{2\beta\Gamma(\frac{1}{\alpha})} e^{-(|y_i|/\beta)^\alpha}. \tag{29}$$

Under this assumption, the negative score function $\varphi_i(y_i)$ can be approximated as [13]:

$$\varphi(y_i) = -\frac{\dot{p}(y_i)}{p(y_i} = sign(y_i)\frac{\alpha}{\beta^\alpha}|y_i|^{\alpha-1}, \tag{30}$$

$$\dot{\varphi}(y_i) = \frac{\alpha(\alpha-1)}{\beta^\alpha}|y_i|^{\alpha-2}. \tag{31}$$

According to the computation in [13], for the generalized gaussian distribution signals, we have the following expressions for the stability conditions:

$$E[y_i^2 \dot{\varphi}_i(y_i)] + 1 = \alpha > 0, \quad (32)$$

$$E[\dot{\varphi}_i(y_i)] = \frac{\lambda_\alpha^2 \alpha(\alpha-1)\Gamma(\frac{\alpha-1}{\alpha})}{\Gamma(\frac{1}{\alpha})} > 0, \quad (33)$$

$$E[y_i^2]E[\dot{\varphi}_i(y_i)] - 1 = \frac{\alpha(\alpha-1)\Gamma(\frac{3}{\alpha})\Gamma(\frac{\alpha-1}{\alpha})}{\Gamma^2(\frac{1}{\alpha})} - 1 > 0. \quad (34)$$

Since we assume that the distributions for the variables are generalized gaussian, these conditions are suitable for all the original and noisy estimations. If all the variables are nongaussian, from (32)-(34), we obtain the stability conditions (21)-(27). The only problem is where a variable is gaussian, it has

$$E[n_l^2]E[\dot{\varphi}_l(n_l)] - 1 = \frac{\alpha(\alpha-1)\Gamma(\frac{3}{\alpha})\Gamma(\frac{\alpha-1}{\alpha})}{\Gamma^2(\frac{1}{\alpha})} - 1 = 0, \quad (35)$$

This means if there are two more gaussian signals for separation (including noisy signals), condition (34) cannot be satisfied, in other words, ICA doesn't work. In this situation we cannot guarantee the stability. However, if the l variables are gaussian, we knew that ICA just cannot separate these gaussian variables from each other, but this is not a problem for our solution. Since the real original sources we wanted are not gaussian, we can estimate all the nongaussian components correctly. The estimates for the gaussian noise will be arbitrary linear combinations of these noise components, and we will just leave them there.

C. Estimate the Original Sources

For the given observations of overcomplete ICA, assume the total number of original signals is known and the number of observed signals m is less than the number of the original signals n. Using the image space reconstruction algorithm in equation (8), the observation $\mathbf{x}_0$ signals are reconstructed. Now we can obtain the estimations of the missing observations. We have the following general steps for the proposed overcomplete ICA algorithm:

Step 1: For the given observations $\mathbf{x}_o = (x_1, x_2, ..., x_m)$, start the image space reconstruction algorithm (equation (8)) to compute mixing matrix $\mathbf{A}_0$ for the initial estimations of the n original sources $\mathbf{A}_o^T \mathbf{x}_o = \hat{\mathbf{s}} = (s_1, s_2, ..., s_n)$;

Step 2: Determine the distribution for all components in the original source vector $\mathbf{s}$ and predict the n-m observations $\mathbf{x}_m = (x_{m+1}, x_{m+2}, ..., x_n)$ using equation (4);

Step 3: Put the predictions in step 2 and the given observations together to form n observations. Choose ICA algorithm (12) to estimate the n component vector $\mathbf{s} = (s_1, s_2, ..., s_n)$ which will be the eventual estimations of the original sources.

IV. Simulations

To test the efficiency of the proposed overcomplete ICA algorithm, we employ a group of the speech signals for the simulation. All the original signals in the experiment are considered as Laplacian model in the experiment. Figure 1 is the three original speech signals generated from The ICALAB Package: for Signal Processing[16]. These three signals are mixed with a 3x3 random mixing matrix. We can choose any two of these mixed signals as the observations. Fig 2 is the two mixed signals of the three original signals. Using the image space reconstruction algorithm and the regression ICA, those two mixtures are first utilized to estimate the third observation. Then by employing the ICA algorithm in (12) with equivalent numbers of mixtures and original sources, these three mixtures are used to estimate the original signals. Fig 3 is the reordered estimations of the original signals. It is clearly, the simulation shows good quality for the algorithm. Of course, the separations critically depends on the predictions generated by the mixtures in figure 2 and the probability distributions of the original sources. Similar to other overcomplete ICA algorithms, the eventual estimations here still have some noise. For this algorithm, the limitation is, comparing to the total number of the originals, the number of the missed observations is small, and also, the distributions for all the original signals should be identical, otherwise, this algorithm may be hard to predict the missing observations for different distributions.

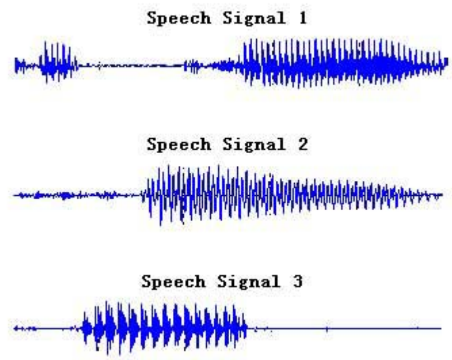

Fig. 1. The Original Speech Signals

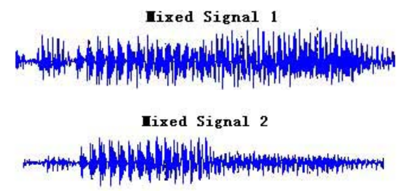

Fig. 2. The Observed Signals

V. Conclusion

In the situation of that the number of observations of mixtures is less than the original signals, by employing regression ICA algorithm and the image space reconstruction algorithm, a new overcomplete ICA algorithm can achieve very good estimations of the original sources. For the new algorithm, it particularly depends on the prediction accuracy in regression ICA. If we have previous knowledge about the distributions of the original signals, we can estimate the n-m unknown mixed

