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

# COMBINING BLIND SOURCE EXTRACTION WITH JOINT APPROXIMATE DIAGONALIZATION: THIN ALGORITHMS FOR ICA

*Sergio Cruces*

Signal Processing Group
Ingenieros, Camino Descubrimientos,
41092-Seville, Spain.
http://viento.us.es/~sergio
*sergio@us.es*

*Andrzej Cichocki*

Brain Science Institute
RIKEN, 2-1 Hirosawa, Wako-shi,
Saitama, 351-0198 Japan.
http://www.bsp.brain.riken.go.jp/
*cia@brain.riken.go.jp*

## ABSTRACT

In this paper a multivariate contrast function is proposed for the blind signal extraction of a subset of the independent components from a linear mixture. This contrast combines the robustness of the joint approximate diagonalization techniques with the flexibility of the methods for blind signal extraction. Its maximization leads to hierarchical and simultaneous ICA extraction algorithms which are respectively based on the thin QR and thin SVD factorizations. The interesting similarities and differences with other existing contrasts and algorithms are commented.

## 1. INTRODUCTION

In the last decades powerful criteria and algorithms has been developed to solve the problem of the analysis of the independent components in a linear mixture [1, 2, 3]. In the history of this problem one may distinguish to different approaches: the first one is usually named blind signal separation (BSS) and consists in the simultaneous estimation of all the latent independent components from the mixture; the second one is known as blind signal extraction (BSE) and consists in the estimation of only a subset of the independent components.

The BSE problem can be regarded as more general and flexible than BSS because of the following two reasons: 1) BSE includes BSS as the particular the case where one is interested in all the independent components, 2) BSE has computational advantages over BSS if only a small subset of the independent components is of interest. These computational savings can be important in applications like MEG and EEG, where the decomposition of the observations considers a large number of possible independent components but only a few of them are of interest.

This work has been supported by the CICYT project of the Government of Spain (Grant TIC2001-0751-C04-04).

The original criteria for blind signal extraction were based in the hierarchical the recovery of the independent components [4, 5] one by one alternating extraction with deflation. These criteria have been extended to allow the simultaneous extraction of an arbitrary number of independent components [6, 7]. Another desired property of any BSS/BSE criterion is its robustness, understood in the sense of its ability to give accurate estimates from the available data. Popular techniques for blind signal separation [8, 9, 10] are robust in the previous sense, having criteria based on the joint approximate diagonalization of several cumulant slices. However, up to our knowledge, no equivalent approach for blind signal extraction has been obtained yet. Thus, the purpose of this contribution is to present algorithms that perform the simultaneously extraction a subset of independent components from the mixture by combining the information of several cumulant slices of the observations process.

## 2. SIGNAL MODEL

As shown in figure 1, the chosen signal model for the observations $\mathbf{x}(t) = [x_1(t), \cdots, x_M(t)]^T$ obey the following equation

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{1}$$

where $\mathbf{s}(t) = [s_1(t), \cdots, s_N(t)]^T$ is the signal vector process of $N$ independent components, $\mathbf{n}(t)$ the noise vector process, and $\mathbf{A}$ is a $M \times N$ mixing matrix. We consider the following assumptions:

**A1** The components of $\mathbf{s}(t)$ are mutually independent, locally stationary and normalized to zero mean and unit variance.

**A2** The noise vector process $\mathbf{n}(t)$ is locally stationary, Gaussian, white ($\mathbf{R_n}(t_2, t_1) = \delta(t_2 - t_1)E[\mathbf{n}(t)(\mathbf{n}(t))^H]$) and with a known correlation matrix $\mathbf{R_n}(t, t) =$

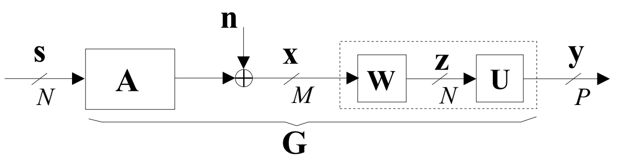


**Fig. 1**. Signal model for the blind extraction of $P$ sources.

$E[\mathbf{n}(t)(\mathbf{n}(t))^H]$ or one that can be accurately estimated from the observations, perhaps using factor analysis or any robust prewhitening technique.

