SIGNAL PROCESSING VII: Theories and Applications
M. Holt, C. Cowan, P. Grant, W. Sandham (Eds.)

A general adaptive algorithm for nonGaussian source separation without any constraint

F. HARROY [(1)], J.L. LACOUME[(1)], M.A. LAGUNAS [(2)]

(1) CEPhAG - ENSIEG INPG CNRS URA 346
BP 46
38 402 Saint-Martin d'Hères Cedex
France
Fax : (33):76:82:63:84

(2) E.T.S.I. Telecom
Apdo 30.002
08080 Barcelona
Spain
Email : harroy@cephag.observ-gr.fr

Abstract. This paper deals with the blind source separation. The task consists in separating some independent and linearly mixed signals called sources. After some general remarks, the model is recalled and our approach based on the Maximum-Likelihood principle and on the higher-order statistics (HOS) is introduced. The main stages of the calculation are presented leading to the criterion of the separation based on a sum of squared cumulants of the sources at the fourth order. The second part is devoted to the adaptive implementation which is in opposition to the block treatment. The procedure using the gradient calculus is described.

Some results obtained in simulations are shown, they correspond to the case of a mixture of two real valued sources. Finally, an example of a possible integration in a communications system based on multidimensional beamformers is briefly shown. But some tests on real data should be carried out beforehand.

INTRODUCTION

In many physical problems, the recovering of signals corrupted by noise and mixed anyway is the main task, generally named: Filtering. This study belongs to the general issue of blind source separation. It means that no knowledge is available concerning the mixing process and the nonGaussian source signals. Then, anyone must take in mind that the sources cannot be reached in a physical way and only the outputs of a sensor array are available.

For a long time, the Gaussianity assumption, as a derivation of the central limit theorem, was adopted as a last resort, but nature rarely delivers such kinds of signals. Moreover, the recent introduction of the higher-order statistics has brought a new renewal of interest to go further in the characterisation of non Gaussian processes. Hence, a variety of methods have appeared with two different levels. The first concerns the theoretical background that can be heuristic [4] or based on statistical ideas, even on algebraic basis [2], whereas the second is relevant to the type of implementation: block treatment or adaptive way. Though, for all kinds of treatment, a performance analysis is useful: the Cramer-Rao bounds for the block treatment and the asymptotic behaviour of the estimators [1] for adaptive treatments.

The first part of this paper is devoted to the explanation of the adopted physical model and the major principles supporting the point of view. Then, the second section is devoted to a method to solve the question. We present a complete adaptive implementation of an algorithm without constraint on the kind of sources (short-tailed or long-tailed random sources). Simulations that have been carried out illustrate the work in the third section. In the last part which concerns the conclusions and the forecasts, an example of a possible integration in a communication system is briefly shown.

1. The theoretical backgrounds

1.1 The physical model

Having an n sensor array and p statistically independent sources mixed linearly, the most general writing of the linear model is the following:

$$x_i(t) = \sum_{j=1}^{p} a_{ij}(t) * s_j(t) + b_i(t) \quad i \in \{1, \ldots, n\}$$

where $\underline{\underline{A}}=(a_{ij})_{i=1,\ldots,n;\ j=1,\ldots,p}$ represent the transfer functions between sources and sensors and b_i, the i^{th} component of the corrupted noise assumed to be spatially white and Gaussian and equi-powerful ($\Rightarrow E[\underline{b}\,\underline{b}^{+}]=\sigma^2 I$). Chosing a treat-ment in time domain for no-delay mixture or in frequency domain for convolutive mixture, the model can be summarized with a linear relation: $\underline{x}_{(n)}= \underline{\underline{A}}\,\underline{s}_{(p)}+\underline{b}_{(n)}$ between the observations $\underline{x}_{(n)}$ and the sources $\underline{s}_{(p)}$. Most of the methods uses the higher-order statistics, more precisely the fourth-order cumulant-tensor. Because of an assumption on the even source probability density function (pdf), the third-order cumulants are null. Thus, the elements of the cumulants at the fourth order in the real case are equal to:

$$\kappa_{i,j,k,l}(\underline{X})=E\{X_i X_j X_k X_l\}-E\{X_i X_j\}E\{X_k X_l\}\ [3]$$

The notation between brackets [3] holds for all permutations over the indexes.

