A Generalized ICA Algorithm

Benoît Stoll, Eric Moreau

A Generalized ICA Algorithm

Benoit STOLL and Eric MOREAU,[1]

MS-GESSY, ISITV, av. G. Pompidou, BP56,
F-83162 La Valette du Var Cedex, France
moreau@isitv.univ-tln.fr
fax: +33 4 94 14 24 48

Submitted to IEEE Signal Processing Letters, November 1999
EDICS number: SPL.SP.3.2

Abstract

In this paper, we consider the source separation problem through a block algorithm based on the maximization of contrast functions. We propose a new contrast with parameterized cross-cumulants. It allows us to put three classical contrast in a common framework. Following the same spirit of the ICA algorithm, we derive the analytical solution for the case of two sources. Finally, a computer simulation is performed to illustrate the behaviour of a Jacobi-like algorithm for the maximization of the new contrast.

[1]Eric Moreau is the corresponding author.

I. Introduction

We consider the source separation problem where several linear spatial mixtures of some independent signals called "sources" are observed as $\boldsymbol{x}(n) = \boldsymbol{G}\boldsymbol{a}(n)$ where $n \in \mathbb{Z}$ is the discrete time, $\boldsymbol{a}(n)$ the $(N,1)$ vector of $N \in \mathbb{N}^* \setminus \{1\}$ inobservable *real* input sources $a_i(n)$, $\boldsymbol{x}(n)$ the $(N,1)$ vector of observations $x_i(n)$ and $\boldsymbol{G}$ the (N,N) square mixing matrix assumed *invertible*. For clarity, we restrict our attention to the real case even though the following derivations could easily be extended to the complex one. The sources $a_i(n)$ are assumed zero-mean, unit power, statistically mutually independent and stationary, *i.e.* the R-th order cumulant $\mathsf{Cum}\,[\underbrace{a_i(n),\ldots,a_i(n)}_{R\times}]$ is an independent function of n, denoted by $\mathsf{C}_R[a_i]$. Moreover, we assume that at most one of the cumulants $\mathsf{C}_R[a_i]$ is zero. The separation problem consists in estimating a matrix $\boldsymbol{H}$ in such a way that the vector $\boldsymbol{y}(n) = \boldsymbol{H}\boldsymbol{x}(n)$ restores one of the different input sources on each of its components. An interesting way to get a solution in the source separation problem, is to perform the optimization of so-called "contrast functions" [1],[2]. They have to be maximized to get a separating solution.

In this paper, we consider a block approach based on high-order (higher than two) statistics. In this field, the ICA algorithm [1], the STOTD algorithm [3] and the JADE algorithm [4] constitute three important contributions. These algorithms rely on the optimization of a specific contrast function. They are Jacobi-like algorithms where, after a whitening of the observations, the final unitary matrix to estimate is decomposed in a product of Givens (or plane) rotations. Hence they process all different possible pairs of outputs. For this task, it has required the analytical derivation of the plane rotation component for the case of two sources.

Our main objective in this paper is to show that the three contrasts optimized by ICA, STOTD and JADE belong in a same family of contrast where the cross-cumulants are parameterized. Furthermore we derive the analytical maximization argument of the generalized contrast for the case of two sources and illustrate the behavior of what we call the "generalized ICA" by a computer simulation.

II. A generalized contrast

Let us recall that contrast functions was introduced for the source separation problem in [1] and recently generalized in [2]. Basically a contrast is a multivariate function defined on a certain set $\mathcal{Y}$ of random vectors $\boldsymbol{y}$ which only depends on the probability density of $\boldsymbol{y}$ and whose global

maxima only correspond to some separation solutions. Hence the (global) maximization of a contrast solves the separation problem. In this paper the set $\mathcal{Y}$ is the set of white vectors, *i.e.* vectors $\boldsymbol{y}(n)$ such that $\mathsf{E}[\boldsymbol{y}\boldsymbol{y}^T] = \boldsymbol{I}$. As a consequence [1] the maximization of a given contrast has to be realized onto the set of unitary matrices.

