Optimal Pairwise Fourth-Order Independent Component Analysis

Vicente Zarzoso, *Member, IEEE*, Juan José Murillo-Fuentes, *Member, IEEE*, Rafael Boloix-Tortosa, and Asoke K. Nandi, *Senior Member, IEEE*

***Abstract*—Blind source separation (BSS) aims at the reconstruction of unknown mutually independent signals, so-called sources, from their mixtures observed at the output of a sensor array. The BSS of instantaneous linear mixtures, which finds application in numerous fields, can be solved through the statistical tool of independent component analysis (ICA). This paper concentrates on the analytic solutions for the fundamental two-signal ICA scenario. A novel estimation class, so-called general weighted fourth-order estimator (GWFOE), is put forward, which is based on the fourth-order statistics of the whitened sensor output. By means of a weight parameter, the GWFOE is able to unify a variety of apparently disparate estimation expressions previously scattered throughout the literature, including the well-known JADE method in the two-signal case. A theoretical asymptotic performance analysis is carried out, resulting in the GWFOE large-sample mean square error and the source-dependent weight value of the most efficient estimator in the class. To extend the pairwise estimators to the general scenario of more than two sources, an improved Jacobi-like optimization technique is proposed. The approach consists of calculating the necessary sensor-output fourth-order statistics at the initialization stage of the algorithm, which can lead to significant computational savings when large sample blocks are processed. Based on this idea, adaptive algorithms are also devised, showing very satisfactory convergence characteristics. Experiments illustrate the good performance of these optimal pairwise ICA strategies, in both off- and on-line processing modes.**

***Index Terms*—Array signal processing, blind source separation, higher order statistics, independent component analysis, performance analysis, unsupervised learning.**

I. Introduction

A. Problem and Motivation

The problem of blind source separation (BSS) consists of recovering a set of unobserved signals, so-called sources, from another set of observed signals which are mixtures of the sources [1]–[3]. The term "blind" signifies that (typically) very few assumptions are made about the sources and the mixing process. By contrast, conventional array processing techniques (e.g., for direction-of-arrival estimation) assume a certain structure for the mixing matrix in terms of the array manifold, or the array response as a function of the arrival angle. Deviations of the assumed structure from reality (calibration errors) can have a significant negative impact on the algorithms' performance. The relative freedom given by BSS methods to the mixing structure makes them very robust to calibration errors [4]. This flexibility and robustness have spurred the interest in the BSS problem over the last decade. Another important motivation has been the vast number of application areas where BSS proves useful [2], [3], [5], ranging from communications [6] to biomedical signal processing (electrocardiogram and electroencephalogram analysis, fMRI, brain imaging) [7]–[9], condition monitoring, image processing [10], financial data analysis, seismic exploration, classification, or data compression and coding [11], among others. Instantaneous linear mixtures, where no time delays occur in the propagation from sources to sensors, is a very accurate signal model in many of those applications. The solution of the more elaborate convolutive-mixture model can often be decoupled into a stage resolving the effects caused by the multipath channel (time equalization) followed by the separation of the remaining instantaneous mixture. The separation of nonlinear mixtures is more involved, and is also receiving attention by some authors (see, e.g., [2], [3], and references therein).

When the time structure cannot be exploited or is simply ignored, the basic approach to instantaneous linear source separation consists of projecting the observation vectors into some basis where the resulting components are statistically independent. This is the independent component analysis (ICA) of the observed data [12], and in its more general form it relies (explictly or not) on higher order statistics. 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) related with the sources through a unitary transformation. ICA is then tantamount to the identification of this unitary matrix.

B. Closed-Form Solutions in the Two-Signal Case

In the fundamental real-valued two-signal case, the problem reduces to the identification of a single parameter, the unknown angle characterizing the Givens-rotation mixing matrix. A variety of closed-form methods for the estimation of this angle have been proposed in the literature. These methods arise from approximations of certain optimality criteria (contrast functions) and provide direct solutions with no iterative search involved. Most of these share the common feature

Manuscript received November 30, 2004; revised September 4, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Peter Handel. The work of V. Zarzoso was supported by the Royal Academy of Engineering, U.K., under a Research Fellowship. The work of J. J. Murillo-Fuentes and R. Boloix-Tortosa was supported by the Spanish Government under Grant MCYT TIC2003-03781.

V. Zarzoso was with the Department of Electrical Engineering and Electronics, The University of Liverpool, U.K. He is now with Laboratoire I3S, Université de Nice - Sophia Antipolis, 06903 Sophia Antipolis Cedex, France (e-mail: zarzoso@i3s.unice.fr).

J. J. Murillo-Fuentes and R. Boloix-Tortosa are with the Departamento de Teoría de la Señal y Comunicaciones, Escuela Superior de Ingenieros, Universidad de Sevilla, 41092 Sevilla, Spain (e-mail: murillo@us.es; rboloix@us.es).

A. K. Nandi is with the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3GJ, U.K. (e-mail: aknandi@liv.ac.uk).

Digital Object Identifier 10.1109/TSP.2006.875391

of being based on the fourth-order statistics of the sensor output. The first expression was obtained in [13] by relating the fourth-order statistics of sources and sensors. Its performance was later shown to depend on the actual value of the unknown parameter [14], [15], thus losing the desirable uniform performance property [16]. A good number of early methods were derived from the maximum likelihood (ML) approach. The truncated Gram–Charlier expansion of the source probability density function (pdf) yielded the solution of [17], restricted to symmetric sources with normalized kurtosis in certain positive range. These validity conditions were broadened through the extended ML (EML) and the alternative EML (AEML) estimators [14], [18], [19]. The EML also generalized the maximum kurtosis (MK) cost function of [20], [21], initially thought to be valid only for sources with same kurtosis sign [14], to source pairs with nonzero source kurtosis sum (sks). The EML and the AEML remain consistent providing the sks and source kurtosis difference (skd) are not null, respectively. This deficiency was overcome in [14], [19], and [22]. In [14] and [19], the choice between the EML or the AEML was made with a simple decision rule as a function of the sensor-output fourth-order statistics. In [22], adopting the ML framework of [17], the two estimators were unified into a single analytic expression, the approximate ML (AML).

The contrast function of [12], which had earlier been reached from the ML principle [23], is itself an approximation of a negentropy maximization principle measuring the deviation of the separator output from Gaussianity. Negentropy can also be readily connected to alternative information-theoretical criteria such as the mutual information (MI) between the separator outputs or the sum of their marginal entropies (ME) [4], [5]. Another major group of two-dimensional closed-form solutions arises from the trigonometric expansion and approximate minimization of the ME contrast criteria developed in [12]. The MaSSFOC (maximum of sum squared fourth-order cumulant) estimator [24] and the recently proposed sinusoidal ICA (SICA) [25], which resemble the AML, are approximate minimizers of the fourth-order contrast function. Further simplifications of this contrast function when the source kurtoses have the same modulus lead to the so-called source kurtosis sum and source kurtosis difference estimators (SKSE, SKDE) [24], very similar to the EML and AEML estimators [14], [18], [19], respectively. The simultaneous exploitation of orders three and four is shown to improve the separation performance when some of the sources present nonsymmetric distributions [26].

The original solution to Comon's fourth-order contrast involved finding the roots of a fourth-degree polynomial (a biquadratic or quartic equation). An analytic procedure for rooting quartics is well known since the sixteenth century (Ferrari's formula), but its calculation can be cumbersome; approximate numerical methods are usually preferred instead. The closed-form estimators that we are concerned with are considerably less elaborate: they consist of simple formulas involving straightforward operations on certain statistics of the whitened sensor output.

The notion of linearly combining estimators was originally put forward in [22]. Through a weight parameter, the EML and AEML are combined together into a single expression, the so-called weighted AML (WAML) estimator. It was suggested that the weight parameter could be adjusted by taking advantage of a priori information on the source pdfs, although no specific guidelines were given on how the actual choice should be made.

C. Scenario of More Than Two Signals

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 [12]. This iterative approach is reminiscent of the Jacobi optimization (JO) technique for matrix diagonalization [27], [28], and can indeed be seen as its extension to higher dimensional tensors [12]. Although no theoretical proof of global convergence has yet been obtained for the pairwise iterations in the tensor case [12], [29], the method remains valid in practice since no experimental or theoretical counterexample of misconvergence has been encountered to date, provided that the validity conditions of the two-dimensional criteria are fulfilled for every signal pair. In the standard JO iteration, the fourth-order statistics used by the closed-form estimators need to be computed for each signal pair at every sweep until convergence. Typically, the statistics are estimated from the signal samples, which may involve extensive computations especially when processing long signal blocks. Adaptive algorithms, such as the so-called adaptive rotation (AROT) [13] and the adaptive EML (adEML) [30], are easily derived from this strategy. However, they sometimes show poor convergence, especially for a large number of sources.

D. Contributions and Outline

Many successful methods are available to perform ICA in the general scenario of more than two sources (see, e.g., [2], [3], and references therein). Nevertheless, the two-signal case remains a scenario of fundamental importance, since it is the most basic and can be considered as the elementary unit for the solution of the general $n > 2$ in the JO approach. Despite this relevance, the relationships between the different analytic solutions have only been explored to a limited extent. The purpose of this paper is to fill the gap in these connections. By means of the complex-centroid notation used in the EML and the AEML [14], [18], [19], [31], we arrive at a compact formulation for the WAML estimator of [22]. It is seen that through different values of the weight parameter, many of the existing fourth-order estimators are obtained, including the well-known JADE method [4] for $n = 2$; hence the more suitable name of *general weighted fourth-order estimator* (GWFOE). The centroid formalism allows a simple derivation of the estimator's large-sample mean square error (MSE), from which the weight parameter of the optimal estimator is determined as a function of the source statistics. Here, "optimal" refers to the asymptotically most efficient estimator in the GWFOE class.

In the general case of more than two signals, we aim to optimize the computational cost of the JO technique. An alternative moment-calculation procedure is proposed, which is less costly in scenarios where the sample size is large relative to the number of sources. The relevant statistics are computed from the sensor-output samples before starting the JO iterations and then modified according to the pairwise rotations. We refer to this

method as *initialized JO* (IJO). By comparing the complexity of the proposed and the conventional moment-estimation procedure, a decision rule is derived to select the most computationally efficient option. This results in the *optimal JO* (OJO). Adaptive algorithms based on IJO can also be designed to improve the convergence properties of previous online approaches. In short, the results presented in this paper unify, generalize, and enhance ICA techniques based on two-dimensional fourth-order contrasts.

This paper encompasses substantially extended as well as thoroughly revised versions of conference publications [32]–[36]. The material is organized as follows. After reviewing the BSS signal model and ICA contrast functions in Section II, Section III derives the GWFOE, highlights its connections with other analytic solutions, performs its asymptotic analysis, and obtains the best estimator of the class. Section IV is devoted to the scenarios of more than two signals, featuring the computationally efficient IJO and OJO procedures. Adaptive implementations are the focus of Section V. Experimental results are reported in Section VI. Section VII concludes herein. The Appendices contain some proofs and other mathematical derivations.

E. Notations

Throughout this paper, vectors and matrices are represented as lowercase and uppercase boldface letters, respectively. Symbols $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{-1}$ indicate the transpose and inverse matrix operators, respectively. $\mathbb{R}$ and $\mathbb{C}$ are the sets of real and complex numbers, respectively; symbol $j = \sqrt{-1}$ is the imaginary unit; $\mathbb{R}\mathrm{e}(\cdot)$ and $\mathbb{I}\mathrm{m}(\cdot)$ denote the real and imaginary part of its complex argument, respectively, whereas function $\angle(\cdot)$ supplies its principal value (i.e., its argument in the interval $]-\pi,\pi]$). $\mathrm{E}[\cdot]$ represents the mathematical expectation. Given a set of signals $\{s_i\}_{i=1}^n$, $\mathrm{M}^s_{i_1\cdots i_r} \stackrel{\mathrm{def}}{=} \mathrm{E}[s_{i_1}\cdots s_{i_r}]$ and $\mathrm{C}^s_{i_1\cdots i_r} \stackrel{\mathrm{def}}{=} \mathrm{Cum}[s_{i_1},\ldots,s_{i_r}]$, $1 \leqslant i_k \leqslant n$, $1 \leqslant k \leqslant r$, denote their rth-order moments and cumulants, respectively, whose mathematical definitions can be found in [37] and [38]. For the pairwise case, we prefer Kendall's notation [37]: $\mu^s_{r-p,p} \stackrel{\mathrm{def}}{=} \mathrm{M}^s_{\underbrace{1\ldots1}_{r-p}\,\underbrace{2\ldots2}_{p}} = \mathrm{E}[s_1^{r-p}s_2^p]$ and $\kappa^s_{r-p,p} \stackrel{\mathrm{def}}{=} \mathrm{C}^s_{\underbrace{1\ldots1}_{r-p}\,\underbrace{2\ldots2}_{p}}$ stand for the rth-order moment and cumulant of the signal pair $\mathbf{s} = [s_1, s_2]^{\mathrm{T}}$.