Now, many approaches have been applied. One among the first was a heuristic one [4] consisting in maximising the inverse of the sum of the squared source cross-cumulants. Afterwards, an original approach has been introduced by J.F. Cardoso, using the multi-linear algebra on the fourth-order cumulant-tensor named Quadricovariance. Theoretically, this method permits to estimate more sources than the number of sensors. Others founded on statistical tools like the Maximum Likelihood (ML) have also appeared [10].

1.2 A method using the ML principle

One of them [Gaeta-Lacoume] use this principle to estimate the parameters of the problem. Before evoking the steps of the calculation, it is important to remember the asymptotic properties of the ML-estimators. They are asymptotically consistent, efficient, unbiased, Gaussian. At first, the mixture matrix $\underline{\underline{A}}$ is decomposed as a general product of three matrices: $\underline{\underline{A}}=\underline{\underline{V}}\ \underline{\underline{\Delta}}\ \underline{\underline{\Pi}}$ where $\underline{\underline{V}}$, $\underline{\underline{\Pi}}$ are the unitary matrices and $\underline{\underline{\Delta}}$ a diagonal matrix. The two first correspond to a pre-whitening of the data and are derived from the eigenvalue-decomposition (EVD) of the spectral matrix (the correlation matrix in a no-delay mixture).

As the third matrix vanishes when the spectral matrix is estimated, the use of the HOS is a mean to solve the question. It is a classical result that the eigen-vector associated to the largest eigen-values of the spectral matrix span the signal subspace and correspond to a maximum for the likelihood function in the Gaussian case.

Now, it is interesting to explain how the Log-likelihood is obtained with the restricted model taking place in the signal subspace: $\underline{x}_{(p)}=\underline{\underline{\Pi}}\ \underline{s}+\underline{b}_{(p)}$. Then $\underline{\underline{\Pi}}$ corresponds to a rotation in the signal subspace and the vectors $\underline{x}_{(p)}$, $\underline{b}_{(p)}$ are the projections of, respectively, the observation $\underline{x}_{(n)}$ and the noise $\underline{b}_{(n)}$. In the sequel, they will be denoted $\underline{x}$ and $\underline{b}$. A scale operation with the inverses of the *p*-largest eigenvalues leads to consider the new fictitious observations as white vectors (with unitary power and uncorrelated components).

So, to obtain the Log-likelihood, three steps are required:

- the probability density function (pdf) of the observations conditioned by the unknown parameters is written as the convolutive product of the pdf's of the independent random vectors: sources $\underline{s}$ and noise $\underline{b}$.
- the Parceval theorem permits the convolutive product to change into a classical one by changing the integration domain passing from the pdf to the characteristic function.
- the Gram-Charlier expansion is used to express the pdf of the unknown sources with respect to a Gaussian one multiplied by a compensating series [5]. Then, the second characteristic function of any random vector is written as:

$$\varphi_{\underline{S}}(\underline{s}) = \varphi_{\underline{SG}}(\underline{s})\,(1 + \delta\varphi(\underline{s}))$$

where $\varphi_{\underline{SG}}$ corresponds to the Gaussian part and $\delta\varphi(\underline{s})$ is the residue depending on the unknown higher-order statistics of the sources. Stopping the expansion at the fourth order, the return in the original space leads to the following formula, in the high SNR case:

$$\text{Log}\ p_{\underline{X}}(\underline{x})\ \#\ -\frac{1}{2}\underline{x}.\underline{x}^T + \sum_{i=1}^{p} \frac{1}{4!}\kappa_4^{\ i,i,i,i}\ H_{i,i,i,i}[\underline{s}(\underline{x})]$$

where $\kappa_4^{\ i,i,i,i}$ is the fourth-order cumulant of the i^{th} source, $H_{i,i,i,i}$ is the Hermite tensor at the fourth order, built with the source vector and $\underline{s}(\underline{x})$ reminds us the linear relation between the sources and the observations. When the Log-likelihood is considered with the average on *N* observation vectors and when the multi-linearity of the 'cumulant operator' is introduced, the remaining part interesting in the estimation of the matrix $\underline{\underline{\Pi}}$, is:

$$ML_4\,(\underline{x}^{(1)}, \ldots, \underline{x}^{(N)} | \underline{\underline{\Pi}}) = \frac{1}{4!} \sum_{i,\alpha,\beta,\gamma,\delta} (\Pi_{\alpha i}\,\Pi_{\beta i}\,\Pi_{\gamma i}\,\Pi_{\delta i}\,\hat{\kappa}_{4\,\alpha\beta\gamma\delta}\,(\underline{x}^{(1)}, \ldots, \underline{x}^{(N)}))^2$$