**A3** The mixing matrix $\mathbf{A}$ is full-column rank.

**A4** For a given subset of $P$ independent components $\{s_1(t), \ldots, s_P(t)\}$, that one wishes to extract, there exist a set $\Omega = \{\tau_k = (t_1^{(k)}, \ldots, t_q^{(k)}), k = 1, \ldots, r : \tau_k \in \mathbb{R}^q$ if $q > 2$, $\tau_k \in \mathbb{R}^2 \setminus \{(t,t), \forall t \in \mathbb{R}\}$ if $q = 2\}$ and a permutation $\sigma$ of the indices $1, \ldots, P$ that sorts the following statistic of the components

$$\psi_\Omega(s_j) = \sum_{\tau\in\Omega} w_\tau \left|Cum(s_j(t_1), \cdots, s_j(t_q))\right|^2$$

in such a way that these inequalities hold true

$$\psi_\Omega(s_{\sigma(1)}) > \cdots > \psi_\Omega(s_{\sigma(P)}) > \psi_\Omega(s_{P+1}) \geq \cdots \geq \psi_\Omega(s_N) \quad (2)$$

From **A1** one obtains $\mathbf{A}\mathbf{A}^H = \mathbf{R_x}(t,t) - \mathbf{R_n}(t,t)$. From **A2** and **A3**, using principal component analysis one can project the observations onto the signal subspace to reduce their dimensionality from $M$ to $N$ and also sphere the resulting signals. The $N \times M$ prewhitenning system $\mathbf{W} = (\mathbf{A}\mathbf{A}^H)^{-1/2}$ gives the vector of preprocessed observations

$$\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t) \quad (3)$$

In order to extract $P$ independent components ($1 \leq P \leq N$) from the mixture we use a $P \times N$ matrix $\mathbf{U}$ which is semi-unitary ($\mathbf{U}\mathbf{U}^H = \mathbf{I}_P$). This matrix multiplies the preprocessed observations to give the outputs or estimated sources $\mathbf{y}(t) = [y_1(t), \cdots, y_P(t)]^T$ as

$$\mathbf{y}(t) = \mathbf{U}\mathbf{z}(t) = \mathbf{G}\mathbf{s}(t) + \mathbf{U}\mathbf{W}\mathbf{n}(t) \quad (4)$$

where $\mathbf{G} = \mathbf{UWA}$ denotes the global transfer function from the independent components to the outputs.

Assumption **A4** guarantees that there exist and order relation among the sources that is maximized by those we want to extract.

## 3. EXTRACTION OF A SINGLE SOURCE

We will first analyze the case of $P = 1$, i.e., the extraction of a single source. Then $\mathbf{U}$ and $\mathbf{G}$ are row vectors both of unit 2-norm and there is a single output $y(t) = \mathbf{U}\mathbf{z}(t)$. We propose to estimate the desired independent component by jointly maximizing a weighted square sum of cumulants of fixed[1] order $q \geq 2$, determined by the tuples $\tau = (t_1, \cdots, t_q)$ contained the set $\Omega$. A contrast function that achieves this objective is given by

$$\psi_\Omega(y) = \sum_{\tau\in\Omega} w_\tau \left|Cum\left(y(t_1), \cdots, y(t_q)\right)\right|^2 \quad \text{subject to } \|\mathbf{U}\|_2 = 1. \quad (5)$$

were $w_\tau$ are positive weighting terms. The chosen notation $t_1, \ldots, t_q$ in (5) is due to the fact that from **A1** and **A2** the observation process may be large term non-stationary, case for which the contrast can exploit the same cumulant slices at different times (using segments of quasi-stationary data).

The problem with this approach is in the difficulty of the optimization of (5), which is highly non-linear with respect to $\mathbf{U}$. The following theorem, whose proof is sketched in the Appendix, shows how to circumvent this difficulty by proposing a similar contrast function to (5) but whose dependence with respect to each of the extracting system candidates is quadratic, and thus, much more easy to optimize using algebraic methods.

**Theorem 1** *Under the assumptions* **A1**-**A4***, particularized for the extraction of only one independent component* $s_1$*, there exist a set* $\Omega$ *for which*

$$\psi_\Omega(s_1) > \psi_\Omega(s_j) \qquad \forall j = 2, \ldots, N. \quad (6)$$

*Considering a set of* $q$ *candidates for the extracting system* $\{\mathbf{U}^{[1]}, \ldots, \mathbf{U}^{[q]}\}$ *and the set of their respective outputs* $\bar{y} = \{y^{[1]}, \ldots, y^{[q]}\}$*, the following multivariate function*

$$\psi_\Omega(\bar{y}) = \sum_{\tau\in\Omega} w_\tau \left|Cum(y^{[1]}(t_1), \cdots, y^{[q]}(t_q))\right|^2 \quad \textit{subject to } \|\mathbf{U}^{[m]}\|_2 = 1,\ m = 1, \ldots, q. (7)$$

