

Adaptive Blind Source Separation for Virtually Any Source Probability Density Function

Vicente Zarzoso, *Student Member, IEEE*, and Asoke K. Nandi, *Senior Member, IEEE*

Abstract—Blind source separation (BSS) aims to recover a set of statistically independent source signals from a set of linear mixtures of the same sources. In the noiseless real-mixture two-source two-sensor scenario, once the observations are whitened (decorrelated and normalized), only a Givens rotation matrix remains to be identified in order to achieve the source separation. In this paper, an adaptive estimator of the angle that characterizes such a rotation is derived. It is shown to converge to a stable valid separation solution with the only condition that the sum of source kurtosis be distinct from zero. An asymptotic performance analysis is carried out, resulting in a closed-form expression for the asymptotic probability density function of the proposed estimator. It is shown how the estimator can be incorporated into a complete adaptive source separation system by combining it with an adaptive prewhitening strategy and how it can be useful in a general BSS scenario of more than two signals by means of a pairwise approach. A variety of simulations assess the accuracy of the asymptotic results, display the properties of the estimator (such as its robust fast convergence), and compare this on-line BSS implementation with other adaptive BSS procedures.

Index Terms—Adaptive algorithms, blind source separation, closed-form estimation, convergence and performance analysis, higher order statistics.

I. INTRODUCTION

SINCE THE LATE 1980's, the problem of *blind source separation* (BSS) has attracted a great deal of attention from the signal processing community. This interest stems from the great number of applications that accept a BSS model and, hence, can be tackled by means of techniques for BSS. The problem arises when a set of unobservable signals (the *sources*) are to be extracted from a set of *sensor outputs* or observed signals, each of which can be regarded as a linear mixture of the sources. The adjective “blind” stresses the fact that very little is known or assumed about the mixing structure. This is precisely the origin of the power and versatility of the BSS model since in many cases, it is extremely difficult to model the transfer functions between sources and sensors, or simply, no *a priori* information is available about the mixture. Applications comprise a variety of different areas such as radar and sonar, communications, speech processing, seismic prospecting, and medicine [3], [5]–[7], [19].

Manuscript received July 31, 1998; revised August 23, 1999. This work was supported by The University of Liverpool. The associate editor coordinating the review of this paper and approving it for publication was Prof. Jonathon A. Chambers.

The authors are with the Signal Processing and Communications Group, Department of Electrical Engineering and Electronics, The University of Liverpool, Liverpool, U.K. (e-mail: vicente@liv.ac.uk; aknandi@liv.ac.uk).

Publisher Item Identifier S 1053-587X(00)01011-4.

In a number of cases, it is required to extract the source signals in real time. For instance, in the biomedical problem of fetal electrocardiogram extraction [19], such adaptive methods would allow on-line monitoring and condition diagnosis of the fetus's heart. In other situations, the parameters of the mixing system or the source signals are not stationary but are subject to variations with time. A typical example is a multiuser mobile communication system [3], where those variations in the mixing pattern occur as the users roam around, thus altering the channel characteristics. Even if the users remain still, the channel is still subject to time variations (fading) due to changing atmospheric conditions and environment. In all these instances, the performance of so-called batch processing or off-line methods would be very poor, and adaptive procedures are necessary for a reasonable separation performance.

Interestingly enough, the first adaptive BSS implementation was also the first successful attempt to tackle the BSS problem [11]. Jutten and Herault proposed a method based on a neural network architecture aiming at the cancellation of certain odd functions of the observations. Consequently, the system is stable and provides good separation solutions only for sources with different symmetric probability density functions (pdf's) [16]. In [5], a closed-form expression is derived to find the orthogonal transformation that remains to be unveiled once the observations have been prewhitened (decorrelated and normalized) in the basic BSS scenario composed of two sources and two sensors. The work of Nandi and Zarzoso [13] benefits from such a result to simplify the solution given in [12], supporting the convenience of a two-step approach to signal separation [18]: first, decorrelation; second, search for higher order independence. The estimator suggested in [5] depends exclusively on the fourth-order cross-cumulants of the whitened sensor outputs. Hence, its adaptive implementation reduces to a simple adaptive cumulant estimation, yielding the so-called *adaptive rotation* (AROT) algorithm. Departing from rather different points of view, Gaeta and Lacoume [7] and Comon [6] arrive separately at the same solution to the BSS problem: a cost function (so-called by the latter author “contrast function”) composed of the fourth-order marginal cumulants of the whitened observations, whose maximization leads to the source extraction. The adaptive algorithm based on this contrast function is developed in [9]. The major drawback of the implementations of [5] and [9] is that, by construction of the respective algorithms, it is difficult to carry out their convergence study.

