4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), April 2003, Nara, Japan

ADAPTIVE INITIALIZED JACOBI OPTIMIZATION IN INDEPENDENT COMPONENT ANALYSIS

† *Juan J. Murillo-Fuentes and* † *Rafael Boloix-Tortosa* ‡ *Francisco J. González-Serrano,*

† ATSC. ESI. Universidad de Sevilla. Paseo de los Descubrimientos sn, 41092 Sevilla, Spain.
‡ DT. EPS, Universidad Carlos III de Madrid. Butarque 15, 28911 Leganés, Madrid, Spain.
E-mail: murillo@esi.us.es

ABSTRACT

In this paper, we focus on the fourth order cumulant based adaptive methods for independent component analysis. We propose a novel method based on the Jacobi Optimization, available for a wide set of minimum entropy (ME) based contrasts. In this algorithm we adaptively compute a moment matrix, an estimate of some fourth order moments of the whitened inputs. Starting from this matrix, the solution to the n-dimensional ME ICA problem may be solved at any time by means of the Jacobi Optimization approach. In the experiments included, we compare this method to previous ones such as the adEML or the EASI, obtaining a better performance.

1. INTRODUCTION

Independent Component Analysis (ICA) [1] involves the task of computing the matrix projection of a set of components onto another set of so called independent component. Here, the objective is to minimize the statistical independence of the outputs. If we know the inputs to the ICA to be linear instantaneous mixture of a set of sources. The ICA process provides an estimates of the original sources. Here, and in the context of this paper, neither the original sources nor the mixture matrix itself are known. This is the Blind Separation of Sources (BSS) [2] where the aim is to obtain a non-observable set of signals, the so-called sources, from another set of observable signals regarded as mixtures.

In this paper we focus on BSS/ICA approaches based on the minimization of the entropy (ME) [1], [3]. Most of them are based on the minimization of a cost function, the contrast, in the two dimensional case. In the n-dimensional problem the Jacobi Optimization (JO) [1] is used, i.e, we operate pairwise minimizing the associated 2-dimensional contrast for every whitened-signal pair in turn over several sweeps until convergence. This process may be carried out adaptively, as the adEML method in [4]. The adEML algorithm is based on the adaptive learning of a different set of parameters for every pair and sweep. These parameters are functions of the fourth order moments of the outputs for that pair and sweep. Thus, the algorithm is, at the same time, learning the system and the solution.

In this paper we propose a new method where the learning of the system and the computation of the solution are decoupled. On the one hand we adaptively learn the moments of the whitened mixtures. On the other hand, we may compute the solution to the ICA/BSS problem at any moment from these moments by using the JO applied to any 2-dimensional contrast.

The paper is organized as follows. We first introduce the matrix model of the BSS problem. In Section 2, we propose a general contrast to be used in the JO. Next, we extend the adEML to be available for these family of functions. We devote Section 3 to the new algorithm proposed in this paper, the Adaptive Initilized Jacobi Optimization (AIJO). Some experimental results are included in Section 4. We end with main conclusions.

1.1. Matrix model in the BSS/ICA

In the BSS/ICA instantaneous model, the entries of the $n \times 1$ mixture vector $\boldsymbol{x}_t$ at time t are instantaneous linear combinations of the n statistically independent sources (components) $\boldsymbol{s}_t$, i.e., $\boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{s}_t$. We assume $\boldsymbol{x}_t$ to be a stationary ergodic random sequence and that the mixing matrix $\boldsymbol{A}$ is non-singular. Under these assumptions, it is possible to estimate a separation matrix $\boldsymbol{B}$ to obtain the sources as $\boldsymbol{y}_t = \boldsymbol{B}\boldsymbol{x}_t$. This separating matrix $\boldsymbol{B}$ can be decomposed into the product of a whitening $\boldsymbol{W}$ and a rotation $\boldsymbol{V}$ matrix. Hence,

$$\boldsymbol{y}_t = \boldsymbol{B}\boldsymbol{x}_t = \boldsymbol{V}\boldsymbol{W}\boldsymbol{A}\boldsymbol{s}_t = \boldsymbol{V}\boldsymbol{z}_t \quad t = 1, 2, \ldots \tag{1}$$