Let us define the following four functions:

$$\mathcal{I}(\boldsymbol{y}) = \sum_{i_1=1}^{N} (\mathsf{C}_4[y_{i_1}])^2 \; ; \qquad \mathcal{C}_1(\boldsymbol{y}) = \sum_{\substack{i_1,i_2=1 \\ i_2 \neq i_1}}^{N} (\mathsf{Cum}[y_{i_1}, y_{i_1}, y_{i_1}, y_{i_2}])^2 \; ;$$

$$\mathcal{C}_2(\boldsymbol{y}) = \sum_{\substack{i_1,i_2=1 \\ i_2 > i_1}}^{N} (\mathsf{Cum}[y_{i_1}, y_{i_1}, y_{i_2}, y_{i_2}])^2 \; ; \qquad \mathcal{C}_3(\boldsymbol{y}) = \sum_{\substack{i_1,i_2,i_3=1 \\ i_3 \neq i_2 \neq i_1 \\ i_3 > i_2}}^{N} (\mathsf{Cum}[y_{i_1}, y_{i_1}, y_{i_2}, y_{i_3}])^2 \; .$$

We propose the following result:

Proposition 1. *If* $(\alpha_1, \alpha_2, \alpha_3) \in (-\infty, 1]^3$*, the function*

$$\mathcal{J}_{\alpha_1,\alpha_2,\alpha_3}(\boldsymbol{y}) = \mathcal{I}(\boldsymbol{y}) + 2(\alpha_1 \mathcal{C}_1(\boldsymbol{y}) + \alpha_2 \mathcal{C}_2(\boldsymbol{y}) + \alpha_3 \mathcal{C}_3(\boldsymbol{y})) \tag{1}$$

is a contrast for white vectors $\boldsymbol{y}$*.*

Proof. Because $\alpha_k \leq 1$ for $k = 1, 2, 3$, we have $\mathcal{J}_{\alpha_1,\alpha_2,\alpha_3}(\boldsymbol{y}) \leq \mathcal{J}_{1,1,1}(\boldsymbol{y})$. Now recalling that $\mathcal{J}_{1,1,1}(\boldsymbol{y})$ is a contrast (see e.g. [4] or [5]) then $\mathcal{J}_{1,1,1}(\boldsymbol{y}) \leq \mathcal{J}_{1,1,1}(\boldsymbol{a})$. Because the sources are statistically independent their cross-cumulants are zero and thus $\mathcal{C}_1(\boldsymbol{a}) = \mathcal{C}_2(\boldsymbol{a}) = \mathcal{C}_3(\boldsymbol{a}) = 0$. Then $\mathcal{J}_{1,1,1}(\boldsymbol{a}) = \mathcal{I}(\boldsymbol{a}) = \mathcal{J}_{\alpha_1,\alpha_2,\alpha_3}(\boldsymbol{a})$ and considering altogether the above results $\mathcal{J}_{\alpha_1,\alpha_2,\alpha_3}(\boldsymbol{y}) \leq \mathcal{J}_{\alpha_1,\alpha_2,\alpha_3}(\boldsymbol{a})$. Then the function $\mathcal{J}_{\alpha_1,\alpha_2,\alpha_3}(\boldsymbol{y})$ has to be maximized. Finally, because $\mathcal{J}_{1,1,1}(\boldsymbol{y})$ is a contrast, it is not difficult to see that the equality holds in $\mathcal{J}_{\alpha_1,\alpha_2,\alpha_3}(\boldsymbol{y}) \leq \mathcal{J}_{\alpha_1,\alpha_2,\alpha_3}(\boldsymbol{a})$ only for separating states. Thus $\mathcal{J}_{\alpha_1,\alpha_2,\alpha_3}(\boldsymbol{y})$ is a contrast. □

Let us make three remarks.