II. BSS and ICA

A. Matrix Model

In its simplest form, the BSS problem accepts the following matrix model. The entries of sensor-output vector $\mathbf{x}(t) = [x_1(t),\ldots,x_m(t)]^{\mathrm{T}}$ are instantaneous linear combinations of a set of unobserved source signals $\mathbf{s}(t) = [s_1(t),\ldots,s_n(t)]^{\mathrm{T}}$

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

where $\mathbf{A}$ represents the *mixing matrix,* with dimensions $(m\times n)$, $m \geqslant n$. In this paper all signals and mixtures are assumed to be real valued. If the mixing matrix is full column rank, the sources are mutually independent, and at most one of them is Gaussian, it is possible to obtain a separation matrix $\mathbf{B}$ and estimate the sources [12], [39] as

$$\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t) = \mathbf{B}\mathbf{A}\mathbf{s}(t) = \mathbf{C}\mathbf{s}(t). \tag{2}$$

Since the scale and order of the components of $\mathbf{s}(t)$ do not affect their statistical independence, a satisfactory separation is characterized by a global matrix $\mathbf{C}$ with a nonmixing structure, that is, with a single nonnull element per row and per column (the product of a invertible diagonal matrix and a permutation matrix). As the source amplitudes are not important, it can be assumed, without loss of generality, that the source variance is unity $\mathrm{E}[\mathbf{s}(t)\mathbf{s}(t)^{\mathrm{T}}] = \mathbf{I}_n$.

Source separation is typically carried out in two steps. First, whitening or standardization [principal component analysis (PCA)] projects the observed vector on the signal subspace and yields a set of second-order decorrelated, normalized signals $\mathbf{z}(t) = [z_1(t),\ldots,z_n(t)]^{\mathrm{T}}$ such that $\mathrm{E}[\mathbf{z}\mathbf{z}^{\mathrm{T}}] = \mathbf{I}_n$. As a result, the source and whitened vectors must be related through a unitary transformation

$$\mathbf{z}(t) = \mathbf{Q}\mathbf{s}(t). \tag{3}$$

The separation problem thus reduces to the computation of unitary matrix $\mathbf{Q}$, which is accomplished in a second step. The ICA approach to BSS consists of computing $\mathbf{Q}$ such that the entries of the separator output $\mathbf{y}(t) = \mathbf{Q}^{\mathrm{T}}\mathbf{z}(t)$ are as independent as possible.[1] Since we consider methods that do not exploit the temporal structure of the source process $\mathbf{s}(t)$, in the sequel, the time index t will be dropped when convenient.

B. Contrast Functions

A contrast function [12] is a mapping $\psi(\mathbf{y})$ from the set of densities $\{p_{\mathbf{y}}, \mathbf{y} \in \mathbb{R}^n\}$ to $\mathbb{R}$ satisfying the following requirements. If $\mathbf{y}$ has independent components, then $\psi(\mathbf{y}) \geqslant \psi(\mathbf{A}\mathbf{y})$, $\forall\mathbf{A}$ nonsingular (domination), with equality if and only if $\mathbf{A}$ is nonmixing (discrimination); also, $\psi(\mathbf{y})$ is unaltered by permutations or scaling of the components of $\mathbf{y}$ (invariance). Thus, the maximization of a contrast function yields the ICA solution. Contrasts are attractive because they allow an optimal processing in the presence of unknown noise and interference, adding robustness to the separation performance.

The ML principle provides the contrast [23]

$$\psi^{\mathrm{ML}}(\mathbf{y}) = \log p_x(\mathbf{x}|\mathbf{A})\Big|_{\mathbf{x}=\mathbf{A}\mathbf{y}} = \sum_{i=1}^{n}\log p_{s_i}(y_i) - \log|\det\mathbf{A}|. \tag{4}$$

[1]This two-step process corresponds to the "hard whitening" approach. Recently, the "soft whitening" concept has been introduced [40], in which the second- and higher order processing is carried out simultaneously.

If this function is maximized for all possible distributions under the whitening constraint, we arrive at the ME contrast [41]

$$\psi^{\mathrm{ME}}(\mathbf{y}) = -\sum_{i=1}^{n} \mathrm{H}[y_i] \tag{5}$$

where $\mathrm{H}[\cdot]$ represents the differential entropy. Using the Edgeworth expansion of the source pdf [37], after second-order whitening, the ME contrast can be approximated as a function of the fourth-order cumulants [12], [41]

$$\psi_{24}^{\mathrm{ME}}(\mathbf{y}) = \sum_{i=1}^{n} \left(\mathrm{C}_{iiii}^{y}\right)^2 \tag{6}$$

where C_{iiii}^{y} is the fourth-order marginal cumulant (kurtosis) of y_i, which, in the zero-mean unit-variance case, reduces to $(\mathrm{E}[y_i^4] - 3)$. This contrast is discriminating over the set of random vectors $\mathbf{y}$ having at most one non-kurtic component [12]. Alternatively, instead of maximizing the ML for all possible distributions, we can also exploit some available information on the source pdf to maximize the ML contrast. In the fourth-order case, if all sources have the same sign of kurtosis, (6) simplifies to [20]

$$\psi_{24}^{\mathrm{ML}}(\mathbf{y}) = \pm \sum_{i=1}^{n} \mathrm{E}\left[y_i^4\right]. \tag{7}$$

Finally, the JADE method [4] is based on the criterion

$$\psi^{\mathrm{JADE}} = \sum_{i,k,l=1}^{n} \left(\mathrm{C}_{iikl}^{y}\right)^2 \tag{8}$$

whose maximization can be efficiently carried out as the joint approximate diagonalization of a set of matrix slices of the whitened cumulant tensor. In the two-signal scenario, approximations to these optimality criteria can be solved in closed form as explained in the next section. In the case of JADE, the associated closed-form estimator that we develop is an exact minimizer of criterion (8).

III. Optimal Analytic Solution in the Two-Signal Case

A. Complex Centroids

In the two-signal case, $\mathbf{Q}$ is a Givens rotation matrix, characterized by an unknown angle $\theta \in]-\pi, \pi]$

$$\mathbf{Q}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{9}$$

ICA then reduces to the estimation of θ from the whitened sensor outputs. Relation (3) accepts a compact complex-valued formulation

$$z_1 + jz_2 = \mathrm{e}^{j\theta}(s_1 + js_2) \tag{10}$$

or $\phi = \phi' + \theta$, where $(z_1 + jz_2) = \rho \mathrm{e}^{j\phi}$ and $(s_1 + js_2) = \rho \mathrm{e}^{j\phi'}$. Geometrically, (10) signifies that the whitened-signal pdf is a rotated version of the source pdf.

Centroids are defined as particular nonlinear averages of the complex points (10) [14], [18], [31]. The following centroids are useful in deriving closed-form expressions for the estimation of θ

$$\xi_\gamma \stackrel{\mathrm{def}}{=} \mathrm{E}[\rho^4 \mathrm{e}^{j4\phi}] = (\kappa_{40}^z - 6\kappa_{22}^z + \kappa_{04}^z) + j4(\kappa_{31}^z - \kappa_{13}^z) \tag{11}$$

$$\xi_\eta \stackrel{\mathrm{def}}{=} \mathrm{E}[\rho^4 \mathrm{e}^{j2\phi}] = (\kappa_{40}^z - \kappa_{04}^z) + j2(\kappa_{31}^z + \kappa_{13}^z) \tag{12}$$

$$\beta \stackrel{\mathrm{def}}{=} \mathrm{E}[\rho^4] - 8 = \kappa_{40}^z + 2\kappa_{22}^z + \kappa_{04}^z. \tag{13}$$

When written as a function of the source statistics, the above centroids yield

$$\xi_\gamma = \gamma \mathrm{e}^{j4\theta}, \quad \xi_\eta = \eta \mathrm{e}^{j2\theta}, \quad \beta = \gamma \tag{14}$$

where symbols $\gamma \stackrel{\mathrm{def}}{=} (\kappa_{40}^s + \kappa_{04}^s)$ and $\eta \stackrel{\mathrm{def}}{=} (\kappa_{40}^s - \kappa_{04}^s)$ represent the sks and the skd, respectively.[2]

B. General Weighted Fourth-Order Estimator (GWFOE)

The EML estimator [18] can be expressed as

$$\hat{\theta}_{\mathrm{EML}} = \frac{1}{4}\angle(\beta\xi_\gamma) \tag{15}$$

Similarly, the AEML [19] reads

$$\hat{\theta}_{\mathrm{AEML}} = \frac{1}{2}\angle\xi_\eta. \tag{16}$$

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 [14], [18]. 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

$$\xi_{\mathrm{GWFOE}} = w\beta\xi_\gamma + (1-w)\xi_\eta^2, \quad 0 < w < 1. \tag{17}$$

Then, parameter θ can also be estimated through

$$\hat{\theta}_{\mathrm{GWFOE}} = \frac{1}{4}\angle\xi_{\mathrm{GWFOE}} \tag{18}$$

which we call the 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 $[w\gamma^2 + (1-w)\eta^2]\mathrm{e}^{j4\theta}$. More importantly, 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 [14], [19];
ii) $w = 1/3$: AML estimator of [22];
iii) $w = 3/7$: SICA estimator of [25], [33];

[2]Note that β is an estimate of γ from the whitened sensor output. Hence, the equality expressed in (14) only holds for the ensemble averages.

iv) $w = 1/2$: MaSSFOC estimator of [24];
v) $w = 1$: EML estimator of [14], [18].

In addition, Appendix I proves that the solution provided by JADE [4] for $n = 2$ sources is equivalent to the GWFOE with $w = 1/2$. Similarly, the fourth-order part of the recently proposed CuBICA method [26] corresponds to the GWFOE with $w = 3/7$. On the other hand, by substituting β with ± 1 in (17)–(18), we also obtain the ML, MK, and SKSE/SKDE estimators of [17], [20], [21], [24], and [41]. These latter methods require advance knowledge of the source kurtosis sign.

Some of the above estimators arise from the ML criterion when the source pdf is approximated by its Gram–Charlier expansion truncated at fourth-order, and the sources are symmetrically distributed. Different solutions are then obtained under additional conditions:

EML estimator: $\eta = 0, \gamma \neq 0$;
AEML estimator: $\gamma = 0, \eta \neq 0$;
AML estimator: $\gamma \neq 0, \eta \neq 0$.

The GWFOE does not directly arise from the ML criterion, but it can be considered as the combination of two solutions (EML and AEML) which are approximate ML estimators under specific assumptions. Even if the validity conditions of an approximate ML solution hold, the use of a different weight w will divert the GWFOE from such a solution. However, the GWFOE variance can be fine-tuned by appropriately selecting w. In this manner the GWFOE can be made more efficient than any of the pairwise ML methods, especially in scenarios where their validity conditions do not hold. This improved efficiency is possible because the other estimators are only *approximate* ML solutions. This interesting feature will be developed in the next section and illustrated by the experiments of Section VI.

The use of the complex-centroid formalism allows us to bring out the connections with other existing closed-form solutions and facilitates the theoretical performance analysis of the estimator (as carried out next). 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 more generic denomination of GWFOE.

C. Performance Analysis: Optimal GWFOE

In this section, we intend to provide specific guidelines for the choice of GWFOE's weight parameter. We search for the value of w that minimizes the asymptotic (large-sample) MSE of the GWFOE class.

The asymptotic MSE of the GWFOE (18) is determined in Appendix II and is given by

$$\mathrm{MSE}_{\mathrm{GWFOE}} = \frac{\mathrm{E}\left\{\left[w\gamma\left(s_1^3 s_2 - s_1 s_2^3\right) + (1-w)\eta\left(s_1^3 s_2 + s_1 s_2^3\right)\right]^2\right\}}{T\left[w\gamma^2 + (1-w)\eta^2\right]^2} \quad (19)$$

where T is the number of samples per signal. It is interesting to note the following.

i) $\mathrm{MSE}_{\mathrm{GWFOE}}$ reduces to the asymptotic MSE of the AEML and EML estimators [14], [15] for $w = 0$ and $w = 1$, respectively. This is not surprising, since the GWFOE becomes such estimators at those weight values (see the previous section).
ii) When $\gamma = 0$ (respectively, $\eta = 0$), GWFOE performance reduces to that of the AEML (respectively, EML) estimator, for any $0 < w < 1$.

