BLIND SOURCE SEPARATION USING A SEMI-PARAMETRIC APPROACH WITH APPLICATION TO HEAVY-TAILED SIGNALS

M. Sahmoudi[1], *K. Abed-Meraim*[2], *M. Lavielle*[3], *E. Kunth*[3] *and Ph. Ciblat*[2]

[1] LSS, CNRS-SUPELEC-UPS, 91192 Gif-sur-Yvette, France
[2]TSI, Telecom-Paris, 37-39, rue Dareau, 75014 Paris, France
[3] Laboratoire de Mathématiques, Paris-Sud University, France

ABSTRACT

In this paper, we propose a new semi-parametric approach for blind source separation (BSS) with application to heavy-tailed signals. The semi-parametric statistical principle is used to formulate the BSS problem as a maximum likelihood (ML) estimation. More precisely, this approach consists of combining the log-spline model for sources density approximation with a stochastic version of the EM algorithm for mixing matrix estimation. The proposed method is truly blind to the particular underlying distribution of the mixed signals and performs simultaneously the estimation of the unknown probability density functions (pdf) of the source signals and the estimation of the mixing matrix. In addition, it is robust against outliers and impulsive effect. To illustrate that, the new method is compared with a previously proposed parametric approach for heavy-tailed signals which is based on alpha-stable parametric model and on the minimum dispersion (MD) criterion.

1. INTRODUCTION

Blind source separation is a rapidly developing technology that attracts a lot of attention in the signal processing literature [6]. Applications can be found in a variety of fields, e.g. multiuser multi-access communications, speech processing, bioengineering and seismology. In order to perform the BSS task, a measure of the amount of independence between signals is required. Many measures of independence exist, based on which, different algorithms have been proposed to solve the BSS problem [6]. A very popular parametric approach for estimating the BSS model is the maximum likelihood method. It is not difficult to derive the likelihood function using a parametric model of source densities. However, the distribution model mismatch between the output pdf and the chosen underlying distribution model is a serious problem in such approaches. Incorrect assumptions on the source distributions can result in poor estimation performance or in a complete failure to achieve the source separation. This issue does not arise in the case of cumulant maximization based solutions [4]. However, these approaches usually rely exclusively on third or fourth order cross-cumulants in order to measure independence, and represent just an approximation of the mutual information minimization principle [6]. Alternative methods that employ a non-parametric density estimation have been introduced in [3]. These methods usually consist in a density estimation technique that alternates with a cost function optimization step in an iterative approximation framework. Although these approaches do not require the definition of a specific model for the density functions, neither their convergence properties, nor their capability of separation arbitrarily distributed sources, have been fully assessed. Thus, finding a compromise between computational complexity, performances and the robustness to the source's pdf mismatch in a blind signal separation framework is still an open and challenging problem. In this paper, we propose a new semi-parametric BSS method using the ML approach and an approximation of the sources densities by the log-spline function. This method can be applied in particular to heavy-tailed signals. Moreover, to assess the problem of parametric versus non-parametric modelling, we compare the proposed method with the (parametric) minimum dispersion based method and discuss the robustness of the two methods against modelization errors.

THIS WORK WAS IN PART SUPPORTED BY THE CNRS WITHIN A MATHSTIC PROJECT ENTITLED 'DETECTION AND ESTIMATION TECHNIQUES FOR IMPULSIVE SIGNALS'.

2. PROBLEM FORMULATION

2.1. Linear instantaneous mixtures

In this paper, we consider the classical linear BSS model with instantaneous mixtures given by:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\varepsilon}(t), \quad t = 1 \dots T \tag{1}$$

where $\mathbf{A}$ is a $n \times m$ *unknown* full column rank mixing matrix. The sources $s_1(t), \cdots, s_m(t)$ are collected in a $m \times 1$ vector denoted $\mathbf{s}(t)$ and are assumed to be mutually independent identically distributed (i.i.d) under the common distribution density π. We also suppose that the components $\varepsilon_1(t), \dots, \varepsilon_n(t)$ of the noise vector $\boldsymbol{\varepsilon}(t)$ are independent and Gaussian distributed with zero mean and unknown variance σ^2.[1]

