International Journal of Physical Sciences Vol. 4 (5), pp. 265-270, May, 2009
Available online at http://www.academicjournals.org/IJPS
ISSN 1992 - 1950 © 2009 Academic Journals

Full Length Research Paper

Blind signal separation using an adaptive weibull distribution

M. El-Sayed Waheed

Department of Mathematics, Faculty of Science, Zagazig University Egypt. E-mail: mewahed@yahoo.com

Accepted 01 May, 2009

We propose an independent component analysis (ICA) algorithm 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

Key words: Independent component analysis, weibull distribution, maximum likelihood, sub- and super-Gaussian, blind signal separation.

INTRODUCTION

The problem of independent component analysis (ICA) has received wide attention in various fields such as biomedical signal analysis and processing (EEG, MEG and ECG), geophysical data processing, data mining, speech recognition, image recognition and wireless communications (Amari and Cichocki, 1998; Amari et al., 1997; Gardner, 1991; El-sayed Wahed and Mohamed, 2006). 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 *S* defined by $S(n) = [S_1, S_2,\ldots,S_m]$ 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 S 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 S (n) as follow X=AS where $X = [X_1, X_2,\ldots,X_m]^T$. The source vector *S* 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 *S* can be recovered from the output vector *Y* defined by *Y=WX* where $Y = [Y_1, Y_2,\ldots,Y_m]^T$. 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 rescaling and permutation of those of the matrix *A*, that is, the solution may be expressed in the form $Y=WX=WAS \rightarrow DPS$ where *D* is a nonsingular diagonal matrix and *P* is a permutation matrix.

Since Jutten and Herault (Karhunen, 1996) proposed a linear feedback network with a simple unsupervised learning algorithem, several methods have been developed.

Cichocki and Unbehauen (1996) and Comon (1994) proposed robust, flexible algorithm with equivariant properties. Comon (Diversi et al., 2005) gave a good insight to ICA problem from the statistical point of view. Bell and Sejnowski (Bell and Sejnowski, 1995) adopted an information maximization principle to find a solution to ICA problem. Maximum likelihood estimation (Alberg et al., 2002; Amari et al., 1997; El-sayed Wahed, 2007) was proposed by Pham et al. and was elaborated in (Abu-Taleb et al., 2006; El-sayed Wahed, 2007). The nonlinear extension of PCA was extensively studied in (Karhunen, 1996; El-sayed Wahed and Mohamed, 2006). Serial updating rule was introduced by Cardoso and Laheld (Cardoso and Laheld, 1996) and the resulting algorithm was shown to have equivariant performance. Independent, natural gradient was proposed and applied

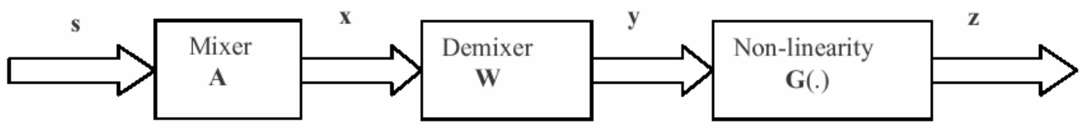

Figure 1. Maximum Entropy Method.

to ICA by (Amari et al., 1996; Gardner, 1991; Gradshteyn et al., 1994). Conditions on cross cumulants for the separation of the source signals were investigated in (Alberg et al., 2002; Amari, 1998; Amari and Cichocki, 1998; Abu-Taleb et al., 2006; Choi et al., 1998; Choi and Cichocki, 1997).

Maximum entropy algorithm

This is an adaptive algorithm based on information theoretic approach and was suggested by Bell and Sejnowski (Bell and Sejnowski, 1995). The block diagram in Figure 1 explains the maximum entropy method for blind source separation.

The demixer operates on the observed data X to produce an output Y = WX, which is an estimate of source S. The output Y is transformed into Z by passing it through a non-linearity G (.), which is invertible and monotonic. For a given non-linearity G (.), the maximum entropy method produces an estimate of source S by maximizing the entropy h (Z) with respect to W. The mathematical representation of the whole process may be given as follows:

$$Z = G(y) = G\,(WAS) \qquad \Rightarrow S = A^{-1}W^{-1}G^{-1}(z) = \psi(z)$$

where G^{-1} is the inverse non-linearity.