If $|\kappa_{40}^s| \neq |\kappa_{04}^s|$, the global minimum of $\mathrm{MSE}_{\mathrm{GWFOE}}$ (19) is obtained at (see Appendix II)

$$w_{\mathrm{opt}} = \frac{1}{2} + \frac{\mu_{40}^s \mu_{04}^s \left[\left(\kappa_{40}^s\right)^2 - \left(\kappa_{04}^s\right)^2\right] + \kappa_{40}^s \kappa_{04}^s \left(\mu_{60}^s - \mu_{06}^s\right)}{2\left[\left(\kappa_{40}^s\right)^2 \mu_{06}^s - \left(\kappa_{04}^s\right)^2 \mu_{60}^s\right]}. \quad (20)$$

If $w_{\mathrm{opt}} \notin [0, 1]$, the derivative of $\mathrm{MSE}_{\mathrm{GWFOE}}$ with respect to w does not change sign and thus $\mathrm{MSE}_{\mathrm{GWFOE}}$ is monotonic in such an interval. In that case, we choose between $w_{\mathrm{opt}} = 0$ (AEML) and $w_{\mathrm{opt}} = 1$ (EML) the value that provides the lowest $\mathrm{MSE}_{\mathrm{GWFOE}}$ in (19). If $|\kappa_{40}^s| = |\kappa_{04}^s|$, case ii) holds. Hence, given the source statistics, one can select the estimator of the GWFOE family with minimum asymptotic MSE. The experiments of Section VI will illustrate the validity of the asymptotic approximation (19) and the performance improvements that can be derived from the use of the optimal weight coefficient.

In the event that nothing is known in advance about the source statistics, a possible simple strategy is to perform an initial separation with any $w \in]0, 1[$. The optimal value of w can then be estimated from the obtained sources, and the separation can be repeated until w_{opt} converges. This iterative estimation of w_{opt} converges very fast (typically within one to two iterations), as will be demonstrated in the experiments of Section VI. Depending on the actual source statistics and the application in hand, the performance gain may compensate the increased cost of performing several separations.

IV. More Than Two Signals Case

A. Standard Jacobi Optimization

Jacobi optimization (JO) techniques have favorable rounding-error properties and high computational parallelism, allowing for numerically stable efficient implementations [28]. In the ICA context, Comon applied a JO-like procedure to extend a two-dimensional contrast $\phi(\theta)$ to the n-dimensional scenario, with $n > 2$. Thanks to its flexibility, the JO approach can easily integrate any valid two-signal solution, such as the GWFOE.

Algorithm (JO-GWFOE)

n-dimensional GWFOE using conventional Jacobi optimization.

1) *Whitening*. Compute the whitened signals as $\mathbf{z} = \mathbf{W}\mathbf{x}$ from a whitening matrix $\mathbf{W}$. Set $\mathbf{y} = \mathbf{z}$ and sweep number $c = 1$.
2) *Sweep* c. For all $g = n(n-1)/2$ pairs, $1 \leq p < q \leq n$, do
   a) Set $[z_p, z_q]^{\mathrm{T}} = [y_p, y_q]^{\mathrm{T}}$ and compute the Givens angle $\theta_{pq} = \hat{\theta}_{\mathrm{GWFOE}}$ from (18).
   b) If $|\theta_{pq}| > \theta_{\min}$, rotate the pair $[y_p, y_q]^{\mathrm{T}}$ by θ_{pq}.
3) *End*? If the number of sweeps c reaches a maximum value $c = K$ or no angle θ_{pq} has been updated, terminate. Otherwise go to Step 2) for another sweep, with $c = c + 1$.

In [41], the algorithm only stops when the whole set of g Givens rotations have been updated by a value under a threshold $\theta_{\min}$, but no limit is set on the number of sweeps c. The value $\theta_{\min}$ is chosen 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 sample size. In [12], the algorithm stops after $K = 1 + \sqrt{n}$ sweeps. This limit is also appropriate in our implementation, due to the existing connection between the contrast of [12] in the basic two-dimensional case and the GWFOE solutions. In a bid to avoid useless computations, we also set a fixed threshold $\theta_{\min} = \pi/360$ rad (0.5°).

B. Initialized Jacobi Optimization

Step 2a) of the JO-GWFOE computes the Givens angle θ_{pq} by using (18). The centroids (11)–(13) are calculated by averaging over the whole set of samples of signal pair $[z_p, z_q]^{\mathrm{T}}$. Since the sample averaging is repeated over several sweeps, this procedure may be computationally very costly for large sample sizes.

A more efficient alternative may be obtained as follows. Centroids (11)–(13) may 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 once at an initial stage and later "rotate" them at each step of the algorithm without reusing the observed signal samples. The relationship between the moments of the whitened sensor output and their rotated counterparts is established below (Appendix III).

Proposition 1: Let $\mathbf{y} = \mathbf{V}\mathbf{z}(t)$, where $\mathbf{V}$ is an arbitrary $(n \times n)$ matrix. Then, there exists a symmetric $(r \times r)$ matrix $\mathbf{M}^z$, with $r = n(n+1)/2$, such that

$$\mathbf{M}^z\left(a(i,j), a(k,l)\right) = \mathrm{M}^z_{ijkl} \tag{21}$$

where

$$a(i,j) = (i-1)\left(n - \frac{i}{2}\right) + j, \qquad 1 \leqslant i \leqslant j \leqslant n. \tag{22}$$

Moreover, there exist vectors $\mathbf{v}_{pp}$, $\mathbf{v}_{pq}$, and $\mathbf{v}_{qq}$ of length r, such that the fourth-order moments of the outputs $[y_p, y_q]^{\mathrm{T}}$ are given by

$$\begin{aligned} \mu^y_{40} &= \mathbf{v}^{\mathrm{T}}_{pp}\mathbf{M}^z\mathbf{v}_{pp}, \quad \mu^y_{31} = \mathbf{v}^{\mathrm{T}}_{pp}\mathbf{M}^z\mathbf{v}_{pq}, \\ \mu^y_{22} &= \mathbf{v}^{\mathrm{T}}_{pp}\mathbf{M}^z\mathbf{v}_{qq}, \quad \mu^y_{13} = \mathbf{v}^{\mathrm{T}}_{pq}\mathbf{M}^z\mathbf{v}_{qq}, \\ \mu^z_{04} &= \mathbf{v}^{\mathrm{T}}_{qq}\mathbf{M}^z\mathbf{v}_{qq}. \end{aligned} \tag{23}$$

The formulation introduced above allows an easy computation of the output statistics for a given rotation matrix, as the entries of $\mathbf{V}$ are easily arranged into the three "rotation vectors" $\mathbf{v}_{pp}$, $\mathbf{v}_{pq}$, and $\mathbf{v}_{qq}$ used in (23). Since only the subset $1 \leqslant i \leqslant j \leqslant k \leqslant l \leqslant n$ is needed in matrix $\mathbf{M}^z$, the number of computed moments reduces to $\binom{n+3}{4} = (n+3)!/((n-1)!4!)$. The resulting ICA algorithm based on this algebraic structure is outlined below.

Algorithm (IJO-GWFOE)

n-dimensional GWFOE using *initialized* Jacobi optimization.

1) *Whitening*. Compute a whitening matrix $\mathbf{W}$ and set $\mathbf{z} = \mathbf{W}\mathbf{x}$.
2) *Moments Initialization*. Compute matrix $\mathbf{M}^z$ in (21) from the sample estimates of M^z_{ijkl}, $1 \leqslant i, j, k, l \leqslant n$. Initialize the accumulated rotation matrix as $\mathbf{V} = \mathbf{I}_n$.
3) *Sweep* c. For all $g = n(n-1)/2$ pairs, $1 \leq p < q \leq n$, do
   a) Compute the moments of current signal pair from (23) and $\mathbf{V}$. Compute the Givens angle $\theta_{pq} = \hat{\theta}_{\mathrm{GWFOE}}$ from (18).
   b) if $|\theta_{pq}| > \theta_{\min}$, update the rotation matrix $\mathbf{V}$ by rotating an angle θ_{pq} the proper coordinates.
4) *End*? If the number of sweeps c satisfies $c = K$ or no angle θ_{pq} has been updated, terminate. Otherwise return to Step 3) for another sweep, with $c = c + 1$.

At convergence, matrix $\mathbf{V}$ is an estimate of $\mathbf{Q}^{\mathrm{T}}$ in (3). The main advantage of the alternative formulation presented in this section is that the whitened sensor samples are directly used only once, for computing matrix $\mathbf{M}^z$ before starting the iterations. The moments of each signal pair at each step of Algorithm IJO-GWFOE are computed as quadratic forms involving simple vector-matrix products. The main drawback of this alternative procedure is that at a large number of components, the number of entries of the moment matrix is of order $\mathrm{O}(n^4)$. However, we will show later in this section that the complexity of the standard JO can be improved if the number of sources is low. Hence, memory problems will not appear. By "initialization" we mean a previous computation of the whitened-signal statistics to simplify subsequent calculations.

The IJO algorithm described above is reminiscent of JADE [4]. Indeed, the GWFOE with $w = 1/2$ is equivalent to JADE in the scenario of $n = 2$ sources, as seen in Section III-B and Appendix I. Moreover, JADE also calculates the cumulant matrix in advance and performs planar rotations in a JO-like fashion. Nevertheless, the equivalence between JADE and GWFOE-based algorithms vanishes in the presence of more than two sources, for JADE's cost function involves cumulants from more than two signals at each Jacobi iteration. On the other hand, JADE updates the cumulant matrix with the Givens angles after each iteration, whereas the IJO algorithm calculates the pertinent signal-pair cumulants from the moment matrix as in Proposition 1, without updating the moment matrix.

C. Computational Complexity: Optimal Jacobi Optimization

This section compares the computational complexity of the initialized and standard JO methods. As in [30], and for the sake of comparison, a floating-point operation (flop) will be considered as a real multiplication followed by an addition. The following values are used: $g = n(n-1)/2$ is the number of signal pairs, $K = 1 + \sqrt{n}$ denotes the maximum number of sweeps in the JO, and $r = n(n+1)/2$ represents the dimension of the moment matrix $\mathbf{M}^z$ in (21).

The computational burden of a fourth-order moment sample estimate is $3T$ flops. The number of flops for the JO-GWFOE algorithm is

$$f_{\mathrm{JO}} = 15gKT + 4gKT = 19gKT. \tag{24}$$

The first term is the computational cost related to the calculation of the moments, whereas the second accounts for the data

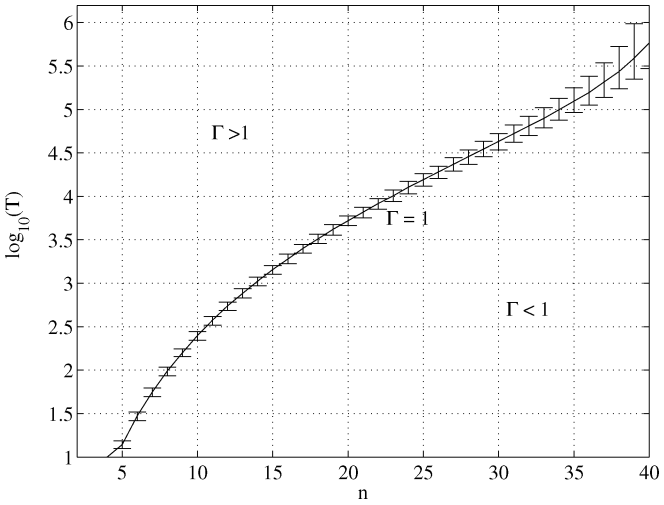

Fig. 1. Computational burden ratio between the JO and the IJO approaches. The vertical lines represent the range of (n, T) with Γ between 0.9 and 1.1.

rotation performed at Step 2b) of Algorithm JO-GWFOE. The number of operations of the IJO-GWFOE algorithm is given by

$$f_{\mathrm{IJO}} = 3T\binom{n+3}{4} + gKr(r+1) \tag{25}$$

where the first term is the number of operations needed to compute the entries of the moment matrix (21). Since some multiplications are repeated in the calculation of the moments (e.g., the product $x_i x_j$ appears in any term of the form $x_i x_j x_k x_l$), this number could be further reduced to $T(\binom{n+3}{4} + \binom{n+1}{2})$. The second term in (25) is the number of operations in computing (23) at each Givens angle. Hence, the relation between f_{JO} and f_{IJO} is

$$\Gamma(n,T) = \frac{19gKT}{T\left(\binom{n+3}{4} + \binom{n+1}{2}\right) + gKr(r+1)}. \tag{26}$$

Fig. 1 plots the loci of $\Gamma(n,T) = 1$. We can draw the following conclusions. Since usually $T > 10^2$, IJO is to be used for a low number of sources, $n \leqslant 5$. As $n \to \infty$, the number of moments μ^z_{ijkl} becomes of the order $\mathrm{O}(n^4)$, making $\Gamma(n,T) < 1$ for any sample size. This outcome takes place at $n \approx 40$. Since the IJO is not to be used for large numbers of components, potential memory problems associated with the storage of matrix $\mathbf{M}^z$ are avoided. As a result of the above decision rule, the following computationally optimal JO algorithm can be devised.