A different approach is adopted in [3] by Cardoso and Laheld. Taking advantage of the implicit multiplicative structure

of the BSS problem and employing the “relative gradient,” they derive a family of adaptive algorithms with serial updating, i.e., the separating matrix is updated at each iteration by multiplying the one obtained at the previous iteration by another matrix. This class of algorithms, which are known as *equivariant adaptive separation via independence* (EASI) is shown to exhibit *uniform performance* characteristics, that is, the quality of the source estimation does not depend on the particular value of the mixing matrix. In contrast with the other two methods cited above, the performance analysis of this latter method is easily accomplished with the conventional tools for the study of stochastic algorithms. On the negative side, the stability of the algorithm is very dependent on the choice and tuning of certain nonlinear functions with respect to the source statistics, which in a genuine blind separation problem are not known *a priori*. In addition, its convergence to nonspurious solutions is not guaranteed.

The objective of this paper is to put forward a new method for adaptive BSS that avoids those two drawbacks, namely, the dependence of the convergence on the source statistics (shown, e.g., by the methods of [3] and [11]) and the difficulties in analyzing the asymptotic performance of the algorithm (as seen, e.g., in the methods of [5] and [9]). By contrast, the new procedure exhibits two very desirable properties: 1) convergence (virtually) independent of the source distribution and 2) straightforward performance analysis. The algorithm is based on the batch-processing method introduced in [20] and [21], from which the first property is inherited. From this fourth-order off-line procedure, its adaptive counterpart is derived, resulting in a very simple updating rule in the typical form of a stochastic adaptive algorithm that has been so widely studied [1]. The analysis of its convergence and asymptotic properties is presented herein as well. In principle, most of the derivations are developed in the simplified noiseless two-source two-sensor BSS scenario, but it is also shown how the procedure can be extended to more than two signals in the pairwise processing fashion already suggested in the literature (for instance, in [5] and [6]). Simulations endorse the effectiveness of this extension. Although some of the results developed next could easily be extended to the complex-signal case, we restrict our study to the treatment of real-valued signals and mixtures.

The paper is organized as follows. In the first place, Section II describes mathematically the problem of BSS and its basic foundations. Next, Section III reviews the off-line method on which the adaptive algorithm presented here is based. The most attractive feature of this batch method is that it is valid for practically any source distribution combination and does not depend on any choice of functions or nonlinearities to achieve the source extraction. The adaptive version is developed in Section IV. Its asymptotic performance and convergence characteristics are analyzed in depth in Section V, whereas its computational complexity is studied in Section VI. Section VII reports on some simulation experiments carried out in order to test the validity of the theoretical results, as well as to compare the proposed method with other adaptive implementations. Section VIII concludes the paper.

II. PROBLEM STATEMENT

Given the set of q source signals $\mathbf{x}_k = [x_1(k), \dots, x_q(k)]^t \in \mathbb{R}^q$, where k represents a discrete time index and $(\cdot)^t$ the vector transpose operator, the purpose of BSS is to extract them from the set of p ($p \geq q$) instantaneous linear mixtures observed at the sensor output, $\mathbf{y}_k = [y_1(k), \dots, y_p(k)]^t \in \mathbb{R}^p$. If matrix $M = (m_{ij}) \in \mathbb{R}^{p \times q}$, which is also called *mixing* or *transfer matrix*, contains the mixture coefficients, then the BSS problem accepts the matrix model

$$\mathbf{y}_k = M\mathbf{x}_k, \quad k = 1, 2, \dots \quad (1)$$

The channel variability in nonstationary environments would be modeled by allowing the mixing matrix to be time-varying, $M_k = (m_{ij}(k))$. In addition, in general, the observations would be noisy, yielding an additive noise vector $\mathbf{n}_k \in \mathbb{R}^p$ on the right-hand side of model (1). However, in the sequel, only noise-free observations will be considered.

Two basic assumptions are made in order to achieve the source estimation and transfer matrix identification:

- A1) The sources are statistically independent.
- A2) The mixing matrix is full column rank.

The first hypothesis is the cornerstone of blind separation and, albeit it could sound too strong, it is actually fairly plausible in many real situations, given the physical independence of the underlying phenomena generating the source signals.

Most approaches to BSS operate in two steps [2], [5], [6], [10], [20], [21]. First, the observations are prewhitened by estimating a *whitening matrix*, B_k^\dagger (symbol \dagger denoting the Moore-Penrose pseudoinverse [8]), which results in a set of uncorrelated and normalized (unit-power) signals $\mathbf{z}_k = B_k^\dagger \mathbf{y}_k$ termed *whitened observations*. The operations necessary to achieve off-line prewhitening correspond to well-known second-order techniques (PCA) and, therefore, will not be discussed here. For on-line prewhitening, in [3] and [5], for instance, two different methods are proposed. The former method is based on the adaptive minimization of the Kullback–Leibler divergence between the whitened output and a random vector with identity covariance matrix when their respective distributions are truncated at the second order, i.e., when a Gaussian approximation is used. By employing the relative gradient of this contrast, we arrive at the following serial updating rule for the whitening matrix:

$$B_{k+1}^\dagger = [I - \lambda_k (\mathbf{z}_k \mathbf{z}_k^t - I)] B_k^\dagger. \quad (2)$$

The point of interest is that once the observations have been prewhitened, only an orthogonal transformation Q_k remains to be estimated in order to identify the mixture [5], [6], [21] and, thus, the sources as $\hat{\mathbf{x}}_k = \mathbf{s}_k = Q_k^t \mathbf{z}$. If the time structure is ignored or cannot be exploited, as in the i.i.d. case, the estimation of this matrix requires the higher order statistical information of the data. Several different methods to estimate Q in a block fashion are reported in [5]–[7], [10], and [21]. In this paper, we will focus on the adaptive estimation of Q . The obtained results will then be combined with existing prewhitening

strategies, such as that in (2), in order to yield a complete separation system. Fig. 1 summarizes graphically the typical setup for adaptive BSS, as well as the terminology and all the matrix and vector relationships referred to above.

