Independent Component Analysis with α-Stable Distributions

Ryszard Szupiluk, Piotr Wojewnik, Tomasz Ząbkowski, Krzysztof Siwek

***Abstract* — In this paper we propose the selection method for choosing the nonlinearities in ICA algorithms for the signals that have alpha-stable distributions. The computer simulations were conducted to confirm the correctness of the approach.**

I. Introduction

INDEPENDENT Component Analysis (ICA) is a very intensively developed field of the science, which has many applications across a variety of industries like telecommunication, electricity, medicine, economics, etc. The ICA methods are especially used for blind signal separation (BSS). There exist many ICA methods and algorithms based on different rules. The classical ICA methods are Natural Gradient algorithms, which have the efficacy directly connected with the used nonlinearities. The selection of the nonlinearities depends on the probability distribution of the processed signals.

In practice, we often have to do with the impulse signals where quite suitable approach can be the application of alpha-stable distributions. Signals with alpha stable distributions have characteristic properties called "fat tails". In this paper we present the concept and the findings for the choice of the nonlinearities for alpha-stable distributions in Natural Gradient algorithm.

We will take into the consideration the standard in ICA linear model of observed variables. It will be called a generative model. Let us assume that the *m*-dimensional observation vector $\mathbf{x}(t)$ is generated by

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

where:

1. $\mathbf{A} \in \Re^{m \times n}$ is the unknown full column rank mixing matrix $m \geq n$,
2. $\mathbf{s}(t)$ is the *n*-dimensional basis nongaussian (except one) vector with mutual independent variables.

To estimate independent components we are looking for linear transformation that will be an inverse operation to mixing. The basis difficulty is lack of knowledge both $\mathbf{A}$ and $\mathbf{s}$ in (9). This fact introduces two ambiguities in the solution. It is impossible to recover original scale and order of $\mathbf{s}$. So, we accept estimated signals rescaled and reordered in comparison with original signals. Let $\mathbf{W}$ describe demixing matrix and

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) \tag{2}$$

We obtain independent components in $\mathbf{y}$ if we find such matrix $\mathbf{W}$ that

$$\mathbf{G} = \mathbf{W}\mathbf{A} = \mathbf{P}\mathbf{D} \tag{3}$$

where $\mathbf{G}$ is the global transformation matrix which combines the mixing and separating system ($\mathbf{G}$ is called the generalized permutation matrix), $\mathbf{P}$ is some permutation matrix, $\mathbf{D}$ is some scaling non-singular diagonal matrix.

To obtain independent components we need to analyze statistical structure of $\mathbf{y}$. The joint probability of independent variables can be factorized by the product of the marginal probabilities

$$\overbrace{p_1(y_1)p_2(y_2)...p_n(y_n)}^{q_{\mathbf{y}}(\mathbf{y})} = \overbrace{p_{1\cdots n}(y_1, y_2, ..., y_n)}^{p_{\mathbf{y}}(\mathbf{y})} \tag{4}$$

As the practical measure of difference between $p_{\mathbf{y}}(\mathbf{y})$ and $q_{\mathbf{y}}(\mathbf{y})$ we take Kullback-Leibler divergence

$$D_{KL}(p_{\mathbf{y}}(\mathbf{y}) \,\|\, q_{\mathbf{y}}(\mathbf{y})) = \int_{-\infty}^{+\infty} p_{\mathbf{y}}(\mathbf{y}) \log \frac{p_{\mathbf{y}}(\mathbf{y})}{q_{\mathbf{y}}(\mathbf{y})} d\mathbf{y} \tag{5}$$

We are looking for a matrix $\mathbf{W}$ that mineralizes (5) so

$$\mathbf{W}_{opt} = \min_{\mathbf{W}} D_{KL}(p_{\mathbf{y}}(\mathbf{Wx}) \,\|\, q_{\mathbf{y}}(\mathbf{Wx})) \tag{6}$$

The natural gradient optimisation method [1] applied to (6) gives us an update rule for matrix $\mathbf{W}$

$$\Delta\mathbf{W}(t) = \mu(t)\left[\mathbf{I} - \mathbf{f}(\mathbf{y}(t))\mathbf{y}^T(t)\right]\mathbf{W}(t) \tag{7}$$

where

$$\mathbf{f}(\mathbf{y}) = [f_1(y_1), ..., f_n(y_n)]^T \tag{8}$$

with

$$f_i(y_i) = \frac{\partial \log(p_i(y_i))}{\partial y_i} = \frac{1}{p_i(y_i)} \frac{\partial p_i(y_i)}{\partial y_i} \tag{9}$$