The goal of a BSS method is to find a *separating matrix* i.e. an $m \times n$ matrix $\mathbf{B}$ such that the recovered sources are as independent as possible. Model (1) admits a unique solution up to scaling and permutation indeterminacy $\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t)$ such that $\mathbf{C} \triangleq \mathbf{B}\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda}$ is a diagonal scaling matrix and $\mathbf{P}$ is a permutation matrix (see [6]). At most one source is allowed to be Gaussian to ensure the identifiability.

[1]The assumptions of a common source pdf and gaussian noise pdf can be relaxed easily at the expense of extra computational cost (in the general case, we would have to approximate m different source pdfs instead of one single pdf common to all sources).

2.2. Heavy-tailed α-stable sources

Recent experimental measurements have demonstrated that many signals are decidedly non-Gaussian due mostly to impulsive phenomena (see [1] and references therein). It has been shown that impulsive signals are heavy-tailed in nature and can be better modeled by using distributions with algebraic tails rather than Gaussian or other exponentially tailed distributions [1]. A popular member of the class of heavy-tailed distributions is the α-stable distribution. Recently, the alpha-stable statistical model of heavy-tailed signals has been proposed for signal processing applications [1]. The class of symmetric alpha-stable (SαS) distributions is best defined by its characteristic function $\varphi(\omega) = \exp(j\mu\omega - \gamma|\omega|^{\alpha})$, where $\alpha \in]0;2]$ is the *characteristic exponent* that determines the shape of the distribution. The smaller α is, the heavier the tails of the α-stable density. $\mu \in \mathbb{R}$ is the *location parameter*, and $\gamma > 0$ is the *dispersion* index that determines the spread of the distribution around its location parameter μ. No closed-form expressions exist for α-stable density other than the cases of $\alpha = 2$ (Gaussian distribution), $\alpha = 1$ (Cauchy distribution) and of $\alpha = 1/2$ (Levy distribution). Alpha-stable densities obey three important properties which further justify their role in data modeling.

Proposition 1
1) ***Stability***: *A weighted sum of independent α-stable random variables is α-stable with the same α.*
2) ***Generalized central limit theorem (GCLT)***: *Without limitation of finite variance, stable models are the only distribution that can be the limit in distributions of i.i.d. random variables.*
3) ***Heavy-tailed asymptotic behavior***: *Let X be an α-stable r.v. with $\alpha < 2$. Then : $P(X > x) \sim \gamma C_{\alpha} x^{-\alpha}$ as $x \to \infty$ where C_{α} is a positive constant depending only on α.*

GCLT implies that if the observed randomness is the result of many cumulative effects and these effects follow a heavy-tailed distribution, then a stable model may be appropriate. In contrast of the Gaussian distribution, which has exponential tails, stable laws have inverse power, i.e. algebraic tails. An important consequence of this property is the *non-existence* of the second and higher order moments of stable distributions, except for the special case $\alpha = 2$. For this reason, most classical BSS methods are inadequate in this context and divergence behaviors may be observed.

3. SEMI-PARAMETRIC SEPARATION OF HEAVY-TAILED SOURCES

Our purpose is to estimate by maximum likelihood the density π, the mixing matrix $\mathbf{A}$ and the noise-variance σ^2. In this section, we present a new semi-parametric method to BSS using maximum likelihood estimation in a log-spline model in order to avoid any assumption of the source distribution. Nevertheless we suppose that all sources are independent and have the same common distribution (as mentioned previously, this assumption is just for simplification and can be alleviated at the cost of extra computational load). We use log-spline models for two reasons: on one hand, they have good functional approximation properties, on the other hand, they are well-adapted to the implementation of the SAEM (Stochastic Approximation version of the Expectation Maximization) algorithm [8] allowing to compute easily our estimator. Moreover, this estimation technique is inherently robust towards outliers and impulsiveness effects. For this reason, we apply this method to impulsive random variables with possibly *heavy-tailed distributions* characterized by infinite second and higher order moments.