The fourth order approximation to the ME contrast [1] yields

$$\phi^{ME}(\boldsymbol{y}) \approx \frac{1}{48}\phi_{24}^{ME}(\boldsymbol{y}) = -\frac{1}{48}\sum_i (C_{iiii}^{\boldsymbol{y}})^2 \tag{2}$$

where, for zero-mean signals, $C_{iiii}^{\boldsymbol{y}} = \mathrm{E}[y_i^4] - 3\mathrm{E}[y_i^2]$ are the marginal cumulants or autocumulants of the ith output and

$\mathrm{E}[\cdot]$ denotes mathematical expectation. Notice that contrast ϕ_{24}^{ME} assumes the outputs are decorrelated. Thus, the problem reduces to the computation of matrix $\boldsymbol{V}$. Several approximations to the contrast in (2) have been proposed.

2. GENERALIZED ADAPTIVE METHOD

In this section we first include a general expression for most of the fourth order moments ME based contrast and then rewrite the adEML method for them.

2.1. A General Contrast

The pair (p, q) of whitened inputs may be writen in polar coordinates as $[z_p(t)\ z_q(t)]^{\mathrm{T}} = [r(t)\cos(\beta(t))\ r(t)\sin(\beta(t))]^{\mathrm{T}}$. As matrix $\boldsymbol{V}$ performs a rotation of θ so that $\rho(t) = \theta + \beta(t)$ is the angle of vector $\boldsymbol{y}$ the product $\boldsymbol{y} = \boldsymbol{V}\boldsymbol{z}$ yields

$$\begin{bmatrix} r(t)\cos(\rho(t)) \\ r(t)\sin(\rho(t)) \end{bmatrix} = \boldsymbol{V}(\theta) \begin{bmatrix} r(t)\cos(\beta(t)) \\ r(t)\sin(\beta(t)) \end{bmatrix} \quad (3)$$

If we denote by $(r(t), \alpha(t))$ the zero-mean unit-variance pair of sources $\bar{\boldsymbol{s}}_t = [\bar{s}_p(t)\ \ \bar{s}_q(t)]^{\mathrm{T}}$, a correct estimate should meet $\rho = \hat{\theta} + \beta(t) = \alpha(t) + k\pi/2, k = 1, 2, 3, ...$ The contrasts and estimates of the rotation angles $\hat{\theta}$ may be written in polar form by means of the following complex-valued linear combinations (*centroids*) [5] of the statistics of the outputs

$$\xi = \mathrm{E}[r^4(t)e^{j4\beta(t)}] \quad (4)$$

$$\zeta = \mathrm{E}[r^4(t)e^{j2\beta(t)}] \quad (5)$$

$$\gamma = \mathrm{E}[r^4(t)] - 8 \quad (6)$$

where $j = \sqrt{-1}$. A wide set of estimates accept a general expression, the so called 'weighted estimators' (WE) [6], [7]. In order to extend the WE estimator to the ML [8] or MK [9] case we propose the generalized weighted estimator (GWE) as

$$\hat{\theta}_{GWE}(\omega_\gamma, \omega_\xi) = \frac{1}{4}\angle(\omega_\xi\omega_\gamma\xi + (1-\omega_\xi)\zeta^2) \quad (7)$$
$$0 < \omega_\xi < 1, \quad \omega_\gamma = \pm 1, \gamma$$

where $\angle(\cdot)$ supplies the principal value of its argument. As described in [6], particular cases for this contrast function has been proposed as approximations to the ME contrast function: $\hat{\theta}_{EML} = \hat{\theta}_{GWE}(\gamma, 1)$ [5], $\hat{\theta}_{AML} = \hat{\theta}_{GWE}(\gamma, 1/3)$ [7] and $\hat{\theta}_{MaSSFOC} = \hat{\theta}_{GWE}(\gamma, 1/2)$ [10]. With the GWE in (7) we may rewrite the estimators in [11], MK [9], [12], SKSE or ML [8] as $\hat{\theta}_{GWE}(\pm 1, 1)$. It also may be proved [13] that the minimization of $\phi_{24}^{ME}(\theta)$ yields the angle

$$\hat{\theta}_{SICA} = \hat{\theta}_{GWE}(\gamma, 3/7) \quad (8)$$

2.2. Extended adEML

We first study the JO to extend the previous GWE contrast to the n-dimensional case. In this sense the adEML [4] is and adaptive algorithm to solve the EML contrast based on the JO. We rewrite this algorithm to solve the GWE.