Algorithm (OJO-GWFOE)

n-dimensional GWFOE using computationally *optimal* Jacobi optimization.

1) Compute the condition $\Gamma(n,T)$ in (26), and decide:
   a) If $\Gamma(n,T) < 1$ then use JO-GWFOE.
   b) Else, use IJO-GWFOE.

V. Adaptive Algorithms

A. Adaptive Jacobi Optimization

The JO procedure is easily extended to operate online, resulting in the adaptive Jacobi optimization (AJO). The AROT [13] and the adEML [30] are methods of this type. This section derives the AJO implementation of the GWFOE pairwise solution. This implementation is referred to as AJO-GWFOE.

The JO computes the two-dimensional estimate $\hat{\theta}_{\mathrm{GWFOE}}$ for each signal pair over several sweeps. Accordingly, centroids (11)–(13) must be calculated for every sweep c and signal pair (p,q). In the design of an adaptive version, such statistics can be updated with a new sample arriving at instant t as

$$\xi_\gamma^{(c,pq)}(t+1) = (1-\lambda)\xi_\gamma^{(c,pq)}(t) + \lambda\rho^4_{c,pq}(t)\mathrm{e}^{j4\phi_{c,pq}(t)} \tag{27}$$

$$\xi_\eta^{(c,pq)}(t+1) = (1-\lambda)\xi_\eta^{(c,pq)}(t) + \lambda\rho^4_{c,pq}(t)\mathrm{e}^{j2\phi_{c,pq}(t)} \tag{28}$$

$$\beta^{(c,pq)}(t+1) = (1-\lambda)\beta^{(c,pq)}(t) + \lambda\left(\rho^4_{c,pq}(t) - 8\right) \tag{29}$$

where λ is the learning or adaption coefficient. Since we estimate the rotation matrix $\mathbf{V}$ under the whitening constraint, we must first update the whitening matrix $\mathbf{W}(t)$. In the following, we will use the relative gradient based whitening algorithm [16]

$$\mathbf{W}(t+1) = \mathbf{W}(t) + \alpha\frac{\mathbf{I}_n - \mathbf{z}(t)\mathbf{z}(t)^{\mathrm{T}}}{1 + \alpha\left|\mathbf{z}(t)^{\mathrm{T}}\mathbf{z}(t)\right|}\mathbf{W}(t) \tag{30}$$

where α is the associated learning rate, which may be different from λ. The adaptive algorithm is then:

Algorithm (AJO-GWFOE)

Adaptive n-dimensional GWFOE using standard Jacobi optimization.

Initial setting. Set $\mathbf{W}(0) = \mathbf{I}_{n\times m}$.

At each sample instant: run Algorithm JO-GWFOE replacing the following steps:

Step 1) Use (30) to update the whitening matrix $\mathbf{W}(t)$. Compute $\mathbf{z}(t) = \mathbf{W}(t)\mathbf{x}(t)$. Set $\mathbf{y}(t) = \mathbf{z}(t)$ and $c = 1$.

Step 2)
   a) Set $[z_p(t), z_q(t)]^{\mathrm{T}} = [y_p(t), y_q(t)]^{\mathrm{T}}$ to update centroid estimates $\xi_\gamma^{(c,pq)}(t)$, $\xi_\eta^{(c,pq)}(t)$ and $\beta^{(c,pq)}(t)$ in (27)–(29). Compute the Givens angle $\hat{\theta}^{(pq)}_{\mathrm{GWFOE}}$ in (18) from those estimates.

Algorithm AJO-GWFOE is the adaptive version of Algorithm JO-GWFOE. From the connections established in Section III-B, it turns out that adEML of [30] is equivalent to the AJO-GWFOE with $w = 1$.

B. Adaptive Initialized Jacobi Optimization

In this section, we develop the adaptive version of the IJO-GWFOE—consequently called *AIJO-GWFOE*—aiming to alleviate the computational burden and covergence problems of

the previous algorithm. The main idea is to adaptively update matrix $\mathbf{M}^z$ of Proposition 1 as

$$\mathbf{M}^z(t+1) = (1-\lambda)\mathbf{M}^z(t) + \lambda\mathcal{M}^z(t) \tag{31}$$

where matrix $\mathcal{M}^z(t)$ is computed as $\mathbf{M}^z$ in (21) but using only $\mathbf{z}(t)$, the whitened-output sample at time instant t. The corresponding adaptive algorithm takes the form:

Algorithm (AIJO-GWFOE)

Adaptive n-dimensional GWFOE using *initialized* Jacobi optimization.

Initial setting. Set $\mathbf{W}(0) = \mathbf{I}_{n\times m}$ and $\mathbf{V}(0) = \mathbf{I}_n$.

- At each sample instant:
  1) *Whitening*. Update the whitening matrix $\mathbf{W}(t)$ as in (30) and obtain the whitened output sample $\mathbf{z}(t) = \mathbf{W}(t)\mathbf{x}(t)$.
  2) *Moment matrix updating*. Adaptively compute matrix $\mathbf{M}^z(t)$ as in (31) using the current whitened output $\mathbf{z}(t)$ to form matrix $\mathcal{M}(t)$.
- Each N samples: set sweep number $c = 1$ and run Steps 3)–4) of Algorithm IJO-GWFOE.

In the conventional AJO-GWFOE algorithm, centroids are updated from samples of the last estimated outputs $[y_p(t), y_q(t)]^{\mathrm{T}}$. However, these outputs depend on the updated statistics of previous pairs of outputs and sweeps, and, in consequence, the statistics of latest sweeps cannot converge until the previous statistics 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, compromising the stability of the algorithm. Since the number of sweeps grows with the dimension of the problem, the AJO method typically shows convergence problems for a high number of components.

By contrast, in the AIJO-GWFOE, the learning of the separation system and the computation of the solution are decoupled. In the first stage, the output moments are updated with the last output sample. In the second stage, a current separating matrix $\mathbf{B}(t)$ is computed. The right solution for $\mathbf{B}(t)$ is obtained if the learning of $\mathbf{M}^z$ has converged. Classical results of adaptive-algorithm analysis [42] show that, if the whitened-output moments are well defined, the equilibrium point of moment matrix update (31) is locally asymptotically stable and corresponds to the ensemble average $\mathbf{M}^z$. Consequently, this two-stage design improves the stability and convergence rate of the conventional AJO. To reduce complexity, the computation of $\mathbf{V}(t)$ can be carried out every N samples, with $N > 1$. In such a case, the algorithm could better be regarded as *semi-online*.

C. Computational Complexity of the Adaptive Algorithms

We now estimate the computational cost of the AIJO-GWFOE and compare it to that of the AJO-GWFOE, AROT [13] and EASI [16]. The authors of [30] estimate the number of flops per iteration for the adEML (an AJO method) and the AROT as $C_{\mathrm{AJO}} = C_{\mathrm{adEML}} = g(18+f)(1+\sqrt{n})$ and $C_{\mathrm{AROT}} = g(14+f)(1+\sqrt{n})$, respectively, where $f = 26$. They also compute it for the EASI as $C_{\mathrm{EASI}} = n^3 + 3n^2 + hn$, where each nonlinearity elements assumed to require h flops (e.g., for cubic nonlinearities $h = 2$). An extra number of flops would have to be added in the normalized version of EASI [16]. Note also that the figures for C_{AJO} and C_{AROT} in [30] do not include the whitening stage, so $n^2(n+1)$ flops must be added.

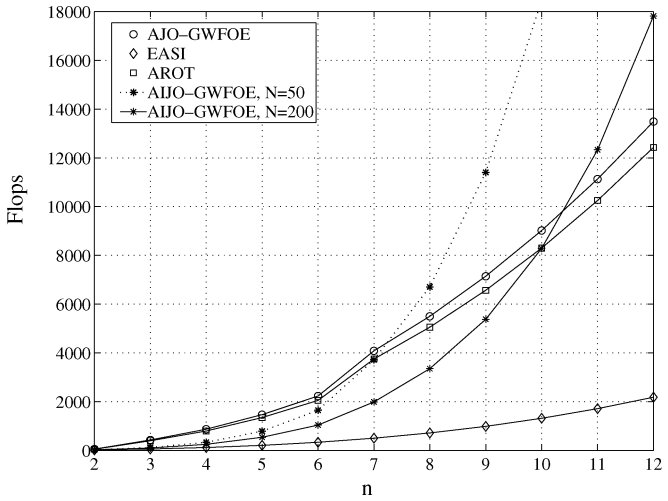

Fig. 2. Computational complexity as a function of the number of sources for AJO-GWFOE, AROT, EASI, and AIJO-GWFOE with $N = 50$ and $N = 200$.

Regarding the AIJO-GWFOE algorithm, at each sample instant this algorithm must perform the following tasks.

1) *Whitening:* The whitening algorithm (30) takes $n^2(n+1)$ flops.
2) *Moment matrix calculation:* As described before, the number of flops necessary to compute $\mathcal{M}^z(t)$ can be reduced to $\binom{n+3}{4} + \binom{n+1}{2}$.
3) *Moment matrix updating:* Adaptively computing matrix $\mathbf{M}(t)$ in (31) takes $r^2 = \binom{n+1}{2}^2$ flops.

On the other hand, each N samples, for each signal pair, we have the following.

Compute the moments: As described before, the number of flops needed to compute (23) is approximately $gKr(r+1)$.
Compute $\hat{\theta}^{(pq)}_{\mathrm{GWFOE}}$*:* Using (18), this task takes about $f = 26$ flops.
Rotate: 4 flops.

This makes $\binom{n+3}{4} + \binom{n+1}{2}[\binom{n+1}{2} + 1] + n^3 + n^2$ flops per iteration plus no more than $(1+\sqrt{n})(n(n-1)/2)[\binom{n+1}{2}^3 + \binom{n+1}{2} + 30]$ flops every N iterations.

Hence, the computational burden of AIJO is always higher than that of AJO, AROT, and EASI when $N = 1$. However, as N increases and for a reduced number of sources, we can force the complexity of AIJO below that of AJO and AROT, and of the order of EASI's. This result is illustrated in Fig. 2, which displays the number of flops per iteration as a function of the number of sources for these four adaptive methods, with $N = 50$ and $N = 200$. When the number of independent sources is $n \leqslant 7$ and $N = 50$ is selected, the complexity of AIJO is lower than AJO and AROT, as evidenced by the dotted line of Fig. 2. Also, when N is increased to 200, AIJO is less costly than AJO and AROT if the number of independent sources is $n \leqslant 10$, as

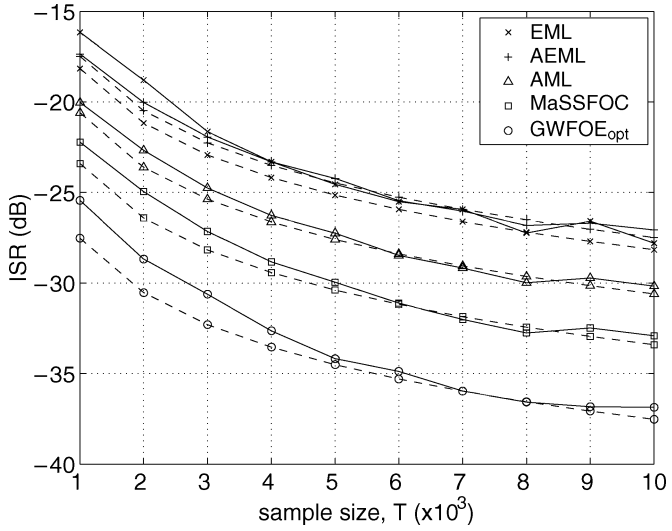

Fig. 3. ISR performance of the GWFOE versus sample size, for different weight coefficients. Uniform-Rayleigh sources, $\theta = 15°$, ν independent Monte Carlo runs, with $\nu T = 5 \cdot 10^6$. Solid lines: average experimental values. Dashed lines: asymptotic MSE (19).

