Optimization Neural Network for Blind Signal Separation Using an Adaptive Weibull Distribution

M. El-Sayed Waheed , Khaled Hashem
Dept of Mathematics, Faculty of Science, Zagazig University (Egypt)

Annotation:

In this papre We propose a neural network optimization algorithm for independent component analysis(ICA) which can separate mixtures of sub- and super- Gaussian source signals with self-adaptive nonlinearities. The ICA algorithem in the framework of natural Riemannian gradient, is derived using the parameterized Weibull density model. The nonlinear function in ICA algorithem is self-adaptive and is controlled by the shape parameter of Weibull density model. Computer simulation results confirm the validity and high performance of the proposed algorithm

Keywords: *Independent component analysis, Weibull distribution, Maximum likelihood, sub- and super- Gaussian.*

1-INTRODUCTION

The problem of independent component analysis (ICA) has received wide attention in various fields such as biomedical signal analysis and processing (EEG, MEG, ECG), geophysical . data processing, data mining, speech recognition, image recognition and wireless communications [4, 6, 17, 27]. In many applications, the sensory signals(Observations obtained from multiple sensors) are generated by a linear generative model which is unknown to us . In other words, the observations are linear instantaneous mixtures of unknown source signals and the objective is to process the observations in such a way that the outputs correspond to the separate primary source signals. The operation starts with a random source vector *U(n)* defined by $U(n) = [U_1,...,U_2,...,U_m]$ where the m components are supplied by a set of independent sources. Temporal sequences are considered here; henceforth the argument n denotes discrete time. The vector U is applied to a linear system whose input-output characterization is defined by a nonsingular m-by-m matrix *A*, called the mixing matrix. The result is an m-by-1 observation vector X(n) related to U(n) as follow X=AU where $\mathrm{X} = [\mathrm{X}_1, \mathrm{X}_2,..., \mathrm{X}_\mathrm{m}]^T$. The source vector *U* and the mixing matrix *A* are both unknown. The only thing available to us is the observation vector *X*. Given *X*, the problem is to find a demixing matrix *W* such that the original source vector *U* can be recovered from the output vector *Y* defined by *Y=WX* where $Y = [Y_1, Y_2,..., Y_m]^T$. This is called the blind source separation. The solution to the blind source separation is feasible, except for an arbitrary scalling of each signal component and permutation of indices. In other words, it is possible to find a demixing matrix *W* whose individual rows are a rescalling and permutation of those of the matrix *A*. that is, the solution may be expressed in the form $Y=WX=WAU \to DPU$ where *D* is a nonsingular diagonal matrix and *P* is a permutation matrix.

Since Jutten and Herault[21] Proposed a linear feedback network with a simple unsupervised learning algorithem, several methods have been developed .

Cichocki el al. [13;14] proposed robust, flexible algorithm with equivariant properties. Comon [15] gave a good insight to ICA problem from the statistical point of view. Bell and Sejnowski[7] adopted an information maximization principle to find a solution to ICA problem. Maximum likelihood estimation[1;6;25] was proposed by Pham et al. an was elaborated in [23;26]. The nonlinear extension of PCA was extensively studied in [21;24]. Serial updating rule was introduced by Cardoso and Laheld[8;27] and the resulting algorithm was shown to have equivariant performance. Independent, natural gradient was proposed and applied to ICA by Amari et al.

[5;17;19]. Conditions on cross cumulants for the separation of the source signals were investigated in [1;2;3;4;23;10;9].

2-DERIVATION OF NATURAL GRADIENT BASIC LEARNING RULES

Let us consider a linear feedforward memoryless neural network wich maps the observation *X(n)* to *Y(n)*

$$Y(n)=WX(n) \tag{1}$$

Where (i,j)th element of the matrix *W*, i.e., w_{ij} represents a synaptic weight between $y_i(n)$ *and* $x_j(n)$. In the limit of zero noise, for the square ICA problem (equal number of sources and sensors, the result can be extended to the case m>n) maximum likelihood or mutual information minimization suggest the following loss function [12]:

$$L(W) = -\log\left|\det W\right| - \sum_{i=1}^{m} \log p_i(y_i(n)), \tag{2}$$

where $p_i(.)$ represent the probability density function. Let us define

$$f_i(y_i(n)) = -\frac{d \log p_i(y_i(n))}{dy_i(n)} \tag{3}$$

With this definition, the gradient of the loss function (2) is

$$\begin{aligned} \nabla L(W(n)) &= \frac{\partial L(W(n))}{\partial W(n)} \\ &= -W^{-T}(n) + \mathbf{f}(y(n))X^T(n), \end{aligned} \tag{4}$$

where **f**(*y(n)*) is the element-wise function whose ith component is $f_i(y_i(n))$.

