ORIGINAL RESEARCH

Parametric independent component analysis for stable distributions

Mohammad Reza Ameri [1, 2], Mona Shokripour[3], Adel Mohammadpour[2], Vahid Nassiri [2, 4]

1. Department of Computer Science and Software Engineering, Concordia University, Montreal, Canada. 2. Faculty of Mathematics and Computer Science, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. 3. Department of Statistics, Shahid Beheshti University, Tehran, Iran. 4. Department of Mathematics, Vrije Universiteit Brussels (VUB), Brussels, Belgium

Correspondence: Adel Mohammadpour. Address: Faculty of Mathematics and Computer Science, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. Email: adel@aut.ac.ir.

Received: July 22, 2012 **Accepted:** October 23, 2012 **Online Published:** March 21, 2013
DOI: 10.5430/air.v2n3p27 **URL:** http://dx.doi.org/10.5430/air.v2n3p27

Abstract

Independent Component Analysis (ICA) is a method for blind source separation of a multivariate dataset that transforms observations to new statistically independent linear forms. Infinite variance of non-Gaussian α-stable distributions makes algorithms like ICA non-appropriate for these distributions. In this note, we propose an algorithm which computes mixing matrix of ICA, in a parametric subclass of α-stable distributions.

Key words

Blind Source Separation, Independent Component Analysis, Stable Distribution.

1 Introduction

Independent component analysis is a solution for blind source separation problem. The goal of Independent Component Analysis (ICA) is to find a linear transformation of multivariate data such as random vectors such that its components becomes statistically independent. Independent components also are called sources and input vectors are known as observations.

If we have a vector of observation $\mathbf{X} = \left(X_1,...,X_M\right)^T$ and a source vector $\mathbf{Y} = \left(Y_1,...,Y_N\right)^T$ the ICA model follows as: $\mathbf{X} = A\mathbf{Y}$, where $A_{M\times N}$ is a mixing matrix. The de-mixing process needs to invert A. In most of the literature, it's supposed the number of mixtures and sources are equal so the ICA can be solved by:

$$\mathbf{Y} = W\hat{\mathbf{X}} = WA\mathbf{Y} = A\hat{A}\mathbf{Y}$$

W is usually an estimation of A^{-1} to make the components as independent as possible. Readers are referred to [1-4] and their references for more details on ICA.

Considering stable random vectors as an input for ICA requires a new assumption rather than Central Limit Theorem that says the standardized sum of Independent and Identically Distributed (IID) random variables converge to a random variable with Gaussian distribution, but Generalized Central Limit Theorem informally states that a normalized sum of a sequence of IID random variables with infinitive variance converges to a non-Gaussian stable random variable [5].

In this work, we consider non-Gaussian stable sources and propose a parametric ICA as an especial case of Kidmose's suggestion described in [6, 7]. Sahmodi et al. [8] introduced a BSS method for the symmetric class of stable distributions. Extension of [8] to the case of random matrix A is given in [9] by a semi-parametric approach.

In section 2 we are introducing stable distributions and an ICA algorithm is proposed in section 3. Simulation results and comparisons are given in section 4. The paper is concluded in section 5.

2 Stable distributions

A random variable X is said to be a stable (α -stable) random variable if its characteristic function, $\Phi_X(t)$ has the following form:

$$\Phi_X(t) = \begin{cases} \exp\left(-\gamma^\alpha |t|^\alpha \left(1 - i\beta \operatorname{sign}(t) \tan\frac{\alpha\pi}{2}\right) + it\mu\right) & \alpha \neq 1 \\ \exp\left(-\gamma|t|\left(1 + i\frac{2}{\pi}\beta \operatorname{sign}(t)\ln|t|\right) + it\mu\right) & \alpha = 1 \end{cases} \quad (1)$$

where $\operatorname{sign}(u) = -1, 0, 1$ if $u <=$, or > 0 , respectively. Univariate stable distributions are parameterized by four parameters, $0 < \alpha < 2$, index of stability, $\gamma \geq 0$, scale, $-1 \leq \beta \leq 1$, skewness, and μ , location. Since stable distributions characterized by these four parameters, it's denoted by $S(\alpha, \beta, \gamma, \mu)$ and we will write $X \sim S(\alpha, \beta, \gamma, \mu)$ to say that a random variable X has an α -stable distribution.