Optimal non-linearities used in (9) need the knowledge of source probability distributions what is impossible in general. To avoid these disadvantages we can use extended version of (7)

Ryszard Szupiluk and Piotr Wojewnik are with Polska Telefonia Cyfrowa Ltd., Al. Jerozolimskie 181, 02-222 Warsaw, Poland and Warsaw School of Economics, Al. Niepodleglosci 162, 02-554 Warsaw, Poland. Tomasz Ząbkowski is with Polska Telefonia Cyfrowa Ltd.. Krzysztof Siwek is with the Institute of Theory of Electrical Engineering, Measurement and Information Systems Information Systems, Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warszawa, Poland, e-mail: ksiwek@iem.pw.edu.pl ,{rszupiluk/pwojewnik/tzabkowski}@era.pl

$$\Delta \mathbf{W}(t) = \mu(t)\left[\mathbf{I} - \mathbf{f}(\mathbf{y}(t))\mathbf{g}^T(\mathbf{y}(t))\right]\mathbf{W}(t) \qquad (10)$$

with adaptive choice of non-linearities [2]

$$f_i(y_i) = \begin{cases} \tanh(\beta_i y_i) & \text{for } \kappa_4(y_i) > \delta, \\ \operatorname{sign}(y_i)\,|y_i|^{r_i - 1} & \text{for others,} \end{cases}$$

and

$$g_i(y_i) = \begin{cases} \operatorname{sign}(y_i)\,|y_i|^{r_i - 1} & \text{for } \kappa_4(y_i) > -\delta, \\ \tanh(\beta_i y_i) & \text{for others,} \end{cases}$$

where $r_i \geq 2$, $\delta \geq 0$.

Term $\kappa_4(y_i)$ means normalized kurtosis:

$$\kappa_4(y_i) = E\{y_i^4\}/E^2\{y_i^2\} - 3 \qquad (11)$$

The main problem with nonlinearities choice based on kurtosis is fact that not always the statistics do exist. There are many distributions without higher order moments as e.g. the family of α-stable distributions, which do not posses statistics of order higher than second.

II. Using Nonlinearities for the α-Stable Distributions

The α-stable distributions can be defined as distributions that satisfied stability property what means that if x_1, x_2, x_3 are independent α-stable random variables

One of the main problems connected with the α-stable distributions is the lack of the close expressed form for their pdf with only small number of exceptions (Cauchy, Pareto, Gaussian). In general case the distribution is given by characteristic function of the form:

$$E(e^{itZ}) = \begin{cases} \exp\{-\sigma^{\alpha}|t|^{\alpha}(1 - i\beta(\operatorname{sign}(t)\cdot\tan\frac{\pi\alpha}{2}) + i\mu t\} \\ \quad \text{if } \alpha \neq 1 \\ \exp\{-\sigma|t|(1 + \frac{2i\beta}{\pi}(\operatorname{sign}(t)\cdot\ln|t|) + i\mu t\} \\ \quad \text{if } \alpha = 1 \end{cases} \qquad (12)$$

where $0 < \alpha \leq 2$ describes “thickness” of the tails, $\sigma > 0$ - dispersion, $-1 \leq \beta \leq 1$ - symmetry and $\mu \in \Re$ - location. The α parameter is crucial for description of the α-stable distributions, because for given distribution there are no moments higher than α. If $\beta = 0$, then distribution is called symmetric α-stable ($S\alpha S$), and this kind of pdfs will be considered in this article.

Using expansion of the pdfs into series we can estimate the shape of the nonlinearities for functions with various α (Fig. 1a-c).

Based on numerical simulation we propose a general form of nonlinearities for α-stable distributions as:

$$f_i(y_i) = \frac{-2 \cdot y_i}{\theta \cdot \sigma^2 + y_i^2 \cdot (2 - \alpha)} \qquad (13)$$

where typically $\theta = \alpha$ or $\theta = 1$. Example theoretical nonlinearities for $\alpha \in \{0{,}5; 1{,}0; 1{,}5\}$ are shown in Fig. 1d-f.