R1: Let us notice that if $\alpha_1 = \alpha_2 = \alpha_3 = 0$ then $\mathcal{J}_{0,0,0}(\boldsymbol{y}) = \mathcal{I}(\boldsymbol{y})$ which is the ICA based contrast. Also if $\alpha_1 = \alpha_2 = \alpha_3 = 1$ then $\mathcal{J}_{1,1,1}(\boldsymbol{y})$ is the JADE based contrast. Finally if $\alpha_1 = 0.5$ and $\alpha_2 = \alpha_3 = 0$ then $\mathcal{J}_{\frac{1}{2},0,0}(\boldsymbol{y})$ is the STOTD based contrast. All other values of α_k, $k = 1, 2, 3$, yields a new contrast.

R2: For simplicity we have only considered fourth-order cumulants. But, using the results in [5], the proposition 1 can be easily generalized to any order of cumulants if it is greater or equal to three.

R3: If a given contrast have cross-cumulants then following the same principle of the above proof,

one can parameterized alike its cross-cumulants. This is done here in Proposition 1 using the contrast $\mathcal{J}_{1,1,1}(\boldsymbol{y})$.

III. The case of two sources

Let us now consider the specific case of $N = 2$ for which we would like to determine an argument of the maximization of $\mathcal{J}_{\alpha_1,\alpha_2,\alpha_3}(\boldsymbol{y})$. For this task we follow now the same derivations as in [1]. Using the following notation $\mathsf{Y}_{i_1,i_2,i_3,i_4} = \mathsf{Cum}[y_{i_1}, y_{i_2}, y_{i_3}, y_{i_4}]$, the generalized contrast in (1) can be written as

$$\mathcal{J}_{\alpha_1,\alpha_2}(\boldsymbol{y}) = \mathsf{Y}^2_{1,1,1,1} + \mathsf{Y}^2_{2,2,2,2} + 2\alpha_1(\mathsf{Y}^2_{1,1,1,2} + \mathsf{Y}^2_{1,2,2,2}) + 2\alpha_2\mathsf{Y}^2_{1,1,2,2} \tag{2}$$

where we have drop the dependency of the contrast w.r.t. α_3 because the term $\mathcal{C}_3(\boldsymbol{y})$ does not exist. Assuming that a first whitening stage have been realized on the observations, we have then to determine a unitary matrix $\boldsymbol{H}$ which we parameterized as $\boldsymbol{H} = \frac{1}{\sqrt{1+\theta^2}}\begin{pmatrix} 1 & \theta \\ -\theta & 1 \end{pmatrix}$. Introducing the notation $\xi = \theta - \frac{1}{\theta}$, the contrast $\mathcal{J}_{\alpha_1,\alpha_2}(\boldsymbol{y})$ in (2) can be written as $\mathcal{J}_{\alpha_1,\alpha_2}(\boldsymbol{y}) = (\xi^2+4)^{-2}\sum_{k=0}^{4} b_k\xi^k$ where

$$\begin{aligned}
b_4 &= A_1 + 2\alpha_1 A_2 + 2\alpha_2 A_3 \\
b_3 &= 4(2-\alpha_1)A_4 + 4(3\alpha_1 - 2\alpha_2)A_8 \\
b_2 &= 2(2+\alpha_1)A_1 + 4(4-\alpha_1+2\alpha_2)A_2 + 4(9\alpha_1 - 2\alpha_2)A_3 + 4(3-3\alpha_1+\alpha 2)A_5 + \\
&\quad 4(6\alpha_1 - 4\alpha_2)A_9 \\
b_1 &= 8(3-\alpha_2)A_4 + 8(1-2\alpha_1+\alpha_2)A_6 + 16(3-3\alpha_1+\alpha 2)A_8 \\
b_0 &= 2(1+2\alpha_1+\alpha_2)A_1 + 16(2+\alpha_1)A_2 + 8(9+\alpha_2)A_3 + 8(3-\alpha 2)A_5 + \\
&\quad 4(1-2\alpha_1+\alpha_2)A_7 + 32(2-\alpha_1)A_9
\end{aligned} \tag{3}$$