Notice that in the JO we operate pairwise computing the two dimensional estimate $\hat{\theta}_{GWE}$ for every signal pair in turn over several sweeps. As this process is carried out at each iteration t, we may adaptively learn the following statistics at sweep (c) for the pair (p, q) as

$$\xi_t^{(c,pq)} = (1-\nu)\xi_{t-1}^{(c,pq)} + \nu\mathrm{E}[r_{c,pq}^4(t)e^{j4\rho_{c,pq}(t)}] \quad (9)$$

$$\zeta_t^{(c,pq)} = (1-\nu)\zeta_{t-1}^{(c,pq)} + \nu\mathrm{E}[r_{c,pq}^4(t)e^{j2\rho_{c,pq}(t)}] \quad (10)$$

$$\gamma_t^{(c,pq)} = (1-\nu)\gamma_{t-1}^{(c,pq)} + \nu(\mathrm{E}[r_{c,pq}^4(t)] - 8) \quad (11)$$

Besides, as we estimate the rotation matrix $\boldsymbol{V}$ under the whitening constraint, we must compute $\boldsymbol{z}_t = \boldsymbol{W}\boldsymbol{x}_t$. Thus, we first should update the whitening matrix $\boldsymbol{W}$. In the following, we will use the relative gradient based [14] whitening algorithm

$$\boldsymbol{W}_{t+1} = \boldsymbol{W}_t + \lambda \frac{\boldsymbol{I} - (\boldsymbol{W}_t\boldsymbol{x}_t)(\boldsymbol{W}_t\boldsymbol{x}_t)^{\mathrm{T}}}{1 + \lambda|(\boldsymbol{W}_t\boldsymbol{x}_t)^{\mathrm{T}}(\boldsymbol{W}_t\boldsymbol{x}_t)|}\boldsymbol{W}_t \quad (12)$$

The adaptive algorithm yields

Algorithm 1 *Adapative JO for the GWE estimates (AJO-GWE).*

At each sample instant,

1. *Whitening. Update the whitening matrix $\boldsymbol{W}_t$ as in (12) and set $c = 1$, $\boldsymbol{y}_t = \boldsymbol{z}_t = \boldsymbol{W}_t\boldsymbol{x}_t$ and $\boldsymbol{V}_t = I$.*
2. *One sweep (c). For all $g = n(n-1)/2$ pairs (y_p, y_q), i.e., for $1 \le p < q \le n$, do*
   - *(a) Compute $\xi_t^{(c,pq)}$, $\zeta_t^{(c,pq)}$, $\gamma_t^{(c,pq)}$ in (9)-(11).*
   - *(b) Compute the Givens angle $\hat{\theta}_{GWE}^{(pq)}$ in (7) by using $\xi_t^{(c,pq)}$, $\zeta_t^{(c,pq)}$, $\gamma_t^{(c,pq)}$.*
   - *(c) If $\hat{\theta}_{GWE}^{(pq)} > \theta_{min}$, do update the rotation matrix $\boldsymbol{V}$ and rotate the pair (y_p, y_q) with rotation angle $\hat{\theta}_{GWE}^{(pq)}$.*
3. *End? If the number of sweep c satisfies $c = 1 + \sqrt{n}$ or no angle $\hat{\theta}_{GWE}^{(pq)}$ has been updated, compute the separation matrix as $\boldsymbol{B}_t = \boldsymbol{V}\boldsymbol{W}_t$ and stop. Otherwise go to step 2 for another sweep with $c \leftarrow c + 1$.*

3. ADAPTIVE INITILIZED JACOBI OPTIMIZATION

In the previous section the 'Jacobi optimization' was introduced to extend the problem to n dimensions. In the step 2.a of Algorithm 1, the Givens angle θ_{pq} is computed by using equations (9)-(11). Simple calculus and trigonometrics show that these statistics $\xi_t^{(c,pq)}$, $\zeta_t^{(c,pq)}$ and $\gamma_t^{(c,pq)}$ may be written as a function of the moments of the outputs $\mathrm{E}[y_p^2 y_q^2], \mathrm{E}[y_p^4], \mathrm{E}[y_q^4], \mathrm{E}[y_p y_q^3]$, and $\mathrm{E}[y_p^3 y_q]$ at sweep c. Bearing this in mind, we will face next the computation of the whole set of moments just one time at an initial stage and then rotate them at each step of the algorithm.