The probability density function of the output Z is defined in terms of that of the source S by

$$f[Z(z)] = \frac{f[S(s)]}{\left|\det(J(s))\right|}\Bigg|_{s = \psi(z)}$$

Where det (J(s)) is the determinant of the Jacobian matrix J(s). The ij-th element of the matrix J(s) is defined by $J_{ij} = \frac{\partial z_i}{\partial s_j}$. Hence, the entropy of the output Z at the output of the non-linearity G (.) is

$$h\,(Z) = -E[\log f_z(z)] = -E\left[\log\left(\frac{f[S(s)]}{\left|\det(J(s))\right|}\Bigg|_{s = \psi(z)}\right)\right] = -D_{fs}\left|\det(j)\right| \text{ evaluated } S = \psi(z).$$

Hence, maximizing the entropy h (Z) is equivalent to minimizing the Kullback-Leibler divergence between $f_s(s)$ and a probability density function of S, defined by $\left|\det[J(s)]\right|$.

If the random variable z_i (ith element of z) is uniformly distributed inside the interval [0, 1] for all i, then the entropy h (z) is equal to zero. Accordingly,

$$h\,(Z) = -E[\log f_z(z)] = -E\left[\log\left(\frac{f[S(s)]}{\left|\det(J(s))\right|}\Bigg|_{s = \psi(z)}\right)\right] \Rightarrow f_s(s) = \left|\det(J(S))\right|$$

Under the ideal condition, $W = A^{-1}$, the above relationship reduces to

$$f_{S_i}(S_i) = \frac{\partial z_i}{\partial y_i}\Bigg|_{z_i = g(s_i)} \quad \text{for all i.}$$

Conversely, the results from Maximum Entropy Method may be stated as follows:

Let the non-linearity at the demixer output be defined in terms of the original source distribution as

$$z_i = g_i(y_i) = \int_{-\infty}^{z_i} fS_i(s_i)\,ds_i, \quad \text{for all i = 1,2,…,n.}$$

Then, maximizing the entropy of the random vector z at the output of the non-linearity G is equivalent to W = A^{-1}, which yields perfect blind source separation. The maximum entropy and maximum likelihood methods for blind source separation are equivalent under the condition that the random variable i z is uniformly distributed inside the interval [0, 1] for all i. This relationship may be proven with the help of chain rule of calculus as

$$J_{ij} = \sum_{k=1}^{n} \frac{\partial z_i}{\partial z_i}\frac{\partial y_i}{\partial x_i}\frac{\partial x_i}{\partial s_i} = \sum_{k=1}^{n} \frac{\partial z_i}{\partial z_i} w_{ik} a_{kj} \quad J$$

The Jacobian matrix J is expressed as J = DWA, where D is a diagonal matrix given by

$$D = \mathrm{diag}\left(\frac{\partial z_1}{\partial y_1}, \frac{\partial z_2}{\partial y_2}, ..., \frac{\partial z_n}{\partial y_n}\right).$$

$$\left|\det(J) = \left|\det(WA)\right|\right| \prod_{i=1}^{n} \frac{\partial g_i(y_i)}{\partial y_i}$$

Hence,

In the light of the above equation, an estimate of the probability density function $f_s(s)$ parameterized by the weight matrix W and the non-linearity G may be written formally as

$$f_s(s/W,G) = \left|\det(WA)\right| \prod_{i=1}^{n} \frac{\partial g_i(y_i)}{\partial y_i} \qquad (1)$$

Therefore, under the above condition, maximizing the log-likelihood function {logf$_s$ (s/W, G)} {}) is equivalent to maximizing the entropy h (Z) for blind source separation. Referring to the expression h (Z) = -E[log $f_z(z)$] = -E $\left[\log\left(\frac{f[S(s)]}{\left|\det(J(s))\right|}\middle| s = \psi(z)\right)\right]$, it is seen that since the source distribution is fixed, maximizing the entropy h(Z) requires maximizing the expectation of the denominator term {log | det(J(s) | } with respect to the separating matrix W.

