

INDEPENDENT COMPONENT ANALYSIS WITH SINUSOIDAL FOURTH-ORDER CONTRASTS

Juan J. Murillo-Fuentes and Francisco J. González-Serrano

Universidad Carlos III. DTC. ATSC.
Butarque, 15, 28911 Leganés, Madrid, Spain.
E-mail : jmurillo@ieee.org <http://alcaudon.tsc.uc3m.es/~murillo/>

ABSTRACT

The authors propose a new solution to the Independent Component Analysis (ICA) problem. In the two-dimensional case, we prove that under the whiteness constraint some fourth-order contrasts may be approximated by a sinusoid. Thus, the minimization of the contrast reduces to computing its phase. The novel approach, called SICA (Sinusoidal ICA), uses the 'Jacobi optimization' to cope with higher dimensions. The method presented has a good performance along with a low computational cost. Some experiments with blind separation of audio and synthetic sources are included to compare the algorithm to other well-known approaches.

1. INTRODUCTION

Consider an $m \times 1$ vector \mathbf{x}_t projected into a space of n components \mathbf{y}_t as statistically independent as possible. The ICA problem consists in finding the change of basis represented by an $n \times m$ matrix B . Let $s_i(t)$ with $i = 1, \dots, n$ be n zero-mean independent unknown sources, where only one of them may be Gaussian. An instantaneous mixture of them may be written as:

$$\mathbf{x}_t = A\mathbf{s}_t \quad t = 1, 2, \dots \quad (1)$$

where A is the mixing matrix. Blind Source Separation (BSS) computes matrix B so that

$$\mathbf{y}_t = B\mathbf{x}_t = BAs_t = C\mathbf{s}_t \quad t = 1, 2, \dots \quad (2)$$

where C is ideally the identity matrix. However, the BSS problem can be solved up to permutations and/or scaling of the outputs [1]. In this sense C is a non mixing matrix if it has one and only one non-zero entry in each column and each row. It can be observed that BSS provides the model (2) to solve the ICA problem.

Matrix B can be decomposed into the product of a whitening W and a rotation V matrix. If $m > n$, the whitening process allows reducing the number of dimensions (we do not consider here the case $n > m$). This spherling stage gives us signals $z_i(t)$:

$$\mathbf{y}_t = B\mathbf{x}_t = VWAs_t = V\mathbf{z}_t \quad t = 1, 2, \dots \quad (3)$$

Most of the off-line solutions to ICA [1] [2] are based on the minimization of one criterion, contrast function, or cancellation of multiple criteria. For instantaneous mixtures of two sources, direct methods, which consist of directly estimating the mixing matrix from the mixtures, are also possible. Direct methods suggested in [3] [4] [5] solve polynomial equations based on cumulants up to

Thanks to spanish CICYT for funding this project (TIC99-0219).

fourth-order. The solutions of the equations provide the entries of the mixing matrix. In this paper we propose to face the problem under the whiteness constraint using the 'Jacobi optimization' [3] [2]. The independent components $y_i(t)$ and $y_j(t)$ in the orthogonal two-dimensional approach may be written as:

$$\mathbf{y}_t = \begin{bmatrix} y_i(t) \\ y_j(t) \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} z_i(t) \\ z_j(t) \end{bmatrix} \quad (4)$$

where $z_i(t)$ and $z_j(t)$ are a rotation α of the normalized sources $\bar{s}_i(t)$ and $\bar{s}_j(t)$.

When the kurtosis of the sources have identical signs, simple orthogonal contrasts may be exhibited [6] [7] [2] (MaxKurt algorithm). In this sense, if all the kurtosis are negative, it is easily proved that minimizing

$$\phi_{Kur}(\theta) = E[y_i^4] + E[y_j^4] \quad (5)$$

subject to $E[\mathbf{y}\mathbf{y}^T] = I$ is achieved only when θ is a demixing angle. If we have no knowledge on the kurtosis, we may use the contrast in [1]:

$$\phi_{ICA}[\mathbf{y}] = -\text{Cum}_{iiii}^2[\mathbf{y}] - \text{Cum}_{jjjj}^2[\mathbf{y}] \quad (6)$$

where $\text{Cum}_{iiii}[\mathbf{y}] = E[y_i^4] - 3E^2[y_i^2]$ are the autocumulants of the outputs. Simple algebra shows that the orthogonal version of this contrast, up to a constant, yields

$$\phi_{ICA}(\theta) \stackrel{c}{=} -f(\theta) + 6\phi_{Kur}(\theta) \quad (7)$$

where $f(\theta) = E^2[y_1^4] + E^2[y_2^4]$.