3.1. Non-parametric estimation of the source distribution

In order to get a non parametric estimate of the source density function π, we propose to use the log-spline models. We denote $\mathcal{S}$ the vector space of spline functions of a given positive order q on an interval $\mathcal{I}$, namely piecewise polynomials function of degree $q - 1$. Given a subdivision of $\mathcal{I}$, the space $\mathcal{S}$ has a finite dimension J and admits a B-spline basis denoted $B_1, \cdots, B_J$ (see [2] for more details) having the following properties: the $(B_j)_{1 \le j \le J}$ are nonnegative and their sum is equal to 1 on $\mathcal{I}$. So for $\theta = (\theta_1, \ldots, \theta_J)^T$ in $\mathbb{R}^J$, we define the density function π_θ for all s in $\mathcal{I}$ by:

$$\pi_\theta(s) = \exp\left[\sum_{j=1}^{J} \theta_j B_j(s) - c(\theta)\right]$$

$$\text{where } \; c(\theta) = \log\left(\int_{\mathcal{I}} \exp\left[\sum_{j=1}^{J} \theta_j B_j(s)\right] ds\right).$$

We choose the dimension J of the log-spline model in function of the sample size T such that $J = o(\sqrt{T})$ (see [7] for more details). Then we consider the value $\hat{\theta}_{T,J}$ of θ that maximizes the log-likelihood of the observations $(\mathbf{x}(t), t = 1 \ldots T)$ and the estimate $\pi_{\hat{\theta}_{T,J}}$ of the density π.

3.2. The SAEM algorithm

To compute the unknown parameters $\eta \triangleq (\theta^T, vec(\mathbf{A})^T, \sigma^2)^T$, we use the SAEM algorithm coupled with a MCMC (Markov Chain Monte-Carlo) procedure presented in [8], [7]. Here we apply this algorithm for estimating the mixing matrix $\mathbf{A}$ and the variance σ^2 using the log-spline model to approach the estimate $\pi_{\hat{\theta}_{T,J}}$. The minimal sufficient statistics used are $\tilde{S}(s) = (\frac{1}{T}\sum_{t=1}^{T} B_j(\mathbf{s}(t)),\ 1 \le j \le J)$ and we implement the k-th iteration of the SAEM algorithm as follows:

- **S-step**: Generate a realization $\mathbf{s}'$ using as proposal distribution the prior distribution π_{θ_k} and take $\mathbf{s}_k$ equal to $\mathbf{s}'$ or to $\mathbf{s}_{k-1}$ according to the value of the acceptance probability.
- **A-step**: Update the minimal sufficient statistics according to the stochastic approximation.
- **M-step**: Update η_k by maximizing the complete log-likelihood of the model evaluated in the observations and in the current value of the minimal sufficient statistics.

This algorithm converges a.s. toward a local maximum of the log-likelihood of the observations under very general regularity conditions (see [8] and [7] for theoretical convergence results). Moreover, in practice, the algorithm is easy to implement and has a relatively low computational cost.

4. A PARAMETRIC BSS METHOD FOR HEAVY-TAILED SOURCES

In this section we assume that the sources $\mathbf{s}(t)$ are impulsive random variables with symmetric α-stable distributions with the same index α. Few algorithms exist for blind heavy-tailed source separation. We briefly describe here a parametric approach using minimum dispersion criterion [9] for comparison with the new semi-parametric approach introduced above.

4.1. Minimum dispersion (MD) based BSS method

The MD method is a two-step separation procedures that achieves the BSS through minimization of a dispersion criterion. The first step is a whitening procedure that 'orthogonalizes' the mixture matrix.