Proposition 1 *Given the model $\boldsymbol{y}_t = \boldsymbol{V}\boldsymbol{z}_t$ in (1), there exist a symmetric $l \times l$, $l = n(n+1)/2$, matrix*

$$\boldsymbol{M}(a(k,l), b(i,j)) = \mu^z_{ijkl}, \tag{13}$$

a diagonal constant matrix $\boldsymbol{S}$ and vectors $\boldsymbol{v}_{pp}, \boldsymbol{v}_{pq}$ and $\boldsymbol{v}_{qq}$ such that the fourth order moments of the outputs, y_p and y_q, yields

$$E[y_p^2 y_q^2] = \boldsymbol{v}_{pp}\boldsymbol{S}\boldsymbol{M}\boldsymbol{S}\boldsymbol{v}_{qq}^{\mathrm{T}} \tag{14}$$

$$E[y_p^4] = \boldsymbol{v}_{pp}\boldsymbol{S}\boldsymbol{M}\boldsymbol{S}\boldsymbol{v}_{pp}^{\mathrm{T}} \tag{15}$$

$$E[y_q^4] = \boldsymbol{v}_{qq}\boldsymbol{S}\boldsymbol{M}\boldsymbol{S}\boldsymbol{v}_{qq}^{\mathrm{T}} \tag{16}$$

$$E[y_p y_q^3] = \boldsymbol{v}_{pq}\boldsymbol{S}\boldsymbol{M}\boldsymbol{S}\boldsymbol{v}_{qq}^{\mathrm{T}} \tag{17}$$

$$E[y_p^3 y_q] = \boldsymbol{v}_{pp}\boldsymbol{S}\boldsymbol{M}\boldsymbol{S}\boldsymbol{v}_{pq}^{\mathrm{T}}, \tag{18}$$

Where $\boldsymbol{M}$ is a $l \times l$, $l = n(n+1)/2$, symmetric matrix whose entries are the fourth-order moments $\mu^z_{ijkl} : 1 \le i \le j \le n, 1 \le k \le l \le n$. The moment μ^z_{ijkl} is stored in the entry $\boldsymbol{M}(a,b)$, where the column a and row b yield

$$b = \sum_{h=n-i+2}^{n} h + (j-i) : \; 1 \le i \le j \le n \tag{19}$$

$$a = \sum_{h=n-k+2}^{n} h + (l-k) : \; 1 \le k \le l \le n \tag{20}$$

Notice that $\mu_{ijkl} = \mu_{jikl} = \mu_{kjil} = \mu_{ikjl}$. Hence, we only estimate the subset of different moments $\mu^z_{ijkl} \quad 1 \le i \le j \le k \le l \le n$. The computation of (14)-(18) may be rewritten by introducing a pair of 'rotation vectors' to left-right multiply matrix $\boldsymbol{M}$. The entries a of these vectors written as a function of the entries of the unitary matrix $\boldsymbol{V}$ in (1) yields

$$\begin{aligned} \boldsymbol{v}_{pp}(a) &= 2\boldsymbol{V}(p,k)\boldsymbol{V}(p,l) \\ \boldsymbol{v}_{pq}(a) &= \boldsymbol{V}(p,k)\boldsymbol{V}(q,l) + \boldsymbol{V}(p,k)\boldsymbol{V}(q,l) \\ \boldsymbol{v}_{qq}(a) &= 2\boldsymbol{V}(q,k)\boldsymbol{V}(q,l) \end{aligned} \tag{21}$$

where the indexes k, l and a are related through (20). Finally, $\boldsymbol{S}$ is a diagonal matrix whose entries $\boldsymbol{S}(a,a)$, a as in (20), yield

$$S(a,a) = 1 \;\; l \neq 1 \tag{22}$$

$$S(a,a) = 1/2; \; l = 1 \tag{23}$$

The formulation introduced above allows an easy computation of the output statistics for a given rotation matrix, as the entries $\boldsymbol{V}(p,q)$ involved are easily arranged in a pair of rotations vectors.

The matrix moment may be easily updated with each new sample as

$$\boldsymbol{M}_t = (1-\nu)\boldsymbol{M}_{t-1} + \nu\mathcal{M}^z_t \tag{24}$$

where $\mathcal{M}^z_t$ is matrix $\boldsymbol{M}$ in (13) computed with just the sample t of the whitened inputs, $\boldsymbol{z}_t$.