*where* $w_\tau > 0$*, is a contrast function whose global maximum leads to the extraction of the desired source, i.e., at this extreme point* $y^{[1]}(t) = \cdots = y^{[q]}(t) = s_1(t)$*.*

We should note that this contrast admits a least squares matching interpretation associated with the rank one approximation of cumulant tensors [11]. A good method to maximize the proposed contrast $\psi_\Omega(\bar{y})$ is to optimize it cyclically with respect to each one of the elements $\mathbf{U}^{[m]}$, $m = 1, \ldots, q$, while keeping fixed the others. In the following, the superindex $()^{[k]}$ will continue denoting the $k$-th variable (cyclic notation) while the superindex $()^{(k)}$ will indicate the variable taken value at the $k$-th iteration (sequential notation). Then, note that at the $(k)$th iteration the $[(k$

[1] The results of the paper also apply for an arbitrary combination of cumulants with different orders but, due to the somewhat more cumbersome notation it needs, this extension will be provided elsewhere.

$\mathrm{mod}\ q)+1]$ variable will be optimized. Due to the invariant property of $\psi_\Omega(\bar{y})$ with respect to permutations in its arguments, the cyclic maximization of the contrast is equivalent to the sequential maximization of the function

$$\begin{aligned}\phi_\Omega(\mathbf{U}^{(k)}) &= \sum_{\tau\in\Omega} w_\tau |Cum(y^{(k)}(t_1), y^{(k-1)}(t_2), \cdots \\ &\qquad \ldots, y^{(k-q+1)}(t_q))|^2 \\ &= \mathbf{U}^{(k)}\mathbf{M}^{(k-1)}\mathbf{U}^{(k)H} \qquad (8)\end{aligned}$$

with respect to the extraction system $\mathbf{U}^{(k)}$ through iterations. Note that $\mathbf{M}^{(k-1)}$ is a constant matrix (as long as $\mathbf{U}^{(k-1)}, \cdots, \mathbf{U}^{(k-q+1)}$ are kept fixed) given by

$$\begin{aligned}\mathbf{M}^{(k-1)} &= \sum_{\tau\in\Omega} w_\tau \mathbf{c}_{\mathbf{z}y}^{(k-1)}(\tau)\left(\mathbf{c}_{\mathbf{z}y}^{(k-1)}(\tau)\right)^H \qquad (9)\\ \mathbf{c}_{\mathbf{z}y}^{(k-1)}(\tau) &= Cum(\mathbf{z}(t_1), y^{(k-1)}(t_2), \cdots, y^{(k-q+1)}(t_q))\end{aligned}$$

Since the vector $\mathbf{U}^{(k)H}$ which maximizes $\phi(\mathbf{U}^{(k)})$ is the eigenvector associated to the dominant eigenvalue of $\mathbf{M}^{(k-1)}$, we can move the extracting vector $\mathbf{U}^{(k-1)}$ towards the solution with one or more iterations of any of the standard eigenpair finding methods. Starting from the previous solution, if one considers to use $L$ iterations of the power method to approximate the dominant eigenvector (in practice $L = 1$ works well), the following extraction algorithm is obtained

$$\underline{\mathbf{U}}^{(0)} = \mathbf{U}^{(k-1)}$$

FOR $l = 1 : L$

$$\underline{\mathbf{U}}^{(l)} = \frac{\sum_{\tau\in\Omega} w_\tau d_y^{(l-1)}(\tau)\left(\mathbf{c}_{\mathbf{z}y}^{(k-1)}(\tau)\right)^H}{\left\|\sum_{\tau\in\Omega} w_\tau d_y^{(l-1)}(\tau)\left(\mathbf{c}_{\mathbf{z}y}^{(k-1)}(\tau)\right)^H\right\|_2} \qquad (10)$$

END

$$\mathbf{U}^{(k)} = \underline{\mathbf{U}}^{(L)}$$

where $d_y^{(l-1)}(\tau) = \underline{\mathbf{U}}^{(l-1)}\mathbf{c}_{zy}^{(k-1)}(\tau)$.

