ECMS 2007 & Doctoral School
8th International workshop on Electronics, Control, Modeling, Measurement and Signals 2007
& Doctoral School (EDSYS, GEET), May 21-23, 2007, Liberec, Czech Republic
----

A Variant of EFICA Algorithm with Adaptive Parametric Density Estimator

J. Málek[1], Z. Koldovský[1], S. Hosseini[2], Y. Deville[2]
[1]Faculty of Mechatronics and Interdisciplinary Engineering Studies, Technical University of Liberec, Czech Republic
[2]Laboratoire Astrophysique de Toulouse-Tarbes (LATT), Observatoire Midi-Pyrénées – Université Paul Sabatier, Toulouse, France
jiri.malek@tul.cz, zbynek.koldovsky@tul.cz, shosseini@ast.obs-mip.fr, ydeville@ast.obs-mip.fr

Abstract: *FastICA is a popular method for Independent Component Analysis used for separation of linearly mixed independent sources. The separation proceeds through optimization of a contrast function that is based on kurtosis or other entropy approximations using a nonlinear function. The EFICA algorithm is a recently proposed version of this algorithm that is asymptotically efficient when all source distributions are from the Generalized Gaussian family. It is known that the optimal nonlinearity for entropy estimation is the score function of the source distribution. For its evaluation the knowledge of the probability density function (pdf) of the source signals is necessary. Because these pdfs are unknown, the EFICA algorithm assumes that the distribution is Generalized Gaussian Distribution, calculates its parameter α for each source and selects one nonlinearity from the available set for fine tuning of the sources. This paper proposes to modify the EFICA algorithm by parametric estimation of the real score function, which is subsequently used as the contrast function in fine tuning iterations. The algorithm is tested on artificial data; the results are compared to original EFICA, FastICA and JADE.*

I. INTRODUCTION

Independent Component Analysis (ICA) [1] is a popular tool for Blind Source Separation (BSS). In this technique, original source signals are retrieved from their mixtures using the assumption of their mutual independence. The simplest data model for ICA is instantaneous noiseless mixture

$$\mathbf{X} = \mathbf{AS}, \tag{1}$$

where $\mathbf{S}$ is a $d \times N$ matrix consisting of the original sources whose i-th row corresponds to an original source that is considered as a sequence of independent and identically distributed (i.i.d.) realizations of a random variable with pdf $p_i(\cdot)$. $\mathbf{A}$ is an unknown $d \times d$ regular mixing matrix, and $\mathbf{X}$ is a $d \times N$ data matrix representing the mixed signals. The goal of ICA is to find a demixing matrix $\mathbf{W}$, $\mathbf{W}=\mathbf{A}^{-1}$, up to unretrievable order and scales of its rows. Only one of the signal densities $p_i(\cdot)$ is allowed to be Gaussian.

There are several algorithms for ICA which differ in the way how the signal densities are estimated, because it is the pivotal task in order to achieve the best possible separation [1, 2, 3, 4]. More precisely, the score function $-p'_i / p_i$ should be known or estimated or replaced by an acceptable nonlinearity. For example, JADE [5] executes the kurtosis approximation via cumulants, or RADICAL [6] uses order statistics-based estimation. The FastICA algorithm [7] applies a suitable nonlinearity in the contrast function.

Recently, a modified version of the FastICA algorithm, called EFICA, has been proposed [8]. This algorithm is asymptotically efficient, i.e. its accuracy given by residual error variance attains the Cramér-Rao lower bound [2, 3, 8]. However, this optimality is only achieved for signals with Generalized Gaussian Distribution (GGD) with parameter $\alpha>2$.

This paper deals with a modification of EFICA in the sense that the choice of the nonlinearity is more adaptive, and the algorithm yields superior performance for a wider family of distributions. This is achieved via parametric score function estimation proposed in [3, 4].

In Section II, the original EFICA algorithm and the modified version proposed in this paper are briefly described. Section III provides several details about the score function estimator, and the Section IV demonstrates results of performed experiments.

II. EFICA ALGORITHM AND ITS PROPOSED MODIFICATION

A. Original EFICA algorithm

EFICA [8] is an elaborated version of FastICA that consists of three steps. In the first step, the original FastICA (the symmetric approach [7]) is applied to the preprocessed data with removed sample mean and correlations. The nonlinearity used in this step is such that it guaranties convergence for a wide spectrum of distributions; a popular choice for this is hyperbolic tangent [7]. The first step is completed by the test of saddle points [2] to ensure the global convergence of FastICA.