To do the computation using an adaptive algorithm that will maximize the objective function, the instantaneous objective function φ may be considered as:

$$\varphi = \log\left|\det(J)\right| \qquad (2)$$

On expanding (2), we get:

$$\varphi = \log\left|\det(A) + \log\left|\det(W)\right|\right| + \sum_{i=1}^{n} \log\left(\frac{\partial z_i}{\partial y_i}\right)$$

and

$$\frac{\partial \varphi}{\partial W} = W^{-T} + \sum_{i=1}^{n} \frac{\partial}{\partial W} \log\left(\frac{\partial z_i}{\partial y_i}\right) \qquad (3)$$

The non-linear function should be judiciously selected to deal with the super-Gaussian, sub-Gaussian, stationary and non-stationary signals. The popular non-linearity used is logistic function and hyperbolic tangent function:

$$z_i = g(y_i) = \frac{1}{1+e^{-y_i}},$$

$$z_i = g(y_i) = \tanh(y_i), \qquad i = 1,2,...,n$$

The non-linear functions should be monotonic and invertible.

Finding out $\frac{\partial \varphi}{\partial W}$ using the above non-linearity, we obtain

$$\frac{\partial \varphi}{\partial W} = W^{-T} + (1-2z)x^T$$

where x is the observed source vector, z is the non-linearly transformed output vector and 1 is a corresponding vector of ones.

Using the steepest ascent method to maximize the entropy h(Z), the change in weight matrix W is given by

$$\Delta W = \eta \frac{\partial \varphi}{\partial W} = \eta\left(W^{-T} + (1-2z)x^T\right)$$

where η is the learning rate parameter. The generalized final version for the update on W or the learning rule is obtained by using the natural gradient, which is equivalent to multiplying the expression for ΔW by W^TW instead of evaluating W^{-T} as given below:

$$\Delta W == \eta\left(W^{-T} + (1-2z)x^T\right)W^TW = \eta\left(I + (1-2z)(Wx)^T\right)W = \eta\left(I + (1-2z)y^T\right)W$$

$$W(k+1) = W(k) + \eta\left(I + (1-2z(k))y^T(k)\right)W(k) \qquad (4)$$

Where y is the output of the demixer before passing through the non-linearity, I is the unity matrix and is a fixed learning rate parameter with value less than 1.

The algorithm gives better result when applied on pre-whitened data. It is sensitive to the learning rate parameter and works better for super-Gaussian signals.

WEIBULL MODEL FOR SOURCES

Optimal nonlinear activation function $f_s(s)$ is calculated by (1). 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 (Bell and Sejnowski, 1995) leads to the nonlinear function given by

$$f_s(s) = \tanh(\beta f_s(s)). \qquad (5)$$

Such sigmodal function was also used in (Bell and Sejnowski, 1995). For sub-Gaussian source signals, cubic nonlinear function $f_s(s)$ - f_s^3 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 (Diversi et al., 2005; Douglas et al., 1997) [For example, either $f_s(s)$ - f_s^3 or $f_s(s) = \tanh(\beta f_s(s))$]. several approaches (Girolami and Fyfe, 1997; Choi et al., 1998; Cichocki et al., 1997) are already available.

This paper present a flexible nonlinear function derived using Weibull density model. It will be shown that the