### 3.1. Convergence analysis

The iterative optimization of the function $\phi_\Omega(\cdot)$ with respect to $\mathbf{U}^{(k)}$ through iterations will result in a monotonous ascent sequence $\phi_\Omega(\mathbf{U}^{(0)}) \leq \phi_\Omega(\mathbf{U}^{(1)}) \leq \ldots \leq \phi_\Omega(\mathbf{U}^{(k)})$ that maximizes $\phi_\Omega(\cdot)$. However, this property by itself does not guarantee the convergence to an extraction solution because deceptive local maxima of $\phi_\Omega(\mathbf{U}^{(k)})$ might exist. The following theorem (whose proof is sketched in appendix B) shows that this is not the case.

**Theorem 2** *Under assumptions* **A1**-**A4**, *the only local maxima of* $\phi_\Omega(\mathbf{U}^{(k)})$ *correspond with solutions that extract one of the independent components of the mixture.*

## 4. EXTRACTION OF SEVERAL SOURCES

In the case of the extraction of $P$ independent components ($1 \leq P \leq N$) the $P \times N$ matrix $\mathbf{U}^{(k)}$ is semi-unitary with rows $\mathbf{U}_{i:}^{(k)},\ i = 1, \ldots, P$. A result proved in [7] states that any non-negative contrast $\psi_\Omega(y^{[1]}, \ldots, y^{[q]})$ designed for the extraction of a single source, that satisfies (2) and (20), can be used to construct a contrast function for the extraction of the independent components $s_1(t), \cdots, s_P(t)$. Particularized to our case, this leads to the sequential optimization, of

$$\Phi_\Omega(\mathbf{U}^{(k)}) = \sum_{i=1}^{P} \phi_\Omega(\mathbf{U}_{i:}^{(k)}) \quad \text{s.t.} \quad \mathbf{U}^{(k)}(\mathbf{U}^{(k)})^H = \mathbf{I}_P. \qquad (11)$$

In table 1 we consider two choices for the optimization of the previous function which lead to the thin ICA algorithms (TICA).

### 4.1. Hierarchical extraction

A first option (see step 5b, 1st choice) is to hierarchically maximize (11) with respect to the rows $\mathbf{U}_{i:}^{(k)},\ i = 1, \ldots, P$ in such a way that each $i$th row satisfies the constraints $\mathbf{U}_{i:}^{(k)}(\mathbf{U}_{j:}^{(k)})^H = \delta_{ij}\ \forall j \leq i$, i.e., the first rows are less constrained than the last ones. Using Householder reflections the new update is expressed in a very compact and simple form, it is $\mathbf{U}^{(k)} = \mathbf{Q}^H$ where $\mathbf{Q}$ is a tall semi-unitary matrix of dimension $N \times P$ which results from the thin QR decomposition[2] of the weighted statistic $\mathcal{C}_{\mathbf{zy}}^{(k-1)}$ defined in step 5a of table 1.

### 4.2. Simultaneous extraction

A second option (see step 5b, 2nd choice) is to simultaneously maximize $\Phi_\Omega(\mathbf{U}^{(k)})$ with respect to all the rows of $\mathbf{U}^{(k)}$. After defining the Hermitian matrix of multipliers $\mathbf{\Lambda}$, the gradient of the Lagrangian function associated to (11) is

$$\nabla L_\Omega(\mathbf{U}^{(k)}) = \nabla\Phi_\Omega(\mathbf{U}^{(k)}) - \mathbf{\Lambda}\mathbf{U}^{(k)} \qquad (12)$$

Noting that $\nabla\Phi_\Omega(\mathbf{U}^{(k)})$ weakly depends with $\mathbf{U}^{(k)}$, only through a set of the diagonal terms, and approximating these diagonal terms by their current estimates $\mathbf{D}_{\mathbf{y}}^{(k-1)}(\tau)$ we obtain $\nabla\Phi_\Omega(\mathbf{U}^{(k)}) \approx (\mathcal{C}_{\mathbf{zy}}^{(k-1)})^H$. The solution of the equations

$$\begin{aligned}(\mathcal{C}_{\mathbf{zy}}^{(k-1)})^H &= \mathbf{\Lambda}\mathbf{U}^{(k)} \qquad (13)\\ \mathbf{U}^{(k)}(\mathbf{U}^{(k)})^H &= \mathbf{I}_P \qquad (14)\\ \mathbf{\Lambda}^H &= \mathbf{\Lambda} \qquad (15)\end{aligned}$$

[2]The thin QR decomposition and the thin Singular Value Decomposition both have, for $P << N$, a computational complexity of $O(NP^2)$ flops. An efficient implementation of them can be found under the MatLab commands qr$(\cdot, 0)$ and svd$(\cdot, 0)$.

