

# NOVEL BLIND SOURCE SEPARATION ALGORITHMS USING CUMULANTS

*Sergio Cruces\* and Luis Castedo†*

Area de Teoría de la Señal  
University of Seville  
41092-Seville, SPAIN  
sergio@viento.us.es

*Andrzej Cichocki*

Lab. for Open Information Systems  
Brain Science Institute, RIKEN  
Wako-Shi, Saitama, 351-01, JAPAN  
cia@brain.riken.go.jp

## ABSTRACT

This paper investigates new algorithms for blind source separation that use cumulants instead of nonlinearities matched to the probability distribution of the sources. It is demonstrated that separation is a saddle point of a cumulant-based entropy cost function. To determine this point we propose two quasi-Newton algorithms whose convergence is isotropic and does not depend on the sources distribution. Moreover, convergence properties remain the same when there is Gaussian noise in the mixture.

## 1. INTRODUCTION

Blind Source Separation (BSS) is the problem of recovering statistically independent signals from the observations of an unknown linear mixture of them. It is a fundamental problem in signal processing with a large number of extremely diverse applications such as array processing, multiuser communications, voice restoration and biomedical engineering [1].

The BSS problem is typically formulated as follows. Let us assume that an array of sensors provides a vector of  $M$  observed signals  $\mathbf{x} = [x_1[n], x_2[n], \dots, x_M[n]]^T$  that are noisy linear mixtures of  $N \leq M$  unobserved random processes  $\mathbf{s} = [s_1[n], s_2[n], \dots, s_N[n]]^T$  termed sources, i.e.,

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{e} \quad (1)$$

where  $\mathbf{A}$  is an unknown  $M \times N$  full-column rank matrix that represents the mixing system and  $\mathbf{e} = [e_1[n], e_2[n], \dots, e_M[n]]^T$  is the noise vector. The exact probability density function (p.d.f.) of the sources is unknown. We

\* This work was supported by grant CICYT (TIC96-0500-C10-08) and by the Frontier Research Program of Japan.

† Supported by CICYT (TIC96-0500-C10-02), Fondos FEDER (1FD97-0082).

The MATLAB code of the presented algorithms is available at <http://viento.us.es/~sergio/CIH.html>

will only assume that they are real-valued, zero-mean, non-Gaussian distributed and mutually independent. Additionally, we will assume that the noise  $\mathbf{e}$  is Gaussian distributed and statistically independent from the sources.

In order to recover the sources, the observations are processed by a  $N \times M$  separating matrix  $\mathbf{B}$  to produce the vector of outputs or sources estimation

$$\mathbf{y} = \mathbf{B}\mathbf{x} \quad (2)$$

When the separation is obtained the overall mixing and separating transfer matrix  $\mathbf{G} = \mathbf{B}\mathbf{A}$  contains a single nonzero element per row and per column. For simplicity reasons we will assume that separation is achieved when  $\mathbf{G} = \mathbf{I}$ .

Recently, several authors have developed an unified interpretation of the BSS criteria that naturally leads to the utilization of nonlinear functions of the outputs [2, 3, 4]. Separation is achieved when these nonlinearities match the sources distribution. It was shown in [3], however, that the convergence properties of BSS algorithms are extremely dependent on the sources distribution and the algorithm nonlinearities. Also, these properties are considerably different when Gaussian noise is present in the mixture.

In this work, we propose a new adaptive algorithm for BSS that uses higher order cumulant functions of the outputs as algorithm nonlinearities. Unlike existing approaches to BSS, the local convergence of our adaptive rule is isotropic and independent on the sources distribution. Additionally, convergence properties remain the same when there exists Gaussian noise in the mixture due to the utilization of higher order cumulants as algorithm nonlinearities.

Along this paper we will use the following notation:  $C_{y_i}^\alpha$  is the  $\alpha$ -order cumulant of the output  $y_i$  and  $\mathbf{C}_{y,y}^{\alpha,\beta}$  is the cross-cumulant matrix whose elements are  $[\mathbf{C}_{y,y}^{\alpha,\beta}]_{ij} = \text{Cum}(\underbrace{y_i, y_i, \dots, y_i}_\alpha, \underbrace{y_j, y_j, \dots, y_j}_\beta)$ ; the op-

erator  $vec(\cdot)$  represents the rearrangement of the elements of a square  $N \times N$  matrix  $\mathbf{M}$  into a vector  $vec(\mathbf{M}) = [\dots, M_{kk}, \dots, \dots, M_{ij}, M_{ji}, \dots]^T$  where  $k, i, j | i < j = 1, \dots, N$ .