This result is important because of its similarity with another obtained independently by P. Comon His theory has reinforced the weight of these 'nice' ML-estimators obtained by Gaeta, by mixing the above mentioned properties and a rigorous theoretical support. Here, the concept of contrast function was introduced [3].

Using this set of tools, the following sections are dedicated to one way to perform the blind source separation where the aim is to maximize the function $\sum_{1\le i\le p}(\kappa_4^{\ i,i,i,i})^2$ (equivalent to ML_4) with respect to the set of unitary matrices.

2. The algorithm without any constraint

In this part, we explain an endeavour to practise adaptively the source separation. The main purpose is to use the previous criterion explained above. Some authors [6,7] have presented adaptive algorithms based on objective functions which are the sum of the absolute value of the source cumulant. The main drawback of this procedure is that it is required a priori knowledge on the sources by imposing the same sign to the cumulants. Using the Comon-Lacoume criterion, this constraint vanishes keeping a whole freedom.

Taking account that the remaining matrix belonging to $R^{\,nxn}$ is orthogonal, it can be expressed as a product of $r=\frac{n(n-1)}{2}$ Givens matrices:

$$\underline{\underline{\Pi}} = \prod_{1\le i\le r} R_i(\theta_i)$$

The determination of $\underline{\underline{\Pi}}$ with the unitary constraint corresponds to the knowledge of a set of angles with a smaller size. The objective function remains:

$$J(\underline{Y}) = \sum_{1\le i\le p}\left(\kappa_4^{\ i,i,i,i}(Y_i)\right)^2$$

knowing that $\begin{cases} E[\underline{Y}\ \underline{Y}^+] = \underline{\underline{1}} \\ \kappa_4^{\ i,i,i,i} = E[|Y_i|^4] - 3\,E^2[|Y_i|^2] \end{cases}$

and becomes:

$$J(\underline{Y}) = \sum_{1 \le i \le p} E[Y_i^4 - 3].$$

The gradient with respect to $(\theta_\alpha)_{i \le \alpha \le r}$ gives:

$$\frac{\partial J(\underline{Y})}{\partial \theta_\alpha} = 8 \sum_{1 \le i \le p} \left(E[Y_i^4] - 3\right) E[\frac{\partial Y_i}{\partial \theta_\alpha} Y_i^3]$$

So, the relation $\underline{Y} = \prod_{i=1}^{p} R^{-1}(\theta_i)\, \underline{X}$ implies:

$$\frac{\partial \underline{Y}}{\partial \theta_\alpha} = \prod_{i=1}^{\alpha-1} R_i^{-1}(\theta_i) \frac{\partial}{\partial \theta_\alpha} R_\alpha^{-1}(\theta_\alpha) \prod_{i=\alpha+1}^{p} R_i^{-1}(\theta_i)\, \underline{X}$$

From this, several choices are possible to write a "stochastic" version of this problem, often without any theoretical support . The way chosen by O. Macchi permits to derive easily the stochastic version and perhaps, to perform the asymptotic study. For our case, among the different possibilities, one of them is considered here by evaluating the different expectations in (A) with moving averages

$$E[X^\alpha Y^\beta](n) = (1-\delta) E[X^\alpha Y^\beta](n-1) + \delta x^\alpha (n) y^\beta (n)$$