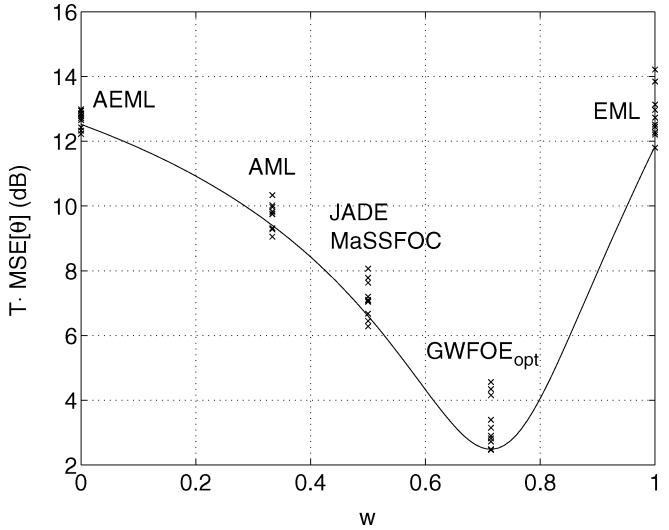

Fig. 4. Performance of the GWFOE as a function of the weight coefficient in the experiment of Fig. 3. Solid line: theoretical MSE (19). "×": experimental values from Fig. 3.

observed in Fig. 2. In such a case ($N = 200$ and $n \leqslant 7$), the computational burden of AIJO is still heavier than EASI's, but they become of the same order of magnitude.

VI. Experimental Results

The interference-to-signal power ratio (ISR) will be used as an objective separation index [1] to illustrate the main results presented in this paper. This performance index reads

$$\mathrm{ISR} = \sum_{i=1}^{n} \left(\frac{\sum_{j=1}^{n} |c_{ij}|^2}{\max\limits_{j} |c_{ij}|^2} - 1 \right) \tag{32}$$

where c_{ij} represents the element (i, j) of the global mixing–unmixing matrix $\mathbf{C}$. The ISR is an objective measure of separation performance, for it is method independent. In the two-signal case, the ISR approximates the MSE of the angle estimates around any valid separation solution (as shown at the end of Appendix II).

A. Performance of the GWFOE

We first demonstrate the potential benefits of the GWFOE and test the goodness of asymptotic approximation (19). Two source signals with independent identically distributed (i.i.d.) uniform and Rayleigh distribution are mixed through a unitary transformation with $\theta = 15°$. According to (20), this source combination provides an optimal weight value of $w_{\mathrm{opt}} = 0.7141$. Centroids are computed from their polar forms (11)–(13). ISR values are averaged over ν independent signal realizations, with $\nu T = 5 \cdot 10^6$. Fig. 3 shows the ISR performance obtained by the EML, AEML, AML, MaSSFOC, and optimal GWFOE, together with the expected asymptotic MSE, for varying sample size. The optimal GWFOE substantially outperforms the other estimators; e.g., it proves five and ten times more efficient than the AML and the AEML, respectively. The fitness of asymptotic approximation (19) is very precise in all cases and improves as T increases, as expected. Fig. 4 shows the variation in the MSE of the GWFOE angle estimates as a function of the weight coefficient. The solid line plots the theoretical values of $T \cdot \mathrm{MSE}_{\mathrm{GWFOE}}$ from (19), whereas the crosses represent the empirical values of $T \cdot \mathrm{ISR}$ obtained in Fig. 3. Remark that a 10-dB gap appears between the maximum and the minimum performance achievable by the GWFOE family in this scenario. These results highlight the substantial impact that the choice of w can have on the separation performance.

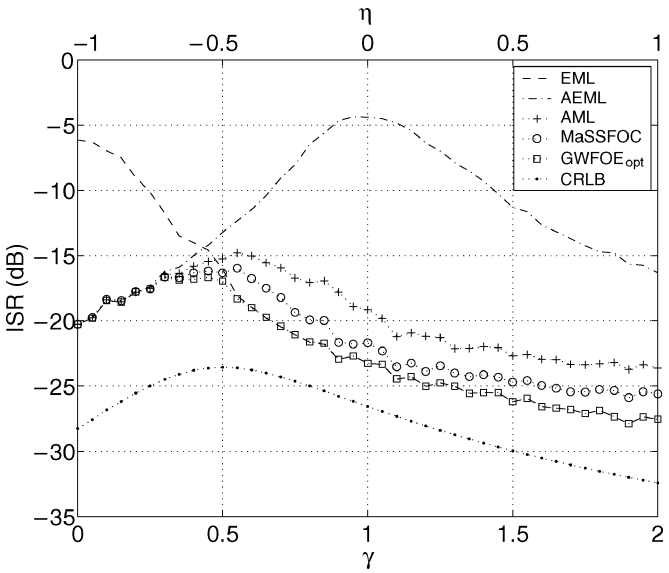

Fig. 5. ISR performance versus sks γ and skd η. GGD sources, $\kappa_{04}^s = 0.5$, $\theta = 15°$, $T = 5 \cdot 10^3$ samples, 10^3 Monte Carlo runs.

The generalized Gaussian distribution (GGD) with shape parameter α, $p(s) \propto \exp(-|s|^\alpha)$, is used as source pdf in the simulation of Fig. 5. We fix $\kappa_{04}^s = 0.5$ and smoothly vary κ_{40}^s to generate a range of sks and skd values. The optimal GWFOE, with w_{opt} calculated as in (20) and shown in Fig. 6, is compared with other analytic solutions and the Cramer–Rao lower bound (CRLB) obtained in [22] for the real case. The optimal

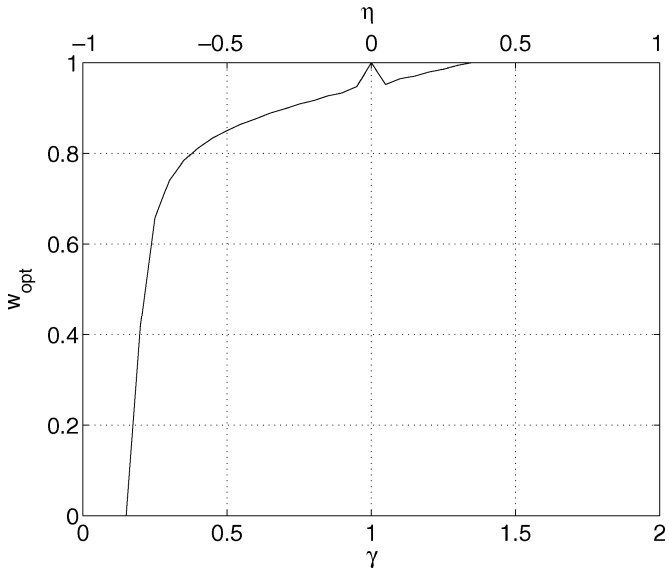

Fig. 6. Optimal value of the GWFOE weight parameter in the separation scenario of Fig. 5.

GWFOE avoids EML's and AEML's performance degradation around $\gamma = 0$ and $\eta = 0$ (respectively) and, though closely followed by MaSSFOC and AML, approaches the CRLB more tightly than any of the other methods.

When the source distribution is unknown, the iterative procedure presented at the end of Section III-C can be used to estimate GWFOE's optimal weight. To illustrate the performance of this iterative method, uniform-Rayleigh source realizations are mixed by a (2×2) mixing matrix with elements drawn from a zero-mean unit-variance Gaussian distribution. The mixture is first whitened via PCA based on the singular value decomposition of the observed data matrix. The GWFOE with initial weight uniformly distributed in [0, 1] is then applied to the whitened signals, resulting in a set of estimated sources. From the sample estimate of the source statistics, $w_{\rm opt}$ is obtained as in (20); then the GWFOE with the new weight is applied to the whitened observations, and so forth. Fig. 7 displays the trajectories of the $w_{\rm opt}$ estimate as a function of the iteration number, for several sample sizes T. The curves have been averaged over ν independent Monte Carlo runs, with $\nu T = 5{\cdot}10^6$. The method typically converges to the theoretical value of the optimal weight within just one to two iterations, the final bias decreasing as the sample size increases.

B. Performance of the OJO-GWFOE

The performance of the n-dimensional OJO-GWFOE using SICA [25] $(w = 3/7)$ is compared to JADE [4], the fourth-order-based ME method by Comon [12], and the FastICA algorithm [43].[3] The same whitening method is used in all algorithms, as the focus is on the computation of the unitary matrix $\mathbf{Q}$. A few changes are introduced in the code by Comon to save up some operations, while FastICA is executed with the parameters by default, including stabilization. In the OJO-GWFOE,

[3]MATLAB code for these methods is available at ftp://sig.enst.fr/pub/jfc/Algo/Jade/jadeR.m, http://www.i3s.unice.fr/~comon/matlab.html, and http://www.cis.hut.fi/projects/ica/fastica/index.shtml.

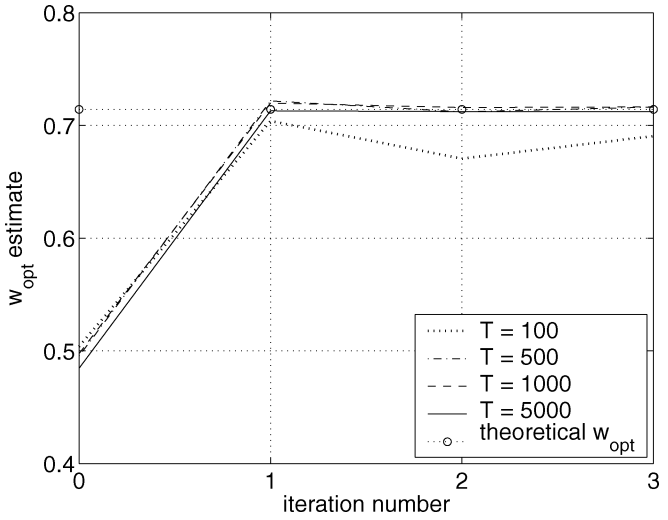

Fig. 7. Iterative estimation of GWFOE's optimal weight from the observed sensor output. Uniform-Rayleigh sources, mixing matrix with normalized Gaussian random elements, initial $w_{\rm opt}$ with uniform random distribution in [0, 1], ν independent Monte Carlo runs, with $\nu T = 5 \cdot 10^6$.

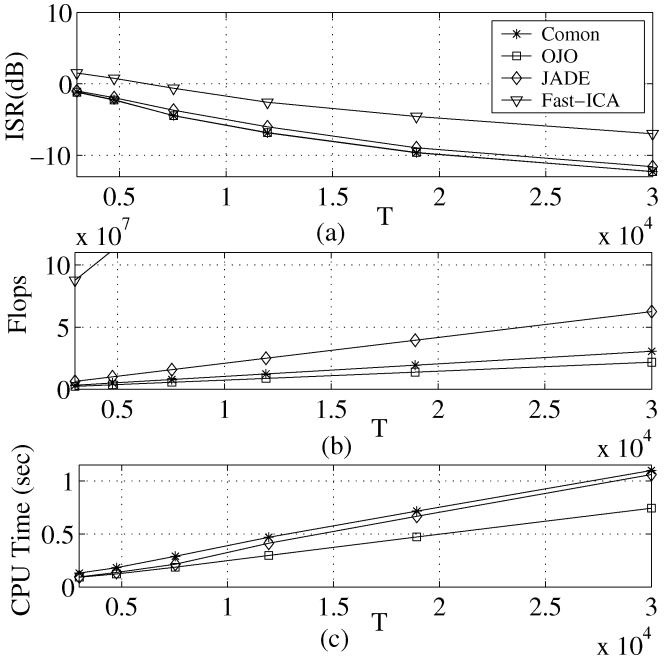

Fig. 8. Comparison of Comon's ME, OJO-GWFOE, JADE, and FastICA in the $n = 6$ dimensional case: (a) mean ISR, (b) flops, and (c) CPU time.