The adaptive algorithm yields

Algorithm 2 *Adapative Initialized JO for the GWE estimates (AIJO-GWE). Set $\boldsymbol{V}_0 = \boldsymbol{I}$. At each sample instant,*

1. *Whitening. Update the whitening matrix $\boldsymbol{W}_t$ as in (12) and $\boldsymbol{z}_t = \boldsymbol{W}_t\boldsymbol{x}_t$.*
2. *Matrix Moment Initialization. Adaptively compute matrix $\boldsymbol{M}_t$ in (24) with $\boldsymbol{z}_t$.*

Each N samples,

1. *Set sweep number $c = 1$.*
2. *One sweep (c). For all $g = n(n-1)/2$ pairs (y_p, y_q), i.e., for $1 \le p < q \le n$, do*
   (a) *Compute moments in (14)-(18) by using $\boldsymbol{M}_t$.*
   (b) *Compute the Givens angle $\hat{\theta}^{(pq)}_{GWE}$ in (7) by using ξ_t, ζ_t, γ_t in (4)-(6) (with $[z_p(t)\;\, z_q(t)]^{\mathrm{T}} = [y_p(t)\; y_q(t)]^{\mathrm{T}}$).*
   (c) *If $\hat{\theta}^{(pq)}_{GWE} > \theta_{min}$, do update the rotation matrix $\boldsymbol{V}_t$ with rotation angle $\hat{\theta}^{(pq)}_{GWE}$.*
3. *End? If the number of sweep c satisfies $c = 1 + \sqrt{n}$ or no angle $\hat{\theta}^{(pq)}_{GWE}$ has been updated, compute the separation matrix as $\boldsymbol{B}_t = \boldsymbol{V}_t\boldsymbol{W}_t$ and stop. Otherwise go to step 2 for another sweep with $c \leftarrow c+1$.*

In Algorithm 2 the learning of the system and the computation of the solution are decoupled. In fact, the AIJO algorithm is divided in two parallel subroutines. On the one hand we update the moments of the outputs with the last sample. On the other hand, we compute the solution $\boldsymbol{B}_t = \boldsymbol{V}_t\boldsymbol{W}_t$ each N samples. The main advantage of this design is that we improve the performance. Notice that in the AJO we update the statistics $\xi^{(c)}$, $\zeta^{(c)}$ and $\gamma^{(c)}$ with samples of the last estimated outputs $\boldsymbol{y}$, and these ones depend on the previous estimations $\xi^{(c)-1}$, $\zeta^{(c-1)}$ and outputs. Thus, convergence in the last sweeps and pairs is conditioned to the behaviour of the first ones. Hence, for large numbers of sources we need to increase the total number of sweeps and the convergence deteriorates.

4. COMPUTATIONAL COMPLEXITY

We now measure the computational burden of the adaptive algorithm presented and compare it with other methods. We will consider a floating point operation (flop) as a real multiplication. At each sample instant algorithm must:

1. Whitening: the whitening algorithm in (12) takes $n^3 + n^2$ flops.

2. Matrix Moment calculation: as described in [15] the number of multiplications and accumulations necessary to compute $\mathcal{M}_t^z$ are approximately $\frac{(n+3)!}{(n-1)!4!}4$. Since there are some duplicated multiplications in the calculation of the moments, this number could be reduced to $\frac{(n+3)!}{(n-1)!4!} + \frac{(n+1)n}{2}$

3. Matrix Moment updating: adaptively computing matrix $\boldsymbol{M}_t$ in (24) takes $(\frac{n(n+1)}{2})^2$ flops.

Each N samples, for each signal pair:

4. Computation of Moments: as described in [15] the number of multiplications and accumulations necessary to compute the necessary moments are approximately $Kg(l^3 + l)$, where $g = \frac{n(n-1)}{2}$, as defined in Algorithm 2, is the number of signal pairs, $l = \frac{n(n+1)}{2}$ is the dimension of the moments matrix and $K \leq 1 + \sqrt{n}$ is the number of sweeps.