A similar orthonormal contrast based on fourth-order cross-cumulants, $\phi_{JADE}[\mathbf{y}]$, was introduced and solved by Cardoso *et al.* [8] using joint diagonalization. We will return to it in Section 5.

We first model contrasts $\phi_{Kur}(\theta)$ and $f(\theta)$, then we will compute $\phi_{ICA}(\theta)$ as a linear combination of them.

2. SINUSOIDAL APPROXIMATIONS

The solution to ϕ_{Kur} was first given in [7]. In [2] the contrast was modeled as a sinusoid whose phase was given as a function of the cumulant matrices. Modelling the contrast as a function of the moments in polar co-ordinates is immediate. The vector $\mathbf{y}(t)$ in polar co-ordinates yields

$$\mathbf{y}_t = \begin{bmatrix} y_i(t) \\ y_j(t) \end{bmatrix} = \begin{bmatrix} r(t) \cos(\theta + \beta(t)) \\ r(t) \sin(\theta + \beta(t)) \end{bmatrix} \quad (8)$$

By ignoring the terms of the contrast function that do not depend on the angle θ , the contrast function yields

$$\phi_{Kur}(\theta) \stackrel{c}{=} \frac{1}{4} E[r^4 \cos(4(\theta+\beta))] = q_1 \cos(4\theta) - q_2 \sin(4\theta) \quad (9)$$

where

$$\begin{aligned} q_1 &= \frac{1}{4} E[r^4 \cos(4\beta)] \\ q_2 &= \frac{1}{4} E[r^4 \sin(4\beta)] \end{aligned} \quad (10)$$

Equation (9) is a sinusoid with phase

$$-4\theta_{Kur} = 4\alpha = \arctan(q_2, q_1) \quad (11)$$

where $\arctan(a, b)$ is the four quadrant inverse tangent of a/b . It can be stated that under whiteness constraint, $E[\mathbf{y}\mathbf{y}^\top] = I$, the contrast function $\phi_{Kur}(\theta)$ is a sinusoid whose phase is four times the angle α involved in (4).

We operate as before to model $f(\theta)$ in (5). By using elementary trigonometrics, function $f(\theta)$ yields

$$\begin{aligned} f(\theta) &\stackrel{c}{=} \frac{3}{16} E[r^4] E[r^4 \cos(4(\theta+\beta))] \\ &\quad + \frac{1}{2} E^2[r^4 \cos(2(\theta+\beta))] \\ &\quad + \frac{1}{32} E^2[r^4 \cos(4(\theta+\beta))] \end{aligned} \quad (12)$$

We will show in the next section how the last term in (12) may be suppressed and how function $f(\theta)$ reduces to a sinusoid.

3. THE SINUSOIDAL CONTRAST

The aim of this section is to provide a sinusoidal function whose solution (phase) is the same to that of contrast ϕ_{ICA} in (7). That is, to model ϕ_{ICA} by a sinusoid.

Lemma 1: Under whiteness constraint, the contrast function $\phi_{ICA}(\theta)$ may be modeled by a sinusoid

$$\phi_{ICA}(\theta) \equiv \phi_{SICA}(\theta) = (-l_1 + 6q_1) \cos(4\theta) + (l_2 - 6q_2) \sin(4\theta) \quad (13)$$

its phase being minus four times the rotation angle θ_{ICA}

$$\theta_{ICA} = \theta_{SIN} = -\frac{1}{4} \arctan(-l_2 + 6q_2, -l_1 + 6q_1) \quad (14)$$

with $q_{1,2}$ as given in (13) and $l_{1,2}$ by

$$\begin{aligned} l_1 &= \frac{3}{16} l_{11} + \frac{1}{4} l_{12}^2 - \frac{1}{4} l_{13}^2 \\ l_2 &= \frac{3}{16} l_{21} + \frac{1}{2} l_{22} l_{23} \end{aligned} \quad (15)$$

where

$$\begin{aligned} l_{11} &= E[r^4] E[r^4 \cos(4\beta)] \\ l_{21} &= E[r^4] E[r^4 \sin(4\beta)] \\ l_{12} = l_{22} &= E[r^4 \cos(2\beta)] \\ l_{13} = l_{23} &= E[r^4 \sin(2\beta)] \end{aligned} \quad (16)$$

Proof: See the Appendix.