III. A BATCH-PROCESSING METHOD

This section briefly reviews, for the sake of completeness, a method proposed in [20] and [21] for the estimation of the orthogonal matrix Q from a batch of whitened observation samples. The main results are summarized in Section III-A for the noiseless BSS scenario composed of two sources and two sensors. A generalization to more than two signals can also be carried out as in Section III-B.

A. The EML: A Fourth-Order Angle Estimator

In the two-signal case, unknown matrix Q becomes an elementary Givens rotation matrix, which can be parameterized as

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

Hence, the estimation of Q reduces to the estimation of angular parameter θ . The following closed-form estimator of angle θ is developed and studied in [20] and [21]:

$$\hat{\theta}_{\text{EML}} = \frac{1}{4} \text{angle}(\xi \cdot \text{sign}(\gamma)) \quad (4)$$

with

$$\xi \triangleq E[(z_1 + jz_2)^4] \quad (5)$$

$$\gamma \triangleq E[(z_1^2 + z_2^2)^2] - 8 \quad (6)$$

where $\mathbf{z} = [z_1, z_2]^t$ denotes the whitened signals, and $j = \sqrt{-1}$ is the imaginary unit. The angle or argument function “angle(\cdot)” provides the phase in $[-\pi, \pi]$ of its complex argument. The term $(z_1 + jz_2)$ can be interpreted, at each time instant, as the complex form of the corresponding whitened-output bidimensional scatter-plot point. Consequently, the sample estimate of (5) can be considered as a centroid in the complex plane calculated as the average value of such points raised to the fourth power. Remark that ξ is a complex-valued quantity, although we are dealing with real mixtures. The term (6) is an estimate of the source kurtosis sum (sks) ($\kappa_{40}^x + \kappa_{04}^x$).

The simple algebraic formalism in terms of complex-valued centroids simplifies the characterization of the statistical properties of the estimator [21]. As we will see in Section V, analogous analysis simplifications will be obtained for its adaptive counterpart. Batch estimator (4) is shown to provide a valid separation solution as long as the sks is not null $\kappa_{40}^x + \kappa_{04}^x \neq 0$. Under this basic condition, it is found to be unbiased for any sample size when there is a symmetrically distributed source and asymptotically unbiased and strongly consistent in the i.i.d. case. An analytic expression of its asymptotic pdf is derived and an interesting geometrical interpretation provided. Remarkable connections with the approximate ML (AML) estimator of [10] are also made. It is shown to generalize such an estimator, which was originally derived under quite restrictive conditions, to virtually any source distribution combination, hence, the acronym

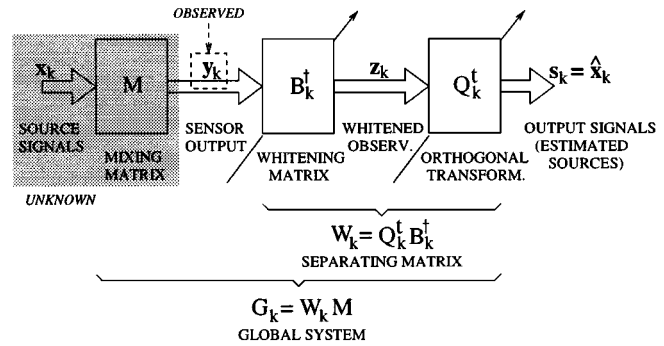


Fig. 1. Graphical depiction of the adaptive BSS, summarizing the nomenclature and notation employed throughout the paper.

Step 1. Obtain whitened signals from given observations (PCA). Consider the whitening-matrix (pseudo)inverse and the whitened signals as a first estimate of the mixing matrix and the sources, respectively: $B \rightarrow \hat{M}$, $\mathbf{z} \rightarrow \hat{\mathbf{x}}$.

Step 2. For all signal pairs $\hat{\mathbf{x}}_{ij} = [\hat{x}_i, \hat{x}_j]^t$:

2.1. Estimate (2×2) -unitary transformation \hat{Q}_{ij} from EML estimator (4) (with $[z_1, z_2]^t = [\hat{x}_i, \hat{x}_j]^t$).

2.2. Counter-rotate: update $\hat{\mathbf{x}}_{ij} := \hat{Q}_{ij}^t \hat{\mathbf{x}}_{ij}$, $\hat{M}_{:[i,j]} := \hat{M}_{:[i,j]} \hat{Q}_{ij}$, where $\hat{M}_{:[i,j]}$ denotes the matrix composed of only the i th and j th columns of \hat{M} .