A random vector $\mathbf{X} = (X_1, ..., X_M)^T$ is said to be stable if there are parameters $\boldsymbol{\mu} = (\mu_1, ..., \mu_M)^T \in \mathbb{R}^M$, and a finite measure Γ , called the spectral measure, on the unit sphere $S_M = \left\{\mathbf{s} = (s_1, ..., s_M)^T \middle| \mathbf{s}^T\mathbf{s} = 1\right\}$ such that:

$$\Phi_{\mathbf{X}}(\mathbf{t}) = \begin{cases} \exp\left(-\int_{S_M} |\mathbf{t}^T\mathbf{s}|^\alpha \left(1 - i\operatorname{sign}(\mathbf{t}^T\mathbf{s}) \tan\frac{\pi\alpha}{2}\right)\Gamma_{\mathbf{X}}(d\mathbf{s}) + i\mathbf{t}^T\boldsymbol{\mu}\right) & \alpha \neq 1 \\ \exp\left(-\int_{S_M} |\mathbf{t}^T\mathbf{s}|\left(1 + i\left(\frac{2}{\pi}\right)\operatorname{sign}(\mathbf{t}^T\mathbf{s})\ln(|\mathbf{t}^T\mathbf{s}|)\right)\Gamma_{\mathbf{X}}(d\mathbf{s}) + i\mathbf{t}^T\boldsymbol{\mu}\right) & \alpha = 1 \end{cases} \quad (2)$$

where $\mathbf{t} = (t_1, ..., t_M)^T$.

Multivariate stable distributions are semi-parametric and can be identified uniquely by the pair of $(\boldsymbol{\mu}, \Gamma)$. A parametric subclass of stable distributions is stable random vectors with discrete spectral measure. A discrete spectral measure of a M -dimensional stable random vector $\mathbf{X}$, with N pair of directions $\{\mathbf{s}_n, -\mathbf{s}_n\}, n = 1, ..., N$, can be considered as follows:

$$\Gamma_{\mathbf{X}}(.) = \sum_{n=1}^{N} \left(\frac{1+\beta_n}{2}\gamma_n^\alpha \delta_{\{-\mathbf{s}_n\}}(.) + \frac{1-\beta_n}{2}\gamma_n^\alpha \delta_{\{\mathbf{s}_n\}}(.)\right) \quad (3)$$

where $\beta_n = \left(\Gamma(\mathbf{s}_n) - \Gamma(-\mathbf{s}_n)\right) \Big/ \left(\Gamma(\mathbf{s}_n) + \Gamma(-\mathbf{s}_n)\right)$, $\left(1 \pm \beta_n / 2\right)\gamma_n^\alpha$'s are the weights, and $\delta_{\{\mathbf{s}_n\}}$'s are point masses (Dirac measures of mass 1) at the points $\mathbf{s}_n \in S_M$. The characteristic function (2) with the spectral measure (3) reduces to:

$$\Phi_{\mathbf{X}}(\mathbf{t}) = \begin{cases} \exp\left\{-\sum_{n=1}^{N} \gamma_n^\alpha \left|\mathbf{t}^T \mathbf{s}_n\right|^\alpha \left(1 - i\beta_n \operatorname{sign}\left(\mathbf{t}^T \mathbf{s}_n\right) \tan\frac{\pi\alpha}{2}\right) + i\mathbf{t}^T \boldsymbol{\mu}\right\} & \alpha \neq 1 \\ \exp\left\{-\sum_{n=1}^{N} \gamma_n \left|\mathbf{t}^T \mathbf{s}_n\right| \left(1 + i\left(\frac{2}{\pi}\right)\beta_n \operatorname{sign}\left(\mathbf{t}^T \mathbf{s}_n\right) \ln\left(\left|\mathbf{t}^T \mathbf{s}_n\right|\right)\right) + i\mathbf{t}^T \boldsymbol{\mu}\right\} & \alpha = 1 \end{cases} \quad (4)$$

A M-dimensional stable random vector $\mathbf{X}$ has independent component if and only if its spectral measure is discrete and concentrated on the intersection of the axes with the unit sphere S_M. For more information on stable distributions, the reader can refer to [10].