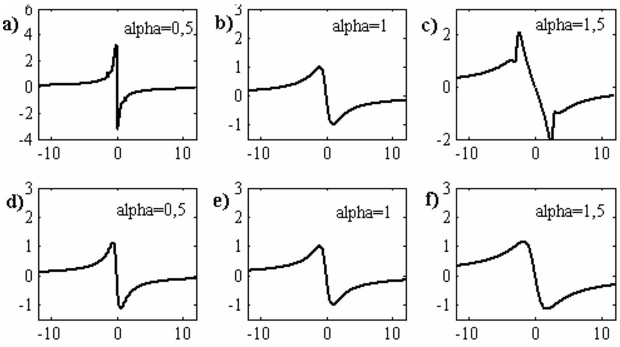

Fig. 1. Nonlinearities obtained empirically (a-c) and theoretically (d-f) for different α

It should be noted that there exist the family of alpha-stable distributions, called sub-gaussian alpha stable distributions, that are always mutually dependent. For such distributions our methods is not addressed of course.

III. Alpha Parameter Estimation

The crucial thing in our considerations is proper alpha parameter estimation. There are many method of estimation of alpha parameter. For example we can utilize following methods for our purpose:

1. For Symmetric Alpha Stable ($S\alpha S$) distributions we can easy compute α from

$$\sigma_z^2 = \frac{\pi^2}{6}\left(\alpha_x^{-2} - 0.5\right), \qquad (14)$$

where σ_z^2 is second moment of $z = \ln(|x|)$. For adaptive learning we can introduce following on-line rule

$$\alpha_x(t) = \sqrt{\frac{6}{\pi^2}\left[(1-\eta)E\{z^2(t-1)\} + \eta\,|z|^2\right] + 5} \qquad (15)$$

2. For signals witch can be treated as fractional Brownian Motion we have

$$\alpha = \frac{1}{H} \qquad (16)$$

where H is Hurst (or Holder) exponent.

IV. Algorithm for Finding H-exponention

For finding the H exponent Hurst proposed the R/S method. Let’s consider a series of N observations. The algorithm is as follows:

1. Divide the series of observations into d sub-series with n observations, where $d \times n = N$;

2. For every series $m = 1,...,d$:
   a. Find the mean E_m and the standard deviation S_m ;
   b. Rescale the observations $Z_{i,m}$ by subtracting the E_m : $X_{i,m} = Z_{i,m} - E_m$;
   c. Cumulate the observations: $Y_{i,m} = \sum_{j=1}^{i} X_{j,m}$ for $i = 1,...,n$;
   d. Calculate the range: $R_m = \max\{Y_{1,m},...,Y_{n,m}\} - \min\{Y_{1,m},...,Y_{n,m}\}$;
   e. Rescale the range by standard deviation R_m / S_m ;
3. For d sub-series of the length n find the mean $(R/S)_n = \frac{1}{d} \cdot \sum_{m=1}^{d} R_m / S_m$ as an estimate for $E(R/S)_n$;
4. Given $(R/S)_n$ for different n you will find H as the coefficient of the following regression:

$$\ln E(R/S)_n = \ln c + H \cdot \ln n . \qquad (17)$$

V. Computer Simulations

In this section we confirm validity of above described method by computer simulation. We mix four computer generated alpha stable signals with $\alpha = 1.5$ by arbitrary matrix **A**. The all the elements of mixing matrix are drawn form uniform distribution (-1,1). Both signals and matrix are assumed to be unknown. To measure of the performance of our method we use performance index defined as

$$\mathrm{PI} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left\{ \left(\sum_{k=1}^{n} \frac{|g_{ik}|}{\max_j |g_{ij}|} - 1 \right) + \left(\sum_{k=1}^{n} \frac{|g_{ki}|}{\max_j |g_{ji}|} - 1 \right) \right\}$$

where g_{ij} is the (i,j) element of global system matrix **G**=**WA**. When performance index is zero the perfect estimation is achieved.

Example of obtained results are presented in Fig.2.