The proposed lemma states that the solution of the new sinusoidal contrast, SICA, is the same one as for ϕ_{ICA} . But the benefits of this contrast stems from its simplicity: it requires less operations and its solution is more accurate.

Computational considerations: by considering Lemma 1 we reduce the computational burden in the minimization of $\phi_{ICA}(\theta)$. First, we avoid computing the whole set of terms involved in the contrast. And secondly, the minimization of $\phi_{SICA}(\theta)$ is immediate as the solution is minus four times the phase of the resulting function, a sinusoid. Recall that SICA does not include the term by $1/32$ in (12) and needed in (7). In [1] a root of a fourth-order polynomial provides the solution to the orthogonal contrast. Solving this polynomial takes in the MATLAB version of the ICA algorithm by Comon in [9] about 1400 floating point operations. The approach presented in this paper takes only 1 floating point operation in computing the function \arctan . Notice that solving $\phi_{SICA}(\theta)$ takes approximately 10 more floating point operations than in the case of contrast $\phi_{Kur}(\theta)$, the contrast solved by the MaxKurt algorithm in [2].

Accuracy: In polynomial based approaches [1], [4], [5], the estimating error in a given coefficient c_k affects all the zeros. Suppose the roots are located at $\sigma = \sigma_i$ for $i = 1, 2, \dots, N$. Furthermore, let us define the roots as $\sigma_i + \Delta\sigma_i$. The error in the location of the i th zero can be expressed in terms of the errors in the coefficients as

$$\Delta\sigma_i = \sum_{k=1}^N \frac{\partial\sigma_i}{\partial c_k} \Delta c_k, \quad i = 1, 2, \dots, N \quad (17)$$

where

$$\frac{\partial\sigma_i}{\partial c_k} = \frac{\sigma_i^k}{\prod_{j=1, j \neq i}^N (\sigma_i - \sigma_j)} \quad (18)$$

That is, as σ_j/σ_i tends to one the results are more sensitive to errors in the estimation of the coefficients of the polynomial. Henceforth, the performance of the polynomial based methods deteriorates significantly. On the other hand, if one root $\sigma_l \gg \sigma_i$ for $i = 1, 2, \dots, N$ $i \neq l$, from (17) it follows that the sensitivity of σ_l to errors in the estimation of the coefficient c_{N-1} and c_N becomes close to one and σ_l itself respectively. In [1] the larger real root ξ of a fourth-order polynomial is the coefficient of the second-order polynomial $P(\sigma) = \sigma^2 - \xi\sigma - 1$. The zero o of this polynomial in the interval $[-1, 1]$ provides the tangent of the rotation angle $o = \tan(\theta_{ICA})$. If $\xi \gg 1$ and ξ is much larger than the rest of the roots, the sensitivity to the estimation of the upper coefficients is high. In addition, this error propagates to the final solution o as the sensitivity of o to changes in ξ is close to one:

$$\Delta o \approx \Delta\xi \approx \Delta c_{N-1} + \xi \Delta c_N \quad (19)$$

As a conclusion, the methods based on polynomials are ill-conditioned and the performance may deteriorate significantly. On the contrary, although we do not analyze the sensitivity of SICA, methods based on the estimation of the rotation by computing the inverse tangent of a function of the moments usually show a better sensitivity.

4. EXTENSION TO N SOURCES

The algorithm described above was designed for 2 dimensions. Comon introduced in [1] the 'Jacobi optimization' to extend the solution of contrast $\phi_{ICA}(\theta)$ to the n -dimensional problem. We

have rewritten the algorithm using the proposed contrast. Such an algorithm can be summarized as

- 1 *Initialization.* Compute a whitening matrix W and set $\mathbf{y} = W\mathbf{x}$
- 2 *One sweep.* For all $n(n - 1)/2$ pairs, i.e., for $1 \leq i < j \leq n$, do
 - (a) Compute the Givens angle θ_{ij} optimizing $\phi_{SICA}(\mathbf{y})$ when the pair (y_i, y_j) is rotated by using (14).
 - (b) if $\theta_{ij} > \theta_{min}$, do rotate the pair (y_i, y_j) by θ_{ij} according to (4).
- 3 *End?* If no pair has been rotated in previous sweeps, end. Otherwise go to 2 for another sweep.