Step 3. Repeat from Step 2 until convergence.

Fig. 2. Batch algorithm for BSS of more than two signals via the EML estimator.

extended ML (EML). In fact, all the previous AML conditions are reduced to the much weaker single restriction on the sks. For a full account on this batch estimator, see [21]. The linkage between the EML method and the procedure of [5], together with a number of other interesting results on closed-form estimators for BSS, are investigated in [22]. Finally, by combining the results of [2] and [21], the uniform performance of the EML estimator can be easily proven.

B. Extension to General BSS Scenario

The estimator presented in Section III-A is designed to operate on two-sensor mixtures. Its extension to the general BSS scenario composed of more than two mixtures of more than two sources can be done by applying expression (4) in turn to each pair of decorrelated measurements until convergence. This heuristic notion was originally suggested in [5] and found mathematical justification in [6] when contrast functions of polynomial form are utilized. Even though the convergence of such iterative scheme still lacks a solid mathematical proof, numerous experiments endorse the empirical validity of this generalization [21]. Fig. 2 summarizes this pairwise algorithm for more than two signals. The number of sweeps over all signal pairs necessary for convergence was empirically found to be roughly the same as the value proposed in [6], that is, on the order of $1 + \sqrt{q}$.

IV. ADAPTIVE SOURCE SEPARATION

A. Adaptive Algorithm in the Two-Signal Case

The derivation of an adaptive algorithm from the batch version depicted in Section III-A reduces to the adaptive estimation of the centroid location (5) and the sks (6). It is this simplicity that will lead to a straightforward analysis of the asymptotic properties of the resulting algorithm.

Let ν_k denote the fourth power of the prewhitened observation scatter-diagram point at time instant k

$$\nu_k = (z_1(k) + jz_2(k))^4. \quad (7)$$

Equations (5) and (6) can then be written, respectively, as $\xi = E[\nu]$ and $\gamma = E[|\nu| - 8]$. Therefore, ξ and γ may be obtained in an adaptive fashion from

$$\xi_k = (1 - \mu_k)\xi_{k-1} + \mu_k\nu_k \quad (8)$$

$$\gamma_k = (1 - \mu_k)\gamma_{k-1} + \mu_k(|\nu_k| - 8) \quad (9)$$

where μ_k is the adaption coefficient at iteration k . This coefficient balances a tradeoff between convergence speed and accuracy, as it is well established from the standard theory of adaptive systems [17]. If $\mu_k = 1/k$, then (8) and (9) become the adaptive equivalents of the sample estimates (time averages) of (5) and (6), respectively. The rotation angle θ is estimated at each iteration from the above two parameters by replacing (8) and (9) into (4)

$$\hat{\theta}_k = \frac{1}{4} \text{angle}(\xi_k \cdot \text{sign}(\gamma_k)). \quad (10)$$

In the following, a constant adaption coefficient will be assumed: $\mu_k = \mu, \forall k$. This choice, in turn, allows the tracking of nonstationarities in the mixing system, which is one of the main rationales behind the use of adaptive separation procedures. In that case, (8) may be written in the form $\xi_k = \mu \sum_{n=1}^k (1 - \mu)^{k-n} \nu_n$. Taking expectations at both sides of the previous equation and assuming i.i.d. mixtures, $E[\xi_k] = [1 - (1 - \mu)^k]E[\nu] = [1 - (1 - \mu)^k]\xi \rightarrow_{k \rightarrow \infty} \xi$ for $0 < \mu < 2$. That is, adaptive estimator (8) of centroid (5) is *asymptotically unbiased*. On the other hand, assuming i.i.d. source processes, the variance of ξ_k becomes $\text{Var}[\xi_k] = (\mu/[2 - \mu])[1 - (1 - \mu)^{2k}] \text{Var}[\nu] \rightarrow_{k \rightarrow \infty} (\mu/[2 - \mu]) \text{Var}[\nu]$ so that it can be made arbitrarily small by choosing μ sufficiently small. Totally analogous conclusions hold for adaptive estimator (9) of the sks. These preliminary asymptotic results lead to $E[\hat{\theta}_k] \rightarrow_{k \rightarrow \infty} \theta$ if the sks is not null since in that case, $\hat{\theta}_{\text{EML}}$ given by (4) is also asymptotically unbiased [21].

Estimator (10) is referred to as *adEML* for *adaptive EML*. A more exhaustive analysis of its convergence and asymptotic properties is carried in Section V. Next, we address the extension of this adaptive scheme to BSS scenarios of more than two signals.

B. Adaptive Algorithm for the General BSS Scenario

The extension of adaptive scheme (10) to the broader BSS scenario in which more than two whitened observations appear is very similar to that for its batch counterpart described in Section III-B. Now, at each time instant, a whitened-signal sample is

At each sample instant k :

Step 1. Obtain whitened-output sample from current observation sample via a suitable adaptive pre-whitening method. Consider the resulting whitening-matrix (pseudo)inverse and the whitened sample as a first estimate of the mixing matrix and the source sample, respectively: $B_k \rightarrow \widehat{M}_k, \mathbf{z}_k \rightarrow \widehat{\mathbf{x}}_k$.

