INDEPENDENT COMPONENT ANALYSIS WITH OPTIMIZED PAIRWISE PROCESSING

Vicente Zarzoso[a*], *Juan José Murillo-Fuentes*[b†], *Rafael Boloix-Tortosa*[b†] *and Asoke K. Nandi*[a]

[a] Dept. of Electrical Engineering and Electronics, The University of Liverpool, Brownlow Hill, Liverpool L69 3GJ, UK
{vicente, aknandi}@liv.ac.uk

[b] Depto. Teoría de la Señal y Comunicaciones, Univ. de Sevilla, Paseo de los Descubrimientos s/n, 41092 Sevilla, Spain
{murillo, rboloix}@us.es

ABSTRACT

The present contribution investigates the solutions to independent component analysis (ICA) based on the pairwise 4th-order statistics of the observed data vector. Previously proposed solutions to the two-signal scenario, including the well-known JADE, are unified under the general weighted fourth-order estimator (GWFOE). A theoretical asymptotic performance analysis enables the selection of the optimal estimator in the GWFOE class, i.e., the solution with minimum mean square error performance. To extend the pairwise estimators to the general scenario of more than two sources, an improved Jacobi-like optimization (JO) approach with reduced computational complexity is put forward. Adaptive versions of the JO methods are also revised, focusing on the enhancement of their convergence properties. The ultimate goal of this paper is to develop general guidelines for an optimized use of the pairwise processing strategy for ICA.

1. INTRODUCTION

The objective of independent component analysis (ICA) is the decomposition of an observed data vector into statistically independent components [1]. Its most common application is arguably the separation of instantaneously-mixed unobservable source signals — a problem known as blind source separation (BSS) — when the time structure of the sensor output cannot be exploited or is simply ignored. Other application areas include projection pursuit, financial data analysis, complexity reduction, and feature extraction.

Many successful methods are available to perform ICA in the general scenario of more than two sources (see, e.g., [1] and references therein). Nevertheless, the two-signal case, being the most basic, remains a scenario of fundamental importance. Despite this relevance, the relationships between the different two-signal solutions have only been explored to a certain extent. The first purpose of this paper is to fill the gap in these connections. By means of a simple expression depending on a weight parameter, many of the existing two-signal estimators based on 4th-order statistics are unified, including JADE [2]. The large-sample mean square error (MSE) of this estimation class is derived, from which the weight parameter of the optimal (i.e., asymptotically most efficient) estimator can be determined as a function of the source statistics.

In the n-dimensional case, $n > 2$, ICA can be carried out by applying the two-signal estimators to each whitened signal pair over several sweeps until convergence [3]. The 4th-order statistics used by the closed-form estimators need to be computed for each signal pair at every iteration. The statistics are typically estimated from the signal samples, which may involve extensive computations, especially when processing long data blocks. Adaptive algorithms such as the adaptive EML (adEML) [4] or the AROT [5], easily derived from this approach, generally show poor convergence, particularly as the number of source components increases. In the second part of this paper, we investigate strategies aiming to alleviate these problems. Ultimately, our goal is to provide the reader with some guidelines on the use of two-signal and JO-based ICA algorithms.

Notations. Given a set of components $\{s_i\}_{i=1}^n$, we denote $\mathrm{M}^s_{i_1\cdots i_r} \triangleq \mathrm{E}[s_{i_1}\cdots s_{i_r}]$ and $\mathrm{C}^s_{i_1\cdots i_r} \triangleq \mathrm{Cum}[s_{i_1},\,\ldots,\,s_{i_r}]$, $1 \leqslant i_k \leqslant n$, $1 \leqslant k \leqslant r$, as their rth-order moments and cumulants, respectively. For the pairwise case, Kendall's notation is preferred: $\mu^s_{r-p,p} \triangleq \mathrm{E}[s_1^{r-p}s_2^p]$ and $\kappa^s_{r-p,\,p} \triangleq \mathrm{C}^s_{\underbrace{1\ldots1}_{r-p}\underbrace{2\ldots2}_{p}}$ stand for the rth-order moment and cumulant of the signal pair $\mathbf{s} = [s_1,\, s_2]^{\mathrm{T}}$. Symbols $\gamma \triangleq (\kappa^s_{40} + \kappa^s_{04})$ and $\eta \triangleq (\kappa^s_{40} - \kappa^s_{04})$ represent the source kurtosis sum (sks) and the source kurtosis diference (skd), respectively. Function $\angle a \in]-\pi,\,\pi]$ supplies the principal value of the argument of $a \in \mathbb{C}$.