Whitening: Here, we search for a matrix $\mathbf{W}$ which transforms mixing matrix $\mathbf{A}$ into a unitary matrix. Classically, for a finite variance signal, the whitening matrix is computed as the inverse square root of the signal covariance matrix. In our case, the impulsive α-stable source signals have infinite variances. However, it is proven in [10] that the normalized covariance matrix converges to a finite matrix with the appropriate structure when the sample size T tends to infinity. More specifically, we have the following result:

Theorem 1 *Under the previous data model assumptions, the normalized covariance matrix of* $\mathbf{x}$ *defined by:*

$$\hat{\mathbf{R}}_x^n \triangleq \frac{\hat{\mathbf{R}}_x}{Trace(\hat{\mathbf{R}}_x)} \quad \textit{with} \quad \hat{\mathbf{R}}_x = \frac{1}{T}\sum_t \mathbf{x}(t)\mathbf{x}(t)^T$$

converges asymptotically (i.e. when the sample size T tends to infinity) to the finite matrix $\mathbf{ADA}^T$, where $\mathbf{D}$ is a positive diagonal matrix.

Hence, the normalized covariance matrix has the appropriate structure and the whitening problem becomes standard. One can compute $\mathbf{W}$ as the inverse square root matrix of $\hat{\mathbf{R}}_x^n$. The latter can be obtained from the eigendecomposition of $\hat{\mathbf{R}}_x^n = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^T$ as $\mathbf{W} = \mathbf{\Sigma}_s^{-1}\mathbf{U}_s^T$ where $\mathbf{\Sigma}_s$ (resp. $\mathbf{U}_s$) corresponds to the diagonal (resp. unitary) matrix of the m largest eigenvalues (resp. eigenvectors) of $\hat{\mathbf{R}}_x^n$.

The Minimum Dispersion Criterion: Let $\mathbf{z}(t) \triangleq \mathbf{B}\overline{\mathbf{x}}(t)$ where $\mathbf{B}$ is unitary, $\overline{\mathbf{x}}$ denotes the whitened data, i.e, $\overline{\mathbf{x}} = \mathbf{W}\mathbf{x}$ and $\mathbf{B}$ is a 'separating' matrix to be estimated. $\mathbf{z}(t)$ is an orthogonal mixture of the sources and can be written as $\mathbf{z}(t) = \mathbf{C}\mathbf{s}(t)$ with $\mathbf{C}$ orthogonal. Let consider the global MD criterion given by the sum of dispersions of all entries of $\mathbf{z}$, i.e.

$$J(\mathbf{B}) \triangleq \sum_{i=1}^{m} \gamma_{z_i}$$

where γ_{z_i} denotes the dispersion of $z_i(t)$ the i-th entry of $\mathbf{z}(t)$. It is shown in [9] that one can estimate the remaining unitary matrix as the minimum argument of the MD criterion $J(\mathbf{B})$:

Theorem 2 $J(\mathbf{B})$ *reaches its minimum value in the set of unitary matrices if and only if* $\mathbf{BW}$ *(*$\mathbf{W}$ *being the previously computed whitening matrix) is a separating matrix.*

Theorem 2 proves that under unitary transform the signal has minimum dispersion if its entries are mutually independent. To minimize a cost function under unitary constraint different approaches can be considered. The one chosen in [9] consists in estimating $\mathbf{B}$ iteratively as a product of Givens rotations in such a way that we estimate by line-search one scalar parameter (the angle of the Givens rotation) at each iteration.

4.2. Parametric versus non-parametric approaches

The MD method is said to be parametric in the sense that it relies on the a priori knowledge of the exact source pdf. In this case, we have a finite set of parameters to estimate. On the other hand, the SAEM method is said to be non-parametric or semi-parametric in the sense that the source pdf is unknown and need to be jointly estimated with the desired parameters (i.e. mixing matrix). Clearly, estimating a pdf is a difficult problem as the number of parameters to be estimated is infinite. In the semi-parametric approach, we estimate a limited number of parameters by replacing the estimation problem by an approximation one (at the cost of certain loss in the estimation accuracy).