Step 2. For each sweep number c repeat Step 3:

Step 3. For all signal pairs $\widehat{\mathbf{x}}_{ij} = [\widehat{x}_i, \widehat{x}_j]^t$:

3.1. Update centroid $\xi_{ij}^{(c)}(k)$ and sks $\gamma_{ij}^{(c)}(k)$ using eqns. (8) and (9) (with $[z_1, z_2]^t = [\widehat{x}_i, \widehat{x}_j]^t$).

3.2. Estimate (2×2) -unitary transformation $\widehat{Q}_{ij}^{(c)}$ from adEML estimator (10).

3.3. Counter-rotate: update

$$\widehat{\mathbf{x}}_{ij} := (\widehat{Q}_{ij}^{(c)})^t \widehat{\mathbf{x}}_{ij},$$

$\widehat{M}_{k, [i, j]} := \widehat{M}_{k, [i, j]} \widehat{Q}_{ij}^{(c)}$, where $\widehat{M}_{k, [i, j]}$ denotes the matrix composed of only the i th and j th columns of \widehat{M}_k .

Fig. 3. Adaptive algorithm for BSS of more than two signals via the adEML estimator.

first obtained through a suitable adaptive prewhitening method. Then, the resulting vector sample is processed in sweeps over all its component pairs. For each sample pair, the associated centroid and sks estimates are updated, as in (8) and (9), and the resulting angle estimate (10) is obtained. From it, a counter-rotation aiming at separation is performed on the sample pair. The process is repeated in turn over all the other pairs and over several sweeps. It is important to remark that the values of the centroid and sks associated with a given signal pair are independently stored for each sweep. If the signals are stationary, the angle values obtained by this adaptive procedure will converge to those of its batch counterpart. This adaptive algorithm is outlined in Fig. 3. This pairwise extension is very similar to that suggested in [5] for the AROT method.

V. ASYMPTOTIC ANALYSIS

The most appealing feature about the adaption scheme (8), (9) is that it accepts the classic adaptive algorithm form [1]

$$\xi_k = \xi_{k-1} + \mu_k H(\xi_{k-1}, \nu_k) \quad (11)$$

where, from (8), function $H(\cdot, \cdot)$ can be identified as

$$H(\xi, \nu) = \nu - \xi \quad (12)$$

and, accordingly, for (9). As a result, the standard tools for the study of stochastic algorithms of this general form can be employed to analyze the convergence properties of the one suggested herein. Our main goal is to obtain the asymptotic pdf of the adEML estimator (10).

A. The ODE Method

In the first place, the ordinary differential equation (ODE) method is employed to study the trajectories of the centroid locations (11). This method approximates the true trajectories of the parameters of interest by the discretized solution of certain

ODE. It is shown [1] that under quite general conditions and in the limit of an arbitrarily small adaption coefficient, the true trajectories followed by ξ_k converge to the continuous function of time $\xi(t)$ discretized at time instants $t_k = k\mu$, $\xi(t)$ being given by the solution of the ODE $\dot{\xi}(t) = h(\xi)$. The term $h(\cdot)$, which is called *mean field*, is obtained from the function $H(\cdot, \cdot)$ as

$$h(\xi) \triangleq E[H(\xi, \nu_k)] = \xi_* - \xi \quad (13)$$

where $\xi_* = E[\nu_k]$ is the population centroid location (5), and symbol $E[\cdot]$ denotes the expectation with respect to the distribution of the state ν_k for a fixed value of ξ when ν_k is asymptotically stationary. The conditions for the validity of this approximation are, essentially, that the adaptation coefficient μ be very small and the function $\xi \mapsto H(\xi, \nu)$ be fairly regular. Having a look at (12), it is clear that the last condition holds. In our case, it is straightforward to obtain

$$\xi(t) = \xi_* + (\xi_0 - \xi_*) e^{-t} \quad (14)$$

where ξ_0 is the initial value, i.e., $\xi_0 = \xi(0)$. For the source kurtosis trajectories, the solution of its ODE is also readily found as

$$\gamma(t) = \gamma_* + (\gamma_0 - \gamma_*) e^{-t} \quad (15)$$

where, again, $\gamma_0 = \gamma(0)$ is the initial value, and $\gamma_* = E[|\nu_k| - 8]$ represents the population sks (6).

B. Stability of the Equilibrium Points

The *equilibrium points* or *attractors* of (11) are those values of ξ such that the mean field $h(\cdot)$ vanishes. Therefore, $\xi = \xi_*$ is the unique attractor of (8) [or (11)], as seen from (13) and taking $\lim_{t \rightarrow \infty} \xi(t)$ in (14). An attractor is said to be *locally asymptotically stable* (LAS) if and only if [1]

$$h_\xi(\xi_*) \triangleq \left. \frac{\partial h}{\partial \xi} \right|_{\xi=\xi_*} \quad (16)$$

has a negative real part. In the case of interest, and according to (13), $h_\xi(\xi_*) = -1 < 0$. Hence, the equilibrium point ξ_* is LAS always, regardless of the source distribution and the mixing structure. Analogously, $\gamma = \gamma_*$ is the unique LAS equilibrium point of its respective ODE. Therefore, the attractors of (10) are LAS as well. On the other hand, since ξ_k and γ_k converge to their population values, the adEML equilibrium points constitute valid separation solutions under the EML basic condition on the sks. In conclusion, under such a condition, the attractors of adaptive estimator (10) provide LAS valid separation solutions.