**Table 1**. Summary of the thin ICA algorithms.

1. Set $P \leq N$ the number of independent components to extract from $\mathbf{x}(t)$.
2. Prewhitening $\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t)$
3. Initialization
$$\mathbf{U}^{(0)} = \mathbf{I}_{P\times N};\ \ \mathbf{y}^{(0)}(t) = \mathbf{U}^{(0)}\mathbf{z}(t);\ \ k = 1;$$
4. Estimate $\forall\tau \in \Omega$ the matrices
$$\mathbf{C}_{\mathbf{zy}}^{(k-1)}(\tau) = [\mathbf{c}_{\mathbf{z}y_1}^{(k-1)}(\tau), \ldots, \mathbf{c}_{\mathbf{z}y_q}^{(k-1)}(\tau)]$$
where $\mathbf{c}_{\mathbf{z}y_i}^{(k-1)}(\tau)$ is defined in (18) and initialize $\underline{\mathbf{U}}^{(0)} = \mathbf{U}^{(k-1)}$
5. FOR $l = 1 : L$
   (a) Compute the diagonal matrices
   $$\mathbf{D}_{\mathbf{y}}^{(l-1)}(\tau) = diag\left(\underline{\mathbf{U}}^{(l-1)}\mathbf{C}_{\mathbf{zy}}^{(k-1)}(\tau)\right)$$
   and the weighted sum:
   $$\mathcal{C}_{\mathbf{zy}}^{(l-1)} = \sum_{\tau\in\Omega} w_\tau \mathbf{C}_{\mathbf{zy}}^{(k-1)}(\tau)(\mathbf{D}_{\mathbf{y}}^{(l-1)}(\tau))^*$$
   (b) *Choice* 1. OR 2.
   $$\begin{cases} \text{1. Hierarchical approach:} \\ [\mathbf{Q}, \mathbf{R}] = \mathrm{qr}(\mathcal{C}_{\mathbf{zy}}^{(l-1)}, 0) \\ \underline{\mathbf{U}}^{(l)} = \mathbf{Q}^H \\ \text{2. Simultaneous approach:} \\ [\mathbf{Q}, \mathbf{\Delta}_{P\times P}, \mathbf{V}] = \mathrm{svd}(\mathcal{C}_{\mathbf{zy}}^{(l-1)}, 0) \\ \underline{\mathbf{U}}^{(l)} = \mathbf{V}\mathrm{sign}(\mathbf{\Delta}_{P\times P})\mathbf{Q}^H \end{cases}$$
   END
6. Update $\mathbf{U}^{(k)} = \underline{\mathbf{U}}^{(L)}$ and estimate $P$ independent components:
$$\mathbf{y}^{(k)}(t) = \mathbf{U}^{(k)}\mathbf{z}(t)$$
7. IF Convergence STOP
   ELSE $k = k + 1$; RETURN TO 4

that maximizes $\Phi_\Omega(\mathbf{U}^{(k)})$ is

$$\begin{aligned} [\mathbf{Q}, \mathbf{\Delta}_{P\times P}, \mathbf{V}] &= \mathrm{svd}(\mathcal{C}_{\mathbf{zy}}^{(k-1)}, 0) \\ \mathbf{U}^{(k)} &= \mathbf{V}\mathrm{sign}(\mathbf{\Delta}_{P\times P})\mathbf{Q}^H \end{aligned} \qquad (16)$$

where $\mathrm{svd}(\cdot, 0)$ is the MatLab command for the thin singular value decomposition. The validity of the approximation can be further improved at iteration $k$, by considering $L$ subiterations ($l = 1, \ldots, L$) of a zigzag procedure which updates the estimates $\mathbf{D}_{\mathbf{y}}^{(l-1)}(\tau)$ of the diagonal terms as the solution changes. This is done in step 5. of table 1.