As a consequence, the parametric approach is preferred whenever a reliable a priori knowledge on the source pdf is available. In the situations where the pdf is only partially or inaccurately known, semi-parametric methods should be used because of their robustness against modelization errors as shown next by simulation results.

5. PERFORMANCE EVALUATION & COMPARISON

In this section we compare our proposed semi-parametric method to the parametric MD algorithm. All simulation results are averaged over 100 Monte-Carlo runs and the mixing matrix $\mathbf{A}$ is generated randomly at each run. In experiments 1 and 2, the considered source signals are heavy-tailed with alpha-stable distribution. To measure the quality of α-stable sources separation, we did use the generalized rejection level criterion defined as follows [10]:

$$I_{perf} = \frac{1}{m}\sum_{i=1}^{m} I_i = \frac{1}{m}\sum_{i=1}^{m}\sum_{j\neq i}\frac{|\ C_{ij}\ |^{\alpha}\ \gamma_j}{|\ C_{ii}\ |^{\alpha}\ \gamma_i}. \tag{2}$$

where γ_l denotes the dispersion of source s_l and $\mathbf{C} \stackrel{\text{def}}{=} \mathbf{BA}$.

• **Experiment 1:** In this experiment, mixtures of three ($m = 3$) heavy-tailed symmetric standard α-stable ($\mu = 0$ and $\gamma = 1$) signals with characteristic exponent $\alpha = 1.5$ are considered. The number of observations is $n = 3$ and the mixture is noise-free. Figure 1 presents the generalized mean rejection level (2) versus the sample size for the two methods. From this results, we can observe that the new proposed non-parametric approach can separate correctly the α-stable mixtures even for short sample sizes but is less accurate (eventhough it relies on the ML principle) than the MD method.

• **Experiment 2:** Here, the same three mixtures considered in experiment 1 are corrupted by additive white Gaussian noise. The sample size is set to $T = 500$.

Figure 2 presents the mean rejection level (2) versus the noise power. Similarly to the first experiment, the MD method outperforms the SAEM based method.

• **Experiment 3:** In this experiment, we consider $m = 3$ impulsive sources with generalized gaussian distribution of parameter $p = 1.5$ (i.e. the source pdf is proportional to $\exp(-|x|^p)$). In that case, the signals are of finite variances and thus the standard rejection level criterion [5] is used as a performance measure. $n = 4$ noise free mixtures are considered.

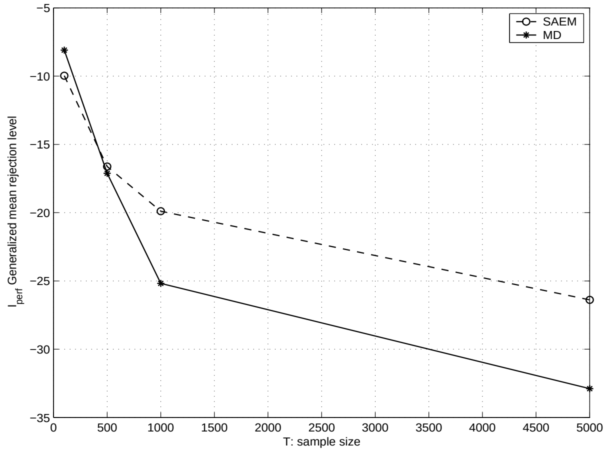

Fig. 1. Generalized mean rejection level versus the sample size.

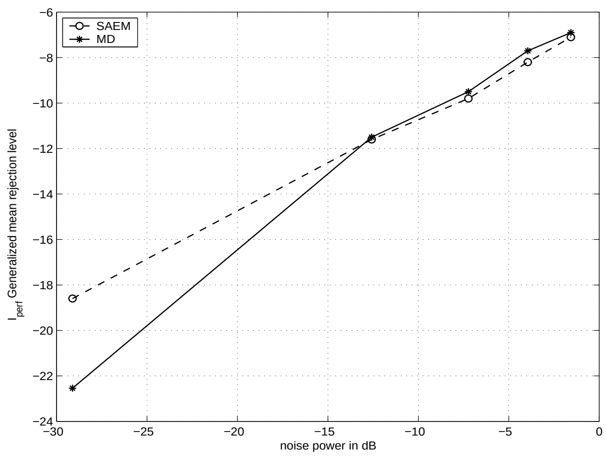

Fig. 2. Generalized mean rejection level versus the noise power.