JADE, and ME by Comon, the Jacobi optimization stops whenever no angle has been updated more than $\pi/360$ rad (0.5°) or it has iterated more than $K = 1 + \sqrt{n}$ times. The flop count and CPU time are used as indices of computational complexity. The mixing matrix entries a_{ij} are random numbers in the range $[-1,1]$. The experiments have been performed using MATLAB on an Intel Pentium 4 2.40-GHz processor and 512 MB RAM.

In this experiment, $n = 6$ zero-mean unit-variance signals with different distributions are mixed: uniform, Laplacian $(p(s) \propto \mathrm{e}^{-|s|})$, Rayleigh $(p(s) \propto s\mathrm{e}^{-s^2/2})$, exponential $(p(s) \propto \mathrm{e}^{-s},\ s \geqslant 0)$, Gaussian $(p(s) \propto \mathrm{e}^{-s^2/2})$, and log-normal $(p(s) \propto \mathrm{e}^{-(\log s)^2/2})$. We study the performance in the sample-size range $3 \cdot 10^3 \leqslant T \leqslant 3 \cdot 10^4$. Each point corresponds to the average of 1000 independent Monte Carlo runs in

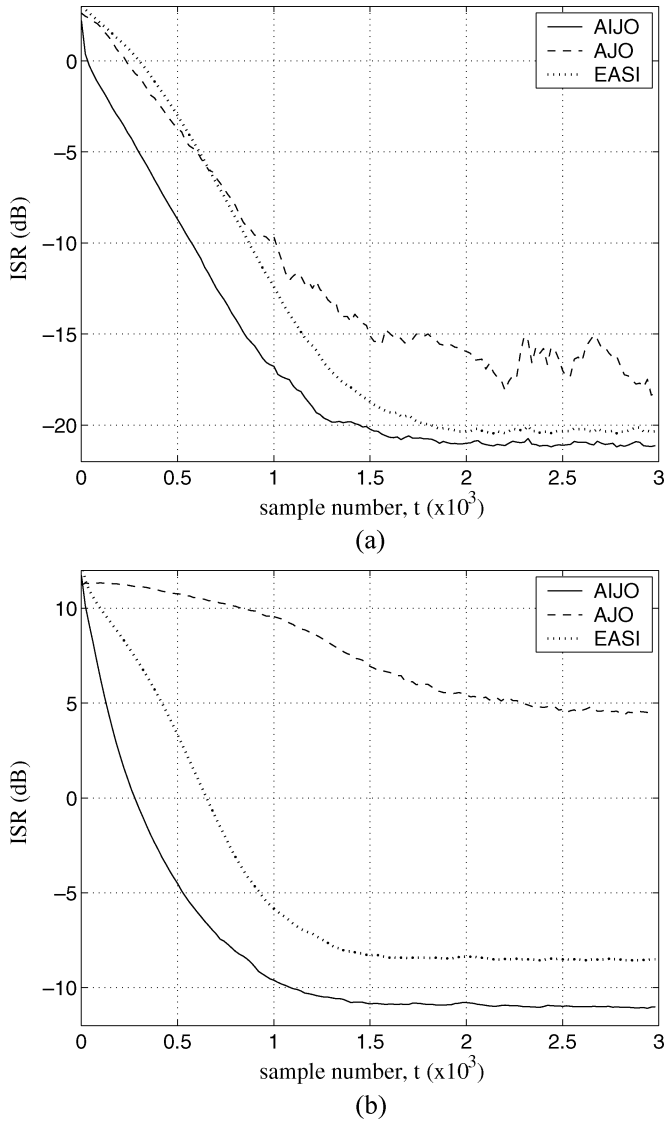

Fig. 9. Performance of the AIJO-GWFOE, AJO-GWFOE, and EASI methods for (a) $n = 3$ (uniform, binary, and sinusoid) and (b) $n = 8$ (six uniform, one binary, and one sinusoid).

which the mixing matrix is randomly chosen. Fig. 8(a) shows that OJO-SICA and ME have nearly identical performance, as expected. JADE also shows a good performance, close to that of the OJO-SICA. The FastICA method exhibits the worst behavior. Regarding the computational cost [Fig. 8(b) and (c)], the OJO-GWFOE method presented in this paper clearly outperforms the other methods. Although JADE takes a larger number of flops than the ME, its CPU time is lower. Similar results may be expected for other mixtures, except for FastICA. Although this latter method usually presents good performance at a low complexity, it may exhibit poor convergence and a high computational burden if its parameters are not properly chosen, as discussed in [25] and [44]. This is evidenced in this experiment, where the parameters by default yield a poor ISR, and a number of flops and CPU time out of the plotted range.

C. Performance of the AIJO-GWFOE

The AIJO-GWFOE method with weight parameter $w = 1$ is compared to other adaptive procedures: AJO-GWFOE with the same value of w (adEML) [30] and EASI [16]. 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. For all methods, the separating matrix is initialized at the identity $\mathbf{B}(0) = \mathbf{I}_n$. Performance curves are averaged over 1000 independent Monte Carlo runs. By default, the solution of the AIJO-GWFOE method is calculated at each sample, $N = 1$. Results for any other N may be easily deduced from the plots for $N = 1$ by holding the value obtained for sample kN until sample $(k+1)N$. The first experiment considers a mixture of three independent sources: a binary sequence, a uniformly distributed process, and a sinusoid with random frequency and phase. Fig. 9(a) shows that AIJO-GWFOE converges to a lower ISR than AJO-GWFOE and EASI. In addition, the stationary state is reached faster than in the two other methods.

To compare the performance of the three algorithms in a more complex separation system, a mixture of eight independent sources is observed in a second setup. All but two of these are uniformly distributed processes; the other two are a binary sequence and a sinusoid with random frequency and phase. The evolution of the performance curves in Fig. 9(b) demonstrates again that the AIJO-GWFOE provides the best final ISR in the lowest number of iterations. By contrast, the AJO-GWFOE algorithm shows a slow poor convergence.

VII. Conclusion

This paper has investigated the approximate closed-form solutions to ICA contrasts in the two-dimensional case. The GWFOE gathers under the same expression many existing analytic solutions based on fourth-order statistics. In particular, for $w = 1/2$ the GWFOE is equivalent to JADE in the two-source scenario. The weight parameter of the most efficient estimator in the GWFOE class has been obtained as a function of the source statistics. Even if these are unknown, a simple iterative procedure allows a fast accurate estimation of the optimal weight. The optimal GWFOE can considerably outperform other analytic solutions, as demonstrated by experimental results.

Analytic solutions can be extended to the general scenario of more than two sources by means of the pairwise JO technique. The algebraic structure of the problem has been exploited through the multilinearity property of moments and cumulants in a bid to optimize the computational complexity of the conventional JO procedure. The resulting IJO computes the necessary statistics before the iteration process, so that the observed signal samples are employed only once. A detailed discussion has concluded that the decision on which method to use (JO or IJO) depends on the relative values of source number and sample size. In our experiments, IJO-GWFOE using SICA has achieved a similar performance than Comon's ME and FastICA, with a reduced complexity.

In the adaptive implementation of IJO, the learning of the system and the computation of the ICA solution are decoupled. This feature enhances the convergence properties (particularly the stability) of the algorithm. With a complexity that can be reduced to the order of EASI's, AIJO presents the advantage of an increased robustness to the source distributions. Experimental

results have shown that the convergence of AIJO is faster than EASI's and adEML's (indeed, its ISR evolution is always below that of the other methods), reaching the best final performance for any number of sources and different source distributions.

Further work includes the development of GWFOE's optimal weight coefficient as a function of the array-output statistics in order to enable a fully blind operation and the incorporation of the optimal GWFOE in the multidimensional JO-based algorithms. The separation performance and convergence characteristics in the presence of additive noise and interference needs to be explored, for both offline and online implementations. Extensions to statistics of orders other than four also deserves to be investigated. The use of characteristic functions [45] might prove helpful in that line of inquiry.

Appendix I
Equivalence Between GWFOE With $w = 1/2$ and JADE for $n = 2$

The maximization of contrast function (8) is associated with the joint approximate diagonalization of the so-called parallel set of cumulant matrices $\{\mathbf{N}_r\}$, whose entries are defined as $(\mathbf{N}_r)_{ij} = \mathrm{C}^z_{ijkl}$ [4]. In this Appendix we prove that, in the two-source scenario, the solution provided by the conventional version of JADE based on the parallel set provides the GWFOE solution (18) with $w = 1/2$.

For $n = 2$, the cumulant matrices of the parallel have the form $\mathbf{N}_r = \begin{bmatrix} a_r & b_r \\ c_r & d_r \end{bmatrix}$

$$\mathbf{N}_1 = \begin{bmatrix} \kappa^z_{40} & \kappa^s_{31} \\ \kappa^z_{31} & \kappa^z_{22} \end{bmatrix}, \quad \mathbf{N}_2 = \mathbf{N}_3 = \begin{bmatrix} \kappa^z_{31} & \kappa^z_{22} \\ \kappa^z_{22} & \kappa^z_{13} \end{bmatrix}$$
$$\mathbf{N}_4 = \begin{bmatrix} \kappa^z_{22} & \kappa^s_{13} \\ \kappa^z_{13} & \kappa^z_{04} \end{bmatrix}. \tag{33}$$

As shown in [4, Sec. 8.1], the joint diagonalization criterion is equivalent to maximizing $\psi = \sum_r |a'_r - d'_r|^2$, where (a'_r, d'_r) are the diagonal elements of $\mathbf{Q}^{\mathrm{T}}\mathbf{N}_r\mathbf{Q}$, matrix $\mathbf{Q}$ denoting the sought Givens rotation of angle θ in (9). Following [4, Sec. 8.1], the criterion can be expressed as $\psi = \mathbf{v}^{\mathrm{T}}\mathbf{H}\mathbf{v}$, with $\mathbf{v} = [\cos 2\theta, \sin 2\theta]^{\mathrm{T}}$ and $\mathbf{H} \stackrel{\mathrm{def}}{=} \mathbf{G}\mathbf{G}^{\mathrm{T}}$, $\mathbf{G} = [\mathbf{g}_1, \ldots, \mathbf{g}_4]$, $\mathbf{g}_r = [a_r - d_r, b_r + c_r]^{\mathrm{T}}$, where (a_r, d_r) and (b_r, c_r) represent the diagonal and off-diagonal entries, respectively, of $\mathbf{N}_r$. Hence, $\mathbf{v}$ is the dominant eigenvector of the symmetric matrix $\mathbf{H} = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$, whose elements are given by

$$\begin{aligned} a &= \sum_{r=1}^{4}(a_r - d_r)^2 \\ &= (\kappa^z_{40} - \kappa^z_{22})^2 + 2(\kappa^z_{31} - \kappa^z_{13})^2 + (\kappa^z_{22} - \kappa^z_{04})^2 \end{aligned} \tag{34}$$

$$\begin{aligned} b &= \sum_{r=1}^{4}(a_r - d_r)(b_r + c_r) \\ &= 2\kappa^z_{31}(\kappa^z_{40} - \kappa^z_{22}) + 4\kappa^z_{22}(\kappa^z_{31} - \kappa^z_{13}) \\ &\quad + 2\kappa^z_{13}(\kappa^z_{22} - \kappa^z_{04}) \end{aligned} \tag{35}$$

$$\begin{aligned} d &= \sum_{r=1}^{4}(b_r + c_r)^2 \\ &= 4(\kappa^z_{31})^2 + 8(\kappa^z_{22})^2 + 4(\kappa^z_{13})^2. \end{aligned} \tag{36}$$

Now, to find the dominant eigenvector of $\mathbf{H}$, we take into account that its eigenvector matrix must be of the form $\tilde{\mathbf{V}} = [\mathbf{v}, \tilde{\mathbf{v}}]$, with $\tilde{\mathbf{v}} = \pm[-\sin 2\theta, \cos 2\theta]^{\mathrm{T}}$. Also, matrix $\hat{\mathbf{V}}$ must diagonalize $\mathbf{H}$. Thus, we force a diagonal structure for matrix $\tilde{\mathbf{V}}^{\mathrm{T}}\mathbf{H}\tilde{\mathbf{V}}$, which leads to two constraints on the resulting off-diagonal elements reducing to $\mathbf{v}^{\mathrm{T}}\mathbf{H}\tilde{\mathbf{v}} = b(\cos^2 2\theta - \sin^2 2\theta) - (a-d)\cos 2\theta \sin 2\theta = 0$. We thus obtain $\tan 4\theta = 2b/(a-d)$. By means of some straightforward algebraic manipulations on (34)-(36), this solution is readily shown to coincide with the GWFOE solution (18) for $w = 1/2$. □