The paper is organized as follows. In Section 2, we demonstrate that the Entropy Maximization (ENTMAX) criterion proposed in [4] does not lead to a separation solution for a certain mismatch between the algorithm nonlinearities and the sources p.d.f. We show that for this specific mismatch the entropy cost function exhibits a saddle point at the separation solution. According to this result, in section 3 we derive two algorithms for BSS and investigate their convergence properties. Section 4 presents the results of computer simulations and, finally, section 5 is devoted to the conclusions.

## 2. ENTROPY OPTIMIZATION BASED ON CUMULANTS

Entropy Maximization (ENTMAX) has been proposed [4] as a guiding principle for BSS. According to this criterion, separation is achieved when the mutual entropy of a nonlinear function of the outputs  $\mathbf{z} = g(\mathbf{y})$  is maximized provided that the nonlinearity  $g(\cdot)$  is equal to the cumulative density function (c.d.f.) of the sources and that there is no noise in the mixture. In this case, the mutual differential entropy of  $\mathbf{z}$  conditioned to the global transfer matrix  $\mathbf{G}$  and to the sources distribution  $p_S = \prod_{i=1}^N p_{S_i}$  is given by

$$h(\mathbf{z}|\mathbf{G}, p_S) = \sum_{i=1}^N E[\log p_{S_i}(y_i)] + \log |\det(\mathbf{G})| + h(\mathbf{s}) \quad (3)$$

where  $h(\mathbf{s})$  is the differential entropy of the sources.

In many practical applications, however, the p.d.f. of the sources is not known *a priori* and in the derivation of practical algorithms we have to use a set of nonlinearities  $\hat{g}_i(\cdot)$ ,  $i = 1, \dots, N$  that do not exactly match the sources c.d.f.  $g_i(\cdot)$ ,  $i = 1, \dots, N$ . As shown in the sequel, the ENTMAX principle no longer yields to source separation. Indeed, let us consider a particular case of nonlinearity misadjustment that results from replacing the term  $E[\log p_{S_i}(y_i)]$  in (3) by a negative function of the outputs cumulants  $-\frac{|C_{y_i}^{1+\beta}|}{1+\beta}$ . The following theorem shows that separation is not a maximum of the function  $h(\mathbf{z}|\mathbf{G}, \hat{p}_S)$  but a saddle point.

**Theorem 1** *Source separation is a saddle point of the following entropy cost function*

$$h(\mathbf{z}|\mathbf{G}, \hat{p}_S) = - \sum_{i=1}^N \frac{|C_{y_i}^{1+\beta}|}{1+\beta} + \log |\det(\mathbf{G})| + h(\mathbf{s}) \quad (4)$$

whenever  $\beta > 1$  is an integer number such that the  $1+\beta$  cumulants of all the sources are not zero.

**Proof** Let us first demonstrate that the gradient of  $h(\mathbf{z}|\mathbf{G}, \hat{p}_S)$  vanishes at the points where source separation is achieved. Defining the sign matrix of the outputs cumulants  $\mathbf{S}_y^\beta = sign(diag(\mathbf{C}_{y,y}^{1,\beta}))$  and assuming that the diagonal cumulants never vanish, the gradient of  $h(\mathbf{z}|\mathbf{G}, \hat{p}_S)$  with respect to  $\mathbf{G}$  is

$$\frac{\partial h(\mathbf{z}|\mathbf{G}, \hat{p}_S)}{\partial \mathbf{G}} = -\mathbf{S}_y^\beta \mathbf{C}_{y,y}^{\beta,1} + \mathbf{G}^{-T} \quad (5)$$

We will assume that the sources are properly scaled and ordered so that  $\mathbf{C}_{s,s}^{\beta,1} = \mathbf{S}_s^\beta$ . Then, it is apparent from eq. (5) that the gradient vanishes at  $\mathbf{G} = \mathbf{I}$ .

