ISSN 2277-3061

Blind Signal Separation Using an Adaptive Generalized Compound Gamma Distribution

M. El-Sayed Waheed,
Department of Computer Science Faculty of Computers and Information
Suez Canal University (Egypt)
E-mail: mewahed@yahoo.com,
Y. A. Amer and A. Moftah Elmabrouk
Dept of Mathematics, Faculty of Science, Zagazig University (Egypt)

Abstract: 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 algorithm in the framework of natural Riemannian gradient, is derived using the parameterized Generalized Compound Gamma Distribution density model. The nonlinear function in ICA algorithem is self-adaptive and is controlled by the shape parameter of Adaptive Generalized Compound Gamma Distribution density model. Computer simulation results confirm the validity and high performance of the proposed algorithm

Keywords: Independent component analysis, Generalized Compound Gamma Distribution, Maximum likelihood, sub- and super- Gaussian., Blind signal separation

Council for Innovative Research

Peer Review Research Publishing System

Journal: INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY

Vol 12, No.3

editor@cirworld.com

www.cirworld.com, member.cirworld.com

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, 24]. 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 $S(n) = [S_1, S_2, ..., S_m]$ 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 nonsigular 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**=A**S** where $X = [X_1, X_2, ..., X_m]^T$. 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**=W**X** 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=WAS→DPS* 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. Maximum Entropy Algorithm:

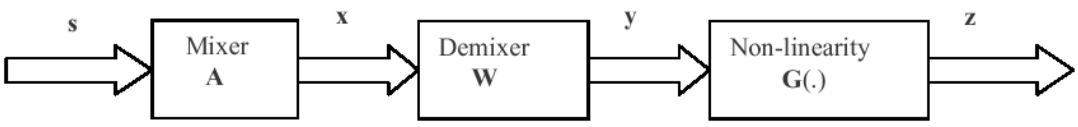


Fig.1: Maximum Entropy Method

This is an adaptive algorithm based on information theoretic approach and was suggested by Bell & Sejnowski [7]. 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 = W**X, 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**) $\Rightarrow$ **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

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

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

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

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 $|\det(J(s))|$.

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

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

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

$$z_i = g_i(y_i) = \int_{-\infty}^{z_i} fS_i(s_i)ds_i,$$ 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

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

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

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

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

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

(1)

Therefore, under the above condition, maximizing the log-likelihood function $\{\log f_s(s/W,G)\}$ {}) is equivalent to maximizing the entropy h(Z) for blind source separation.

Referring to the expression h(z) = -E[log $f_z(Z)$] = -E $\left[\log\left(\frac{f[S(s)]}{|\det(J(s))|}\Big| s = \psi(z)\right)\right]$, 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:

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

(2)

On expanding (2), we get:

$$\varphi = \log\left|\det(A) + \log\left|\det(W)\right|\right| + \sum_{i=1}^{n} \log\left(\frac{\partial z_i}{\partial y_i}\right) \text{ and } \frac{\partial \varphi}{\partial W} = W^{-T} + \sum_{i=1}^{n} \frac{\partial}{\partial W}\log\left(\frac{\partial z_i}{\partial y_i}\right) \quad (3)$$

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's used are logistic function and hyperbolic tangent function:

$$z_i = g(y_i) = \frac{1}{1+e^{-y_i}}, \quad z_i = g(y_i) = \tanh(y_i), \quad i = 1,2,\dots,n$$

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 $W^T W$ instead of evaluating W^{-T} as given below:

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

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

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.

3. Generalized Compound Gamma Distribution 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 [7] leads to the nonlinear function given by

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

Such sigmodal function was also used in [7]. For sub-Gaussian sorce 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 [15;16]. (for example, either $f_s(s) = f_s^3$ or $f_s(s) = \tanh(\beta f_s(s))$. Several approaches [18;10;11] are already available.

This paper presents a flexible nonlinear function derived using Generalized Compound Gamma density model. It is shown that the nonlinear function is self-adaptive and controled by Generalized Compound Gamma shape parameter. It is not a form of fixed nonlinear function.

3.1. Generalized Compound Gamma Distribution

The probability density function of the Generalized Compound Gamma Distribution is given by :

$$f(\chi;\alpha,\theta,\lambda,b) = \frac{1}{b\beta(\alpha,\theta)}\left(\frac{\chi-\lambda}{b}\right)^{\alpha-1}\left(1+\left[\frac{\chi-\lambda}{b}\right]\right)^{-(\alpha+\theta)},$$

$$0 < \lambda < \chi < \infty, \qquad \alpha,\theta,b > 0$$

(6)

where α and θ are the shape parameters, λ is the location parameter and b is the scale parameter and $\beta(.,.)$ is well known beta function.