2. ICA MODEL

Given a random vector $\mathbf{x} = [x_1,\,\ldots,\,x_m]^{\mathrm{T}}$, the purpose of ICA is to find a linear transformation $\mathbf{B}$ such the output vector $\mathbf{y} = \mathbf{Bx}$ contains statistically independent components. ICA implicitly assumes a generative linear model of the form

$$\mathbf{x} = \mathbf{As} \tag{1}$$

where $\mathbf{A}$ represents the so-called mixing matrix, with dimensions $(m\times n)$, $m \geqslant n$, and $\mathbf{s} = [s_1,\,\ldots,\,s_n]^{\mathrm{T}}$ are the mutually independent source signals. All signals and mixtures are assumed to be real valued herein. If the mixing matrix is full column rank, it is possible to obtain a separation matrix such that $\mathbf{B} = \mathbf{CA}^{-1}$, where the global matrix $\mathbf{C}$ is the product of a invertible diagonal matrix and a permutation matrix. As the source amplitudes are usually not important, it can be assumed, without loss of generality, that the sources are unit-variance, $\mathrm{E}[\mathbf{ss}^{\mathrm{T}}] = \mathbf{I}_n$.

In its more general form, ICA relies on higher-order statistics (HOS). The use of HOS requires that at most one of the sources be Gaussian. A previous spatial whitening process (entailing second-order decorrelation and power normalization) helps to reduce the number of unknowns, resulting in a set of normalized uncorrelated components (whitened signals) $\mathbf{z} = [z_1,\,\ldots,\,z_n]^{\mathrm{T}}$ such that $\mathrm{E}[\mathbf{zz}^{\mathrm{T}}] = \mathbf{I}_n$. These are related to the sources through an orthogonal transformation

$$\mathbf{z} = \mathbf{Qs}. \tag{2}$$

ICA is then tantamount to the identification of orthogonal matrix $\mathbf{Q}$.

A contrast function [3] is a mapping from the set of output probability density functions to the real field whose optimization yields the ICA solution. Contrasts constitute one of the most attractive approaches to ICA because they allow an optimal processing in the presence of unknown noise and interference, adding robustness to the source separation performance. Several families of contrasts have been proposed

* Royal Academy of Engineering Research Fellow.

† Supported through the Spanish MCYT Agency (TIC2003-03781).

to date, based on information-theoretical principles such as maximum likelihood (ML), mutual information or marginal entropy (ME). Although apparently different, these contrasts are all related [6]. In the two-signal scenario, these optimality criteria accept (in exact or approximate form) a variety of analytical solutions, whose connections are developed in the next section. The general scenario of more than two source components is addressed in Section 4.

3. OPTIMAL SOLUTION FOR TWO SOURCES

3.1 Existing Pairwise Solutions

In the fundamental real-valued two-signal case, $\mathbf{Q}$ is a Givens rotation matrix

$$\mathbf{Q}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{3}$$

characterized by an unknown angle $\theta \in]-\pi, \pi]$. ICA then reduces to the identification of parameter θ from the whitened sensor outputs. A variety of closed-form methods for the estimation of this angle have been proposed in the literature. These methods, mostly based on 4th-order statistics, provide direct solutions with no iterative search involved.

The first expression was obtained in [5] by relating the 4th-order statistics of sources and sensors. Its performance was later shown to depend on the actual value of the unknown parameter [7], thus losing the desirable uniform performance property. Other early methods were derived from the ML approach by using the truncated Gram-Charlier expansion of the source probability density function (pdf) [8]. Their restricted validity conditions were broadened through the extended ML (EML) and the alternative EML (AEML) estimators [9, 10]. In turn, these latter presented consistence problems, which were overcome with the hybrid approach of [10]. The two estimators were joined into a single analytic expression in [11], yielding the the approximate ML (AML). The MaSSFOC estimator [12] is derived from the approximate maximization of the contrast function of [3]. Relying on a trigonometric approximation to the same contrast, a similar closed-form solution, so-called sinusoidal ICA (SICA), was recently proposed in [13].