C. Asymptotic Normality of the Centroid Locations

Under the above conditions, it is proved in [1, Th. 2, ch. 3] that the asymptotic distribution of ξ_k is Gaussian. Specifically $\tilde{\xi}_k \triangleq \mu^{-1/2}(\xi_k - \xi(t_k)) \rightarrow_{\mu \rightarrow 0} N(0, P)$, where the covariance

$P = \text{Cov}[\tilde{\xi}_k]$ is given by the solution of the Lyapunov equation $2h_\xi(\xi_*)P + R(\xi_*) = 0$, with

$$R(\xi) \triangleq \sum_{k=-\infty}^{\infty} \text{Cov}[H(\xi, \nu_k), H(\xi, \nu_0)]. \quad (17)$$

In conclusion, the distribution of the centroid ξ_k is asymptotically normal $\xi_k \rightarrow_{\mu \rightarrow 0} N(\xi_*, \sigma_\xi^2)$ with mean ξ_* and variance

$$\sigma_\xi^2 = \mu P = \mu R(\xi_*)/2. \quad (18)$$

The pdf of ξ_k for k large enough and μ sufficiently small is then $p_\xi(\xi) = (1/2\pi\sigma_\xi^2) \exp(-\|\xi - \xi_*\|^2/(2\sigma_\xi^2))$, which, as a function of the real and imaginary part of $\xi = \omega_1 + j\omega_2$, becomes the product of two uncorrelated real Gaussian pdf's $p_\xi(\omega_1, \omega_2) = (1/2\pi\sigma_\xi^2) \exp(-[(\omega_1 - \omega_{1*})^2 - (\omega_2 - \omega_{2*})^2]/(2\sigma_\xi^2))$, with $\xi_* = \omega_{1*} + j\omega_{2*}$. Now, we are ready to develop an expression for the asymptotic pdf of the adaptive angle estimator (10).

D. Asymptotic Pdf of the adEML Estimator

First, if k is sufficiently large and μ sufficiently small, it can be reasonably assumed that the sign of the sks is accurately estimated so that it becomes a constant, and then, the “ $\text{sign}(\cdot)$ ” function in (10) may be neglected. In those conditions, and assuming, with no loss of generality, that such a sign is positive, we have

$$\delta \triangleq 4\hat{\theta} = \text{angle}(\xi). \quad (19)$$

This last expression naturally induces the change of variables $\omega_1 = r \cos \delta$ and $\omega_2 = r \sin \delta$. Integrating the joint pdf of (r, δ) with respect to r , the marginal pdf of δ , $p_\delta(\delta)$, is obtained. With the help of the symbolic mathematics package MAPLE™ [4] and after some tedious algebraic simplifications, p_δ turns out to be

$$\begin{aligned} p_\delta(\delta) &= \frac{1}{2\pi} \exp\left(\frac{-\kappa^+}{2\sigma_\xi^2}\right) + \frac{1}{\sigma_\xi \sqrt{8\pi}} \kappa^+ \cos(\delta - 4\theta) \\ &\times \exp\left(\frac{-\kappa^+ \sin^2(\delta - 4\theta)}{2\sigma_\xi^2}\right) \\ &\times \left\{ 1 + \text{erf}\left(\frac{1}{\sigma_\xi \sqrt{2}} \kappa^+ \cos(\delta - 4\theta)\right) \right\} \quad (20) \end{aligned}$$

where the symbol κ^+ denotes the absolute value of the sks

$$\kappa^+ \triangleq |\gamma_*| = |\kappa_{40}^x + \kappa_{04}^x|. \quad (21)$$

With κ^+ defined as above, it is easy to prove that pdf (20) is also applicable to the case of negative sks. Finally, from (19)

$$p_{\hat{\theta}}(\hat{\theta}) = 4 p_\delta(4\hat{\theta}) \quad (22)$$

which is the pdf of $\hat{\theta}_k$ given by (10) when $k \rightarrow \infty$ and $\mu \rightarrow 0$.

Pdf $p_{\hat{\theta}}$ is an even function of $(\hat{\theta} - \theta)$, and hence, $E[\hat{\theta}] = \theta$, that is, adaptive estimator (10) is *asymptotically unbiased*.