The standard form of the distribution will have $\lambda = 0$ and $b = 1$ so that the standard density function is

$$f(\chi;\alpha,\theta,\lambda,b) = \frac{1}{\beta(\alpha,\theta)}\chi^{\alpha-1}(1+\chi)^{-(\alpha+\theta)} \qquad \chi > 0, \alpha > 0 \quad (7)$$

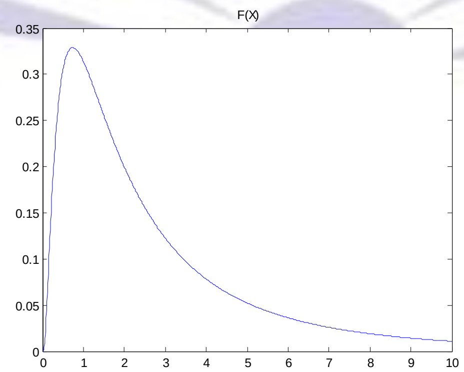


Fig.2: The plot of pdf of the generalized compound gamma distribution for $\alpha = 5, \theta = 5, \lambda = 0.02, b = 8.$

ISSN 2277-3061

If $y = b\left(\frac{\nu_2}{\nu_1}\left[\frac{x-\lambda}{b}\right]\right)^{\gamma}$, $\alpha = \frac{\nu_1}{2}$ and $\theta = \frac{\nu_2}{2}$, then GB2(6) can be transformed to the following generalized F-distribution $\left(\gamma F^{b}_{\nu_1,\nu_2}\right)$ given by Malik (1967)

$$f(y;\gamma,b) = \frac{(\nu_1/\nu_2)^{\nu_1/2}(\gamma b)^{-1}(y/b)^{(\nu_1/2\gamma)-1}}{\beta(\nu_1/2,\nu_2/2)\left(1+[\nu_1/\nu_2][y/b]^{1/\gamma}\right)^{(\nu_1+\nu_2)/2}}, \qquad , \qquad \gamma,b,y>0 \tag{8}$$

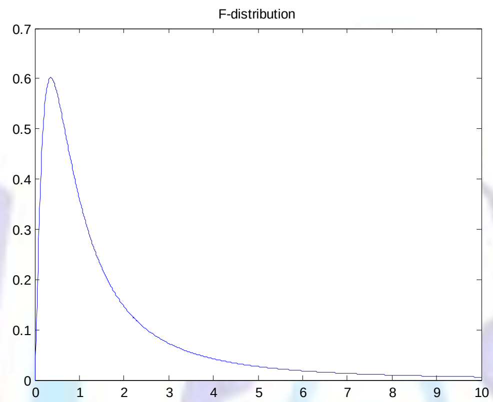


Fig.3: The plot of f-distribution.

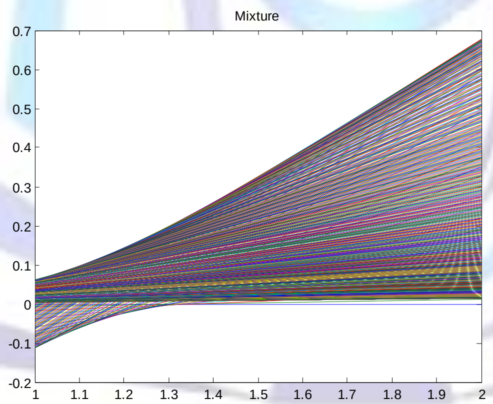


Fig.4: The plot of the mixture of generalized compound gamma distribution and f-distribution.