Practically, as at the step n, the observation x (n) and the latter estimate θ (n-1) are known, we decide to use the source signal being equal to:

$$y(n|n-1) = \underline{\underline{\Pi}}(\theta(n-1)) x(n).$$

As a summary, the algorithm has been implemented with the general way :

$$\theta(n) = \theta(n-1) + \mu\, F(\theta(n-1), X_n)$$

where $F = -\nabla J$ and X_n is the state vector equal, here, to the new observation $\underline{x}$ (n). It means:

$$F(\theta(n-1), \underline{x}(n)) = 8 \sum_{1 \le i \le p} (E[Y_i^4](n) - 3)\, E[\frac{\partial Y_i}{\partial \theta_\alpha} Y_i^3](n) \Big|_{1 \le \alpha \le r}$$

$$E[Y_i^4](n) = (1-\delta) E[Y_i^4](n-1) + \delta\, y(n|n-1)_i^4$$

$$E[\frac{\partial Y_i}{\partial \theta_a} Y_i^3](n) = (1-\gamma) E[\frac{\partial Y_i}{\partial \theta_a} Y_i^3](n-1) + \gamma \frac{\partial y(n|n-1)}{\partial \theta_a} y(n|n-1)_i^3$$

The study with the technique of the ODE seems too difficult since the ODE appears as a trigonometric polynomial.

Moreover, this part has been introduced in a whole algorithm including the pre-whitening stage. This latter uses a procedure implemented by Ortigueira-Lagunas [8]. It is based on the updating of the EVD of the correlation matrix. So, the step using at the second-oder using the correlation matrix has the following way:

$$R_n = (1-\beta) R_{n-1} + \beta\, \underline{x}(n)\, \underline{x}(n)^+$$

EVD of R_{n-1}: $R_{n-1} := V_{n-1} \Delta_{n-1} V_{n-1}^+$

$\Rightarrow R_n = V_{n-1} M_n V_{n-1}^+$

$\Rightarrow M_n = (1-\beta)\Delta_{n-1} + \beta V_{n-1}^+ \underline{x}(n)\, \underline{x}(n)^+ V_{n-1}$

EVD of M_n: $M_n := U_n D_n U_n^+$

$\Rightarrow R_n := (V_{n-1} U_n) D_n (V_{n-1} U_n)^+$

$\Rightarrow V_n = V_{n-1} U_n$ and $\Delta_n = D_n$

We carried out some simulations yielding to promising results. That is why it can be attempted to include this algorithm in an operationnal system.

3. Results in simulations

The tests that were carried out, stress the question of the simultaneous regulation of the different parameters. Generally, the evolution can be based on local dynamics or can be chosen to obtain first a fast convergence and be changed afterwards to fluctuate slowly.

On Figure 1, we show results obtained for a mixture of a Laplacian source (κ_4=3) and a B-PSK signal (κ_4=-1,5). One can see the mixture as a product of a first rotation (25°) with a diagonal matrix and with another rotation which can not be observed at the second order. Curves (A) and (B) on the second graph correspond to the sets of parameters (A) and (B) (in left top of the figure, $\alpha=\beta=\gamma$). The parameter α must be greater than μ to obtain a convergence of the corresponding estimated term faster than one for the searched angle (37°). Of course lower are the gain (α, μ) ((B) < (A)), better and smoother is the result. (compare curve (B) with (A)). We have noted that the parameter β leading to the uncorrelation of signals has not a great influence except at the beginning of the tracking. We can note the **small dispersion** obtained with the algorithm at the second order. Here, the curve fluctuates rapidly with very small dynamics. The histograms of the residual errors plotted with the 20.000 last estimated values look like to a Gaussian shape only for the first one which illustrates the behaviour of the algorithm at the second order.

As different but equivalent attractors exist, we have forced the estimates in the interval [0, 90°]. Here, the problem of the regulation of the whole parameter set is emphasised. As no theoretical result exist to guide the choice , we move forward by trial.

CONCLUSION and FORECASTS

After having spoken about generalities, a method has been explained and tested to perform blind source separation. It is founded on the use of adaptive algorithm which maximizes the criterion found independently by P. Comon and J.L. Lacoume. Simulations have been performed with satisfying results. Now, we can still improve this algorithm by testing another implementations and by carrying out some experiments with real signals. Moreover, its behaviour must be checked in a noisy environment to try afterwards an integration in a whole system of communications as explained now. Figure 2 illustrates the different parts of the process. The fundamental basis of this interesting issue is explained at greater length in another paper [9]. To roughly summarize, the problem appears when several users ask for the use of the channel at the same time. Then, it is necessary to separate their own contributions to adapt another treatment of the procedure in the best way.