### 4.3. Projection onto the symmetric subspace

From theorem 1 one a priori knows that the solutions $\mathbf{U}^{[m]}$, $m = 1, \ldots, q$ which extract the independent components belong to the symmetric subspace $\mathbf{U}^{[1]} = \ldots = \mathbf{U}^{[q]}$. This information can be exploited to improve the convergence of TICA algorithms just adding, at the end of each iteration, a projection step of $\mathbf{U}^{(k)}, \ldots, \mathbf{U}^{(k-q+1)}$ onto this subspace

$$\mathbf{U}^{(k)}, \mathbf{U}^{(k-1)}, \ldots, \mathbf{U}^{(k-q+1)} \rightarrow \mathbf{U}^{(k)}, \ldots, \mathbf{U}^{(k)} \qquad (17)$$

This is obtained using definition (b) instead of (a)

$$\mathbf{c}_{\mathbf{z}y_i}^{(k-1)}(\tau) = \begin{cases} \text{a) Without projection:} \\ Cum(\mathbf{z}(t_1), y_i^{(k-1)}(t_2), \cdots, y_i^{(k-q+1)}(t_q)) \\ \text{b) With projection:} \\ Cum(\mathbf{z}(t_1), y_i^{(k-1)}(t_2), \cdots, y_i^{(k-1)}(t_q)) \end{cases} \qquad (18)$$

## 5. SIMULATIONS

In this section we illustrate how the algorithm (in similarly with [9]) can be applied to obtain accurate estimates from reduced set of observations. In our example an array of 20 sensors registers 250 snapshots of the observations. These are a random instantaneous mixture of 10 independent signals, in presence of white additive Gaussian noise, and with a maximum signal to noise ratio of 15dB. The desired independent components are the three correlated signals that can be obtained, after normalization, from the filtering of three binary processes by the corresponding systems $F_1(z^{-1}) = (1+0.4z^{-1}+0.9z^{-2}+.5z^{-3})^{-1}$, $F_2(z^{-1}) = (1+0.6z^{-1} - 0.3z^{-2})^{-1}$ and $F_3(z^{-1}) = (1 - 0.7z^{-1})^{-1}$. The other seven independent components are samples of temporally i.i.d. uniform processes. We chose second order statistics $q = 2$ because for short data records like this they are usually the most reliable, and we set $\Omega = \{(t_1, t_1 - 1), (t_1, t_1 - 2), \cdots, (t_1, t_1 - 7)\}$ because these seven pairs guarantee that the considered independent components can be ordered according to (2). We run the algorithm in one hundred random experiments (with different random matrices, sources and noise samples). In each experiment we applied the simultaneous TICA algorithm, with $P = 3$ and $L = 2$, which extracted in all the cases the desired subset sources. As can be observed in figure 2 the convergence of the TICA algorithm is quite fast, requiring between 3 and 6 iterations. Unfortunately, no proper comparison can be given with other algorithms because SOBI and JADE cannot extract a subset of signals while Fast-ICA did not perform well for such a small data set.

## 6. DISCUSSION

Recently, we notice that the contrast function proposed in theorem 1 admits a least squares cumulant matching interpretation associated with the rank one approximation of a set of cumulant tensors. A closely related contrast and an alternating Least Squares technique similar to that we use were previously proposed in [11] to solve the Blind Source Separation problem. The results of this paper bring a complementary insight for the simultaneous extraction problem and lead to the proposal of the thin ICA algorithms. In sequel, we comment some of the interesting links of these algorithms with other existing approaches:

- For $P = 1$, $q = 4$ and $\Omega = \{(t,t,t,t)\}$ the TICA algorithms with projection extract a single independent component by maximizing the modulo of the kurtosis of the output. In this case, both TICA algorithms particularize to the fixed point algorithm (Fast-ICA with cubic non-linearity) proposed in [5]. For arbitrary $P \leq N$ the thin ICA implementation based on the thin QR decomposition is equivalent to the hierarchical application of the fixed point algorithm with deflation. Similarly, the thin ICA implementation based on the SVD reduces to the fixed point algorithm with symmetric orthogonalization when $P = N$ (BSS), and provides a novel extension of it for $P < N$ (BSE).

- For $P = N$ and $q = 2$ (alternatively $q = 4$) the TICA algorithms extract all the independent components using second order statistics (fourth order statistics) and, when it is possible, using also any non-stationarity of the independent components. The criterion (7) is equivalent to that of the SOBI [9] (JADE [8]) algorithm based on the joint approximate diagonalization of a certain set of cumulant slices, although, the implementation differs.

- For $1 < P < N$ and arbitrary $q$ the TICA algorithms extract $P$ independent components using the joint optimization criteria (11). In this case, none of the previously cited algorithms can solve this problem: fixed point algorithms (Fast-ICA) does not perform a joint optimization of several statistics, while SOBI and JADE implementations are not suitable for extraction because the extended Jacobi plane rotations they use are not the most adequate technique for the estimation of a subset of eigenvectors.