ISSN 2277-3061

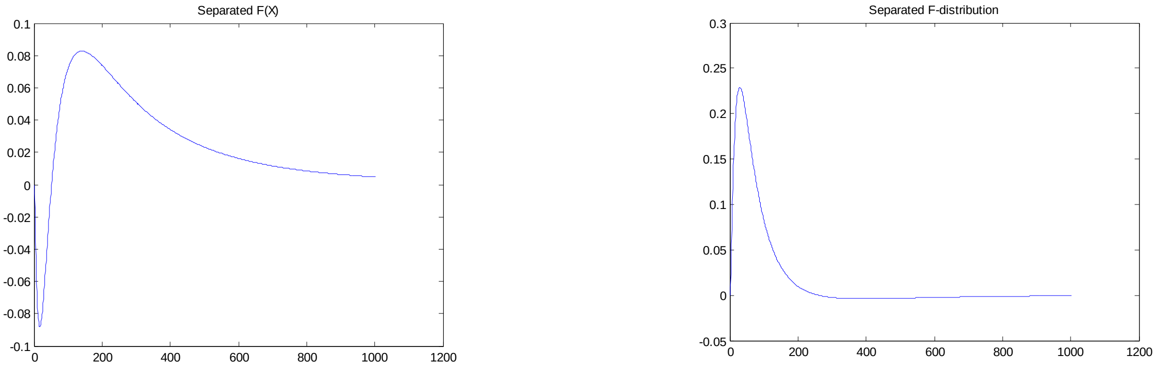


Fig.5: The plots of separated generalized compound gamma distribution and separated f-distribution.

3.2 The Moments of Generalized Compound Gamma Distribution

In order to fully understand the generalized compound gamma distribution, it is usefull to look at its moments (specially 2nd and 4th moments which give the kurtosis). The r^{th} moment about zero for the compound density (6) can be derived as:

$$\mu_r = \frac{b^r}{\Gamma(\alpha)\Gamma(\theta)} \sum_{j=0}^{r} \binom{r}{j} (-1)^j \left(\frac{\alpha}{\theta-1}\right)^j \Gamma(\alpha+r-j)\Gamma(\theta-r+j),$$

$$\theta > r + j, \qquad r = 1,2,3,\ldots$$

(9)

3.3. Kurtosis and shape parameter

The kurtosis is a nondimensional quantity. It 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

$$\alpha_4 = \frac{\mu_4}{\mu_2^2} = \frac{\left[\frac{\alpha(\alpha+1)(\alpha+2)(\alpha+3)}{(\theta-1)(\theta-2)(\theta-3)(\theta-4)} - \frac{4\alpha^2(\alpha+1)(\alpha+2)}{(\theta-1)^2(\theta-2)(\theta-3)} + \frac{6\alpha^3(\alpha+1)}{(\theta-1)^3(\theta-2)} - \frac{3\alpha^4}{(\theta-1)^4}\right]}{\left[\frac{\alpha(\alpha+(\theta-1))}{(\theta-1)^2(\theta-2)}\right]^2},$$

$$\theta > 4.$$

(10)

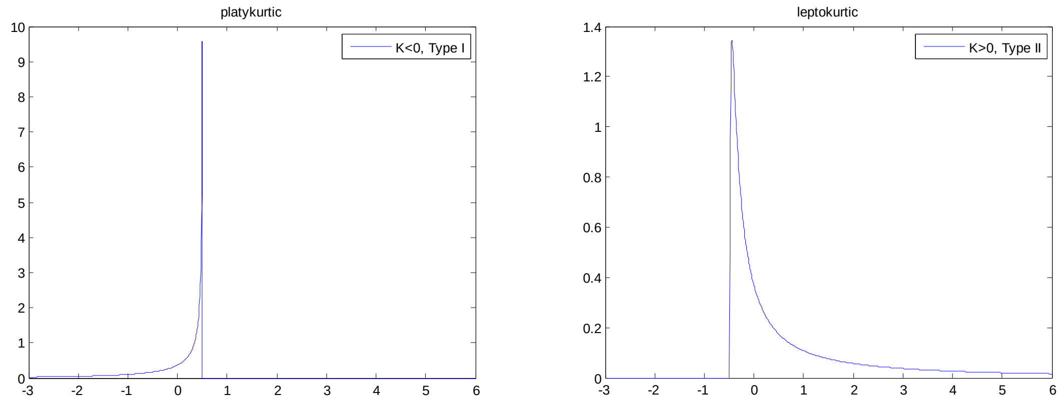


Fig.6: The plots of kurtosis α_4 for platykurtic and leptokurtic signals for generalized compound gamma distribution.