5. Computation of $\hat{\theta}^{(pq)}_{GWE}$: using equation (8) it would take about $f = 26$ flops.

6. Rotation: four flops.

This makes $\frac{(n+3)!}{(n-1)!4!} + \frac{(n+1)n}{2} + (\frac{n(n+1)}{2})^2 + n^3 + n^2$ flops per iteration plus less than $(1+\sqrt{n})(\frac{n(n-1)}{2})[(\frac{n(n+1)}{2})^3 + \frac{n(n+1)}{2} + 26 + 4]$ flops each N iterations. This figures can be compared with those of other adaptive methods, such as adEML, AROT [16] and EASI [14]. In [4] authors estimate the number of flops per iteration for these three methods obtaining the following computational complexities: $C_{adEML} = (18 + f)(1 + \sqrt{n})g$, $C_{AROT} = (14 + f)(1 + \sqrt{n})g$, where $f = 26$ when using equation (8), and $C_{EASI} = n^3 + 3n^2 + ln$, where each nonlinearity element takes l flops (e.g., for cubic nonlinearities $l = 2$). An extra number of flops would have to be added in the normalized version. C_{adEML} and C_{AROT} do not include the number of flops in the whitening stage, so $n^3 + n^2$ must be added to those complexities figures.

Hence, the computational burden of Algorithm 2 is always higher than for the adEML, the AROT and the EASI methods when $N = 1$. However, as N increases and for a reduced number of sources we can force the complexity of Algorithm 2 to be below the complexity of the adEML and the AROT, and of the order of the EASI method. This is illustrated in Fig. 1, where we display the number of flops per iteration versus the number of sources needed for these four adaptive methods for values of $N = 50$ and $N = 200$. We can see how for a number of independent sources equal or below $n = 7$ when $N = 50$, or a number of independent sources equal or below $n = 10$ when $N = 200$, the complexity of Algorithm 2 is lower than for those of the adEML and the AROT methods. It can also be observed in Fig. 1 that, when $N = 200$ or higher and $n \leq 8$, the computational burden of Algorithm 2 is of the order of that for the EASI, leading however to a better solution, as described in the next section.

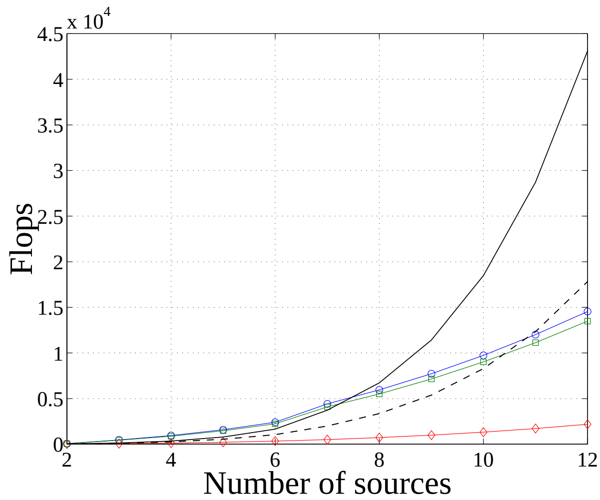

Fig. 1. Computational complexity as a function of the number of sources for adEML(○), AROT(□), EASI(◇) and algorithm 2 with N=50 (solid) and N=200 (dashed).

5. EXPERIMENTAL RESULTS

In the present section the performance of the initialized JO for the GWE estimates (AIJO-GWE) is to be illustrated in a variety of simulations. We will use the estimates in (8), therefore in the following the AIJO-GWE for this estimate will be referred as AISICA. In the experiments we compare AISICA method with other adaptive procedures: the adEML [4] and the EASI [14]. The adaptation coefficient was selected for the whitening stage and the EASI method as $\lambda = 0.005$. The learning rate in this paper was set to $\nu = 0.001$. The solution of the AISICA method was calculated at each sample, i.e. $N = 1$. The performance index

$$Q = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{|p_{ij}|}{\max_k |p_{ik}|} + \sum_{j=1}^{n}\sum_{i=1}^{n} \frac{|p_{ij}|}{\max_k |p_{kj}|} - 2n \quad (25)$$

where $\boldsymbol{P} = (p_{ij}) = \boldsymbol{BA}$, is used as a measure of separation and for the sake of comparing the performance of each method.

As first experiment we face the mixture of three independent sources: a binary sequence, a uniformly distributed process and a sinusoid with random frequency and phase. A random regular 3×3 matrix whose entries are uniformly distributed in $[-1, 1]$ is chosen on each realization. We display in Fig. 2 the trajectories of the global system entries obtained by (a) AISICA, (b) adEML and (c) EASI, respectively, averaged over 1000 independent realizations.