## 7. CONCLUSIONS

We have proposed a multivariate contrast function for the extraction of a subset of desired independent components from a linear mixture. This contrast function jointly optimizes several statistics of the same order and have no spurious maxima. We have suggested the thin ICA algorithms for the maximization of the contrast because they combine, at the same time, several of the advantages of other powerful techniques like Fast-ICA, JADE and SOBI.

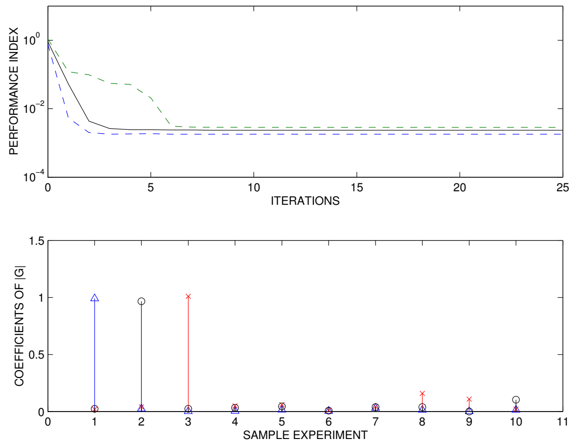


**Fig. 2**. Upper fig.: performance index $P_{index}(\mathbf{G}) = (PN)^{-1}\sum_{i=1}^{P}\left(\|G_{i:}\|_2^2/\|G_{i:}\|_\infty^2 - 1\right)$ versus iterations. Continuous line is the median curve of convergence for 100 experiments, the dashed lines denote the $5^{th}$ and $95^{th}$ percentiles. Lower figure presents the coefficients of the different rows of a $3 \times 10$ matrix $\mathbf{G}$ for one sample experiment.

## A. PROOF OF THEOREM 1

The proof starts observing that under assumptions **A1-A4** the contrast (7) is theoretically unaffected by the additive noise. Then, using Cauchy-Schwarz's inequality one obtains that each square cumulant within the summation is upper bounded by

$$\left|Cum(y^{[1]}(t_1), \cdots, y^{[q]}(t-\tau_q))\right|^2 \leq \sum_{j=1}^{N} |G_{1j}^{[p]}|^2 |Cum(s_j(t_1), \cdots, s_j(t_q))|^2 \cdot \left(\sum_{j=1}^{N} \prod_{m \neq p} |G_{1j}^{[m]}|^2\right)$$

Since the global transfer vectors are normalized

$$\sum_{j=1}^{N} \prod_{m \neq p}^{q} |G_{1j}^{[m]}|^2 \leq \prod_{m \neq p}^{q} \sum_{j=1}^{N} |G_{1j}^{[m]}|^2 = 1 \qquad (19)$$

therefore,

$$\left|Cum(y^{[1]}(t_1), \cdots, y^{[q]}(t_q))\right|^2 \leq \sum_{j=1}^{N} |G_{1j}^{[p]}|^2 \cdot |Cum(s_j(t_1), \cdots, s_j(t_q)|^2$$

Substituting these terms in (7) results in

$$\psi_\Omega(\bar{y}) \quad \leq \quad \sum_{j=1}^{N} |G_{1j}^{[p]}|^2 \psi(s_j) \qquad p = 1, \ldots, q. \quad (20)$$

But from the ordering $\psi_\Omega(s_1) > \psi_\Omega(s_j)\ \forall j = 2, \ldots, N$ in **A4**, and the constraint $\|\mathbf{G}^{[p]}\|_2 = 1$ we finally obtain

$$\psi_\Omega(y^{[1]}, \ldots, y^{[q]}) \quad \leq \quad \psi_\Omega(s_1)\,, \qquad (21)$$

which means that the upper bound coincides with the extraction of the desired independent component. Noting that the equality between both sides of (19) only holds when the row vectors are equal $\mathbf{G}^{[1]} = \cdots = \mathbf{G}^{[q]}$ and aligned with one of the axis $\mathbf{e_j}^T,\ j = 1, \cdots, N$, one can conclude that the global maximum of $\psi_\Omega(\bar{y})$ is only attained at the extraction of the desired source, i.e., when $\mathbf{U}^{[1]} = \cdots = \mathbf{U}^{[q]} = (\mathbf{W}\mathbf{A}\mathbf{e_1})^H$ where $\mathbf{e_1} = (1, 0, \ldots, 0)^T$. □