The notion of linearly combining estimators was originally put forward in [11], giving rise to the so-called weighted AML (WAML). Based on this idea, a general estimation expression is presented next which unifies most of the existing solutions summarized above. Specific guidelines for the choice of the optimal estimator are also derived.

3.2 Unification: the General Weighted Estimator

Relation (2) accepts a compact complex-valued formulation:

$$z_1 + jz_2 = \mathrm{e}^{j\theta}(s_1 + js_2) \tag{4}$$

where $(z_1 + jz_2) = \rho \mathrm{e}^{j\phi}$ and $(s_1 + js_2) = \rho \mathrm{e}^{j\phi'}$. Geometrically, expression (4) signifies that the whitened-signal pdf is a rotated version of the source pdf, $\phi = \phi' + \theta$. Now, *centroids* are defined as particular non-linear averages of complex points (4). The following centroids are useful in deriving closed-form expressions for the estimation of θ:

$$\xi_\gamma \triangleq \mathrm{E}[\rho^4 \mathrm{e}^{j4\phi}] = (\kappa_{40}^z - 6\kappa_{22}^z + \kappa_{04}^z) + j4(\kappa_{31}^z - \kappa_{13}^z) \tag{5}$$

$$\xi_\eta \triangleq \mathrm{E}[\rho^4 \mathrm{e}^{j2\phi}] = (\kappa_{40}^z - \kappa_{04}^z) + j2(\kappa_{31}^z + \kappa_{13}^z) \tag{6}$$

$$\beta \triangleq \mathrm{E}[\rho^4] - 8 = \kappa_{40}^z + 2\kappa_{22}^z + \kappa_{04}^z. \tag{7}$$

When written as a function of the source statistics, the above centroids yield $\xi_\gamma = \gamma \mathrm{e}^{j4\theta}$, $\xi_\eta = \eta \mathrm{e}^{j2\theta}$ and $\beta = \gamma$. The EML estimator [9] can be expressed as $\hat{\theta}_{\mathrm{EML}} = \frac{1}{4}\angle(\beta\xi_\gamma)$. Similarly, the AEML [10] reads $\hat{\theta}_{\mathrm{AEML}} = \frac{1}{2}\angle\xi_\eta$. Under mild conditions, the sample versions of centroids ξ_γ, ξ_η and β are consistent estimators of $\gamma \mathrm{e}^{j4\theta}$, $\eta \mathrm{e}^{j2\theta}$ and γ, respectively, so that $\hat{\theta}_{\mathrm{EML}}$ and $\hat{\theta}_{\mathrm{AEML}}$ consistently estimate θ as long as $\gamma \neq 0$ and $\eta \neq 0$, respectively [9]. The lack of consistency for certain values of source kurtosis is precisely the main drawback of these two estimators. In order to circumvent this deficiency, let us form the compound centroid

$$\xi_{\mathrm{GWFOE}} = w\beta\xi_\gamma + (1-w)\xi_\eta^2, \qquad 0 < w < 1. \tag{8}$$

Then, parameter θ can also be determined through:

$$\hat{\theta}_{\mathrm{GWFOE}} = \frac{1}{4}\angle\xi_{\mathrm{GWFOE}} \tag{9}$$

which we call the *general weighted fourth-order estimator (GWFOE)*. The relevance of the GWFOE lies in the fact that it is a consistent estimator of θ for *any* source distribution, since the GWFOE centroid consistently estimates the complex number $[w\gamma^2 + (1-w)\eta^2]\mathrm{e}^{j4\theta}$. In addition, the GWFOE unifies many of the analytic solutions already proposed in the literature, which are simply obtained for different values of the weight parameter w:

(i)	$w = 0$:	AEML estimator of [10]
(ii)	$w = 1/3$:	AML estimator of [11]
(iii)	$w = 3/7$:	SICA estimator of [13]
(iv)	$w = 1/2$:	MaSSFOC estimator of [12]
		JADE method of [2] (proof: see [14])
(v)	$w = 1$:	EML estimator of [9].