Gaussian Approximation: Similarly, as occurred for the batch EML method (see [22]), asymptotic pdf (20) admits a Gaussian approximation. If μ is small enough, the ratio σ_ξ/κ^+ becomes small too [since σ_ξ is proportional to $\sqrt{\mu}$; see (18)],

and hence, $\exp(-\kappa^+/(2\sigma_\xi^2)) \approx 0$. In addition, the pdf is mostly concentrated around $\delta = 4\theta$ so that $\cos(\delta - 4\theta) \approx 1$, $\sin(\delta - 4\theta) \approx (\delta - 4\theta)$, and $\text{erf}(\cdot) \approx 1$. Consequently, pdf (20) reduces to

$$p_\delta(\delta) \underset{\mu \rightarrow 0}{\approx} \hat{p}_\delta(\delta) = \frac{1}{\sqrt{2\pi}(\sigma_\xi/\kappa^+)} \exp\left\{-\frac{(\delta - 4\theta)^2}{2(\sigma_\xi/\kappa^+)^2}\right\}. \quad (23)$$

That is

$$\delta \underset{\mu \rightarrow 0}{\overset{d}{\rightarrow}} N(4\theta, (\sigma_\xi/\kappa^+)^2) \quad (24)$$

in which $\overset{d}{\rightarrow}$ indicates convergence in distribution [14]. From this key result and (19), the asymptotic variance of the adaptive EML is readily obtained as

$$\text{Var}[\hat{\theta}_k] \underset{\mu \rightarrow 0}{\overset{k \rightarrow \infty}{\rightarrow}} \sigma_{\text{adEML}}^2 = \frac{\sigma_\xi^2}{16\kappa^+} \quad (25)$$

Fig. 4 checks the quality of this approximation. The standard deviation of $\hat{\theta}$ obtained by numerical integration of pdf (20) is compared with the one obtained from the Gaussian approximation (23) [the square root of (25)]. The difference between both is within 1% for $\sigma_\xi/\kappa^+ < 0.138$. Note also that from (20) and (22), when the ratio σ_ξ/κ^+ becomes large (which occurs when either the adaption coefficient is large too or the sks is close to zero), the distribution of $\hat{\theta}_k$ tends to a uniform pdf $U(-(\pi/4), (\pi/4))$ with standard deviation 25.98° . This agrees with the behavior observed in the solid line of Fig. 4.

E. Remarks

In order to bring this analysis to an end, there are a couple of remarks worth pointing out. In the first place, due to the relationship between the source and whitened-output scatter plots [20], [21], it can be proved that R [see (17)] does not depend on the unknown parameter of interest but only on the source signals. In fact, assuming i.i.d. processes, R is easily shown to reduce to

$$R(\xi_*) = \text{Var}[\nu] = \text{Var}[(z_1 + jz_2)^4] = \text{Var}[(x_1 + jx_2)^4]. \quad (26)$$

Observe also that in the case of i.i.d. sources, the variance σ_ξ^2 [see (18)] obtained from this value of R coincides with the result for this kind of sources initially given at the end of Section IV when $\mu \rightarrow 0$. Consequently, the variance [see (25)] and, hence, the performance of the adEML estimator does not depend on the mixture but only on the source signals and, more specifically, on their statistics up to the eighth order. The uniform performance property of batch estimator (4) is therefore inherited by its adaptive counterpart (10).

Finally, it is also interesting to realize the striking resemblance between the asymptotic pdf of the adaptive estimator (10) and the large-sample asymptotic pdf found in [21] for its block-processing version (4).

VI. COMPUTATIONAL COMPLEXITY

We now measure the computational burden of the adaptive scheme presented in Section IV. Since additions are negligible relative to products, we consider a floating point operation (flop)

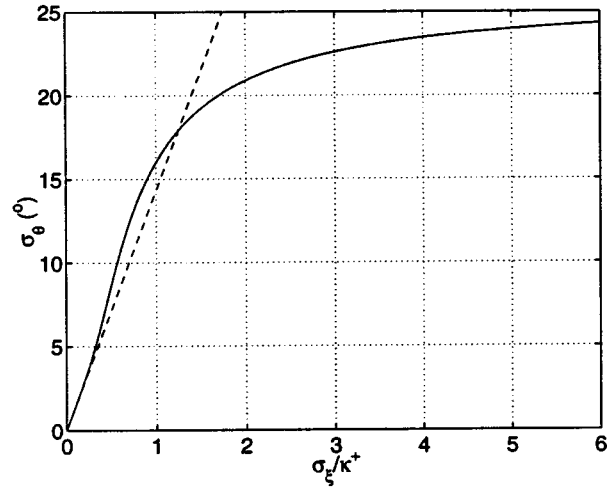


Fig. 4. Asymptotic standard deviation of the adEML estimator versus parameter σ_ξ/κ^+ . Solid line: From analytical pdf (20). Dashed line: From Gaussian approximation (23).

as a real multiplication. A complex product, hence, takes four flops and the multiplication of a real by a complex quantity (real-complex product) two flops. At each time instant (iteration) k and for each signal pair $[\hat{x}_i, \hat{x}_j]^t$, the source-extraction algorithm described in Fig. 3 can be broken down in the following elementary steps:

- 1) Computation of ν_k (7): two complex products = 8 flops. The computation of $|\nu_k|$ takes no additional flops since it can be evaluated as $|\nu_k^{1/2}|^2$ by using interim results of the computation of ν_k from $\nu_k^{1/2}$.
- 2) Centroid updating (8): two real-complex products = 4 flops. Sks updating (9): two real products = 2 flops.
- 3) Computation of $\hat{\theta}_k$ (10) and rotation matrix set-up (3): This would be platform dependent as it could be done through a look-up table or even with fast built-in functions. Say it takes f flops.
- 4) Rotation: four flops.