References

[1] Benveniste A., Metivier M., Priouret P., *Adaptive algorithms and stochastic approximations*, Ed Springer-Verlag, 1990

[2] Cardoso.J-F. , Localisation et Identification par la quadricovariance, Traitement du Signal, Special Issue on nonGaussian signals and non-Linear systems, 1990, Vol 7, n°5

[3] Comon P. , Blind identification in presence of noise, Proc. EUSIPCO'92, Brussels (Bel.)

[4] Lacoume J-L, Ruiz P., Sources Identification : A Solution based on Cumulants, Proc. IEEE ASSP, Minneapolis, August 1988

[5] Mac Cullagh P., *Tensor methods in statistics*, Ed. Chapman & Hall, 1987

[6] Moreau E., Macchi O., New self-adaptive algorithms for source separation based on contrast functions, Proc. Workshop on H.O.S., Tahoe Lake (Cal.), june 1993, pp 215-219

[7] Moreau E., Macchi O., Séparation auto-adaptative de sources sans blanchiment préalable, Proc. 14ième GRETSI, Juans-les-pins (Fr), Sept 1993, pp 325-328

[8] Ortigueira M. D., Lagunas M. A., Eigendecomposition versus singular value decomposition in adaptive array signal processing, Signal Processing, Vol 25, 1991, pp 35-49

[9] Perez A., Lagunas M. A, Simultaneous multiuser blind demodulation based on time reference array beamformers, submission for the EUSIPCO'94

[10] Pham D.T., Garat P., Jutten C., Separation of a mixture of independent sources through a maximum likelihood approach, Proc. EUSIPCO'92, Brussels (Bel.)

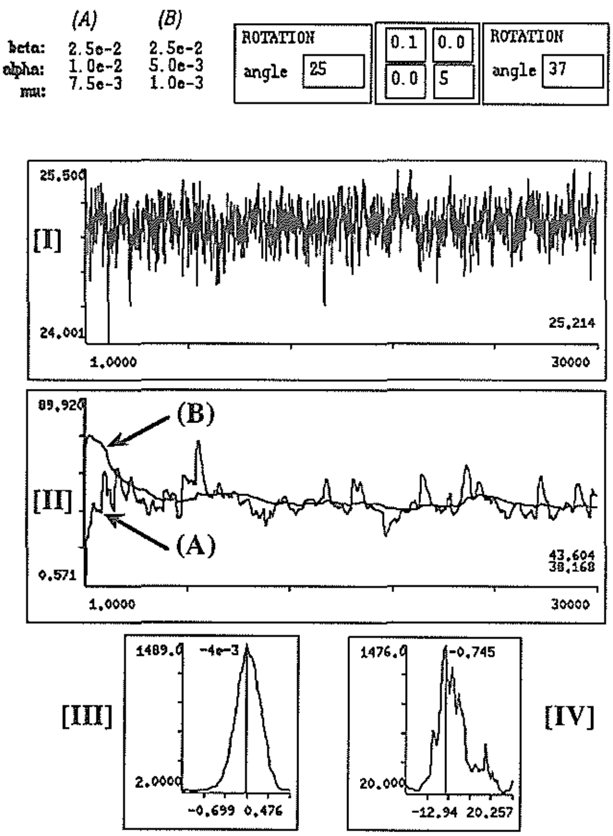

Figure 1. Curve [I] is for the estimation of the angle 25° at the second order, curve [II] is for the second angle 37° obtained with the criterion. [III] and [IV] are the corresponding histograms of the errors.

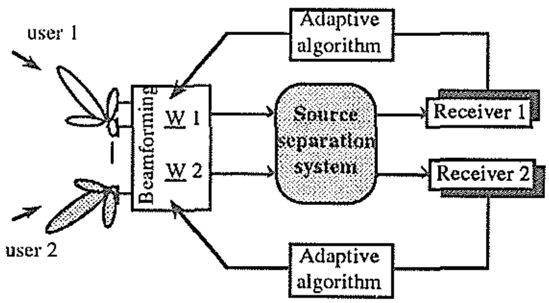

Figure 2. the whole system for the application in communications