Next, let us examine the nature of the Hessian at this point. Arranging the matrix  $\mathbf{G}$  in the following vector  $\mathbf{g} = vec(\mathbf{G})$ , the Hessian of  $h(\mathbf{z}|\mathbf{G}, \hat{p}_S)$  at  $\mathbf{G} = \mathbf{I}$  is

$$\mathcal{H} = \frac{\partial}{\partial \mathbf{g}} \frac{\partial h(\mathbf{z}|\mathbf{G}, \hat{p}_S)}{\partial \mathbf{g}} = - \left( \begin{array}{c|c} (1+\beta)\mathbf{I} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{E} \end{array} \right) \quad (6)$$

where  $\mathbf{E}$  is a  $N(N-1) \times N(N-1)$  matrix of the form

$$\mathbf{E} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix} \quad (7)$$

The eigenvalues of this Hessian matrix are  $-(1+\beta)$  and 1 and, as a consequence, the Hessian is neither positive definite or negative definite. This shows that the separating point  $\mathbf{G} = \mathbf{I}$  is always a saddle point of  $h(\mathbf{z}|\mathbf{G}, \hat{p}_S)$ .

## 3. QUASI-NEWTON ALGORITHMS

In the previous section we demonstrated that source separation is not a maximum of the cost function  $h(\mathbf{z}|\mathbf{G}, \hat{p}_S)$  and, as a consequence, gradient ascent algorithms cannot be used to select the separating system. Nevertheless, we showed that the gradient of  $h(\mathbf{z}|\mathbf{G}, \hat{p}_S)$  vanishes at source separation. Thus, algorithms that solve the equation

$$\left( \frac{\partial h(\mathbf{z}|\mathbf{G}, \hat{p}_S)}{\partial \mathbf{G}} \right)^T = -\mathbf{C}_{s,y}^{1,\beta} \mathbf{S}_y^\beta + \mathbf{G}^{-1} = \mathcal{F}(\mathbf{G}^{-1}) = \mathbf{0}$$

still are adequate for BSS. More specifically, in this section we will derive two pre-conditioned quasi-Newton

algorithms that exhibit desirable convergence properties. Let us start rewriting the above system of equations in vector form

$$\mathbf{f}(\mathbf{g}^{-1}) = \text{vec}(\mathcal{F}(\mathbf{G}^{-1})) = \mathbf{0} \quad (8)$$

where  $\mathbf{g}^{-1} = \text{vec}(\mathbf{G}^{-1})$ . Quasi-Newton algorithms that solve (8) have the form

$$\mathbf{g}^{-1(n+1)} = \mathbf{g}^{-1(n)} - \mathcal{J}^{-1} \mathbf{f}(\mathbf{g}^{-1(n)}) \quad (9)$$

where  $\mathcal{J}$  denotes the Jacobian matrix. It is straightforward to show that the Jacobian at separation is

$$\mathcal{J} = - \left( \begin{array}{c|c} (1+\beta)\mathbf{I} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{I}_{N(N-1)} \end{array} \right) \quad (10)$$

where  $\mathbf{I}_{N(N-1)}$  is the identity matrix of dimension  $N(N-1) \times N(N-1)$ . In order to rewrite the algorithm in a compact matrix form, we can approximate the inverse of the Jacobian matrix by  $\mathcal{J}^{-1} \approx \mu \mathbf{I}_{N^2}$ , where  $\mu$  is an adequate step size and  $\mathbf{I}_{N^2}$  is the identity matrix of dimension  $N^2 \times N^2$ . As demonstrated in [6], this modification will not change the local convergent character of the algorithm as long as the following condition holds true

$$\|\mathbf{I}_{N^2} - \mu \mathcal{J}^{-1}\|_p < 1 \quad (11)$$

where  $\|\cdot\|_p$  is the  $p$  norm of a matrix. Returning to the matrix notation it is obtained the following algorithm

$$\mathbf{G}^{-1(n+1)} = \mathbf{G}^{-1(n)} (\mathbf{I} + \mu(\mathbf{C}_{y,y}^{1,\beta} \mathbf{S}_y^\beta - \mathbf{I})) \quad (12)$$

Note that this algorithm asymptotically exhibits the equivariance property for sufficiently long data sets. Moreover, condition (11) is satisfied whenever  $\mu < \frac{2}{1+\beta}$ .