Thus, the Jacobi approach considers a sequence of 2-dimensional ICA problems. In [2] the value θ_{min} is selected in such a way that rotations by a smaller angle are not 'statistically significant'. Typically $\theta_{min} = 10^{-2}/\sqrt{T}$ where T is the number of samples. In [1] this number is $\theta_{min} = 1/T$ but if the algorithm goes through step 2 more than k times with $k \leq 1 + \sqrt{n}$ it stops.

5. EXPERIMENTAL RESULTS

The performance index

$$Q = \sum_{i=1}^n \left(\sum_{j=1}^n \frac{|p_{ij}|}{\max_k |p_{ik}|} - 1 \right) + \sum_{j=1}^n \left(\sum_{i=1}^n \frac{|p_{ij}|}{\max_k |p_{kj}|} - 1 \right) \quad (20)$$

where $P = (p_{ij}) = BA$, is used as a measure of separation in Fig. 1. Two audio signals were mixed. Each point corresponds to the average of 1000 experiments. In which the mixing matrix is randomly chosen: The matrix entries a_{ij} are random numbers in the range $[-1, +1]$. Fig.1.a shows that contrast functions $\phi_{SICA}(\theta)$ in (13), $\phi_{ICA}(\theta)$ in (13), and $\phi_{JADE}[\mathbf{y}]$ (this last one solved with the JADE algorithm [8]) have similar performance. Regarding the computational cost, the method SICA presented in this paper reduces the number of floating point operations (flops), in computing V and W , in approximately a 52% and 19% with respect to the ICA [9] and the JADE [10] algorithms. Fig.1.b shows the evolution of this feature as the number of samples increases. Fig.1.c includes the CPU time required along the number of samples. This parameter becomes reduced in a 41% and a 67% at $5 \cdot 10^4$ samples. The CPU time is required in this analysis as the algorithms must be optimized taking into account both of the aspects. Notice that an algorithm may require less CPU time using more flops and less accesses to memory.

The experiment in Fig.2 consists in the separation of 10 sources: 5 sources with uniform probability density function (p.d.f.) distributed signals, 1 with gaussian p.d.f. and 4 generated as the cube of the samples of a gaussian distribution. We averaged the results of 100 experiments. Again, the mixing matrix is randomly chosen as before. The sources were also randomly generated at each iteration. Fig.2.a shows the performance index in (20). At low numbers of samples ($[0, 1500]$) $\phi_{SICA}(\theta)$ clearly outperforms the other methods. Besides, the ICA method is unable to perform as accurately as SICA and JADE. Fig.2.d depicts the performance index in the case of a source with uniform p.d.f. and another with the cube of a gaussian. The mixed data used in the figure makes $\xi \approx 10^3$ while the rest of the roots are $\sigma_i \approx 10^{-3}$. Thus, the high sensibility of the method to the estimation of the moments given in

(19), makes the result inaccurate. Fig.2.b show how SICA requires a lower number of flops. Finally, in Fig.2.c it can be observed how the CPU time divides by 2.

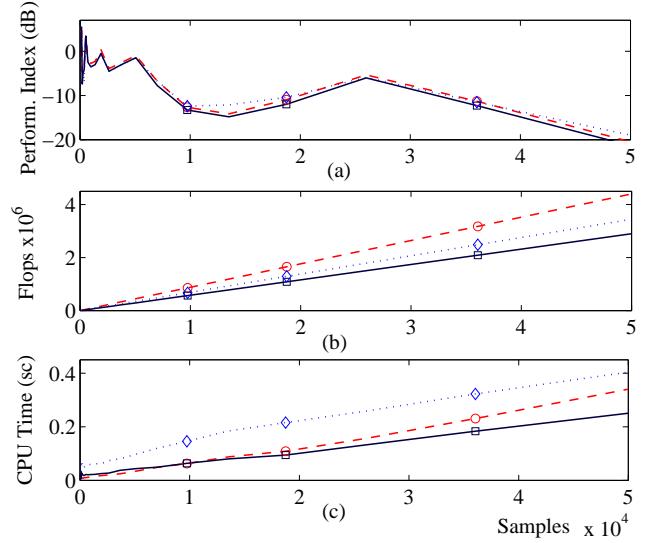


Fig. 1. Mixture of 2 audio signals with the JADE (◊), ICA (○) and SICA (□): (a) Performance Index, (b) Flops, (c) CPU Time.