Solutions derived from the ML approach which assume some knowledge on the source kurtosis can also be included in the GWFOE with minor changes [13, 14]. In effect, the ML, MK and SKSE/SKSD estimators of [8, 12, 15, 16, 17] are obtained with $w = 1$ and substituting β by ± 1 in the GWFOE expression. Eqn. (9) is essentially the WAML estimator [11] written in centroid form. The use of the complex-centroid formalism allows us to evidence the connections with other existing closed-form solutions. Since some of these solutions (such as MaSSFOC or SICA) were originally obtained as approximations to optimality criteria other than ML, we prefer to adhere to the denomination of GWFOE. Moreover, the complex-centroid formalism facilitates the theoretical performance analysis of the estimator, leading to an optimal choice of w.

3.3 Optimal Estimator

The optimal GWFOE corresponds to the the value of w which minimizes the asymptotic (large-sample) mean square error (MSE) of the GWFOE class. The asymptotic MSE of the GWFOE (9) is given by [14, 18]

$$\mathrm{MSE}[\hat{\theta}_{\mathrm{GWFOE}}] = \frac{\mathrm{E}\Big\{\big[w\gamma(s_1^3 s_2 - s_1 s_2^3) + (1-w)\eta(s_1^3 s_2 + s_1 s_2^3)\big]^2\Big\}}{T\big[w\gamma^2 + (1-w)\eta^2\big]^2} \tag{10}$$

where T is the sample size. If $|\kappa_{40}^s| \neq |\kappa_{04}^s|$, the local minimum of (10) is obtained at:

$$w_{\mathrm{opt}} = \frac{1}{2} + \frac{\mu_{40}^s\mu_{04}^s\big[(\kappa_{40}^s)^2 - (\kappa_{04}^s)^2\big] + \kappa_{40}^s\kappa_{04}^s(\mu_{60}^s - \mu_{06}^s)}{2\big[(\kappa_{40}^s)^2\mu_{06}^s - (\kappa_{04}^s)^2\mu_{60}^s\big]}. \tag{11}$$

If $w_{\mathrm{opt}} \notin [0, 1]$, we are left to choose between $w_{\mathrm{opt}} = 0$ (AEML) and $w_{\mathrm{opt}} = 1$ (EML) the value that provides the lowest MSE in (10). Given the source statistics, this simple procedure allows one to select the estimator of the GWFOE family with asymptotic minimum MSE (MMSE) performance.

In a blind problem, the source statistics needed for the computation of $w_{\rm opt}$ in (11) are typically unknown. To circumvent this difficulty, one may perform an initial separation with any $w \in]0,\,1[$ and obtain an initial estimate of $w_{\rm opt}$ from the estimated sources. Then, the source separation and the computation of $w_{\rm opt}$ may be repeated until convergence. For sufficient sample size, this iterative estimation of $w_{\rm opt}$ shows high accuracy and fast convergence rate (1–2 iterations) [14].

4. MORE THAN TWO SOURCES

4.1 Off-line Jacobi Optimization

A Jacobi-like optimization (JO) procedure to extend a two-signal ICA contrast to the n-dimensional case, with $n > 2$, was introduced in [3]. Such an extension can easily be applied to the GWFOE. At each component pair $[z_p,\, z_q]^{\rm T}$, $1 \leqslant p < q \leqslant n$, Givens angle θ_{pq} is computed by using (9), and the output components are updated by applying the corresponding rotation. This rotation also updates the estimate $\widehat{\mathbf{Q}}$ of orthogonal matrix $\mathbf{Q}$. This process is repeated over several sweeps until convergence. Since centroids (5)–(7) are calculated by averaging over the whole set of samples of every signal pair, this procedure may be computationally very costly for large sample sizes.