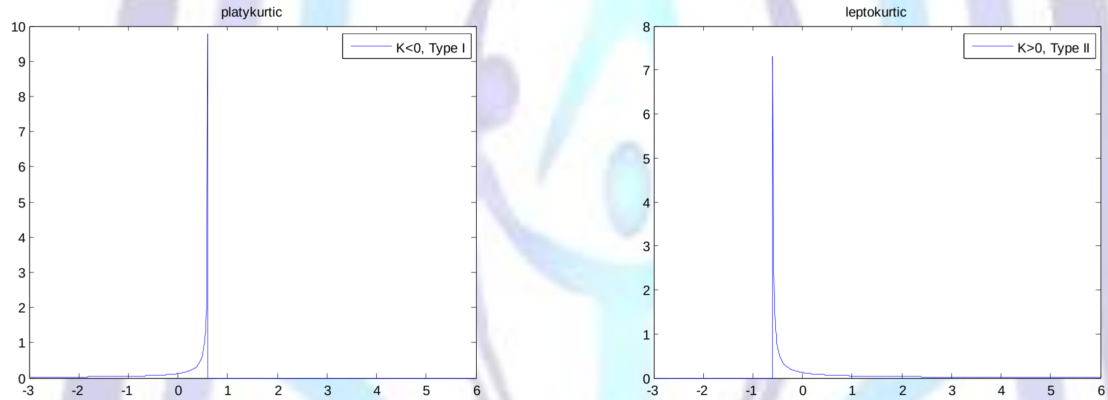


Fig.7: The plots of kurtosis α_4 for platykurtic and leptokurtic signals for f-distribution.

ity

The activation function for generalized compound gamma distribution in (3) is given by

$$f_i(x_i) = \frac{\alpha - 1 - \alpha x^{\alpha}}{x}$$

(11)

$$\frac{\partial L}{\partial \alpha_i} = \frac{\partial \ln f(x_i;\alpha_i)}{\partial \alpha_i} = \ln x_i - \ln(1 + x_i) + \sum_{n=1}^{\infty}(\alpha_i + \theta - n) \qquad (12)$$

$$\Delta \alpha_i = -\eta_i \frac{\partial L}{\partial \alpha_i} = -\eta_i \left(\ln x_i - \ln(1 + x_i) + \sum_{n=1}^{\infty}(\alpha_i + \theta - n) \right) \qquad (13)$$

4. COMPUTER SIMULATION RESULTS

Example1: To illustrate the method applicability, we consider a simulation example which consists of analyzing a mixture of three sources. The mixture is obtained by constructing three synthetic spectra and considering nineteen measures with mixing coefficients chosen in such a way to have a realistic evolution. A Gaussian noise is added to have a signal to noise ratio equal to 50 dB. Figure 1 shows the resulting mixture. To discuss the result accuracy, we use the global system matrix G = $\hat{A}^{-1}$ A that indicates the separation performance. The empirical source covariance matrix is:

$$A = \begin{bmatrix} 1.000 & 0.516 & .386 \\ 0.516 & 1.000 & -0.105 \\ 0.386 & -0.105 & 1.000 \end{bmatrix}$$

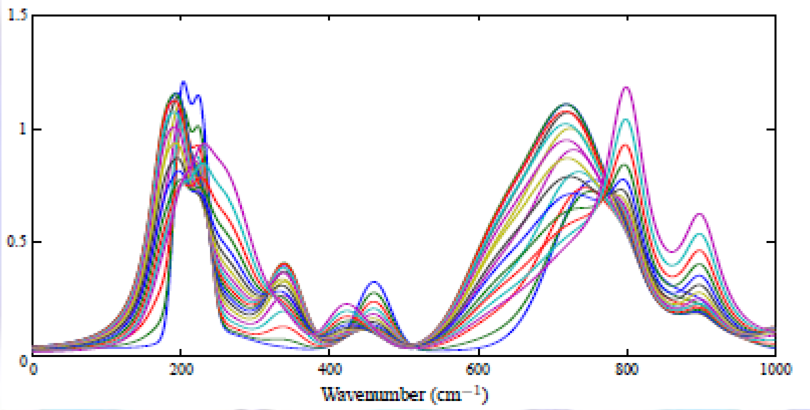


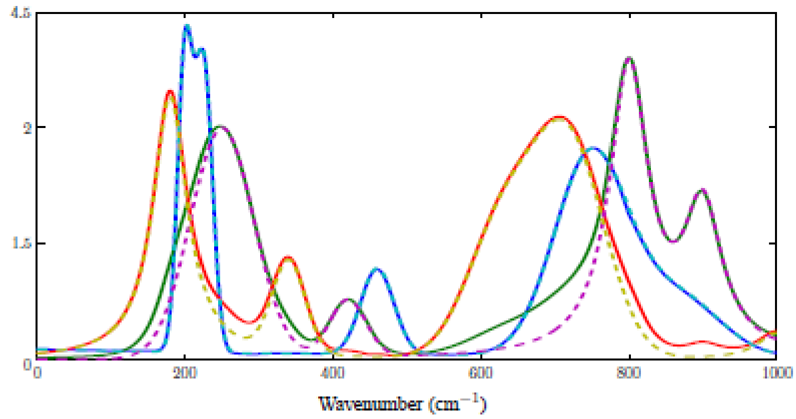
Fig. 8: Mixture synthesis

When analyzing this covariance matrix we note that the available samples of the sources are spatially correlated, so the independence assumption is not sufficient for the spectra reconstruction. This explains the failure in applying directly an ICA algorithm. To give an illustration of this aspect, the global system matrix resulting from the analysis by JADE algorithm [1] is shown:

$$G = \begin{bmatrix} -0.499 & 0.836 & 1.030 \\ 1.263 & -0.412 & -0.280 \\ -0.127 & 0.856 & -0.480 \end{bmatrix}$$