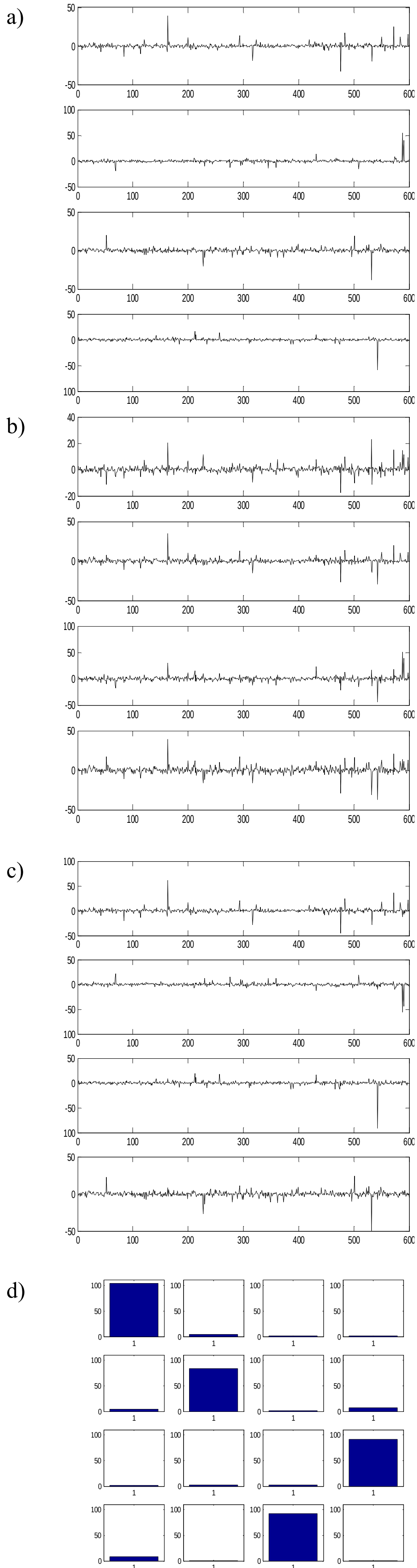

Fig. 2 Source a), mixed b), estimated c) signals and performance matrix **G** of considered example.

VI. CONCLUSIONS

The application of the nonlinearities in the form presented above allows us for practical adaptation for wide range of alpha-stable distributions. The computer simulations confirmed the efficacy. One of the key issues is estimation improvement of the alpha parameter. In this work the most popular estimators of alpha were used. It is still the open question the way of conducting the preprocessing stage in case the extreme values. In such situation the most appropriate approach is to eliminate these samples with regard to numerical stability of the algorithms.

REFERENCES

[1] Amari S.: Natural gradient works efficiently in learning, Neural Computation, No 10, pp. 271-276, 1998.
[2] Cichocki A., Amari S.: Adaptive Blind Signal and Image Processing, John Wiley, Chichester 2002.
[3] Cichocki A., Sabala I., Choi S., Orsier B., Szupiluk R.: Self adaptive independent component analysis for sub-Gaussian and super-Gaussian mixtures with unknown number of sources and additive noise, (NOLTA-97), vol. 2, Hawaii, U.S.A. 1997, pp. 731-734.
[4] Comon P.: Independent component analysis, a new concept? Signal Processing, Elsevier 1994, No 3, vol. 36, pp. 287-314.
[5] Hyvärinen A., Karhunen J., Oja E.: Independent Component Analysis, John Wiley 2001.
[6] Nikias C.L., Shao M.: Signal Processing with Alpha-Stable Distributions and Applications, John Wiley&Son, 1995.
[7] Samorodnitskij G., Taqqu M. S.: Stable non-Gaussian random processes: stochastic models with infinitive variance, New York, London, Chapman and Hall, 1994.
[8] Kidmose P., Independent Component Analysis Using the Spectral Measure for Alpha-Stable Distributions [0.4Mb], Proceedings of IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing, NSIP2001, http://isp.imm.dtu.dk/staff/kidmose/documents/kidmose01-1.pdf
[9] Cruces S., Cichocki A., Castedo L., An iterative inversion method for blind source separation, Proceedings of the 1st International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'99), Assois, France, January 1999, pp. 307-312.
[10] Hurst H.E.: Long term storage capacity of reservoirs, Trans. Am. Soc. Civil Engineers 116, pp. 770-799.