The second step of EFICA takes advantage of having provided preliminary estimates of the original signals from the previous step. This is used for adaptive choice of the nonlinear function, which is separately done for each to-be-estimated signal. Based on the sample fourth moment of the signal, the original EFICA selects the nonlinearity that is appropriate for GGD signals.

In the third step, few fine-tuning iterations of one-unit FastICA using the selected nonlinearities are executed on the preliminary estimate of the sources (obtained in the first step). This improves estimates of individual signals. The algorithm is finished by a refinement that is designed to minimize the residual inter-signal interference.

B. Modification proposed in this paper

The modification proposed in this paper consists in connecting the second "adaptive" step with the fine-tuning iterations of the third step. To achieve improved accuracy and flexibility, the nonlinearity choice is replaced by a parametric score function estimator (for each signal separately) before each fine-tuning iteration. The rest of the algorithm remains unchanged.

III. PARAMETRIC SCORE FUNCTION ESTIMATOR

A. Structure of the estimator:

The marginal score function $\Psi(s)$ of a signal s is defined in the following way:

$$\Psi(s) = -\frac{p'(s)}{p(s)}, \tag{2}$$

where $p(s)$ and $p'(s)$ is the probability density function of s and its derivative, respectively.

The mean-square parametric estimator, denoted by $\mathbf{h}(s \mid \boldsymbol{\theta})$, of the marginal score function is defined as one that minimizes

$$\mathrm{E}[(\Psi(s) - \mathbf{h}(s \mid \boldsymbol{\theta}))^2]. \tag{3}$$

Here, $\boldsymbol{\theta}$ is a vector of parameters subject to minimization, and E stands for the expectation value operator.

Equation (3) can be expressed without dependence on the unknown score function using the following theorem [3, 4], provided that $\lim_{s\to\pm\infty} p(s) = 0$:

$$E[h(s) \cdot \Psi(s)] = E[h'(s)]. \tag{4}$$

By expanding (3), leaving the constant term, and using (4), we get for the minimizing vector Θ

$$\Theta = \arg\min_{\boldsymbol{\theta}} E[\mathbf{h}(s \mid \boldsymbol{\theta})^2] - 2E[\mathbf{h}'(s \mid \boldsymbol{\theta})]. \tag{5}$$

To simplify the optimization problem, we consider $\mathbf{h}(s\,|\,\boldsymbol{\theta})$ as a linear combination of *basis* functions $h_1(s),...,h_k(s)$, i.e.

$$\mathbf{h}(s\,|\,\boldsymbol{\theta}) = \sum_{i=1}^{k} \theta_i \cdot h_i(s). \tag{6}$$

Then, (5) simplifies to

$$\Theta = \mathrm{E}[\mathbf{h}^T(s) \cdot \mathbf{h}(s)]^{-1} \cdot \mathrm{E}[\mathbf{h}'(s)], \tag{7}$$

where $\mathbf{h}(s) = [h_1(s),...,h_k(s)]^T$, and $\mathbf{h}'(s) = [h_1'(s),...,h_k'(s)]^T$.

B. *Choice of the basis functions*

General information regarding selection of basis functions $h_1(s),...,h_k(s)$ can be found in [3, 4, 9, 10]. The functions used in the modified EFICA are given by (8) and (9). These sets of function have been chosen following the results of experiments performed during the development of the algorithm. Different sets of basis functions are used when separating sub-Gaussian or super-Gaussian sources. The selection of the basis is performed by considering the value of the sample fourth moment $\hat{m}_4$, that is calculated from the pre-estimated source signals.

For sub-Gaussian sources we use the score functions of the GGD distributions with α=4,8,12,16, i.e.,

$$h_i(s) = sign(s) \cdot |s|^{\alpha_i - 1}, i = 1..4, \alpha_i = 4,8,12,16 \qquad \hat{m}_4 < 3. \tag{8}$$

It is obvious that such choice is efficient for sub-Gaussian GGD distributions, but our experiments show that this choice is useful for other sub-Gaussian signals (bimodal distributions) as well.

For super-Gaussian sources the score function for GGD α =1.5 and the exponential function from [8] has been chosen, i.e.,

$$h_1(s) = sign(s) \cdot |s|^{0.5}, h_2(s) = s \cdot \exp(-\eta|s|), \eta = 3.348 \qquad \hat{m}_4 > 3. \tag{9}$$

The score functions of GGD for α ≤ 1 are not continuous, however, in EFICA iterations, the derivative of the score functions is needed. For instance, the exponential function represents a continuous approximation of these discontinuous score functions.