The results obtained by applying the proposed method for the mixture analysis are presented in figure 9. We can see that source spectra and mixing coefficients are estimated without apparition of negative values. Concerning the separation performances, the global system matrix associated to the reconstruction is:

$$\begin{bmatrix} 1.028 & -0.027 & -0.011 \\ 0.014 & 0.996 & 0.137 \\ -0.018 & 0.089 & 1.020 \end{bmatrix} \text{G =}$$



True (dashed line) and estimated (continuous line) mixing coefficients

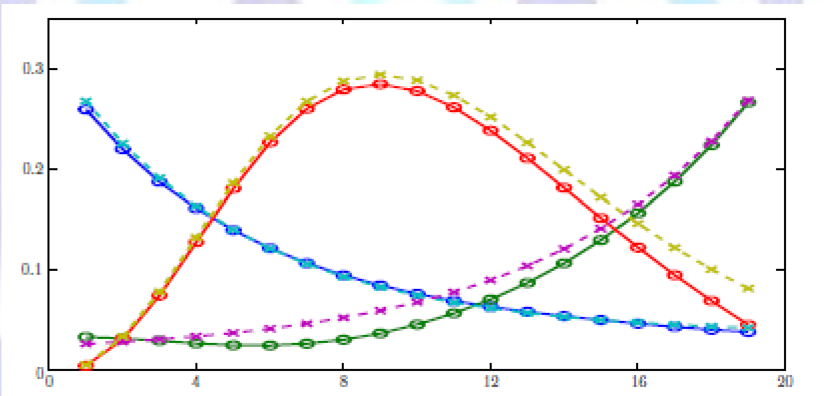


Fig. 9: Mixture analysis results

CONCLUSION

We have proposed an ICA algorithm (in the framework of natural Riemannian gradient) where the self-adaptive nonlinear function eas derived using Adaptive Generalized Compound Gamma Distribution 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 Adaptive Generalized Compound Gamma Distribution. Finally we apply our algorithm on a mixture of other distributions and signal separation, which give a good results.

REFERENCES

[1] F. Alberg, P. Duhamel, M. Nikolova, Adaptive solution for blind identification/equalization using deterministic maximum likelihood, IEEE Trans. Signal Process. 50 (4) ,pp923–936, (April 2002).

[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,” *Neural Computation*, vol 11, pp. 355-367. 1998.

[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, vol. 8, pp. 757-763, MIT press., Eds. 1996

[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, pp. 349-353, 1997,.

[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,* pp. 731-734, *1997*.

[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, pp. 406-411, 1994.

[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., pp. 436-445,

1997.

[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*, pp. 1788-1791, 1997..

[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*, pp. 249-266, , 1996.

[22] D. J. C. MacKay, “maximum likelihood and covariant algorithms for independent component analysis,”, University of cambridge, Cavendish Laboratory, Draft 3.7, 1996.

[23] Abu-Taleb A.S.,Zayed E.M.E., M. El-sayed Wahed, " Blind Source Recovery Using an Adaptive Generalized Gamma Source Function " Georgian Electronic Scientific: Computer Science and Telecommunications 2006, No. 2(9). Pp. 80-89.

[24] M. El-sayed Wahed, Mohamed O. A. " Generalized Gaussian Model for Multi Channel Image Deconvolution" Georgian Electronic Scientific: Computer Science and Telecommunications 2006, No. 3(10). Pp. 127-133.

[25] M. El-sayed Wahed, " Image enhancement using second generation wavelet super resolution " Interational Journal of Physical Science Vol. 2(6), pp. 149-158, june, 2007.

[26] M. El-sayed Wahed, " Registration of totally aliased signals " Interational Journal of Physical Science Vol. 2(9), pp. 149-158, september, 2007.

[27] M. El-sayed Wahed, " Blind signal separation using an adaptive weibull distribution" International Journal of Physical Sciences Vol. 4 (5), pp. 265-270, May, 2009.