## B. PROOF OF THEOREM 2

Any critical point $\mathbf{U}'$ of $\phi_\Omega(\cdot)$ has associated a global extraction system $\mathbf{G}' = \mathbf{U}'\mathbf{W}\mathbf{A}$. Following [12] we define the set of indices for which the elements of $\mathbf{G}'$ are nonzero

$$I = \{m : G'_{1m} \neq 0, m = 1, \ldots, N\}, \qquad (22)$$

From the proof of theorem 1 one can observe that any solution which extracts one of the independent components ($I$ has cardinality one) is a local maximum of the contrast. Thus, the deceptive local maximum of $\phi_\Omega(\cdot)$, if they exit, should correspond with $\mathbf{G}'$ having at least two nonzero elements, however, these kind of points cannot be a maximum of the function $\phi(\cdot)$ because we can always find local perturbation of them for which the function increases. For a local perturbation $\boldsymbol{\alpha} \in \mathbb{R}^N$ where: $\|\boldsymbol{\alpha}\|$ sufficient small, $\sum_{j=1}^{N} \alpha_j = 0$, $\alpha_j = 0 \Leftrightarrow j \notin I$, and such that

$$|G_{1j}|^2 = |G'_{1j}|^2 + \alpha_j, \ \ j = 1, \ldots, N; \qquad (23)$$

the function $\phi(\cdot)$, at the perturbed point $\mathbf{U}$, is written as

$$\phi(\mathbf{U}) \quad = \quad (1 + \gamma(\boldsymbol{\alpha}))\phi(\mathbf{U}') + \beta(\boldsymbol{\alpha}) + o(\|\boldsymbol{\alpha}\|^2)$$

where

$$\gamma(\boldsymbol{\alpha}) \quad = \quad \frac{q(q-2)}{4} \sum_{j \in I} \left| \frac{\alpha_j}{G'_{1j}} \right|^2 \qquad (24)$$

$$\beta(\boldsymbol{\alpha}) \quad = \quad \sum_{\tau \in \Omega} w_\tau \left| \frac{q}{2} \sum_{j \in I} \alpha_j \frac{(G'_{1j})^q}{|G'_{1j}|^2} c_{s_j}(\tau) \right|^2 \qquad (25)$$

Since $\gamma(\boldsymbol{\alpha}) \geq 0$ and $\beta(\boldsymbol{\alpha}) > 0$ for all $q \geq 2$ we have that

$$\phi(\mathbf{U}) \quad > \quad \phi(\mathbf{U}')\,, \qquad (26)$$

and we conclude that any solution for which $I$ has cardinality greater than 1 cannot be a local maximum of $\phi_\Omega(\cdot)$. □

## C. REFERENCES


[1] P. Comon, "Independent component analysis, a new concept?," *Signal Proc.*, vol. 3(36), pp. 287–314, 1994.

[2] J. F. Cardoso, "Blind signal separation: Statistical principles," *Proceedings of the IEEE*, vol. 86(10), pp. 2009–2025, 1998.

[3] A. Cichocki, S.-i. Amari, *Adaptive Blind Signal and Image Processing,* John Wiley & Sons, 2002.

[4] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: A deflation approach," *Signal Processing*, 1995, vol. 45, pp. 59–83.

[5] A. Hyvarinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, 1997, vol. 9, pp. 1483–1492.

[6] S. Amari, "Natural gradient learning for over- and under-complete bases in ICA," *Neural Computation*, vol. 11, pp. 1875–1883, 1999.

[7] S. Cruces, A. Cichocki, S-i. Amari, "On a new blind signal extraction algorithm: different criteria and stability analysis", *IEEE Signal Proc. Letters*, vol. 9(8), pp. 233–236, 2002.

[8] J.-F. Cardoso and A. Solumiac, "Blind beamforming for non Gaussian signals," *IEE Proceedings-F*, vol. 140(6), pp. 362–370, 1993.

[9] A. Belouchrani, K. Abel-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. on Signal Processing*, vol. 45(2), pp. 434–444, 1997.

[10] L. De Lathauwer, B. De-Moor, and J. Vandewalle, "Independent component analysis and (simultaneous) third-order tensor diagonalization", *IEEE Trans. on Signal Processing*, vol. 49(10), pp. 2262-2271, 2001.

[11] L. De Lathauwer, P. Comon, B. De-Moor, and J. Vandewalle, "Higher-order power method - application in independent component analysis," in *International Symposium on Nonlinear Theory and Applications NOLTA*, Dec. 1995.

[12] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution," *IEEE Trans. on Information Theory*, vol. 36(2), pp. 312–321, 1990.