Appendix II
Asymptotic Analysis of the GWFOE

In this Appendix, we analyze the asymptotic performance of the GWFOE estimator (18) for i.i.d. sources. Our main objective is an analytic expression for its large-sample MSE. The estimator reads

$$\tilde{\theta} = \frac{1}{4}\angle\left\{w\hat{\beta}\hat{\xi}_\gamma + (1-w)\hat{\xi}^2\right\} \tag{37}$$

where

$$\hat{\xi} = \frac{1}{T}\sum_{k=1}^{T}\rho_k^4 \mathrm{e}^{j4\phi_k}, \quad \hat{\xi}_\eta = \frac{1}{T}\sum_{k=1}^{T}\rho_k^4 \mathrm{e}^{j2\phi_k}, \quad \hat{\beta} = \frac{1}{T}\sum_{k=1}^{T}\rho_k^4 - 8 \tag{38}$$

are the sample estimates of centroids (11)-(13). Note that

$$\hat{\xi}_\gamma = \hat{\xi}'_\gamma \mathrm{e}^{j4\theta}, \quad \hat{\xi}_\eta = \hat{\xi}'_\eta \mathrm{e}^{j2\theta} \tag{39}$$

with

$$\hat{\xi}'_\gamma = \frac{1}{T}\sum_{k=1}^{T}\rho_k^4 \mathrm{e}^{j4\phi'_k}, \quad \hat{\xi}'_\eta = \frac{1}{T}\sum_{k=1}^{T}\rho_k^4 \mathrm{e}^{j2\phi'_k}. \tag{40}$$

Hence, $\Delta\hat{\theta} \stackrel{\mathrm{def}}{=} (\hat{\theta} - \theta) = (1/4)\angle\hat{\xi}'$, where $\hat{\xi}' = w\hat{\beta}\hat{\xi}'_\gamma + (1-w)\hat{\xi}'^2_\eta$. By virtue of the law of large numbers, the combined source centroid $\hat{\xi}'$ is a consistent estimator of a positive real number

$$\mathrm{E}[\hat{\xi}'] \xrightarrow[T\to\infty]{} w\gamma^2 + (1-w)\eta^2. \tag{41}$$

It follows that the GWFOE is also consistent and, in particular

$$\mathrm{E}[\Delta\tilde{\theta}] \xrightarrow[T\to\infty]{} 0. \tag{42}$$

Now, $\mathrm{MSE}_{\mathrm{GWFOE}} \stackrel{\mathrm{def}}{=} \mathrm{MSE}[\hat{\theta}] = \mathrm{E}[\Delta\hat{\theta}^2]$. At large T, and since the estimator is consistent, $\Delta\hat{\theta}$ will be close to zero; thus $4\Delta\hat{\theta} \approx \tan(4\Delta\hat{\theta})$. Also, since $\Delta\hat{\theta}$ is small and, according to (41), typically $\mathrm{E}[\mathbb{R}\mathrm{e}(\hat{\xi}')] \gg \mathrm{E}[\mathbb{I}\mathrm{m}(\hat{\xi}')]$, the variations of $\hat{\theta}$ will

be mainly due to fluctuations in the imaginary part of ξ'. As a result, we can approximate

$$16\mathrm{E}[\Delta\tilde{\theta}^2] \approx \frac{\mathrm{E}\left[\mathbb{I}\mathrm{m}^2(\hat{\xi}')\right]}{\mathrm{E}^2\left[\mathbb{R}\mathrm{e}(\hat{\xi}')\right]}. \tag{43}$$

Being real valued, $\hat{\beta}$ does not alter the argument of $\hat{\xi}'_\gamma$. Moreover, on the grounds of consistency, it can be further assumed that $\hat{\beta} \approx \gamma$. Then

$$\begin{aligned}\mathbb{R}\mathrm{e}(\hat{\xi}') \approx\, & w\gamma\mathbb{R}\mathrm{e}^2\left(\hat{\xi}'_\gamma\right) \\ & + (1-w)\left\{\mathbb{R}\mathrm{e}^2\left(\hat{\xi}'_\eta\right) - \mathbb{I}\mathrm{m}^2\left(\hat{\xi}'_\eta\right)\right\} \qquad (44)\\ \mathbb{I}\mathrm{m}(\hat{\xi}') \approx\, & w\gamma\mathbb{I}\mathrm{m}^2\left(\hat{\xi}'_\gamma\right) \\ & + 2(1-w)\mathbb{R}\mathrm{e}\left(\hat{\xi}'_\eta\right)\mathbb{I}\mathrm{m}\left(\hat{\eta}'_\gamma\right) \qquad (45)\end{aligned}$$

with

$$\begin{aligned}\hat{\xi}'_\gamma =\, & \frac{1}{T}\sum_{k=1}^{T}\left(s_{1k}^4 - 6s_{1k}^2 s_{2k}^2 + s_{2k}^4\right) \\ & + j\frac{4}{T}\sum_{k=1}^{T}\left(s_{1k}^3 s_{2k} - s_{1k}s_{2k}^3\right) \qquad (46)\\ \hat{\xi}'_\eta =\, & \frac{1}{T}\sum_{k=1}^{T}\left(s_{1k}^4 - s_{2k}^4\right) + j\frac{2}{T}\sum_{k=1}^{T}\left(s_{1k}^3 s_{2k} + s_{1k}s_{2k}^3\right) \qquad (47)\end{aligned}$$

where, to ease the notation, we have written $s_{pk} = s_p(k)$, $p = 1, 2$. The denominator of (43) can be easily obtained by invoking the consistency of the real part of $\hat{\xi}'$ in (41). The calculation of the numerator is slightly more involved. From (45)–(47), we have (48) as shown at the bottom of the page. Taking into account the i.i.d. assumption, the most significant parts of terms A–C turn out to be

$$\begin{aligned}\mathrm{A} =\, & T\mathrm{E}\left[\left(s_1^3 s_2 - s_1 s_2^3\right)^2\right] \\ \mathrm{B} =\, & T(T-1)(T-2)\eta^2\mathrm{E}\left[\left(s_1^3 s_2 + s_1 s_2^3\right)^2\right] \\ \mathrm{C} =\, & T(T-1)\eta\mathrm{E}\left[\left(s_1^3 s_2 - s_1 s_2^3\right)\left(s_1^3 s_2 + s_1 s_2^3\right)\right]. \qquad (49)\end{aligned}$$

Gathering and rearranging terms, we arrive at

$$\begin{aligned}\mathrm{E}\left[\mathbb{I}\mathrm{m}^2(\hat{\xi}')\right] \xrightarrow[T\to\infty]{} & \frac{16}{T}\mathrm{E}\left\{\left[w\gamma\left(s_1^3 s_2 - s_1 s_2^3\right)\right.\right. \\ & \left.\left. + (1-w)\eta\left(s_1^3 s_2 + s_1 s_2^3\right)\right]^2\right\}. \qquad (50)\end{aligned}$$

Finally, the combination of (41), (43), and (50) yields the asymptotic MSE of the GWFOE shown in (19).

The derivation of w_{opt} is simplified with the substitutions $\tilde{\gamma} = \gamma(s_1^3 s_2 - s_1 s_2^3)$ and $\tilde{\eta} = \eta(s_1^3 s_2 + s_1 s_2^3)$, in which case $\mathrm{MSE}_{\mathrm{GWFOE}}$ can be written as

$$\mathrm{MSE}_{\mathrm{GWFOE}} = \frac{\mathrm{E}\left\{\left[w(\tilde{\gamma}-\tilde{\eta}) + \tilde{\eta}\right]^2\right\}}{T\left[w(\gamma^2-\eta^2) + \eta^2\right]^2}. \tag{51}$$

This function of w becomes constant if $\gamma\eta = 0$, i.e., $|\kappa_{40}^s| = |\kappa_{04}^s|$. Performance then reduces to that of the EML (when $\eta = 0$) or AEML (when $\gamma = 0$) estimators for any w. Otherwise, in the interval of interest, $0 \leqslant w \leqslant 1$, we have that $w(\gamma^2 - \eta^2) + \eta^2 \neq 0$. The derivative of (51) then cancels at

$$w_{\mathrm{opt}} = \frac{\mathrm{E}\left\{\tilde{\eta}(\gamma^2\tilde{\eta} - \eta^2\tilde{\gamma})\right\}}{\mathrm{E}\left\{\gamma^2\tilde{\eta}^2 + \eta^2\tilde{\gamma}^2 - (\gamma^2+\eta^2)\tilde{\gamma}\tilde{\eta}\right\}}. \tag{52}$$

Some tedious but straightforward algebraic simplifications then show that the above expression reduces to (20). In addition, it is simple to check that $\partial^2\mathrm{MSE}_{\mathrm{GWFOE}}/\partial w^2|_{w_{\mathrm{opt}}} > 0$, so that w_{opt} defines a minimum.

To conclude this asymptotic study, it is interesting to realize the connection between $\mathrm{MSE}[\hat{\theta}]$ and the ISR performance parameter (32). Assuming a unitary mixture in the two-signal case, the global transformation $\mathbf{C} = \mathbf{V}\mathbf{Q} = \hat{\mathbf{Q}}^{\mathrm{T}}\mathbf{Q}$ is a rotation of angle $(\theta - \hat{\theta})$. Any angle estimate of the form $(\hat{\theta} - \theta) = \Delta\tilde{\theta} + k\pi/2$, with small $\Delta\tilde{\theta}$ and integer k, provides a valid separation solution up to the inherent separation indeterminacies mentioned in Section II. This angle estimate produces $\mathrm{ISR} = \tan^2\Delta\tilde{\theta} \approx \Delta\tilde{\theta}^2$. As a result, in the vicinity of a valid separation solution, the average ISR approximates $\mathrm{MSE}[\hat{\theta}]$ without the potential bias introduced by the admissible $(k\pi/2)$-rad rotations.

$$\begin{aligned}\mathrm{E}\left[\mathbb{I}\mathrm{m}^2(\hat{\xi}')\right] = & \frac{16w^2\gamma^2}{T^2}\underbrace{\sum_i\sum_j \mathrm{E}\left[\left(s_{1i}^3 s_{2i} - s_{1i}s_{2i}^3\right)\left(s_{1j}^3 s_{2j} - s_{1j}s_{2j}^3\right)\right]}_{\mathrm{A}} \\ & + \frac{16(1-w)^2}{T^4}\underbrace{\sum_i\sum_j\sum_k\sum_l \mathrm{E}\left[\left(s_{1i}^4 - s_{2i}^4\right)\left(s_{1j}^4 - s_{2j}^4\right)\left(s_{1k}^3 s_{2k} + s_{1k}s_{2k}^3\right)\left(s_{1l}^3 s_{2l} + s_{1l}s_{2l}^3\right)\right]}_{\mathrm{B}} \\ & + \frac{32w(1-w)\gamma}{T^3}\underbrace{\sum_i\sum_j\sum_k \mathrm{E}\left[\left(s_{1i}^3 s_{2i} - s_{1i}s_{2i}^3\right)\left(s_{1j}^4 - s_{2j}^4\right)\left(s_{1k}^3 s_{2k} + s_{1k}s_{2k}^3\right)\right]}_{\mathrm{C}} \qquad (48)\end{aligned}$$

APPENDIX III
PROOF OF PROPOSITION 1

Expression (23) is a particular case of

$$\mu^y_{pqrs} = \sum_{ij} \mathbf{V}(p,i)\mathbf{V}(q,j) \sum_{kl} \mathbf{V}(r,k)\mathbf{V}(s,l)\mu^z_{ijkl}. \quad (53)$$

Let us denote $\mathbf{M}^z$ the $(r \times r), r = n(n+1)/2$, symmetric matrix containing the fourth-order moments $\{\mu^z_{ijkl}\}^n_{i,j,k,l=1}$. Moment μ^z_{ijkl} is stored in the entry $\mathbf{M}^z(a(i,j), a(k,l))$, where a is given by (22). In order to exploit the symmetry of the whitened-output moment tensor, only the moments with $i \leqslant j$ and $k \leqslant l$ are kept. The computation of (53) can be expressed as a quadratic form involving matrix $\mathbf{M}^z$ and a pair of column vectors related to matrix $\mathbf{V}$

$$\mu^y_{pqrs} = \mathbf{v}^{\mathrm{T}}_{pq}\mathbf{M}^z\mathbf{v}_{rs}. \quad (54)$$

To guarantee the equivalence between this quadratic form and (53), vectors $\mathbf{v}_{pq}$ and $\mathbf{v}_{rs}$ must be constructed by arranging the entries of $\mathbf{V}$ in accordance with the structure of $\mathbf{M}^z$

$$\mathbf{v}_{pq}\left(a(i,j)\right) = \begin{cases} \mathbf{V}(p,i)\mathbf{V}(q,j) + \mathbf{V}(p,j)\mathbf{V}(q,i), & i < j \\ \mathbf{V}(p,i)\mathbf{V}(q,i), & i = j \end{cases} \quad (55)$$

where indexes (i,j) and a are related through (22). □

ACKNOWLEDGMENT

V. Zarzoso wishes to thank P. Comon for his kind hospitality.

REFERENCES

[1] V. Zarzoso and A. K. Nandi, "Blind source separation," in *Blind Estimation Using Higher-Order Statistics*, A. K. Nandi, Ed. Boston, MA: Kluwer Academic, 1999, ch. 4, pp. 167–252.