where, using $\mathsf{X}_{i_1,i_2,i_3,i_4} = \mathsf{Cum}[x_{i_1}, x_{i_2}, x_{i_3}, x_{i_4}]$, the values of A_k, $k = 1, \ldots, 9$, are

$$\begin{aligned}
A_1 &= \mathsf{X}^2_{1,1,1,1} + \mathsf{X}^2_{2,2,2,2} & A_6 &= \mathsf{X}_{1,1,1,2}\mathsf{X}_{2,2,2,2} - \mathsf{X}_{1,1,1,1}\mathsf{X}_{1,2,2,2} \\
A_2 &= \mathsf{X}^2_{1,1,1,2} + \mathsf{X}^2_{1,2,2,2} & A_7 &= \mathsf{X}_{1,1,1,1}\mathsf{X}_{2,2,2,2} \\
A_3 &= \mathsf{X}^2_{1,1,2,2} & A_8 &= \mathsf{X}_{1,1,2,2}(\mathsf{X}_{1,2,2,2} - \mathsf{X}_{1,1,1,2}) \\
A_4 &= \mathsf{X}_{1,2,2,2}\mathsf{X}_{2,2,2,2} - \mathsf{X}_{1,1,1,1}\mathsf{X}_{1,1,1,2} & A_9 &= \mathsf{X}_{1,1,1,2}\mathsf{X}_{1,2,2,2} \\
A_5 &= \mathsf{X}_{1,1,2,2}(\mathsf{X}_{2,2,2,2} + \mathsf{X}_{1,1,1,1})
\end{aligned}$$

Now the derivative of $\mathcal{J}_{\alpha_1,\alpha_2}$ w.r.t. ξ is $\frac{d\mathcal{J}_{\alpha_1,\alpha_2}}{d\xi} = (\xi^2+4)^{-3}\sum_{k=0}^{4} c_k\xi^k$ where

$$c_4 = -b_3 \; ; \quad c_3 = 16b_4 - 2b_2 \; ; \quad c_2 = 12b_3 - 3b_1 \; ; \quad c_1 = 8b_2 - 4b_0 \; ; \quad c_0 = 4b_1 \; .$$

Now one has to calculate the values of ξ such that $\sum_{k=0}^{4} c_k\xi^k = 0$. This can be realized in closed-form since it is the zeros of a polynomial of degree 4. Then we keep the ξ leading to the maximal value of the contrast. Finally, the value of θ is determined from the zeros of $\theta^2 - \xi\theta - 1$ in keeping the one in $(-1, 1]$. The Jacobi-like algorithm based on the above developments is denoted gICA(α_1, α_2) for "generalized ICA" of parameters (α_1, α_2). The implementation follows the same lines as the original ICA [1].

IV. Computer simulation

In order to illustrate the behaviour of the gICA algorithm for different values of α_1 and α_2 a computer simulation is presented in the case of five sources. We consider discrete i.i.d. signal called PS(τ) [5] which takes its values in the set $\{-1\ ,\ 0\ ,\ \tau\}$ with the respective probability $\{\frac{1}{1+\tau}\ ,\ \frac{\tau-1}{\tau}\ ,\ \frac{1}{\tau(1+\tau)}\}$. This gives a simple way to parameterize the fourth-order cumulant of the source which is $\mathsf{C}_4[a] = \tau^2 - \tau - 2$. The first two sources are PS(1.5) (negative fourth-order cumulant), the following two are PS(2.5) (positive fourth-order cumulant) and the last one is Gaussian (i.i.d.). We use blocks of $N_d = 200$ data. The components of the mixing matrix $\boldsymbol{G}$ are chosen randomly with an uniform law between -1 and $+1$ and is kept constant. The condition number of the considered matrix is 5.02. The performances of the algorithm are associated to a non-negative index ind$(\cdot)$ [5] defined on the global matrix $\boldsymbol{S} = \boldsymbol{H}\boldsymbol{G}$. A small value of ind$(\cdot)$ indicates the proximity to one separating solution. The mean and standard deviation (STD) of the estimated index ind$(\cdot)$ over 100 Monte-Carlo runs are plotted with a solid line for gICA$(0,0)$, a dotted line for gICA$(0,1)$, a dashdot line for gICA$(1,0)$ and a dashed line for gICA$(1,1)$ as a function of sweeps[2]. The figure shows that the performance of the gICA algorithm can depend on the parameters (α_1, α_2). In the presented case, gICA$(1,0)$ has the best performances.