3 Parametric ICA algorithm for stable distributions

Using the next theorem, that will be proved in the Appendix, we propose a blind source separation method for de-mixing the observation to independent components. The converse of this theorem is also true [6, 10].

Theorem: Let $\mathbf{X} = \left(X_1, ..., X_M\right)^T$ be a stable random vector with discrete spectral measure (3), location parameter $\boldsymbol{\mu} \in \mathbb{R}^M$ and $Y_n \sim S\left(\alpha, \beta_n, 1, 0\right)$, $n = 1, ..., N$ be independent random variables. Then $\mathbf{X}$ can be decomposed into $\mathbf{X} = A\mathbf{Y} + \mathbf{b}$, where $\mathbf{Y} = (Y_1, ..., Y_N)^T$ and $\mathbf{b}$ defined in (5). $A_{M \times N}$ is the mixing matrix with $\mathbf{a}_{.n} = \gamma_n \mathbf{s}_n$, $n = 1, ..., N$ as its columns,

$$\mathbf{b} = \boldsymbol{\mu} + \delta_{\{1\}}(\alpha)\frac{2}{\pi} A\mathbf{c} \text{ and } \mathbf{c} = (\beta_1 \ln(\gamma_1), ..., \beta_N \ln(\gamma_N))^T. \quad (5)$$

So in the case where the number of sources $N = M$ the de-mixing matrix yields by inverting A, to get an exact solution for ICA. In the other word, if the dimension of an observation vector is equal to the number of directions of the spectral measure then we can use the following parametric ICA Algorithm to recover independent components.

Algorithm 1:

Input: A random sample of M dimensional stable vector $\mathbf{X}: \mathbf{X}_1, ..., \mathbf{X}_L$.

Output: A random sample of N dimensional stable vector with independent components $\widehat{\mathbf{Y}}: \widehat{\mathbf{Y}}_1, ..., \widehat{\mathbf{Y}}_L$.

1) Estimate spectral measure of $\mathbf{X}$ and its location parameter by the random samples $\mathbf{X}_1, ..., \mathbf{X}_L$.

2) Construct the matrix $\widehat{A}_{M \times N}$ using N larger estimated directions of the spectral measure with maximum weights.

3) Calculate $\mathbf{b}$ in (5) and then calculate

$$\widehat{\mathbf{Y}}_l = \widehat{A}^{-1}(\mathbf{X}_l - \mathbf{b}), l = 1, ..., L. \quad (6)$$

The theorem and the proposed algorithm in this section, actually is a precise description and complement of the Kidmose's works [6, 7]. Although it may not be straightforward to have a precise theoretical solution in over-complete and under-complete case but we can do this in a way that compute independent components by choosing an arbitrary number of spectral measure point masses.

The under-complete $M > N$, mixing matrix A is produced by degenerate vector $\mathbf{X}$. The spectral measure of a degenerate M dimensional stable distribution concentrates on S_N instead of S_M [10]. In this situation, we first could find the degenerate components of $\mathbf{X}$, and then eliminate M-N degenerate components. Finally, independent sources can be recover through Algorithm I.

In the over-complete case, $M < N$, we can compute several inverses, one possible choice is to use generalized inverse, Moore–Penrose pseudoinverse [11]. The algorithm would be the same as above but because we do not know the number of source components we can select N (the number of point masses of the spectral measure) based on a threshold level on the mass of the spectral measures. Then we choose a larger N point masses of the spectral measure. If the number of larger point masses is greater than M, it is necessary to use pseudoinverse in Algorithm 1.