IV. SIMULATIONS

In this section, performance of the proposed modified EFICA is presented and compared with that of other ICA algorithms. For this purpose, FastICA [7] with nonlinearity "tanh", JADE [5], and the original EFICA [8] were taken into account. In order to demonstrate the improved performance of the modified algorithm, signals with other distributions than GGD are considered in our examples. Other examples not shown here due to lack of space confirm good performance of the modified algorithm compared to the original EFICA when separating sources from GGD family.

Let **W** be an estimated de-mixing matrix after the original order of its rows was retrieved. The interference-to-signal ratio of the k-th signal, denoted by ISR_k, is defined by

$$\mathrm{ISR}_k = \frac{\sum_{l=1, l\neq k}^{d} \mathbf{G}_{kl}^2}{\mathbf{G}_{kk}^2}, \tag{9}$$

where **G=WA**. We use this quantity to evaluate the separation performance achieved by a corresponding algorithm.

Example 1: Separation of binary phase shift keying signals (BPSK) with absorbed Gaussian noise.

A BPSK signal with absorbed Gaussian noise is one that is generated from distribution of a random variable $\sqrt{1-\varepsilon^2}b + \varepsilon \cdot n$, where b is a binary random variable equal to *1* or *-1* with equal probabilities, and *n* is a standard Gaussian variable. The probability density of such random variable, for different values of parameter ε, is shown in Fig. 1. Note that extreme values of the parameter ε are 0 and 1, respectively, corresponding to the BPSK and Gaussian distribution.

Five signals of length *N=2500* were generated in *100* independent trials, mixed with a random mixing matrix, and subsequently separated. Averaged ISRs over all separated signals and trials are shown in Fig. 2. as a function of the parameter ε. The enhanced adaptability of the nonlinearity choice in the proposed method compared to the original EFICA is most apparent for ε≈0.2, where the distribution of the signals is rather bimodal.

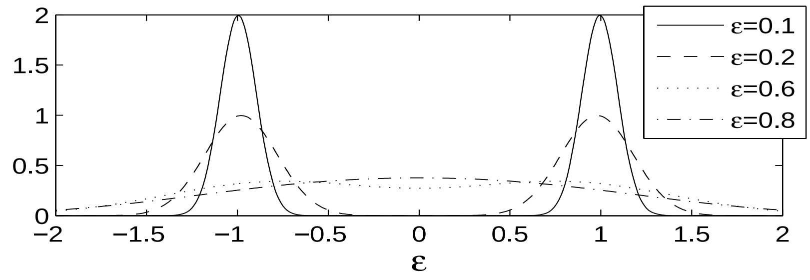

Fig. 1. Probability densities of signals from Example 1, ε=0.1, 0.2, 0.6, 0.8.

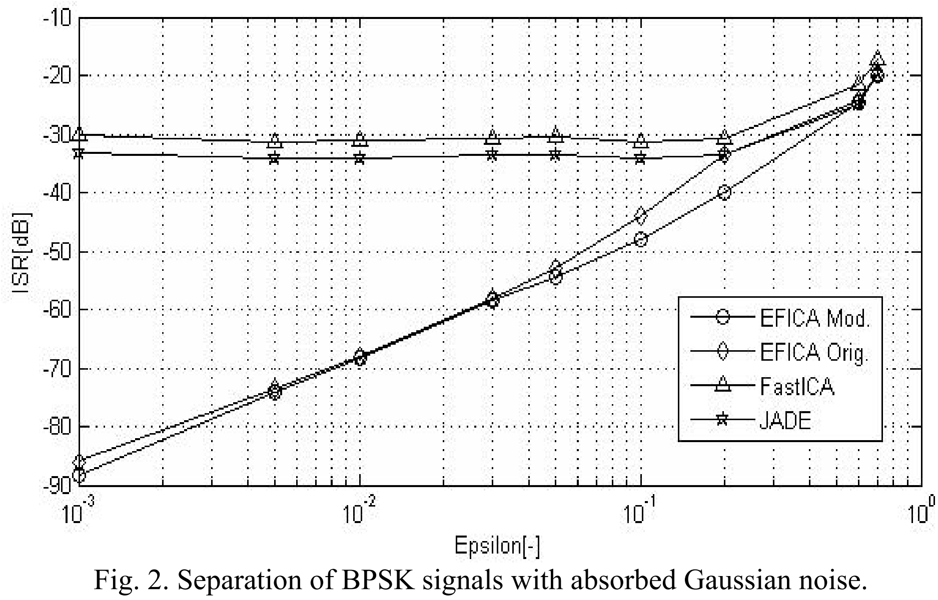

Fig. 2. Separation of BPSK signals with absorbed Gaussian noise.