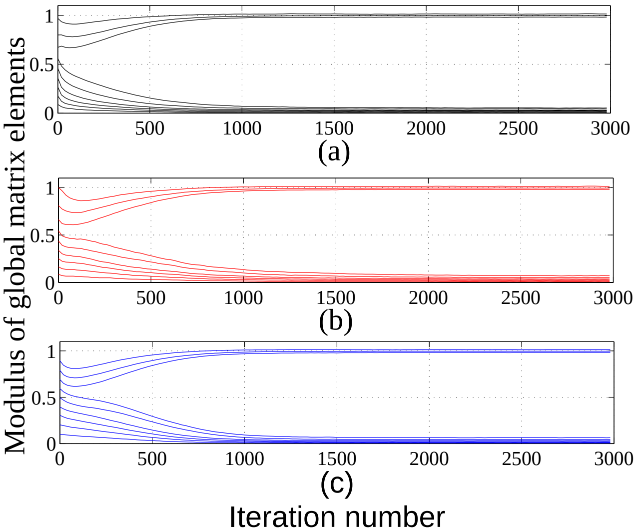

Fig. 2. Modulus of global system coefficients averaged over 1000 mixture realizations for (a) AISICA method, (b) adEML method and (c) EASI method. Number of sources: n=3 (uniform, binary and sinusoid).

In Fig. 5(a) we display the performance index of the three methods. We can see that the performance index for the AISICA method is always lower than those for adEML and EASI, and also the stationary state is reached faster than in the other two cases.

As second experiment, to see the the performance of the three algorithms as the number of sources increases, we face the mixture of eight independent sources. All of them but two were uniformly distributed process, the other two sources were a binary sequence and a sinusoid with random frequency and phase. The values of the parameters in this case are the same that in the first experiment. We can see at Fig. 5(b) the evolution of the performance index for all three methods and see again that the one for the AISICA is always lower than those for adEML and EASI cases, also the stationary state is reached faster than in the other two methods. The results displayed in Fig. 5(b) also illustrate the slow convergence speed of adEML algorithm.

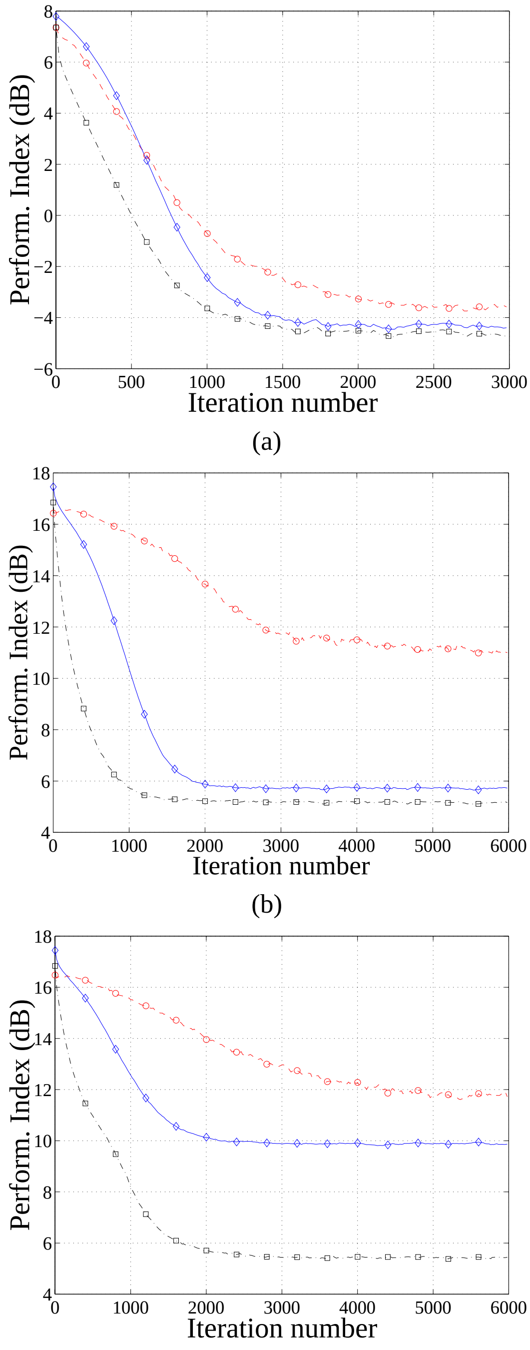

Fig. 3. Performance Index for the AISICA(□), adEML(○) and EASI(◇) methods for (a) n=3 (uniform, binary and sinusoid), (b) n=8 (6 uniform, binary and sinusoid) and (c) n=8 (5 uniform, Laplacian, binary and sinusoid).