As can be observed from figure 3, the MD method fails to separate correctly the sources as it relies on the SαS source pdf assumption that is not verified in this example. This illustrate the robustness of the SAEM compared to the MD method with respect to the pdf modelization errors.

6. CONCLUSION

In this work, we developed a new semi-blind BSS method using the SAEM algorithm. The proposed method is applied for the blind separation of linear instantaneous mixtures of heavy-tailed sources. The SAEM based method is compared with the minimum dispersion (MD) method and shown to be more general (as it can be applied to a larger class of source signals) and outperforms the MD method in terms of robustness against modelization errors.

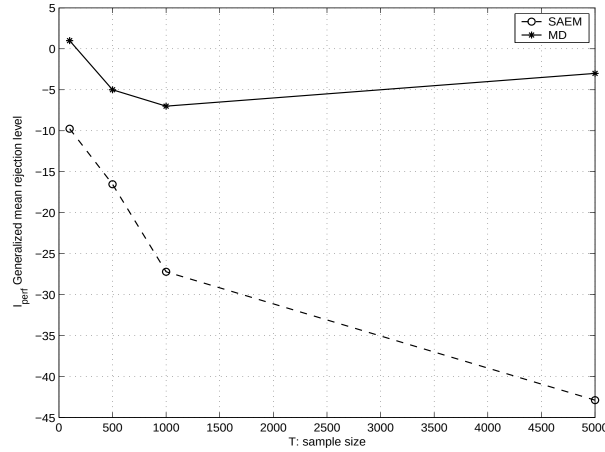

Fig. 3. Generalized mean rejection level versus sample size.

7. REFERENCES

[1] R.J. Adler, R. Feldman and M.S. Taqqu, Eds *A practical Guide to Heavy tails : Statistical Techniques and Applications*, Birkhauser, Boston, 1998.

[2] C. de Boor, *A practical guide to splines, Applied Mathematical Sciences*, Springer-Verlag, New York, 1978.

[3] R. Boscolo et al, "Independent Component Analysis Based on Nonparametric Density Estimation", *IEEE Trans. on Neural Networks*, vol. 15, No. 1, Jan. 2004.

[4] J.F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals", *Radar and Signal Processing, IEE Proceedings F*, Dec. 1993.

[5] A. Belouchrani, K. Abed Meraim, J.-F Cardoso, E. Moulines, "A blind source separation technique using second order statistics," *IEEE Tr. on Signal Processing*, pp. 434–444, Feb. 1997.

[6] A. Hyvarinen, J. Karhunen and E. Oja, *Independent Component Analysis* , Wiley, 2001.

[7] E. Kuhn, "Maximum Likelihood Estimation in Nonlinear Inverse Problems", *Ph.D., Paris-Sud Univ.*, France, Dec. 2003.

[8] E. Kuhn, and M. Lavielle, "Coupling a stochastic approximation version of EM with a MCMC procedure", *to appear in ESAIM P&S*, France, September 2004.

[9] M. Sahmoudi, K. Abed-Meraim and M. Benidir, "Blind Separation of Instantaneous Mixtures of alpha-stable sources", *in Proc. of ISISPA'2003; 3rd IEEE Int. Symposium on Image and Signal Processing and Analysis*, Italy, 2003.

[10] M. Sahmoudi, K. Abed-Meraim and M. Benidir "Blind Separation of Heavy-Tailed Signals Using Normalized Statistics", *n Proc. of ICA'2004; 5th Int. Conf. on Independent Component Analysis*, Granada Spain, Sept. 22-24, 2004.