The natural Riemannian gradient (denoted by $\tilde{\nabla} L(W(n))$)learning algorithm for *W(n)* is given by [13;8;2]

$$\begin{aligned} W(n+1) &= W(n) - \eta_t \tilde{\nabla} L(W(n)) \\ &= W(n) - \eta_t \frac{\partial L(W(n))}{\partial W(n)} W^T(n)W(n) \\ &= W(n) + \eta_t [I - \mathbf{f}(y(n)]y^T(n) \end{aligned} \tag{5}$$

3-WEIBULL MODEL FOR SOURCES

Optimal nonlinear activation function $f_i(y_i(n))$ is calculated by (3). However, it required the knowledge of the probability distribution of source signals which are not available to us. A variety of hypothesized density model has been used. For example, for the supper-Gaussian source signals, unimodal or hyperbolic-Cauchy distribution model [7] leads to the nonlinear function given by

$$f_i(y_i(n)) = \tanh(\beta\, y_i(n)). \tag{6}$$

Such sigmodal function was also used in [7]. For sub-Gaussian sorce signals, cubic nonlinear function $f_i(y_i(n)) = y_i^3(n)$ has been a favorite choice. For Mixtures of Sub- and super-Gaussian source signals, according to the estimated kurtosis of the expected signals, nonlinear function can be elected from two different choices [15;16]. (for example, either $f_i(y_i(n)) = y_i^3(n)$ or $f_i(y_i(n)) = \tanh(\beta\, y_i(n))$.). Several approaches [18;10;11] are already available.

This paper present a flexible nonlinear function derived using Weibull density model. It will be shown that the nonlinear function is self-adaptive and controled by Weibull shape parameter. It is not a form of fixed nonlinear function.

3.1. THE WEIBULL DISTRIBUTION

The weibull probability distribution is a set of distributions parameterized by a positive real number c which is usually referred to as the shape parameter of the distribution. The shape parameter c controls the peakiness of the distribution. The probability density function (PDF) for Weibull is described by

$$p(y;c,\zeta_0,\alpha)=\frac{c}{\alpha}\left(\frac{y-\zeta_0}{\alpha}\right)^{c-1}\exp\left(-\left(\frac{y-\zeta_0}{\alpha}\right)^{c}\right)\qquad y\succ\zeta_0,\ c(\succ 0),\alpha\ (\succ 0) \qquad (7)$$

It is necessary c be greater than –1, for otherwise the integral of $p(y;c,\zeta_0,\alpha)$ between $y=\theta$ *and* $y=\hat{\theta}\succ\theta$ will be infinite. The standard form of the distribution will have $\zeta_0=0$ and $\alpha=1$ so that the standard density function is

$$p(y;c,\zeta_0,\alpha)=\frac{c}{\alpha}y^{c-1}\exp\left(-y^{c}\right)\qquad y\succ 0,\ c(\succ 0) \qquad (8)$$

The distribution of y now depend on the shape parameter c alone. The plots of the standard density function in (8) for c=0.25, 0.5, 1, 1.5, 2, 3, 4, 5 are presented in figure 1 and 2 respectively.