6. CONCLUSIONS

In this paper we present a new approach to independent component analysis based in fourth-order moments under whitening constraint. The starting point is the ICA contrast given by Comon in [1]. We rewrite the contrast in polar form to show how this contrast reduces to a sinusoidal function, the SICA contrast. Its phase being the solution to the contrast. This strategy results in a simple method with accurate results at a low computational cost. The experiments included show how this new method clearly outperforms the ICA and the JADE [2] methods for different number and types of inputs.

The contrast function, $\phi_{ICA}(\theta)$ may be expressed using (7), (9) and (12) as

$$\begin{aligned} \phi_{ICA}(\theta) &= \Psi_1(\theta) + \Psi_2(\theta) + \Psi_3(\theta) + \Psi_4(\theta) \text{ with} \\ \Psi_1(\theta) &= -\frac{3}{16}E[r^4]E[r^4 \cos(4(\theta + \beta))] \\ \Psi_2(\theta) &= -\frac{1}{2}E^2[r^4 \cos(2(\theta + \beta))] \\ \Psi_3(\theta) &= -\frac{1}{32}E^2[r^4 \cos(4(\theta + \beta))] \\ \Psi_4(\theta) &= +\frac{3}{2}E[r^4 \cos(4(\theta + \beta))] \end{aligned} \quad (21)$$

We now analyze these terms. The term $\Psi_4(\theta)$ corresponds to 6 times contrast $\phi_{Kur}(\theta)$ in (5) and (9). Since its frequency is $4 \cdot 2\pi$, there is a minimum every $\pi/2$ radians. In clear concordance with matrix C in (2) performing a set of permutations and changes of scales and signs of the outputs [1]. Being $\phi_{Kur}(\theta)$ a sinusoid, its maximum points are at $\pi/4$ from the minimum θ_{Kur} . This term

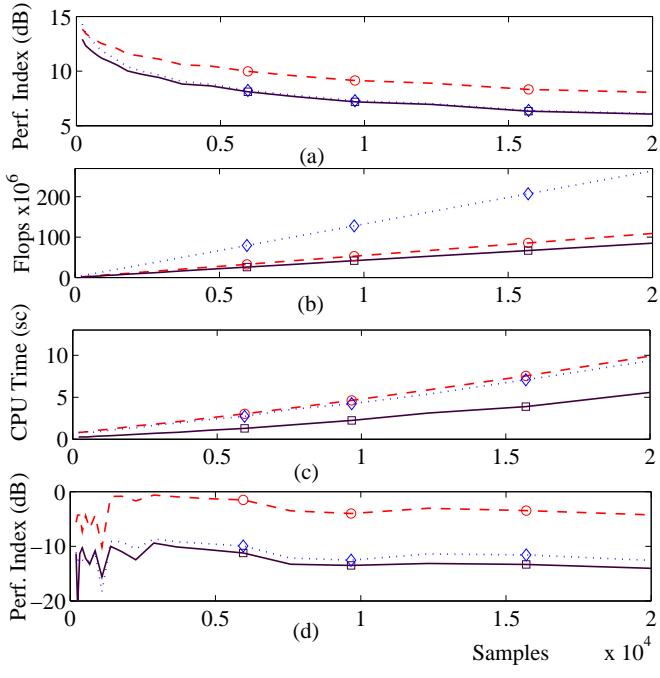


Fig. 2. Mixture of 10 sources wit the JADE (\diamond), ICA (\circ) and SICA (\square): (a) Performance Index, (b) Flops, (c) CPU Time, (d) Performance Index for the mixture of two of the sources

can be written as

$$\Psi_4(\theta) = \psi_4 \cos(4\theta - 4\theta_{Kur}) \quad (22)$$

On the other hand, $\Psi_1(\theta)$ is a sinusoidal function with a π phase shift respect to the term described above,

$$\Psi_1(\theta) = \psi_2 \cos(4\theta - 4\theta_{Kur} + \pi) \quad (23)$$

Next we compare $\Psi_2(\theta)$ to $\Psi_4(\theta)$. It is straight forward to conclude that both of the functions are, up to a constant, sinusoidal functions with the same frequency $4 \cdot 2\pi$. The problem arises when computing the phase shift between them. Elementary trigonometrics show that