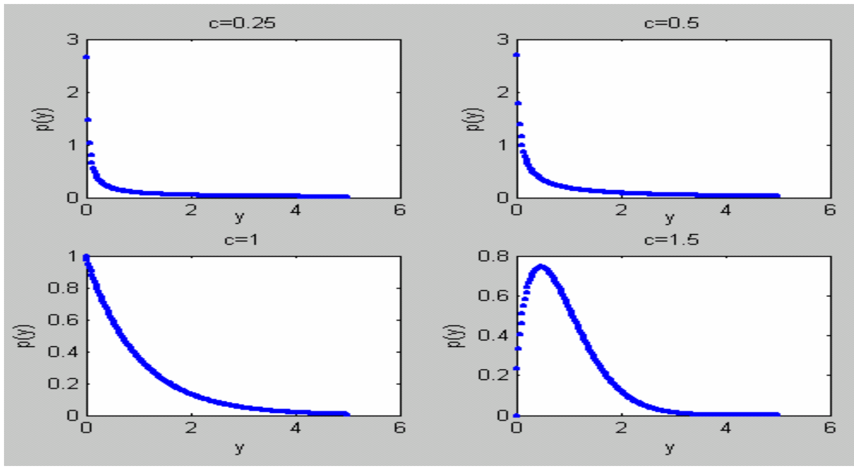

Figure 2. The plots of the standard Weibull density function for c = 0.25, 0.5, 1, 1.5.

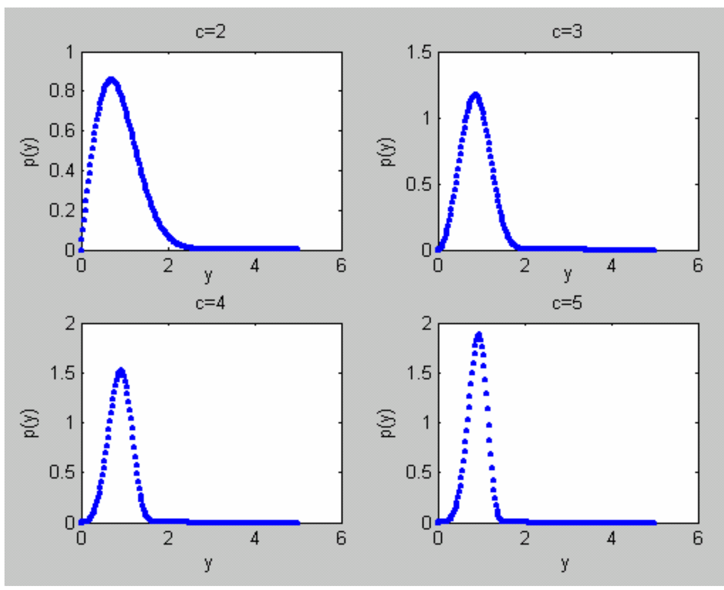

Figure 3. The plots of the standard Weibull density function for c = 2, 3, 4, 5.

nonlinear function is self-adaptive and controled by Weibull shape parameter. It is not a form of fixed nonlinear function.

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

$$p(y;c,\zeta_0,\alpha)=\frac{c}{\alpha}\left(\frac{y-\zeta_0}{\alpha}\right)^{c-1}\exp\left(-\left(\frac{y-\zeta_0}{\alpha}\right)^{c}\right) \quad y\ \phi\ \zeta_0,\ c(\phi\ 0), \alpha\ (\phi\ 0) \tag{6}$$

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

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

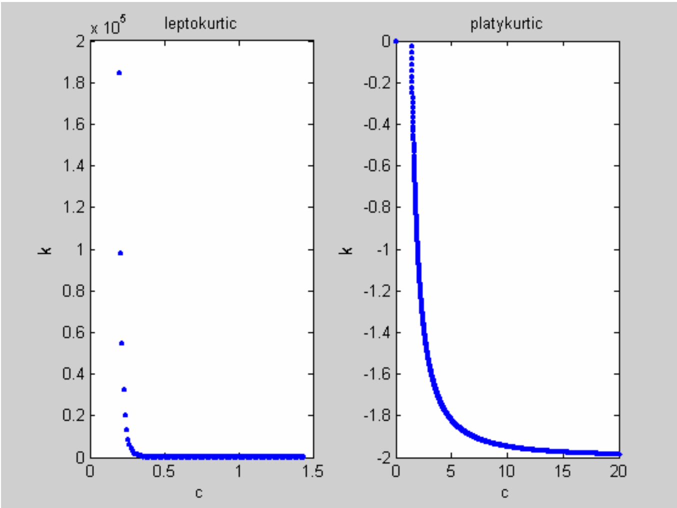

Figure 4. The plot of kurtosis $k(c)$ versus the shape parameter c for leptokurtic and platykurtic signals.