Example 2: Separation of GGD signals with absorbed Gaussian noise.

This example is nearly similar to the previous one, but, here, the distribution of the random variable *b* is GGD with varying parameter α from 0.5 to 20. The value of the parameter ε is fixed to *0.2,* thus, the signal distribution is not from the GGD family. Results of this experiment are shown in Fig. 3. The proposed method yields improved performance for $\alpha<2$, where the signals are super-Gaussian. For $\alpha>2$, the results of both variants of EFICA are comparable.

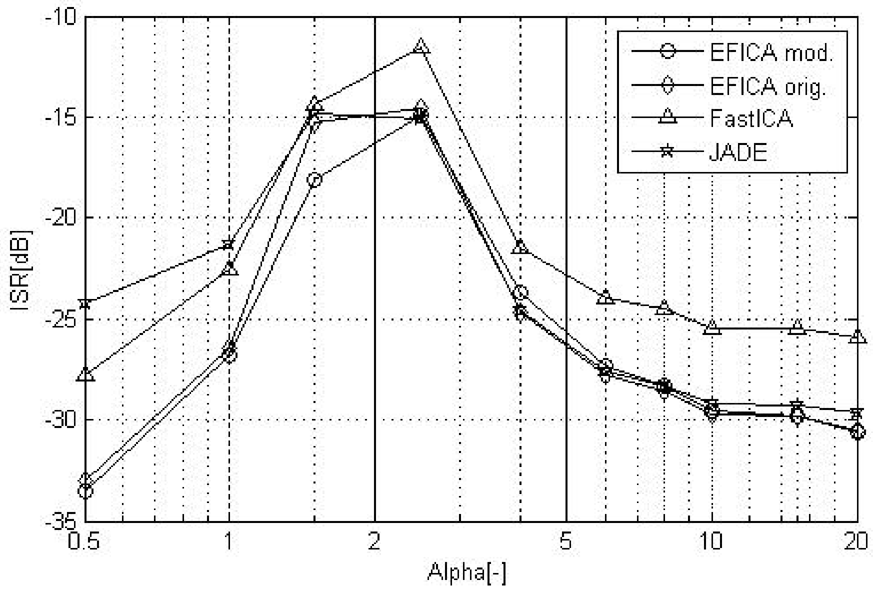

Fig. 3. Separation of GGD sources with absorbed Gaussian noise.

Example 3: Computational burden

Five GGD sources of various data lengths with α equal to *0.5, 1, 1.5, 5* and *10* were mixed with a random matrix and separated. The average CPU times over *50* trials required by the compared algorithms are shown in Fig. 4. The same was done for a fixed $N=5000$ but varying number of signals taking random α from the set of values *0.5, 1, 1.5, 5* and *10.* Corresponding results are shown in Fig. 5.

As can be seen, the proposed algorithm is about three times more computationally demanding compared to the original EFICA. Example with varying data dimension shows that JADE is the fastest algorithm for small dimensions.

These experiments were performed in Matlab on a laptop AMD Turion 64 x2, 1.6 GHz, 1 GB RAM.

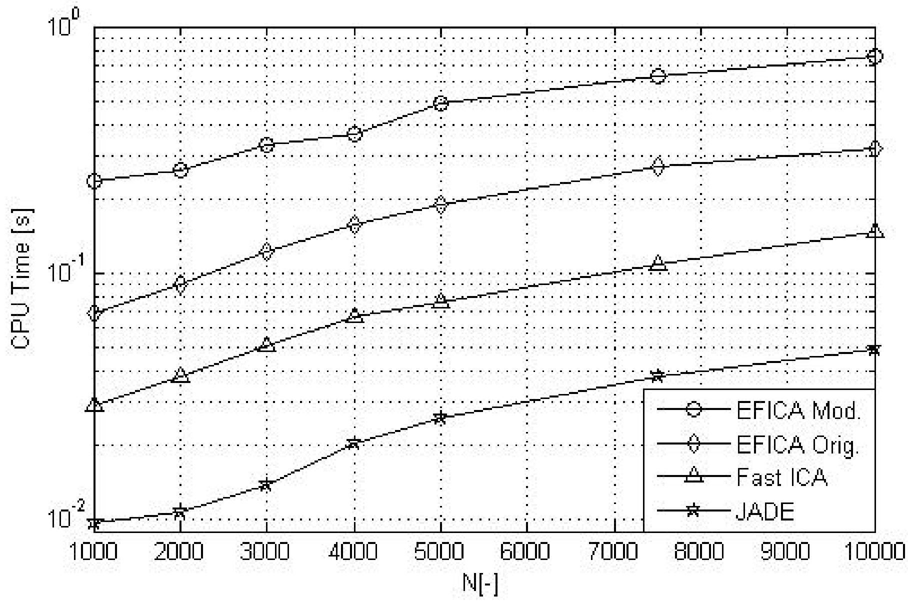

Fig. 4. CPU Time required for separation of GGD data for various sample sizes.