A more efficient alternative may be obtained as follows. Centroids (5)–(7) may readily be written as a function of the moments of the current output pair $\{\mu^y_{4-i,i}\}_{i=0}^4$. The idea is to compute the whole set of whitened-signal moments just one time at an initial stage, and later 'rotate' them at each step of the algorithm without using the signal samples. Define the symmetric $(r \times r)$ matrix

$$\mathbf{M}^z\big(\mathbf{A}(i,j), \mathbf{B}(k,l)\big) = \mathrm{M}^z_{ijkl} \qquad (12)$$

with $r = n(n+1)/2$ and $\mathbf{A}(i,j) = \mathbf{B}(i,j) = n(i-1) + \frac{i}{2}(1-i) + j$, $1 \leqslant i \leqslant j \leqslant n$. If $\mathbf{y} = \mathbf{V}\mathbf{z}$, for any $(n \times n)$ matrix $\mathbf{V}$, then [14, 19]

$$\begin{aligned} \mu^y_{40} &= \mathbf{v}_{pp}^{\rm T}\mathbf{M}^z\mathbf{v}_{pp}, & \mu^y_{31} &= \mathbf{v}_{pp}^{\rm T}\mathbf{M}^z\mathbf{v}_{pq} \\ \mu^y_{22} &= \mathbf{v}_{pp}^{\rm T}\mathbf{M}^z\mathbf{v}_{qq}, & \mu^y_{13} &= \mathbf{v}_{pq}^{\rm T}\mathbf{M}^z\mathbf{v}_{qq} \\ \mu^y_{04} &= \mathbf{v}_{qq}^{\rm T}\mathbf{M}^z\mathbf{v}_{qq} \end{aligned} \qquad (13)$$

where

$$\mathbf{v}_{pq} = \begin{cases} \mathbf{V}(p,i)\mathbf{V}(q,j) + \mathbf{V}(p,j)\mathbf{V}(q,i), & i \neq j \\ \mathbf{V}(p,i)\mathbf{V}(q,i), & i = j. \end{cases} \qquad (14)$$

Hence, moment matrix $\mathbf{M}^z$ can be computed before starting the Jacobi iterations and then the output moments recalculated at each step of the JO algorithm using $\mathbf{V} = \widehat{\mathbf{Q}}^{\rm T}$ in (13)–(14). This procedure, which uses the whitened-output samples only once, is referred to as *initialized JO (IJO)*. Remark that IJO produces the same ICA results as the standard JO; they only differ in their computational complexity.

Denote by $K = 1 + \sqrt{n}$ the typical maximum number of sweeps in the JO iterations, and by $g = n(n-1)/2$ the number of signal pairs. The analysis of the relative computational burden, in floating point operation (flops), between of the standard JO to the IJO approaches [14, 19] draws the following conclusions:

- IJO is to be used for a low number of sources, $n \leqslant 5$.
- For $5 < n \leqslant 25$, resort to the decision rule:

$$\frac{19gKT}{T\big(\binom{n+3}{4} + \binom{n+1}{2}\big) + gKr(r^2+1)} \underset{\rm JO}{\overset{\rm IJO}{\gtrless}} 1. \qquad (15)$$

- For $n > 25$, the standard JO method should be selected.

Since IJO is not to be used for large numbers of components, potential memory problems associated with the storage of a large matrix $\mathbf{M}^z$ are avoided.

4.2 On-line Jacobi Optimization

The JO procedure may be implemented adaptively, that is, on a sample-by-sample basis, resulting in the *adaptive JO (AJO)* algorithm. If the GWFOE is used as basic pairwise separator, a new observed sample updates centroids (5)–(7) for each pair in turn over consecutive sweeps. Since centroids at a given pair and sweep depends on previous updated centroids, the statistics of the latest pairs and sweeps cannot converge until those of previous pairs and sweeps do. Furthermore, fluctuations around the convergence point of the statistics in the first sweeps make those in the final stages fluctuate as well, in a manner difficult to predict, thus compromising the stability of the algorithm. Since the required number of sweeps grows with the source dimension, the method only converges for a low number of components. The AROT [5] and adEML [4] adaptive methods, which are also implemented in AJO's fashion, present analogous deficiencies.