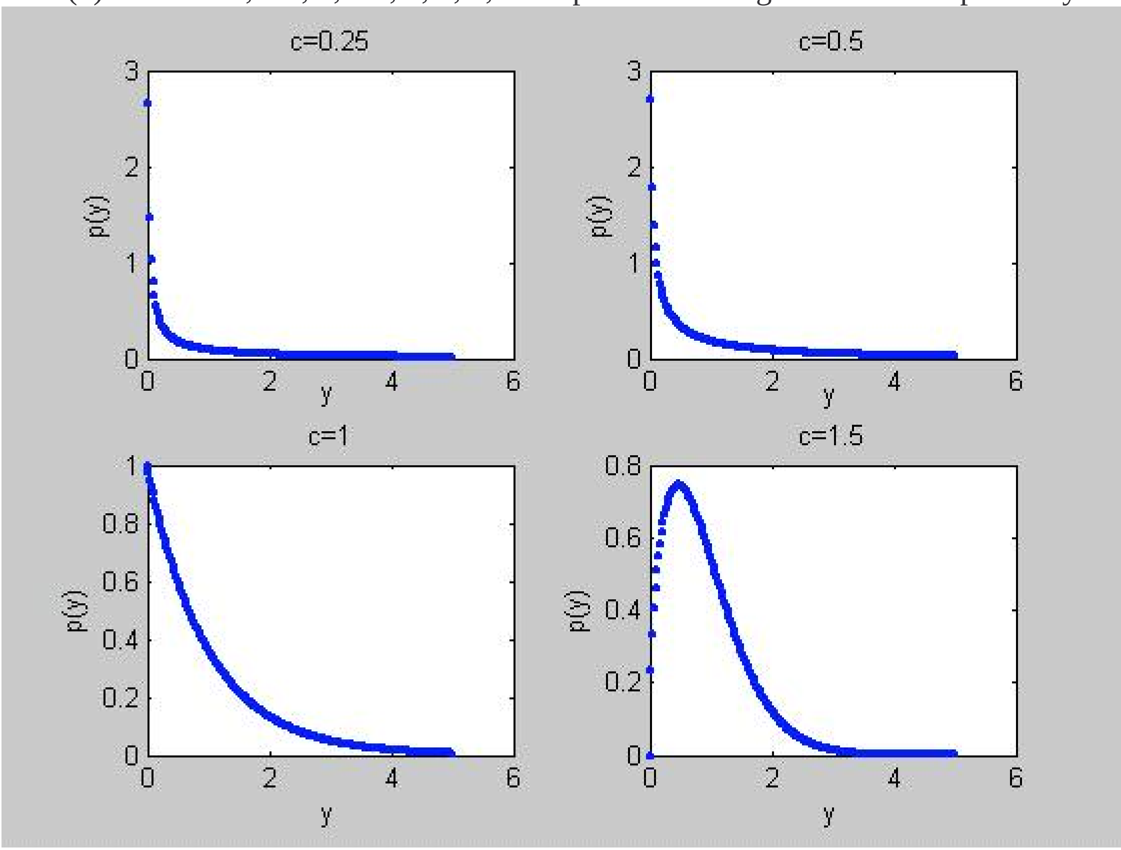

Fig.1:The plots of the standard Weibull density function for c=0.25, 0.5, 1, 1.5

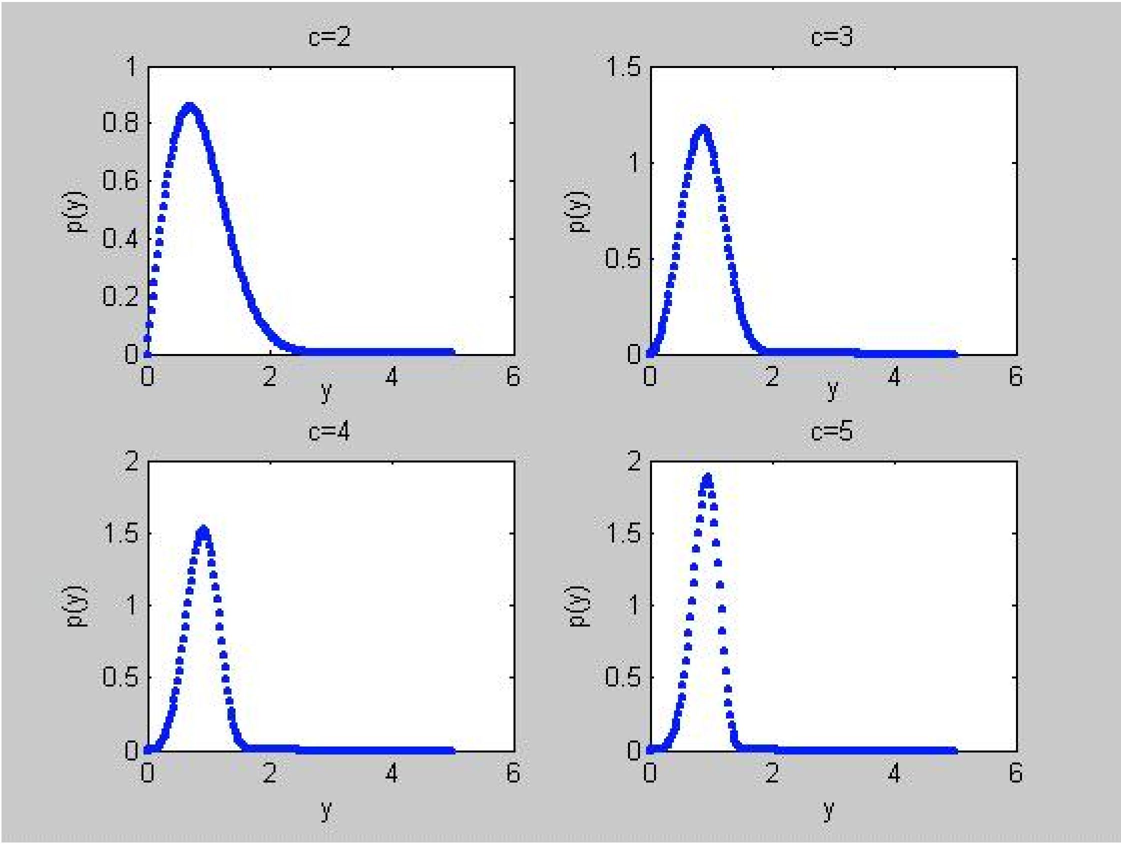

Fig2.: The plots of the standard Weibull density function for c=2, 3, 4, 5

3.2 The Moments Of Weibull Distribution

In order to fully understand the Weibull distribution, it is usefull to look at its moments (specially 2nd and 4th moments which give the kurtosis). The nth moment of Weibull distribution is given by

$$E(y^r) = \int_0^\infty y^r q(y;\, c)\, dy = \Gamma\left(\frac{r}{c}+1\right) \qquad (9)$$

then

$$\mu_2' = \Gamma\left(\frac{2}{c}+1\right)$$
$$\mu'4 = \Gamma\left(\frac{4}{c}+1\right) \qquad (10)$$
$$\mu_{2k}' = \Gamma\left(\frac{2k}{c}+1\right)$$

the moment ratios, coefficient of variation, and standard cumulants $\frac{k_r}{k_2^{r/2}}$ of the standard distribution in (8) are of course the same as those of the distribution in (7).

3.3 Kurtosis And Shape Parameter