Rewriting equation (12) in terms of the separating system  $\mathbf{B}^{(n)} \equiv \mathbf{B}^{(n)}(\mathbf{A}\mathbf{A}^+) = \mathbf{G}^{(n)}\mathbf{A}^+$  where  $(\cdot)^+$  denotes the Moore-Penrose pseudo-inverse, we obtain the following algorithm

$$\mathbf{B}^{(n+1)} = (\mathbf{I} + \mu(\mathbf{C}_{y,y}^{1,\beta} \mathbf{S}_y^\beta - \mathbf{I}))^{-1} \mathbf{B}^{(n)} \quad (13)$$

This recursion can be interpreted as a quasi-Newton algorithm that iteratively inverts the matrix  $\mathbf{C}_{x,y}^{1,\beta} \mathbf{S}_y^\beta$  (see [5] for more details on iterative inversion algorithms). For this reason it will be termed First Cumulant-based Iterative Inversion (CII1) algorithm.

Since the CII1 algorithm is of the quasi-Newton type, we should avoid that it operates in regions where discontinuities of the function  $\mathcal{F}(\cdot)$  appear. In our case, discontinuities occur when matrix  $\mathbf{B}$  is not full row-rank. However, taking into account the triangular inequality  $\|\mathbf{C}_{y,y}^{1,\beta} \mathbf{S}_y^\beta - \mathbf{I}\|_p \leq 1 + \|\mathbf{C}_{y,y}^{1,\beta} \mathbf{S}_y^\beta\|_p = 1 + \|\mathbf{C}_{y,y}^{1,\beta}\|_p$  it is sufficient to choose

$$\mu^{(n)} = \min \left( \frac{2\eta}{1+\eta\beta}, \frac{\eta}{1+\eta\|\mathbf{C}_{y,y}^{1,\beta}\|_p} \right) \quad (14)$$

with  $\eta < 1$  to avoid these singularities. This way to choose  $\mu$  enables us to expand the CII1 algorithm in a power series, truncate the expansion and obtain the recursion

$$\mathbf{B}^{(n+1)} = \left( \mathbf{I} - \mu^{(n)} (\mathbf{C}_{y,y}^{1,\beta} \mathbf{S}_y^\beta - \mathbf{I}) \right) \mathbf{B}^{(n)} \quad (15)$$

which we will denote as the CII2 (Second Cumulant-based Iterative Inversion) algorithm. The primary advantage of the above algorithms is that their convergence satisfy the following two theorems.

**Theorem 2** *The local convergence of the CII1 and CII2 algorithms is isotropic and does not depend on the sources p.d.f. as long as their  $(1+\beta)$ -order cumulants do not vanish.*

The proof of this theorem can be found in [6].

**Theorem 3** *The presence of Gaussian noise in the mixture does not change asymptotically (i.e., for a sufficiently long data set) the convergence properties of the CII1 and CII2 algorithms.*

This theorem is a consequence of using higher order cumulants as nonlinearities in the algorithms.

#### 4. SIMULATIONS

Computer simulations were carried out to illustrate the performance of the proposed algorithms. In our computer experiment, we chose three different sources of normalized unit power that can be seen in figure 1-a). The first source has a Laplacian p.d.f. and, therefore, positive kurtosis. The other two sources are deterministic signals of negative kurtosis. We considered the following mixing system  $\mathbf{A} = [1, -0.5, 1; -0.5, 1, 0.5; -0.5, 0.5, 1]$  and a poor input signal to noise ratio of 5 dB. The observations are shown in figure 1-b). The algorithms performance is measured in terms of the following index

$$P_{index} = \sum_{i=1}^N \left( \sum_{j=1}^N \frac{|G_{ij}|^2}{\max_l \{|G_{il}|^2\}} - 1 \right) + \sum_{j=1}^N \left( \sum_{i=1}^N \frac{|G_{ij}|^2}{\max_l \{|G_{lj}|^2\}} - 1 \right) \quad (16)$$

The algorithms CII1 and CII2 have been implemented in a batch way using a data block of 5000 observations to estimate the statistical averages. The parameters were set to  $\eta = 0.9$ ,  $p = 1$  and  $\beta = 3$ . The convergence results of the CII2 algorithm are shown in figure 2. The signal component of the outputs reveals the

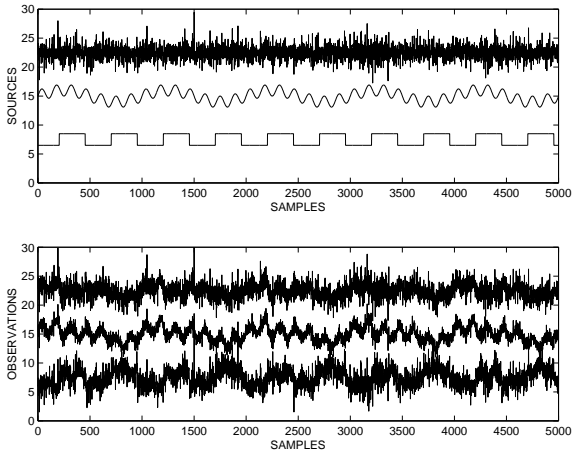


Figure 1: a) Sources (upper fig.). b) Observations (bottom fig.).