4 Simulation results

In this section, we compare the precision of the proposed algorithm through three simulation studies. In the first one, the blind source separation results of the proposed algorithm are compared with some well-known ICA algorithms. To do the simulation we start with making independent components by generating independent α-stable random variables. A bivariate source vector with two independent components has the following parameters in two cases 1) symmetric, $\beta = 0,$ and 2) asymmetric, $\beta = 1$. That is, the source vector $\mathbf{Y}$ has components $Y_n \sim S(\alpha, \beta, 1, 0)$, $n = 1, 2$. We also consider α varying in the range of $1 \le \alpha \le 1.9$, for each case. With these parameters then we generate 10000 samples of our desired vector. Having $\mathbf{Y}$ as sources in BSS problem, we mixed them in following manner $\mathbf{X} = A\mathbf{Y}$, where $A = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}$. $\mathbf{X}$ is the mixed source we used as input for the algorithm. The mixing matrix A could be any nonzero matrix. We choose this A because it does not scale the data; however it is possible to do a simulation for non-normalized mixing matrix. We show the estimated source vectors (of BSS problem) by $\widehat{\mathbf{Y}}_1, ..., \widehat{\mathbf{Y}}_L$. For generating stable random vectors and estimating the spectral measure of random vectors the methods in [12, 13] are used, respectively. The algorithms that are compared, are FastICA [14], minimizing mutual information [15], ComonICA [3], JADE [16], Temporal Predictability ICA [17] and Infomax [18]. We compute two sample errors for each simulation. The first one is the sum of absolute value of the distance between the original source and the estimated one and the other is the first one raised by the power of $(\alpha - 0.1)$. These criteria are convergent for stable distributions with an index of stability greater than 1 and $(\alpha - 0.1)$, respectively. Two criteria for computing errors we used formulates as follows:

$$E_{Abs} = \frac{1}{NL}\sum_{n=1}^{N}\sum_{l=1}^{L}\left|Y_{nl} - \widehat{Y}_{nl}\right|,$$

$$E_{\alpha-0.1} = \frac{1}{NL}\sum_{n=1}^{N}\sum_{l=1}^{L}\left(\left|Y_{nl} - \widehat{Y}_{nl}\right|\right)^{(\alpha-0.1)}.$$

For each set of parameters (symmetric and asymmetric with ten different α ranged in $1 < \alpha < 1.9$) we simulate experiment 10000 times. Then we compute the errors which are defined (N=2) and the average value of errors. Figure 1 – Figure 4 are comparisons between our algorithm and the other methods. Figure 1 and Figure 2 illustrates the average value of errors in the symmetric case β = 0 and ten different α value ($1 < \alpha < 1.9$) for EAbs and Eα-0.1, respectively. Fig. 3 and

Figure 4 shows the same comparison for asymmetric case, $\beta = 1$, for EAbs and Eα-0.1, respectively. The results of our proposed algorithm are called myStable in the figures which have the smallest value of errors compared with other methods.

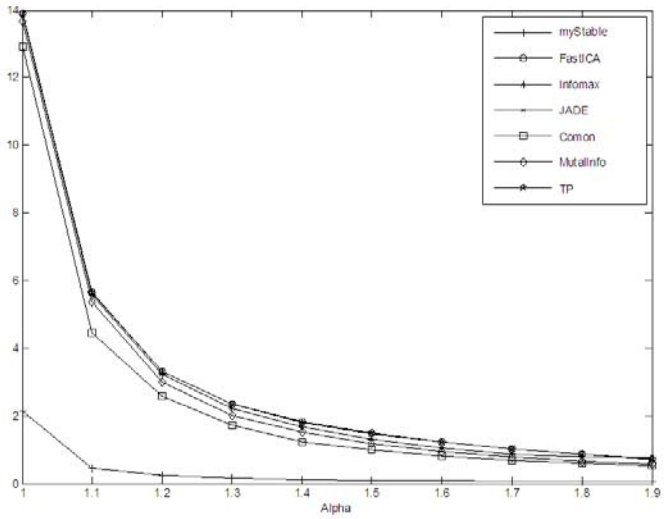

Figure 1. Symmetric case ($\beta = 0$ and $1<\alpha<1.9$): Showing the average value of E_{Abs} error-computed in 10000 simulations. It compares performance of different algorithms, our algorithm labeled as mystable