The kurtosis is anondimensional quantity. It is measures the relative peakdness or faltness of a distribution. A distribution with positive kurtosis is termed leptokurtic(super-Gaussian). A distribution with negative kurtosis is termed platykurtic(sub-Gaussian). The kurtosis of the distribution is defined in termsof the 2nd-and 4th –order moment as

$$\mathrm{k(y)} = \frac{\mu'_4 - 4\mu'_1\mu'_3 + 6\mu'^2_1\mu'_2 - 3\mu'^4_1}{\left[\mu'_2 - \mu'^2_1\right]^2} - 3 \qquad (11)$$

where the constant term –3 makes the value zero for the standard normal distribution. For Weibull distribution, the kurtosis can be expressed in terms of the shape parameter, given by

$$\mathrm{k(c)} = \frac{\Gamma\left(\frac{4}{c}+1\right) - 4\Gamma\left(\frac{1}{c}+1\right)\Gamma\left(\frac{3}{c}+1\right) + 6\left[\Gamma\left(\frac{1}{c}+1\right)\right]^2\Gamma\left(\frac{2}{c}+1\right) - 3\left[\Gamma\left(\frac{1}{c}+1\right)\right]^4}{\left(\Gamma\left(\frac{2}{c}+1\right) - \left[\Gamma\left(\frac{1}{c}+1\right)\right]^2\right)^2} - 3 \quad (12)$$

The plot of kurtosis $k(c)$ versus the shape parameter c for leptokurtic and platykurtic signals are shown in figure 3.