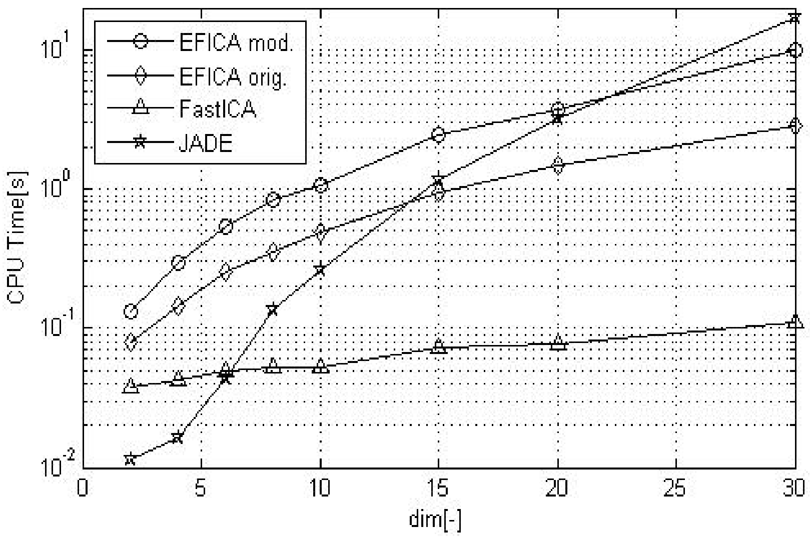

Fig. 5. CPU Time required for separation of GGD data for various numbers of signals.

V. CONCLUSION

In this paper, a novel variant of EFICA algorithm utilizing a computationally affordable parametric score function estimator has been proposed. The experimental results show that this modification yields improved separation performance when the distribution of the original signals is not from the GGD family.

ACKNOWLEDGEMENT

This work was supported by the Czech Science Foundation (GA ČR) through the project 102/07/P384 and by the internal grant of the Faculty of Mechatronics and Interdisciplinary Engineering Studies.

REFERENCES

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley-Interscience, New York, 2001.

[2] P. Tichavský, Z. Koldovský, and E. Oja, “Asymptotic Performance Analysis of the Fixed-Point Algorithm (Fast-ICA) for Independent Component Analysis”, *IEEE Statistical Signal Processing Workshop*, Bordeaux, France, July 2005.

[3] D.T. Pham, Garat, P., “Blind separation of mixture of independent sources through a quasi-maximum likelihood approach”, *IEEE Transactions on Signal Processing*, Vol. 45, No. 7, July 1997, pp. 1712 - 1725.

[4] S. Hosseini, R. Guidara, Y. Deville, Ch. Jutten, Markovian Blind Image Separation, *ICA 2006, LNCS3889,*Springer-Verlag, pp.106-114, 2006.

[5] J.-F. Cardoso and A. Souloumiac, “Blind Beamforming from non-Gaussian Signals”, *IEEE Proc.-F*, 140(6):362-370, Dec. 1993.

[6] E. G. Learned-Miller and J. W. Fisher III, “ICA using spacing estimates of entropy”, *Journal of Machine Learning Research*, vol. 4, no. 7-8, pp. 1271-1295, 2004.

[7] A. Hyvärinen,”Fast and Robust Fixed-Point Algorithms for Independent Component Analysis”. *IEEE Transactions on Neural Networks*, vol. 10, no.3, 626-634, 1999.

[8] Z. Koldovský, P.Tichavský, E. Oja, “Efficient Variant of Algorithm FastICA for Independent Component Analysis Attaining the Cramér-Rao Lower Bound”, *IEEE Transactions on Neural Networks*, vol. 17,no. 5, 1265-1277, 2006.

[9] Ch. L. Brown, R.F. Brcich, A.Taleb, Suboptimal robust estimation using rank score functions, *ICASSP 2003, pp. 753-756,* 2003.

[10] R.F. Brcich, F.M. Zoubir, Robust estimation with parametric score function estimation, *IEEE Proc.* 1149-1152, 2002