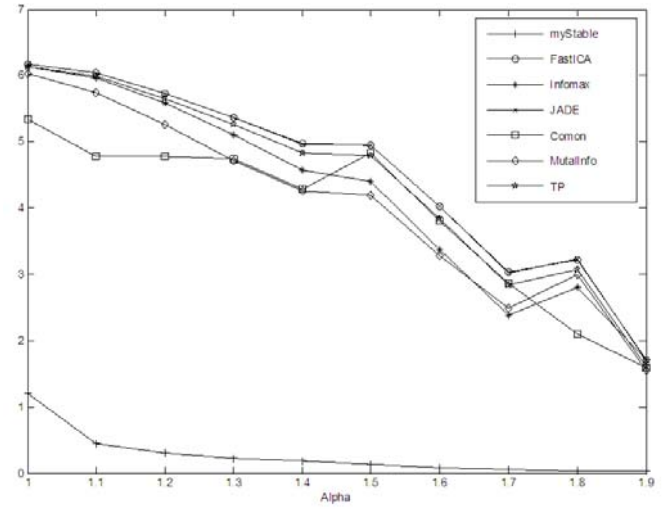

Figure 2. Symmetric case ($\beta = 0$ and $1<\alpha<1.9$): Showing the average value of Eα-0.1error computed in 10000 simulations. It compares performance of different algorithms, our algorithm labeled as mystable

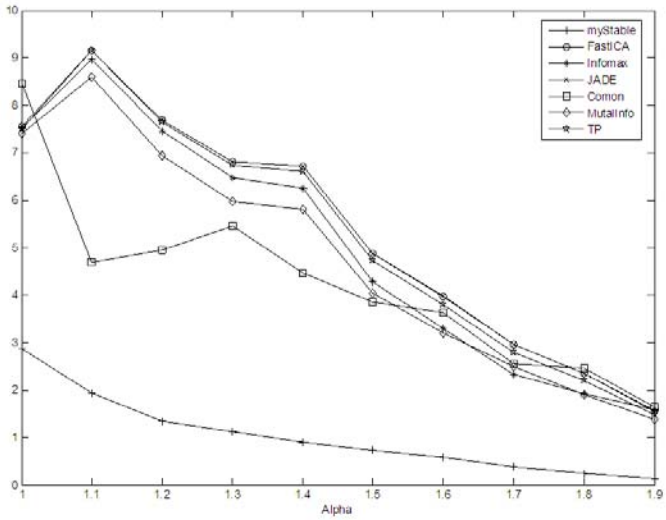

Figure 3. Asymmetric case ($\beta = 1$and $1<\alpha<1.9$): Showing the average value of E_{Abs} error computed in 10000 simulations. compares performance of different algorithms, our algorithm labeled as mystable

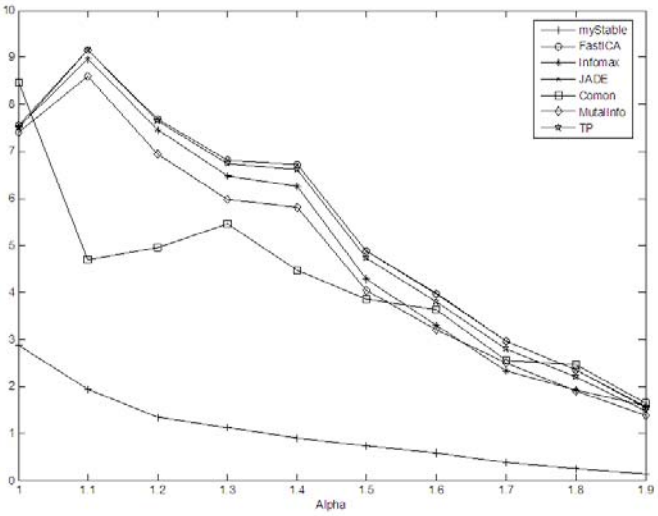

Figure 4. Asymmetric case (β = 1and 1<α<1.9): Showing the average value of $E_{\alpha\text{-}0.1}$ error computed in 10000 simulations. It compares performance of different algorithms, our algorithm labeled as mystable

The second simulation experiment dedicates to the ICA problem of speech observation in chapter 6 of [15], called 'speech_1' and 'speech_2'; see the first row of Fig. 5. The observations are added to two independent simulated stable distributions with parameters $S\left(1.5,0,1,0\right)$. Then, they mixed by the matrix A which is defined above. These signals are plotted in rows 2 and 3 of Fig. 5, respectively. ICA results using the proposed method (myStable) and FastICA are plotted in the two last rows of Fig. 5. The absolute errors are 1.3397 and 0.0569, respectively.