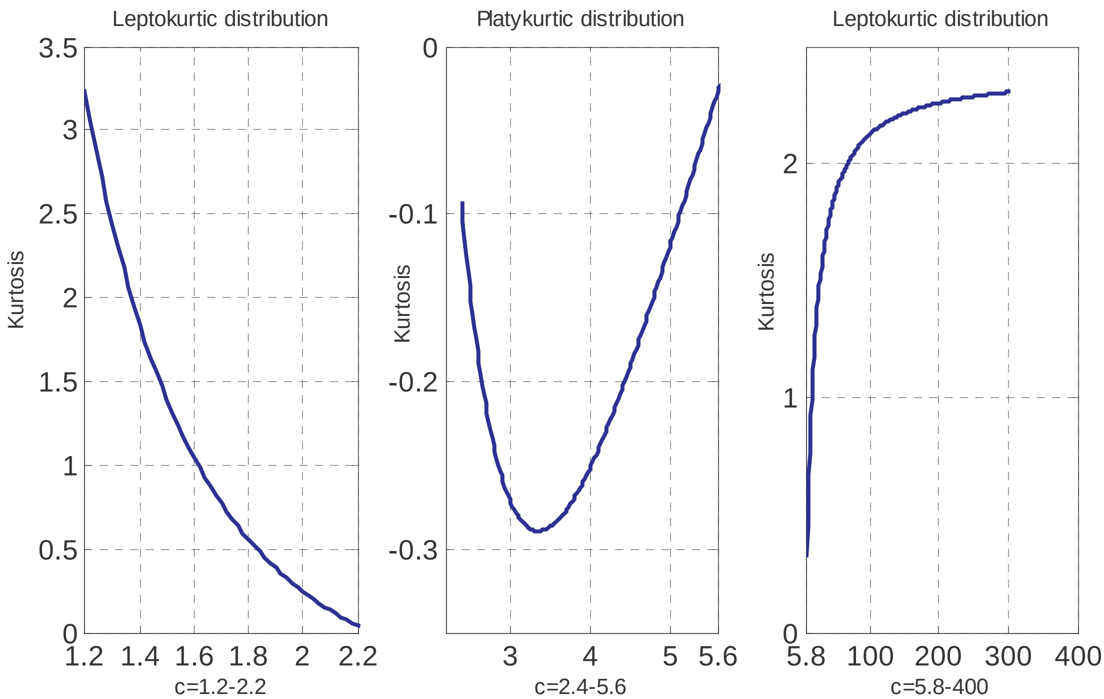

Fig 3:The plot of kurtosis $k(c)$ versus the shape parameter c for leptokurtic and platykurtic signals.

The activation function for Weibull distribution in (8) is given by

$$.\,f_i(y_i) = \frac{c - 1 - cy^c}{y} \qquad (13)$$

$$\frac{\partial L}{\partial c_i} = \frac{\partial \ln p(y_i; c_i)}{\partial c_i} = \frac{1}{c_i} + \ln(y_i) - y_i^{c_i}\ln(y) \qquad (14)$$

$$\Delta c_i = -\eta_i \frac{\partial L}{\partial c_i} = -\eta_i\left(\frac{1}{c_i} + \ln(y_i) - y_i^{c_i}\ln(y)\right) \qquad (15)$$

4. COMPUTER SIMULATION RESULTS

Consider the system involving the following three independent sources

$$U_1(n) = 0.1\sin(400n)\cos(30n)$$

$$U_2(n) = 0.01\,\mathrm{sgn}(\sin(500n + 9\cos(40n))) \qquad (16)$$

$$U_3(n)\ \textit{noise uniformly distributed in the range}\ [-1,1]$$

The mixing matrix **A** is

$$\mathbf{A} = \begin{bmatrix} 0.56 & 0.79 & -0.37 \\ -0.75 & 0.65 & 0.86 \\ 0.17 & 0.32 & -0.48 \end{bmatrix} \tag{17}$$

The algorithm was implemented using the following conditions:

- Initialization. The weights in the demixing matrix **W** were picked from a random number generator with a uniform distribution inside the range [0.0,0.5].
- The learning rate parameter was fixed at $\eta = 0.1$
- Signal duration. The time series produced at the mixer output had a sampling period 10^{-4}s and contained N =65,000 samples.

Figure (4) displays the waveforms of the source signals and the signals produced at the output of the demixer. It can be observed that after 3000 iterations, source signals are well separated.