To overcome these convergence problems the *adaptive initialized JO (AIJO)* method is proposed [14, 20], which can be considered as the adaptive implementation of the IJO algorithm. The AIJO algorithm consists of two stages. The first stage updates the output moments by introducing the latest output sample:

$$\mathbf{M}^z(t+1) = (1-\lambda)\mathbf{M}^z(t) + \lambda\mathcal{M}^z(t+1) \qquad (16)$$

where $\mathcal{M}^z(t)$ is matrix $\mathbf{M}^z$ in (12) computed with just $\mathbf{z}(t)$, the whitened-output sample at time instant t, and λ is a suitable positive step size. The second stage sweeps over the signal pairs to compute the current estimate of the separating matrix $\mathbf{V}(t)$, as in the IJO method described in the previous section.

Hence, the AIJO algorithm dissociates the learning of the pertinent statistics and the computation of the separation solution. As illustrated by the following experiments, this decoupled design presents better convergence speed and accuracy than the AJO approach and other adaptive ICA algorithms.

5. ILLUSTRATIVE EXPERIMENTS

A few computer experiments illustrate the potential benefits of the optimized pairwise strategies developed in this paper. In the first simulation, two source signal with i.i.d. uniform and Rayleigh distribution are mixed with $\theta = 15^{\rm o}$. The solid line in Fig. 1 plots the theoretical values of $T \cdot {\rm MSE}[\hat{\theta}]$, calculated from (10), as a function of the GWFOE parameter. The crosses represent the values obtained by the EML, AEML, AML, MaSSFOC, JADE and optimal GWFOE [$w_{\rm opt} = 0.7141$, from (11)] for sample sizes $T_k = k \cdot 10^3$, $k = 1, 2, \ldots, 10$. Empirical MSE estimates are averaged over ν_k independent signal realizations, with $\nu_k T_k = 5 \cdot 10^6$. The fitness of asymptotic approximation (10) is very precise. More importantly, the optimal GWFOE substantially outperforms the other estimators, with MSE differences of up to 10 dB. Due to the connection between the estimated-angle MSE and separation performance indices such as the interference-to-signal ratio (ISR) [1], these gaps can have a significant impact on the ICA separation quality. This outcome evidences the importance of an appropriate choice of the weight parameter.

Next, the AIJO-GWFOE method with $w = 1$ is compared to other adaptive procedures: the AJO-GWFOE with the same value of w (adEML) [4] and the well-known EASI [21]. The adaptation coefficient for both the whitening stage and the EASI method is selected as $\alpha = 5 \cdot 10^{-3}$, whereas the learning rate is set to $\lambda = 10^{-3}$ for the two other methods. A mixture of eight independent sources is observed: six uniformly distributed processes, a binary sequence and a

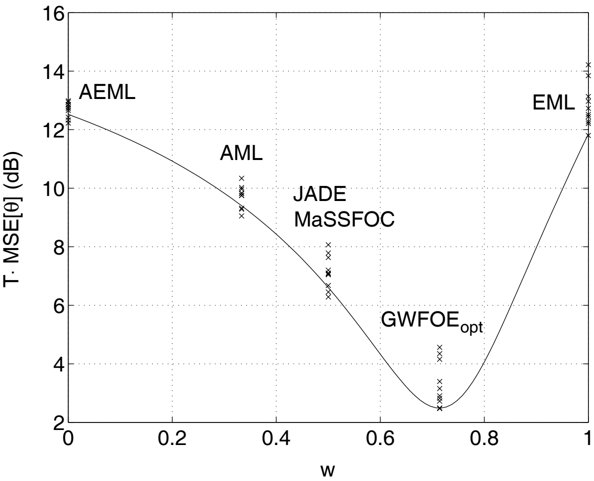

Figure 1: Performance of the GWFOE as a function of the weight coefficient for uniform-Rayleigh sources. Solid line: theoretical MSE (10). '×': empirical values.