V. Conclusion and discussion

We have proposed a new contrast function in parameterizing some cross-cumulants and derived the analytical solution in the case of two sources. A computer simulation illustrate the behavior of the algorithm for four different values of the parameters. Clearly works remain to be done to

[2] A sweep is one "iteration" consisting in processing the outputs through all the possible pairs.

validate the usefulness of the proposed "generalized" algorithm. The simulation show (onto one example) that one could take advantage of the parameters e.g. to derive an optimal contrast. On the other hand, as exemplified in [6] for the ICA and JADE algorithms, performances of this kind of approach can significantly differ for sources with different fourth-order cumulant and with a non-Gaussian background noise. This can be a direction of future works.

REFERENCES

[1] P. Comon, "Independent component analysis, a new concept?", *Signal Processing*, Vol. 36, pp 287-314, 1994.

[2] E. Moreau and N. Thirion-Moreau, "Nonsymmetrical contrasts for source separation", *IEEE Trans. Signal Processing*, Vol. 47, No. 8, pp 2241-2252, August 1999.

[3] L. De Lathauwer, B. De Moor and J. Vandewalle, "Blind source separation by simultaneous third-order tensor diagonalization", in Proc. *EUSIPCO'96*, Trieste, Italy, pp 2089-2092, Sept. 1996.

[4] J.F. Cardoso and A. Souloumiac, "Blind Beamforming for non Gaussian Signals", *IEE Proceedings-F*, Vol. 40, pp 362-370, 1993.

[5] E. Moreau, "A generalization of joint diagonalization criteria for source separation", *IEEE Trans. Signal Processing*, submitted, March 1999.

[6] P. Comon, P. Chevalier and V. Capdevielle, "Performance of contrast-based blind source separation", in Proc. *SPAWC'97*, Paris, France, pp 345-348, April 1997.

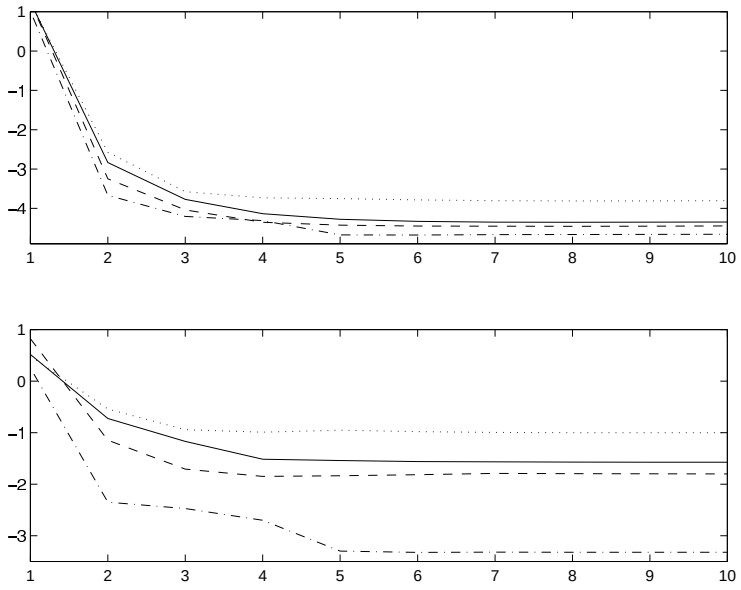

Fig. 1. For $N = 5$, the mean and STD of the index over 100 independent runs is plotted for gICA$(0,0)$ (solid line), gICA$(0,1)$ (dotted line), gICA$(1,0)$ (dashdot line) and gICA$(1,1)$ (dashed line) as a function of sweeps.