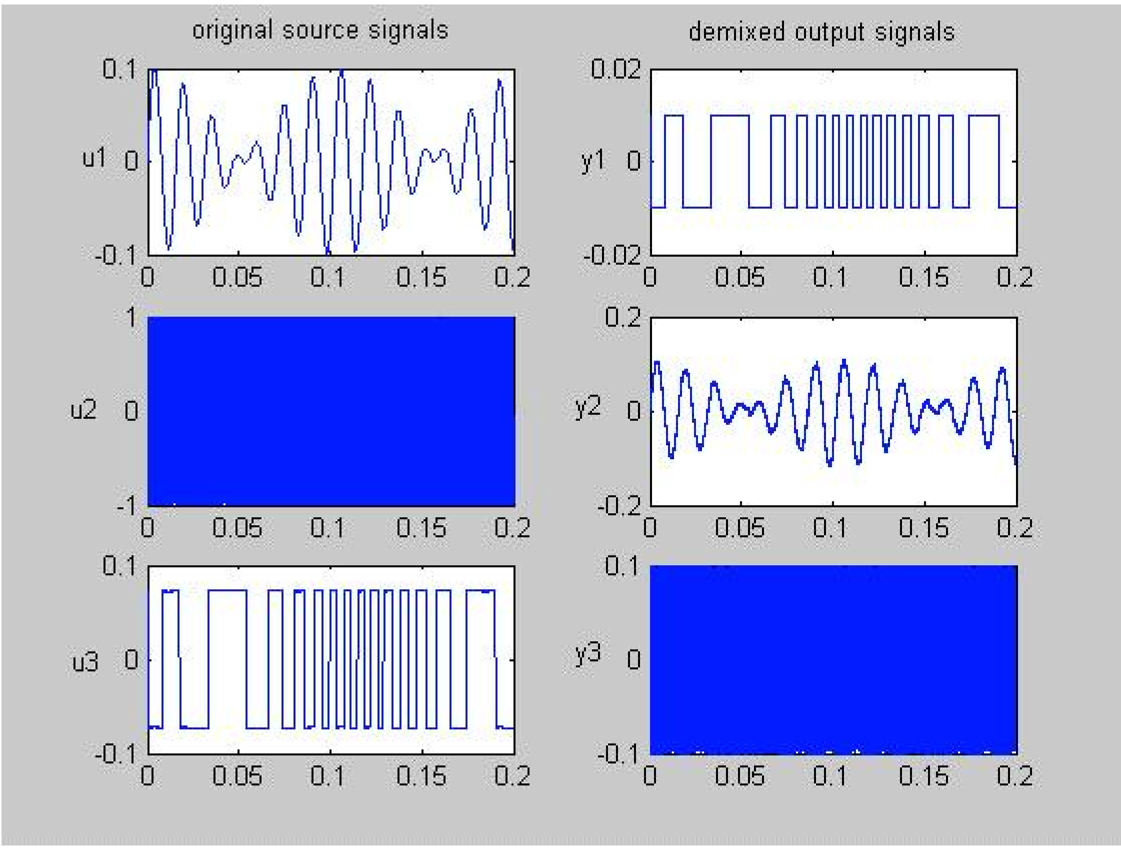

Fig.4: the original and demixed signals

5. CONCLUSION

We have proposed ICA algorithm (in the framework of natural Riemannian gradient) where the self-adaptive nonlinear function eas derived using Weibull density model for the probability distributions of the source signals. We have shown that the proposed ICA algorithm can separate the mixtures of sub-and super-gaussian signals with self adaptive nonlinearities which is controlled by Weibull shape parameter.

REFERENCES:

1. F. Alberg, P. Duhamel, M. Nikolova, Adaptive solution for blind identification/equalization using deterministic maximum likelihood, IEEE Trans. Signal Process. 50 (4) (April 2002) 923–936.
2. S. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, Vol. 10, pp. 251-276,1998.
3. S. Amari, "Natural Gradient for over- and under-Complete bases in ICA," 1998 *Neural Computation*, vol 11, pp. 355-367. .
4. S. Amari and A. Cichocki. "Adaptive blind signal processing –neural network approaches." *Proceedings IEEE,* 86:1186-1187,1998.
5. S. Amari, A. Cichocki, and H. H. Yang , "A new learning algorithem for blind signal separation," in *Advance in Neural Information Processing Systems*, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, Eds. 1996, vol. 8, pp. 757-763, MIT press.
6. S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang."novel on-line adaptive learning algorithms for blind deconvolution using the natural gradient approach." In *proc. 11th IFAC Symposium on system Identification*, Vol(3) ,PP. 1057-1062, Kitakyushu City, Japan,july 1997.
7. A. Bell and T. Sejnowski, "An information maximisation approach to blind separation and blind deconvolution," Neural Computation, vol. 7, pp. 1129-1159, 1995.
8. J. F. Cardoso and B. H. Laheld, "Equivarent adaptive source separation," *IEEE Trans. Signal Processing*, Vol. 44, pp. 3017-3030, 1996.
9. S. Choi and A. Cichocki, "A linear feedforward neural network with lateral feedback connections for blind source separation ," in *IEEE signal processing Workshop on Higher-Order Statistics*, Banff, Canada, 1997, pp. 349-353.
10. S. Choi, R. Liu, And A. Cichocki, "A spurious equilibris-free learning algorithm for blind separation of non-zero skewness signals," *Neural Processing Letters*, vol. 7, pp. 61-68, 1998.
11. A. Cichocki, I. Sabala, S. Choi, B. Orsier,and R. Szupiluk,"self-adaptive independent component analysis for sub-gaussian and super-gaussian mixtures with unknown number of source signals," In *International Symposium on Nonlinaer Theory and Applications, 1997*, pp. 731-734.
12. A. Cichocki, R. Unbehauen, L. Moszczynski, and E. Rummert, "A new on-line adaptive learning algorithm for blind separation of source signals," in International Joint Conference on Neural Networks. 1994, pp. 406-411.
13. A. Cichocki and R. Unbehauen, "Robust neural networks with on-line learning for blind identification and blind separation of sources," IEEE Trans. Circuits and Systems- I: Fundamental Theory and Applications, Vol. 43, pp. 894-906, 1996.
14. P.Comon, "Independent component analysis, a new concept ?," Signal Processing, vol. 36, pp. 287-314, 1994.
15. R. Diversi, R. Guidorzi and U. Soverini, "Blind identification and equalization of two-channel FIR systems in unbalanced noise environments" *Signal Processing* , Vol 85 ,PP.215–225, 2005.
16. S. C. Douglas, A. Cichocki, and S. Amari, "Multichannel blind separation and deconvolution of sources with arbitrary distributions," in *Neural Networks for signal processing 7*, J. Principe, L. Gile. N. Morgan, and E. Wilson, Eds., 1997, pp. 436-445.
17. W. A. Gardner. "A new method of channel identification." *IEEE Trans. Communications,* 39(6):813-817, June 1991.
18. M. Girolami and C. Fyfe, "Generalized Independent component analysis through unsupervised learning with emergent bussgang properties," in *IEEE international Conference on Neural Networks*, 1997, pp. 1788-1791.
19. I. S. Gradshteyn, I. M. Ryzhik, and A. Jeffrey, *Table of Integrals, Series, and Products*, Academic Press, 1994.
20. C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture, " *Signal Processing*, vol. 24, pp. 1-10, 1991.
21. j. Karhunen, "Neural approaches to independent component analysis," in *European symposium on Artificial Neural Networks*, 1996, pp. 249-266.
22. D. J. C. MacKay, "maximum likelihood and covariant algorithms for independent component analysis," 1996, University of cambridge, Cavendish Laboratory, Draft 3.7.
23. J. P. Nadal and N. Parga "Redundancy reduction and independent component analysis: Conditions on cumulants and adaptive approaches," *Neural Computation*, vol. 9, pp. 1421-1456, 1997.
24. E. Oja, "The nonlinear PCA learning rule and signal separation- mathematical analysis," 1995.
25. D. T. Pham, P. Garrat, C. Jutten, "Separation of mixtures of independent sources through a maximum likelihood approach," in *European signal processing Conference*, 1992, pp. 771-774.
26. B. Pearlmuter and L. Parra, "Maximum likelihood blind source separation: A context-sensitive generalization of ICA," in. *advances in Neural information processing System,* M. C. Mozer, M.I. Jordan, and T. Petsche, Eds., 1997, pp.613-619.
27. D. T. M. Slock. "Spatio-temporal training-sequence-based channel equalization and adaptive interference cancellation." In *Proc. ICASSP 96 Conference*, Atlanta, Georgia, May 1996.

Article received: 2006-11-01