The distribution of y now depends on the shape parameter c alone. The plots of the standard density function in (7) for c = 0.25, 0.5, 1, 1.5, 2, 3, 4, 5 are presented in Figures 2 and 3 respectively.

The moments of weibull distribution

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

$$E(y^r) = \int_0^\infty y^r\, p(y;\, c)\; dy = \Gamma\left(\frac{r}{c}+1\right) \tag{8}$$

Then

$$M_2 = \Gamma\left(\frac{2}{c}+1\right)$$

$$M4 = \Gamma\left(\frac{4}{c}+1\right)$$

$$M_{2k} = \Gamma\left(\frac{2k}{c}+1\right) \tag{9}$$

The moment ratios, coefficient of variation, and standard cumulants $\frac{k_r}{k_2^{r/2}}$ of the standard distribution in (7) are of course the same as those of the distribution in (6).

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 terms of the 2nd-and 4th –order moment as

$$k(y) = \frac{M_4}{M_2{}^2} - 3 \tag{10}$$

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

$$k(c) = \frac{\Gamma\left(\frac{4}{c}+1\right)}{\left(\Gamma\left(\frac{2}{c}+1\right)\right)^2} - 3 \tag{11}$$

The plots of kurtosis $k(c)$ versus the shape parameter c for leptokurtic and platykurtic signals are shown in Figure 4. The activation function for Weibull distribution in (3) is given by

$$f_i(y_i) = \frac{c-1-cy^c}{y} \tag{12}$$

$$\frac{\partial L}{\partial c_i} = \frac{\partial \ln p(y_i;c_i)}{\partial c_i} = \frac{1}{c_i} + \ln(y_i) - y_i^{c_i}\ln(y) \tag{13}$$

$$\Delta c_i = -\eta_i \frac{\partial L}{\partial c_i} = -\eta_i\left(\frac{1}{c_i} + \ln(y_i) - y_i^{c_i}\ln(y)\right) \tag{14}$$

COMPUTER SIMULATION RESULTS

Consider the system involving the following three independent sources

$U_1(n)$ = a square wave of amplitude a, and fundamental frequency w0

$U_2(n)$ = a triangular wave of amplitude, and fundamental frequency w0

$$U_3(n) = 0.1\sin(400n)\cos(30n) \tag{15}$$

The mixing matrix A is

$$\mathbf{A} = \begin{bmatrix} 0.56 & 0.79 & -0.37 \\ -0.75 & 0.65 & 0.86 \\ 0.17 & 0.32 & -0.48 \end{bmatrix} \tag{16}$$

The algorithm was implemented using the following conditions:

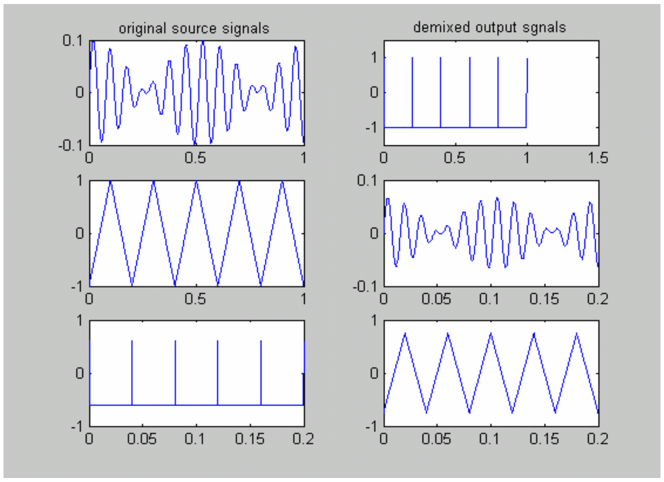

Figure 5. The original and demixed signals.

i.) 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].
ii.) The learning rate parameter was fixed at $\eta = 0.1$
iii.) Signal duration. The time series produced at the mixer output had a sampling period 10^{-4}s and contained N = 65,000 samples.

Figure (5) 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.

Conclusion

We have proposed ICA algorithem (in the framework of natural Riemannian gradient) where the self-adaptive nonlinear function easy 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. Finally we apply our algorithm on a mixture of other distributions and image separation, which give a good result.