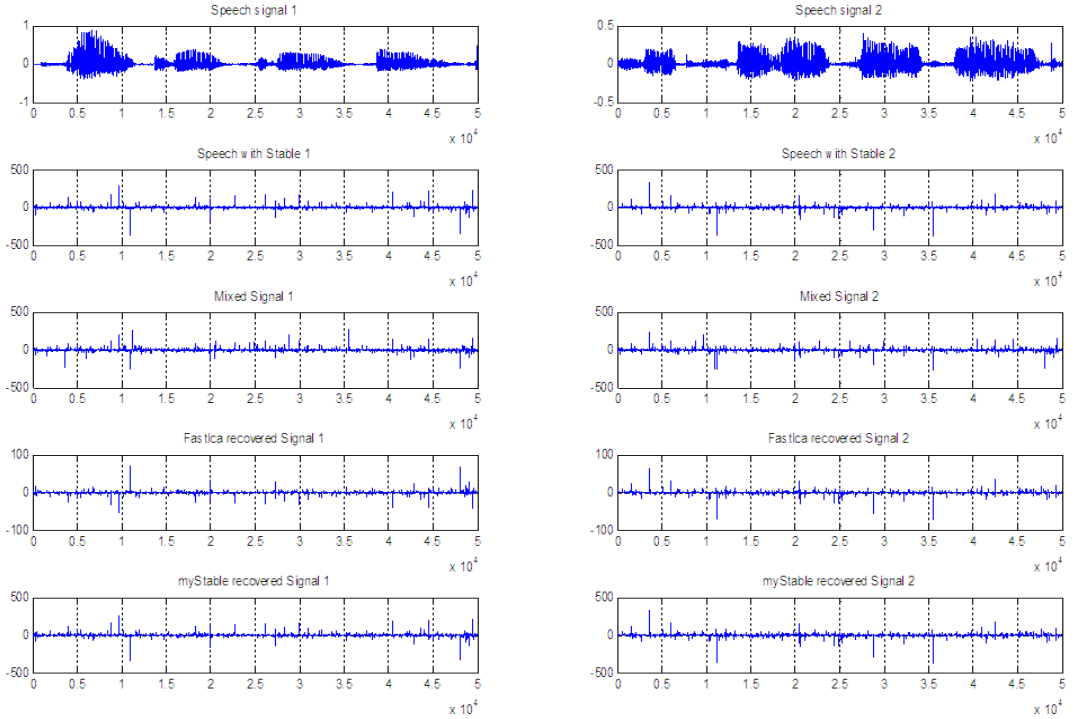

Figure 5. First row speech signals, Second row speech signals added to two independent simulated stables, third row mixed signals, fourth row FastICA recovered signals, last row myStable recovered signals

The third simulation study refers to over-complete case where the dimension of observations is smaller than the dimension of sources. In this situation, we consider three sources with two dimensional observations with the mixing matrix

$A = \begin{bmatrix} \sqrt{2}/2 & 0 & -1 \\ \sqrt{2}/2 & 1 & 0 \end{bmatrix}$. The average and standard deviation (sd) of E_{Abs} for 10000 iterations are computed for three different indices of stability and it shows in Table 1 for symmetric and asymmetric cases. Since, the well known mentioned ICA methods are not appropriate in this situation, we only consider the proposed method. In comparison with the case of M=N it has almost two times error, although these errors are reduced when the index of stability increase.

Table 1. The average and sd of E_{Abs} for three different values of α in the symmetric and asymmetric case.