$$\begin{aligned} E[r^4 \cos(4(\theta + \beta))] &= \cos(4\theta) E[r^4 \cos(4\beta)] \\ &\quad - \sin(4\theta) E[r^4 \sin(4\beta)] \end{aligned} \quad (24)$$

$$\begin{aligned} E^2[r^4 \cos(2(\theta + \beta))] &= \\ &\cos(4\theta) \left(\frac{1}{2} (E^2[r^4 \cos(2\beta)] - E^2[r^4 \sin(2\beta)]) \right) \\ &\quad - \sin(4\theta) E[r^4 \cos(2\beta)] E[r^4 \sin(2\beta)] \end{aligned} \quad (25)$$

If we prove the terms multiplying $\sin(4\theta)$ in (24) and (25) to cancel at the same θ , then $\Psi_2(\theta)$ and $\Psi_4(\theta)$ have a phase shift of $\rho_2 = 0$ or $\rho_2 = \pi$. If we rewrite them as a function of the fourth-order moments of the outputs it follows that

$$\begin{aligned} E[r^4 \sin(4(\alpha))] &= 2E[y_i^3 y_j] + 2E[y_i y_j^3] \\ E[r^4 \sin(2(\alpha))] &= 6E[y_i^3 y_j] - 2E[y_i y_j^3] \end{aligned} \quad (26)$$

As equations in (26) cancel at separation $\Psi_2(\theta)$ yields

$$\Psi_2(\theta) = \psi_2 \cos(4\theta - 4\theta_{Kur} + \rho_2) \quad (27)$$

where $\rho_2 = 0, \pi$. Besides, the sum of the functions in (22), (27) and (26) may be written as

$$\Psi_{124}(\theta) = \Psi_1 + \Psi_2 + \Psi_4 = \psi_{124} \cos(4\theta - 4\theta_{Kur} + \rho_{124}) \quad (28)$$

where $\rho_{124} = 0, \pi$.

Finally, the $\Psi_3(\theta)$ is a sinusoid with double the frequency that $\Psi_{124}(\theta)$, i.e., $8 \cdot 2\pi$:

$$\Psi_3(\theta) = \psi_3 \cos(8\theta - 8\theta_{Kur} + \pi) \quad (29)$$

Function $\Psi_3(\theta)$ in (29) evaluated at the possible minimums of (28) results in the same value. Consequently, $\Psi_3(\theta)$ has the effect of a bias at the candidate points. Thus, it can be removed from (21). It follows that

$$\phi_{ICA}(\theta) \equiv \phi_{SICA}(\theta) = \Psi_{124}(\theta) \quad (30)$$

The solution to contrast $\phi_{ICA}(\theta)$ can also be computed as minus four times the phase of $\phi_{SICA}(\theta)$ in (30) where $\Psi_4 = 6\phi_{Kur}(\theta)$ was developed in (9) and $\Psi_1(\theta) + \Psi_2(\theta)$, after elementary trigonometrics, yields

$$\Psi_1(\theta) + \Psi_2(\theta) = -l_1 \cos(4\theta) + l_2 \sin(4\theta) \quad (31)$$

where l_1 and l_2 were given in (15).

A. 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, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, no. 1, pp. 157–192, January 1999.
- [3] P. Comon, "Separation of stochastic processes," in *Workshop Higher Order Spectral Anal.*, Va, CO, June 1989.
- [4] A. Mansour and C. Jutten, "A direct solution for blind separation of sources," *IEEE Tran. on Signal Processing*, vol. 44, no. 3, pp. 746–748, March 1996.
- [5] R.M. Clemente and J.I. Acha, "Blind separation of sources using a new polynomial equation," *Electronic Letters*, vol. 33, no. 3, pp. 176–177, January 1997.
- [6] 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, January 1996.
- [7] P. Comon and E. Moreau, "Improved contrast dedicated to blind separation in communications," in *ICASSP'97*, Munich, Germany, 1997, vol. V, pp. 3453–3456.
- [8] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non gaussian signals," *IEE Proceedings F*, vol. 140, no. 6, pp. 362–370, December 1993.
- [9] P. Comon, "Web page: Matlab source codes," <http://www.i3s.unice.fr/~comon/matlab.html>.
- [10] J. F. Cardoso, "Web page: Jade for real-valued data," <ftp://sig.ensf.fr/pub/jfc/Algo/Jade/jadeR.m>.