REFERENCES

Alberg F, Duhamel P, Nikolova M (2002). Adaptive solution for blind identification/equalization using deterministic maximum likelihood, IEEE Trans. Signal Process 50 (4): 923–936.

Amari S (1998). "Natural gradient works efficiently in learning," Neural Computation. 10: 251-276.

Amari S (1998). "Natural Gradient for over- and under-Complete bases in ICA," Neural Computation 11: 355-367.

Amari S, Cichocki A (1998). "Adaptive blind signal processing –neural network approaches." Proceedings IEEE 86: 1186-1187.

Amari S, Cichocki A, Yang HH (1996). "A new learning algorithem for blind signal separation," in Advance in Neural Information Processing Systems, Touretzky DS, Mozer MC, and Hasselmo ME 8: 757-763, MIT press., Eds.

Amari S, Douglas SC, Cichocki A, Yang HH (1997)."Novel on-line adaptive learning algorithms for blind deconvolution using the natural gradient approach." In proc. 11th IFAC Symposium on system Identification (3): 1057-1062, Kitakyushu City, Japan.

Bell A, Sejnowski T (1995). "An information maximisation approach to blind separation and blind deconvolution," Neural Computation 7: 1129-1159.

Cardoso JF, Laheld BH (1996). "Equivarent adaptive source separation," IEEE Trans. Signal Processing. 44: 3017-3030.

Choi S, Cichocki A (1997). "A linear feedforward neural network with lateral feedback connections for blind source separation," in IEEE signal processing Workshop on Higher-Order Statistics, Banff, Canada pp. 349-353.

Choi S, Liu R, Cichocki A (1998). "A spurious equilibris-free learning algorithm for blind separation of non-zero skewness signals," Neural Processing Letters 7: 61-68.

Cichocki A, Sabala I, Choi S, Orsier B, Szupiluk R (1997)."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 pp. 731-734.

Cichocki A, Unbehauen R, Moszczynski L, Rummert E (1994). "A new on-line adaptive learning algorithm for blind separation of source signals," in International Joint Conference on Neural Networks pp. 406-411.

Cichocki A, Unbehauen R (1996). "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, 43: 894-906.

Comon P (1994). "Independent component analysis, a new concept ?," Signal Processing 36: 287-314.

Diversi R, Guidorzi R, Soverini U (2005). "Blind identification and equalization of two-channel FIR systems in unbalanced noise environments" Signal Processing , 85: 215–225.

Douglas SC, Cichocki A, Amari S (1997). "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. pp. 436-445.

Gardner WA (1991). "A new method of channel identification." IEEE Trans. Communications, 39(6): 813-817.

Girolami, Fyfe C (1997). "Generalized Independent component analysis through unsupervised learning with emergent bussgang properties," in IEEE international Conference on Neural Networks, pp. 1788-1791.

Gradshteyn IS, Ryzhik IM, Jeffrey A (1994). Table of Integrals, Series, and Products, Academic Press.

Jutten C, Herault J (1991). "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture, " Signal Processing, vol. 24: 1-10. Karhunen J (1996). "Neural approaches to independent component analysis," in European symposium on Artificial Neural Networks pp. 249 - 266.

MacKay DJC (1996). "Maximum likelihood and covariant algorithms for independent component analysis,", University of cambridge, Cavendish Laboratory, Draft 3. 7.

Abu-Taleb AS, Zayed EME, M. El-sayed W (2006). " Blind Source Recovery Using an Adaptive Generalized Gamma Source Function " Georgian Electronic Scientific: Computer Science and Telecommunications 2(9): 80-89.

El-sayed W, Mohamed OA (2006). Generalized Gaussian Model for Multi Channel Image Deconvolution" Georgian Electronic Scientific: Computer Science and Telecommunications. 3(10). 127-133.

El-sayed W (2007). " Image enhancement using second generation wavelet super resolution " Int. J. Physical Sci. 2(6): 149-158.

El-sayed W (2007). "Registration of totally aliased signals " Int. J. Physical Sci. 2(9): 149-158.