β	$\alpha = 1.1$		$\alpha = 1.5$		$\alpha = 1.9$	
	average	**sd**	**average**	**sd**	**average**	**sd**
$\beta = 0$	12.505	1.494	4.421	0.588	2.857	0.498
$\beta = 1$	11.684	2.126	2.959	0.376	2.303	0.269

5 Conclusions

A parametric ICA for stable observations was proposed. An advantage of the proposed method is its precision and its disadvantage is a limitation on the number of point masses of the spectral measure of stable observed data, (i.e., M=N in (3)).This limitation can be relaxed [9] which is a modified semi-parametric principal component analysis of stable distributions to distinguish more important components. We proposed an alternative parametric approach to distinguish important components for stable distributions based on spectral measure.

Acknowledgment

The authors would like to thank the editors and the anonymous reviewers for their valuable comments and suggestions to improve the manuscript.

References

[1] Cichocki, S.-I. Amari, Adaptive Signal and Image Processing: Learning Algorithms and Applications, Wiley, New York, 2002.
[2] Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley-Interscience, New York, 2001. http://dx.doi.org/10.1002/0471221317
[3] P. Comon, Independent Component Analysis, a new concept? Signal Processing. 1994; 36(3): 287–314. http://dx.doi.org/10.1016/0165-1684(94)90029-9
[4] Z. Koldovský, J. Málek, P. Tichavský, Y. Deville, S. Hosseini, Blind Separation of Piecewise Stationary Non-Gaussian Sources, Signal Processing. 2009; 89: 2570-2584. http://dx.doi.org/10.1016/j.sigpro.2009.04.021
[5] J. P. Nolan, Stable Distributions - Models for Heavy Tailed Data, Birkhauser, Boston, 2010.
[6] P. Kidmose, Blind separation of heavy tail signals, Ph.D. thesis, Department of Informatics and Mathematical Modelling. Technical University of Denmark, 2001.
[7] P. Kidmose, Independent Component Analysis Using the Spectral Measure for Alpha-Stable Distributions, in: Proceedings of IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing, 2001.
[8] M. Sahmoudi, K. Abed-Meraim, M. Benidir, Blind Separation of Impulsive α-stable Sources Using Minimum Dispersion Criterion, IEEE Signal Processing Letters. 2005; 12: 281-284 http://dx.doi.org/10.1109/LSP.2005.843771
[9] M. Sahmoudi, K. Abed-Meraim, M. Benidir, Blind Separation of Heavy-Tailed Signals Using Normalized Statistics, in: Proceeding of ICA 2004, Granada, Spain, September 2004.
[10] G. Samorodnitsky, M. S. Taqqu, Stable non-Gaussian Random Processes, Chapman and Hall, New York, 1994.
[11] R. A. Penrose, A generalized inverse for matrices, in: Proceedings of Cambridge Philosophy Society. 1955: 406-413.
[12] R. Modarres, and J. P. Nolan, A Method for Simulating Stable Random Vectors, Computational Statistics. 1994; 9(1): 11.
[13] J. P. Nolan, A. K. Panorska, and J. H. McCulloch, Estimation of stable spectral measures, Mathematical and Computer Modelling. 2001; 34(9-11): 1113-1122. http://dx.doi.org/10.1016/S0895-7177(01)00119-4

[14] Hyvärinen, and E. Oja. A fast fixed-point algorithm for independent component analysis, Neural Computation. 1997; 9(7): 1483-1492. http://dx.doi.org/10.1162/neco.1997.9.7.1483

[15] F. Wang, H. Li, and R. Li, Data Mining with Independent Component Analysis, Springer-Verlag, Berlin, 2006.

[16] J. F. Cardoso, High-order contrasts for independent component analysis, Neural Computation. 1999; 11(1): 157-92. PMid:9950728 http://dx.doi.org/10.1162/089976699300016863

[17] J. V. Stone, Blind Source Separation Using Temporal Predictability, Neural Computation. 2001; 13(7): 1559-74. PMid:11440597 http://dx.doi.org/10.1162/089976601750265009

[18] J. Bell, and T. J. Sejnowski. An Information-Maximization Approach to Blind Separation and Blind Deconvolution, Neural Computation.1995; 7 (6): 1129-1159. PMid:7584893 http://dx.doi.org/10.1162/neco.1995.7.6.1129