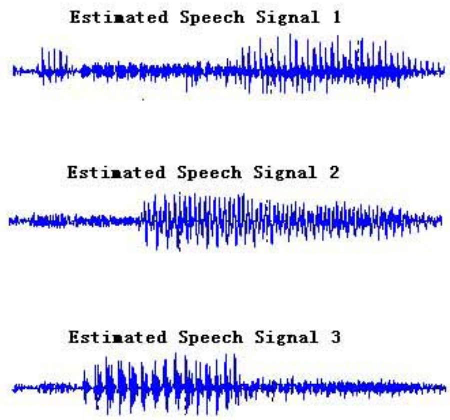

Fig. 3. The Estimated Signals

signals for the overcomplete ICA, and then the overcomplete ICA can be transformed into standard ICA. Comparing with standard ICA, most currently overcomplete ICA algorithms can not separate mixtures as good as Standard ICA did since we don't have enough observations. The advantages of this algorithm are, for some particular signals, if their distributions are all identical, the computation is simple and the estimations have very good results. One of the most important results is if the predication of $\mathbf{x}_m$ is accurate, then our new algorithm can estimate the original sources with very high quality. The simulation result in Figure 3 shows that the estimations of the original sources are not exactly the permutation of the original sources and they still have some noise in the estimated results. This indicates that currently the overcomplete ICA algorithm can not achieve the best estimations. How to make the overcomplete ICA have more efficient estimations is one of the most interesting topics which needs to discuss further.

ACKNOWLEDGMENT

This work was supported by Chinese 863 High-Tech Program Under Grant 2007AA01Z321.

REFERENCES

[1] P. Comon, Independent Component Analysis-a New Concept? Signal Process, Vol. 36, pp. 287-314, 1994.
[2] A. J. Bell and T. J. Sejnowski, An Information-maximization Approach to Blind Separation and Blind Deconvolution, Neural Comput., Vol. 7, pp. 1129-1159, 1995.
[3] P. Comon and B. Mourrain, Decomposition of Quantics in Sums of Powers of Linear Forms, Signal Process, Vol. 53, pp. 93-107, Sept. 1996.
[4] J.-F. Cardoso, Blind Signal Separation: Statistical Principles, Proc. IEEE, Vol. 86, pp. 2009-2025, Oct. 1998.
[5] T.-W. Lee, Independent Component Analysis: Theory and Applications. Boston, MA: Kluwer, 1998.
[6] Amari, S.-I., Cichocki, A., Yang, H., A New Learning Algorithm for Blind Source Separation, Advances in Neural Information Processing System, vol. 8. MIT Press, Cambridge, MA, (1996) pp. 757-763.
[7] Lewicki, M.S., Sejnowski, T. J., Learning Overcomplete Representations. Neural Comput. 12(2), (2000) 337-365.
[8] Girolami, M.: A variational Method for Learning Sparse and Overcomplete Representations. Neural Comput. 13 (11), 2001 2517-2532.
[9] Li, Y., Cichocki, A., Amari, S., Sparse Component Analysis for Blind Source Separation with Less Sensors Than Sources. Fourth Internat. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003) Japan, pp. 89-94.
[10] Te-Won Lee, Michael S. Lewicki, Mark Girolami, and Terrence J. Sejnowski, Blind Source Separation of More Sources Than Mixtures Using Overcomplete Representations. IEEE Signal Processing Letters, Vol. 6, No. 4, April 1999 87-90.
[11] Shi, Z., Tang, H., Liu, W., Tang, Y., Blind Source Separation of More Sources Than Mixtures Using Sparse Model Models. Neurocomputing 61, (2004) 461-469.
[12] A. Cichocki, S.C. Douglas, S. Amari, Robust Techniques for Independent Component Analysis (ICA) with Noisy Data. Neurocomputing 22 (1998) 113-129.
[13] Jianting Cao, Noboru Murata, Shun-ichi Amari, Andrzej Cichocki, and Tsunehiro Takeda., A Robust Approach to Independent Component Analysis of Signals With High-Level Noise Measurements. IEEE Transactions on Nueral Networks, Vol. 14, No. 3, May 2003 631-645.
[14] Heinz Mathis, Marcel Joho., Blind Signal Separation in Noisy Environments Using a Three-step Quantizer. Neurocomputing 49 (2002) 61-78.
[15] Aapo Hyvarinen, Ella Bingham., Connection Between Multilayer Perceptrons and Regression Using Independent Component Analysis. Neurocomputing 50 (2003) 211-222.
[16] Andrzej Cichocki, Shun-ichi Amari, Krzysztof Siwek, Toshihisa Tanaka, The ICALAB Package: for Signal Processing, RIKEN Brain Science Institute, Wako shi, Saitama, Japan. Version 2.2, February 2004.
[17] Cichocki, A., Zdunek, R., and Amari, S., Csiszar's Divergences for Non-Negative Matrix Factorization: Family of New Algorithms", 6th International Conference on Independent Component Analysis and Blind Signal Separation, Charleston SC, USA, Springer LNCS 3889, pp. 32-39, 2006.
[18] Lee, D.D., Seung, H.S., Learning of the parts of objects by non-negative matrix factorization. Nature 401 (1999) 788-791.
[19] Byrne, C., Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative (RBI) methods. IEEE Transactions on Image Processing 7 (1998) 100-109.