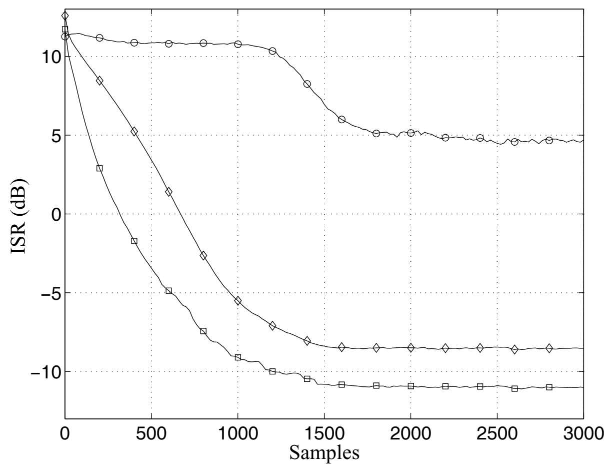

Figure 2: Separation performance for $n = 8$ sources (5 uniform, 1 Laplacian, 1 sinusoid, 1 binary). '○': AJO-GWFOE, $w = 1$; '□': AIJO-GWFOE, $w = 1$; '◇': EASI.

sinusoid with random frequency and phase. The evolution of the ISR performance curves in Fig. 2 demonstrates that the AIJO-GWFOE reaches the best final ISR in the lowest number of iterations. This technique also presents a more robust behaviour near convergence, without the oscillations shown by the AJO-GWFOE.

6. CONCLUSIONS

In the two-signal case, the GWFOE gathers under the same expression many existing 4th-order analytic ICA solutions, including JADE for $n = 2$. The optimal GWFOE in the MMSE sense can be selected by exploiting prior knowledge on the source distribution or by using a simple fast iterative procedure. The algebraic structure of the general scenario of more than two sources has allowed us to optimize the computational complexity of the conventional JO procedure. As a by-product, the proposed approach improves the stability and convergence rate of the associated adaptive implementation. Future research should look into the incorporation of the optimal GWFOE in the general ICA scenario of more than two signals, and the extensions of these methods to statistics of orders other than four.

Acknowledgement. V. Zarzoso carried out part of this work while on leave at Laboratoire I3S, Sophia Antipolis, France. He is indebted to Dr. Pierre Comon for his kind hospitality.

REFERENCES

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, 2001.

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

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

[4] V. Zarzoso and A. K. Nandi, "Adaptive blind source separation for virtually any source probabiliy density function," *IEEE Trans. on Sig. Proc.*, vol. 48, no. 2, pp. 477–488, 2000.

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

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

[7] V. Zarzoso and A. K. Nandi, "Unified formulation of closed-form estimators for blind source separation in real instantaneous linear mixtures," in *Proc. ICASSP*, Istanbul, Turkey, June 5–9, 2000, vol. V, pp. 3160–3163.

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

[10] V. Zarzoso, A. K. Nandi, F. Herrmann, and J. Millet-Roig, "Combined estimation scheme for blind source separation with arbitrary source PDFs," *Electronics Letters*, vol. 37, no. 2, pp. 132–133, Jan. 2001.

[11] M. Ghogho, A. Swami, and T. Durrani, "Approximate maximum likelihood blind source separation with arbitrary source pdfs," in *Proc. SSAP'00*, Pocono Manor Inn, Pennsylvania, Aug 2000.

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

[13] J. J. Murillo-Fuentes and F. J. González-Serrano, "A sinusoidal contrast function for the blind separation of statistically independent sources," *IEEE Trans. on Signal Processing.*, vol. 52, no. 12, pp. 3459–3463, Dec 2004.

[14] V. Zarzoso, J. J. Murillo-Fuentes, R. Boloix-Tortosa, and A. K. Nandi, "Optimal pairwise fourth-order independent component analysis," *IEEE Trans. on Sig. Proc.*, Nov. 2004, submitted.

[15] 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.

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

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

[18] V. Zarzoso, F. Herrmann, and A. K. Nandi, "Weighted closed-form estimators for blind source separation," in *Proc. SSP01*, Singapore, Aug 2001, vol. 4, pp. 2141–2146.

[19] J. J. Murillo-Fuentes, R. Boloix, and F. J. González-Serrano, "Initialized jacobi optimization in independent component analysis," in *Proc. ICA2003*, Nara, Japan, Apr 2003, pp. 1053–1058.

[20] J. J. Murillo-Fuentes, R. Boloix, and F. J. González-Serrano, "Adaptive initialized jacobi optimization in independent component analysis," in *Proc. ICA2003*, Nara, Japan, Apr 2003, pp. 1065–1070.

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