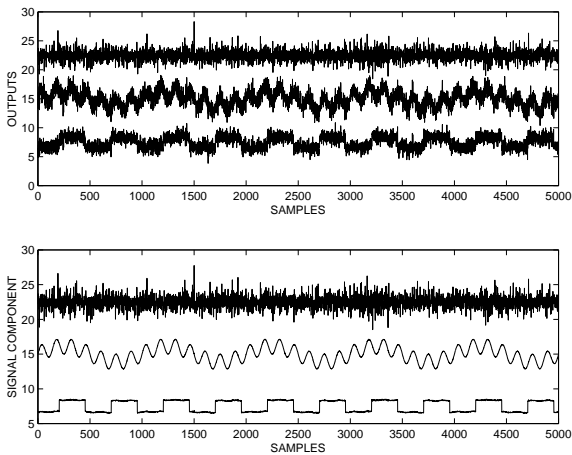


Figure 2: a) Outputs of the CII2 algorithm (upper fig.). b) Signal component of the outputs (bottom fig.).

success of the separation. Additionally, we can verify that although the three sources have different kurtosis sign the algorithm converges to the separation.

A comparison between the convergence of algorithms CII1 and CII2 can be observed in figure 3 for the noisy and noiseless cases. Note that convergence is achieved in only 15 iterations. This fast convergence is due to the quasi-Newton nature of the algorithms. For a finite set of data the involved statistics of the outputs are, up to some extent, sensible to the Gaussian noise, resulting in a slightly poorer convergence for the noisy case. However, this sensitivity asymptotically disappears as the number of samples increases.

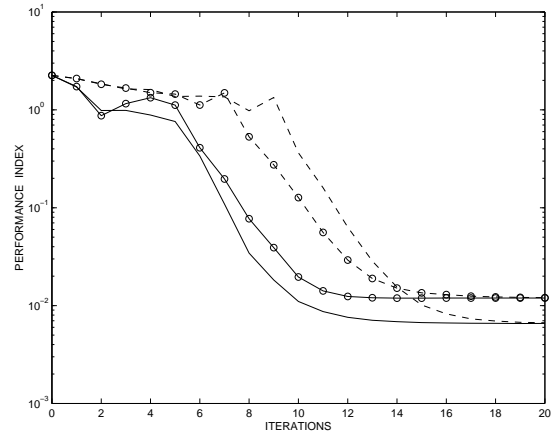


Figure 3: Performance index versus iterations. Noiseless case: CII2 algorithm (continuous line), CII1 algorithm (dashed line). Noisy case: CII2 algorithm (“-o-”), CII1 algorithm (“- -o-”).

## 5. CONCLUSIONS

It is well known that ENTMAX criterion fails to obtain separation when the p.d.f. of the sources is unknown. We have shown that when using cumulants as algorithm nonlinearities the separation solution is always a saddle point of the resulting entropy. Exploiting this property we have proposed two novel asymptotically equivariant algorithms for blind source separation whose local convergence is isotropic and independent on the source statistics, even in the presence of Gaussian noise.

## 6. REFERENCES

- [1] Cardoso, J.-F. (1998) Blind Signal Separation: Statistical Principles *Proceedings of the IEEE* **86**(10):2009–2025.
- [2] Comon, P. (1994) Independent component analysis, a new concept? *Signal Processing* **3**(36):287–314.
- [3] Amari, S.I. & Cardoso, J.F. (1997) Blind source separation – semiparametric statistical approach. *IEEE Transactions on Signal Processing* **45**(11):2692–2697.
- [4] Bell, A.J. & Sejnowski, T.J. (1996) An information maximization approach to blind separation and blind deconvolution, *Neural Computation* **7**:1129–1159.
- [5] Cruces, S. (1999) *An unified view of blind source separation algorithms*. Ph.D. Thesis, University of Vigo, Spain.
- [6] Cruces, S., Castedo, L., Cichocki, A. (1999) *Equivariant Blind Source Separation Using Cumulants*. Submitted to the IEEE Transactions on Signal Processing, 1999.