We have shown in the previous section that the computational complexity of the AISICA algorithm is higher than that of the EASI method. However, it is important to point out that AISICA has not that limitation concerning which *pdf* the source signal have, that is found in the EASI case. To consider that, we replaced in the previous simulation one of the uniformly distributed sources and introduce a Laplacian distributed one. In Fig.5(c) we display the performance index of the three methods under the described circumstances. The index for the EASI method reaches a higher value in the stationary state in this case than in Fig.5(b), since EASI is not able to separate correctly all the sources, while the index for the AISICA reaches almost the same values as in Fig.5(b).

6. CONCLUSIONS

Based on the adEML method, we have propose a novel algorithm to adaptively compute the solution to te ICA problem. The method is based upon the Jacobi Optimization algorithm and is available for a wide set of fourth order ME contrast. The main advantage of the method is that the learning of the system and the computation of the ICA solution have been decoupled into two different routines. At each sample instant, we update a matrix with moments of the inputs. On the other hand, the ICA solution may be computed by using this matrix at any time. With this new scheme we provide a better performance than the adEML or the EASI algorithms as shown in the experiments. Regarding the computational burden, the complexity of this novel method can be reduced to the order of that of the EASI. Besides, the method is available for virtually any source probability density function.

7. REFERENCES

[1] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[2] J. F. Cardoso, "Blind signal separation: Statistical principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct 1998.

[3] J. F. Cardoso, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, no. 1, pp. 157–192, Jan 1999.

[4] Vicente Zarzoso and Asoke K. Nandi, "Adaptive blind source separation for virtually any source probabiliy density function," *IEEE Transactions on Signal Processing*, vol. 48, no. 2, pp. 477–488, February 2000.

[5] Vicente Zarzoso and Asoke K. Nandi, "Blind separation of independent sources for virtually any source probability density function," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2419–2432, September 1999.

[6] Vicente Zarzoso, Frank Herrmann, and Asoke K. Nandi, "Weighted closed-form estimators for blind source separation," in *11th International Workshop on Statistical Signal Processing*, Singapore, August 2001.

[7] M. Ghogho, A. Swami, and T. Durrani, "Approximate maximum likelihood blind source separation with arbitrary source pdfs," in *IEEE Workshop on Statistical Signal and Array Processing (SSAP'00)*, Pocono Manor Inn, Pennsylvania, Aug 2000.

[8] F. Harroy and J.-L Lacoume, "Maximum likelihood estimators and cramer-rao bounds in source separation," *Signal Processing*, vol. 55, no. 2, pp. 167–177, 1996.

[9] E. Moreau and O. Macchi, "Higher order contrast for self-adaptive source separation," *International Journal of Adaptive Control and Signal Processing*, vol. 10, no. 1, pp. 19–46, Jan 1996.

[10] F. Herrmann and A.K. Nandi, "Blind separation of linear instantaneous mixture using close forms estimators," *Signal Processing*, vol. 81, pp. 1537–1556, 2001.

[11] P. Comon and E. Moreau, "Improved contrast dedicated to blind separation in communications," in *IEEE International Conference on Acoustics, Speach and Signal Processing*, Munich, Germany, 1997, vol. V, pp. 3453–3456.

[12] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non gaussian signals," *Proceedings IEE F*, vol. 140, no. 6, pp. 362–370, Dec 1993.

[13] J.J. Murillo-Fuentes and F. González-Serrano, "Independent component analysis with sinusoidal fouth-order contrast," in *International Conference on Audio, Speech and Signal Processing*, Salt Lake City, USA, May 2001, vol. V, pp. 2785–2788.

[14] J. F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3017–3030, Dec 1996.

[15] Rafael Boloix-Tortosa Juan J. Murillo-Fuentes and Francisco J. González-Serrano, "Initialized jacobi optimization in independent component analysis," in *ICA2003, Submitted*, Nara, Japan, April 2003.

[16] P. Comon, "Separation of stochastic processes," in *Workshop Higher Order Spectral Anal.*, Va, CO, June 1989.