[2] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

[3] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. Chichester, U.K.: Wiley, 2002.

[4] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *Proc. Inst. Elect. Eng. F*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[5] J. F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, pp. 2009–2025, Oct. 1998.

[6] A. Caamaño-Fernandez, R. Boloix-Tortosa, J. Ramos, and J. J. Murillo-Fuentes, "Hybrid higher-order statistics learning in multiuser detection," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 34, no. 4, pp. 417–424, Nov. 2004.

[7] S. Makeig, T.-P. Jung, D. Ghahremani, A. Bell, and T. Sejnowski, "Blind separation of auditory event-related brain responses into independent components," in *Proc. Nat. Acad. Sci.*, 1997, pp. 10979–10984.

[8] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Fetal electrocardiogram extraction by blind source subspace separation," *IEEE Trans. Biomed. Eng. (Special Topic Section Advances in Statistical Signal Processing for Medicine)*, vol. 47, pp. 567–572, May 2000.

[9] V. Zarzoso and A. K. Nandi, "Noninvasive fetal electrocardiogram extraction: Blind separation versus adaptive noise cancellation," *IEEE Trans. Biomed. Eng.*, vol. 48, pp. 12–18, Jan. 2001.

[10] J. Murillo-Fuentes, H. Molina-Bulla, and F. González-Serrano, "Independent component analysis applied to digital image watermarking," in *Proc. ICASSP'01*, Salt Lake City, UT, May 2001, vol. III, pp. 1997–2000.

[11] T. Lee, M. Lewicki, and T. Sejnowski, "Unsupervised classification with non-Gaussian mixture models using ICA," in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 1999, vol. 11, pp. 58–64.

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

[13] ——, "Separation of stochastic processes," in *Proc. Workshop Higher-Order Spectral Anal.*, Vail, CO, Jun. 28–30, 1989, pp. 174–179.

[14] V. Zarzoso, "Closed-form higher-order estimators for blind separation of independent source signals in instantaneous linear mixtures," Ph.D. dissertation, Univ. of Liverpool, Liverpool, U.K., Oct. 1999.

[15] V. Zarzoso and A. K. Nandi, "Unified formulation of closed-form estimators for blind source separation in real instantaneous linear mixtures," in *Proc. ICASSP'00*, Istanbul, Turkey, Jun. 5–9, 2000, vol. V, pp. 3160–3163.

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

[17] F. Harroy and J.-L. Lacoume, "Maximum likelihood estimators and Cramer-Rao bounds in source separation," *Signal Process.*, vol. 55, no. 2, pp. 167–177, 1996.

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

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

[20] E. Moreau and O. Macchi, "Higher order contrast for self-adaptive source separation," *Int. J. Adapt. Contr. Signal Process.*, vol. 10, no. 1, pp. 19–46, Jan 1996.

[21] P. Comon and E. Moreau, "Improved contrast dedicated to blind separation in communications," in *Proc. ICASSP'97*, Munich, Germany, 1997, vol. V, pp. 3453–3456.

[22] M. Ghogho, A. Swami, and T. Durrani, "Approximate maximum likelihood blind source separation with arbitrary source pdfs," in *Proc. SSAP'00*, Pocono Manor, PA, Aug. 14–16, 2000, pp. 368–372.

[23] M. Gaeta and J.-L. Lacoume, "Source separation without a priori knowledge: the maximum likelihood solution," in *Proc. EUSIPCO'90*, Barcelona, Spain, 1990, vol. V, pp. 621–624.

[24] F. Herrmann and A. Nandi, "Blind separation of linear instantaneous mixture using close forms estimators," *Signal Process.*, vol. 81, no. 7, pp. 1537–1556, Jul. 2001.

[25] J. J. Murillo-Fuentes and F. J. González-Serrano, "A sinusoidal contrast function for the blind separation of statistically independent sources," *IEEE Trans. Signal Process.*, vol. 52, pp. 3459–3463, Dec. 2004.

[26] T. Blaschke and L. Wiskott, "CuBICA: Independent component analysis by simultaneous third- and fourth-order cumulant diagonalization," *IEEE Trans. Signal Process.*, vol. 52, pp. 1250–1256, May 2004.

[27] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[28] A. Bunse-Gerstner, R. Byers, and V. Mehrmann, "Numerical methods for simultaneous diagonalization," *SIAM J. Matrix Anal. Applicat.*, vol. 14, no. 4, pp. 927–949, 1993.

[29] P. Comon, "Remarques sur la diagonalisation tensorielle par la méthode de Jacobi," in *Proc. XIVème Colloque GRETSI*, Juan-les-Pins, France, Sep. 13–16, 1993, pp. 125–128.

[30] V. Zarzoso and A. K. Nandi, "Adaptive blind source separation for virtually any source probability density function," *IEEE Trans. Signal Process.*, vol. 48, pp. 477–488, Feb. 2000.

[31] I. J. Clarke, "Direct exploitation of non-Gaussianity as a discriminant," in *Proc. EUSIPCO'98*, Rhodes, Greece, Sept. 8–11, 1998, vol. IV, pp. 2057–2060.

[32] V. Zarzoso, F. Herrmann, and A. K. Nandi, "Weighted closed-form estimators for blind source separation," in *Proc. SSP-2001, 11th IEEE Workshop Statist. Signal Process.*, Singapore, Aug. 6–8, 2001, pp. 456–459.

[33] J. Murillo-Fuentes and F. González-Serrano, "Independent component analysis with sinusoidal fourth-order contrast," in *Proc. ICASSP'01*, Salt Lake City, UT, May 2001, vol. V, pp. 2785–2788.

[34] J. J. Murillo-Fuentes, R. Boloix-Tortosa, and F. J. González-Serrano, "Initialized Jacobi optimization in independent component analysis," in *Proc. ICA-2003 4th Int. Symp. Indep. Comp. Anal. Blind Signal Separat.*, Nara, Japan, Apr. 1–4, 2003.

[35] ——, "Adaptive initialized Jacobi optimization in independent component analysis," in *Proc. ICA-2003 4th Int. Symp. Indep. Comp. Anal. Blind Signal Separat.*, Nara, Japan, Apr. 1–4, 2003.

[36] J. J. Murillo-Fuentes, R. Boloix-Tortosa, S. Hornillo-Mellado, and V. Zarzoso, "Independent component analysis based on marginal entropy approximations," in *Proc. ISIAC'04 5th Int. Symp. Intell. Autom. Contr.*, Seville, Spain, Jun. 28–Jul. 1, 2004.

[37] A. Stuart and J. K. Ord, *Kendall's Advanced Theory of Statistics*, 6th ed. London, U.K.: Edward Arnold, 1994, vol. I.
[38] P. McCullagh, *Tensor Methods in Statistics*, ser. Monographs on Statistics and Applied Probability. London, U.K.: Chapman & Hall, 1987.
[39] F. J. Theis, "A new concept for separability problems in blind source separation," *Neural Comput.*, vol. 16, pp. 1827–1850, 2004.
[40] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Process.*, vol. 50, pp. 1545–1553, Jul. 2002.
[41] J. F. Cardoso, "High-order contrasts for independent component analysis," *Neural Comput.*, vol. 11, no. 1, pp. 157–192, Jan 1999.
[42] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. Berlin, Germany: Springer-Verlag, 1990.
[43] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, 1999.
[44] X. Giannakopoulos, J. Karhunen, and E. Oja, "An experimental comparison of neural ICA algorithms," in *Proc. ICANN'98*, Skovde, Sweden, Sep 1998, pp. 651–656.
[45] J. Eriksson and V. Koivunen, "Characteristics-function based independent component analysis," *Signal Process.*, vol. 83, pp. 2195–2208, 2003.

Vicente Zarzoso (S'94–M'03) was born in Valencia, Spain, in 1973. He graduated (with highest distinction) in telecommunications engineering from the Universidad Politécnica de Valencia in 1996. The beginning of his Ph.D. studies were partly funded by a scholarship from the University of Strathclyde, Glasgow, U.K., and the Defence Evaluation and Research Agency (DERA) of the United Kingdom. He received the Ph.D. degree from the University of Liverpool, U.K., in 1999.

He spent five years with the University of Liverpool under a Research Fellowship from the Royal Academy of Engineering, U.K. Since September 2005, he has been a Lecturer with the Université de Nice - Sophia Antipolis and a Researcher with the Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis, France. His research interests include blind statistical signal and array processing and its application to biomedical problems and communications.

Juan José Murillo-Fuentes (M'99) was born in Sevilla, Spain, in 1973. He received the telecommunications engineering degree from the Universidad de Sevilla in 1996 and the Ph.D. degree in telecommunication engineering in 2001 from the Universidad Carlos III de Madrid, Spain.

He is currently an Associate Professor in the Department of Signal Theory and Communication, Universidad de Sevilla. His research interests lie in algorithm development for blind source separation and other signal-processing tools and their application to digital communications and image processing.

Rafael Boloix-Tortosa received the M.Eng. degree in telecommunications engineering and the Ph.D. degree from the Universidad de Sevilla, Spain, in 2000 and 2005, respectively.

He joined the School of Engineering there in 1999 as Research Assistant with the Department of Electronic Engineering. Currently, he is an Assistant Professor with the Department of Signal Theory and Communications. His research interests include blind source separation and higher order statistics and their application to digital communications.

Asoke K. Nandi (SM'96) received the Ph.D. degree from Trinity College, University of Cambridge, Cambridge, U.K., in 1979.

He held research positions with Rutherford Appleton Laboratory, U.K.; the European Organisation for Nuclear Research, Switzerland; the Department of Physics, Queen Mary College, London, U.K.; and the Department of Nuclear Physics, Oxford, U.K. In 1987, he joined Imperial College London as the Solartron Lecturer in the Signal Processing Section of the Electrical Engineering Department. In 1991, he jointed the Signal Processing Division of the Electronic and Electrical Engineering Department, University of Strathclyde, Glasgow, U.K., as a Senior Lecturer; subsequently, he became a Reader in 1995 and a Professor in 1998. In 1999, he joined the University of Liverpool, Liverpool, U.K., as the David Jardine Chair of Signal Processing in the Department of Electrical Engineering and Electronics. In 1983, he was a member of the UA1 team at CERN that discovered the three fundamental particles known as W^{+}, W^{-} and Z°, providing the evidence for the unification of the electromagnetic and weak forces, which was recognized by the Nobel Committee for Physics in 1984. Currently, he is Head of the Signal Processing and Communications Research Group, with interests in the areas of nonlinear systems, non-Gaussian signal processing, and communications research. With his group he has been carrying out research in blind source separation, blind deconvolution, machine condition monitoring, signal modelling, system identification, communication signal processing, time-delay estimation, biomedical signals, underwater sonar, and applications and development of machine learning. He has authored or coauthored more than 130 journal papers and more than 300 technical publications, including *Automatic Modulation Recognition of Communications Signals* (Boston, MA: Kluwer Academic, 1996) and *Blind Estimation Using Higher-Order Statistics* (Boston, MA: Kluwer Academic, 1999).

Prof. Nandi is a Fellow of the Cambridge Philosophical Society, the Institution of Engineering and Technology, U.K., the Institute of Mathematics and its Applications, the Institute of Physics, and the Royal Society of Arts. He received the Mounbatten Premium Division Award from the Electronics and Communications Division, the Institution of Electrical Engineers, in 1998 and the Water Arbitration Prize from the Institution of Mechanical Engineers, U.K., in 1999.