This makes $(18 + f)$ flops per signal pair per iteration. With N sweeps over the $q(q - 1)/2$ signal pairs, the overall complexity per iteration of the adEML procedure becomes $C_{\text{adEML}} = (18 + f)Nq(q - 1)/2$ flop/iteration. Usually, $N = O(\sqrt{q})$ (Section III-B), yielding $C_{\text{adEML}} = O(q^{5/2})$.

These figures can be compared with the values of other adaptive methods, such as AROT [5] and EASI [3]. The former procedure entails a higher complexity since it involves divisions¹ and square roots. However, with the equivalent centroid-based formalism proposed in [22] for the AROT estimator, we obtain $C_{\text{AROT}} = (14 + f)Nq(q - 1)/2$. AROT's computational cost is thus very similar to adEML's. By analogous reasoning, the complexity of EASI's orthogonal-matrix serial update plus source-sample extraction can be determined as $C_{\text{EASI}} = q^3 + 3q^2 + lq = O(q^3)$, where each nonlinearity element takes l flops (e.g., for cubic nonlinearities: $l = 2$). An extra overhead would have to be added in the normalized version of the algorithm. Typically, hence, EASI is more involved than the other two methods.

¹Divisions by powers of two are neglected since they can be efficiently accomplished by simple register shifting.

VII. EXPERIMENTAL RESULTS

Salient features of the theoretical results presented in the previous sections are to be illustrated through a variety of simulations. First, the theoretic asymptotic results of Section V are validated. Then, a comparison with other adaptive methods is established in several contexts. These include stationary parameters, nonstationary parameters, and abrupt changes of the mixing system, and, finally, the whole separation system comprising the prewhitening stage in scenarios of two and more than two signals.

A. Asymptotic Performance

ODE Solution: In the first place, it is seen how the ODE solution is indeed a good approximation of the adaptive-system actual evolution. From i.i.d. samples of two independent unit-power uniformly distributed sources and an arbitrary fixed value for the rotation angle θ , the solid line represented in Fig. 5 shows the trajectory of the sks adaptive estimate (9) for $\mu = 10^{-3}$. The dashed line corresponds to the ODE solution (15), discretized according to $t_k = \mu k$, k being the iteration (or sample) number. Both curves are remarkably similar. The estimate settles down, roughly, from iteration $k = 5 \cdot 10^3$, corresponding to $t_k = 5$, for which the exponential function in (15) has dropped below 1% of its initial value.

Validation of Asymptotic pdf: In the second place, the accuracy of the asymptotic pdf found in Section V [see (20) and (23)] is examined. We compare the empirical with the expected analytic results for the pdf $p_{\Delta\hat{\theta}}$ of the adEML error $\Delta\hat{\theta} \triangleq \hat{\theta} - \theta$, where $\hat{\theta}$ is given by (10) and $p_{\Delta\hat{\theta}}(\Delta\hat{\theta}) = p_{\hat{\theta}}(\Delta\hat{\theta} + \theta)$. Again, two independent uniformly distributed signals act as sources, with a fixed rotation angle of $\theta = 30^\circ$, yielding a hypothetical set of two whitened observations. From 10^5 independent samples of such processes, the value of $R(\xi_*)$ is evaluated by means of the sample estimate of (26) (it was verified that identical results are obtained when computed from the sources or the whitened signals), giving $R = 61.943$ (population value 62.537). The true value of sks gives $\kappa^+ = |-1.2 - 1.2| = 2.4$. With these parameters and $\mu = 10^{-3}$, the centroid standard deviation expected from the theoretical results [see (18)] is $\sigma_\xi = \sqrt{\mu R/2} = 0.176$, providing a ratio $\sigma_\xi/\kappa^+ = 0.073$. Now, adaptive estimator (10) is used on the first 10^4 samples of the whitened observations. In the previous paragraph, it was explained how the ODE solution stabilizes at $\mu k \approx 5$. This remark leads us to consider the values of $\hat{\theta}_k$ from $k = 5/\mu = 5 \cdot 10^3$, i.e., the last $5 \cdot 10^3$ iterations, in order to estimate the adEML-bias empirical pdf. To this end, the kernel pdf estimation method is employed [15] with function width $w = 0.909^\circ$. This empirical pdf is displayed in the solid-line curve of Fig. 6(a). The dotted line represents the analytical pdf obtained from (20) and the parameters κ^+ and σ_ξ above. The dashed line corresponds to Gaussian approximation (23) obtained from the population values. All three curves are very similar.

Next, the experiment is repeated with a new adaption coefficient $\mu = 10^{-4}$ and processing the whole of the initial 10^5 source samples. Now, $\sigma_\xi = 0.056$ and $\sigma_\xi/\kappa^+ = 0.023$. From the last $5 \cdot 10^4$ iterations, the pdf of $\hat{\theta}_k$ is estimated, again using

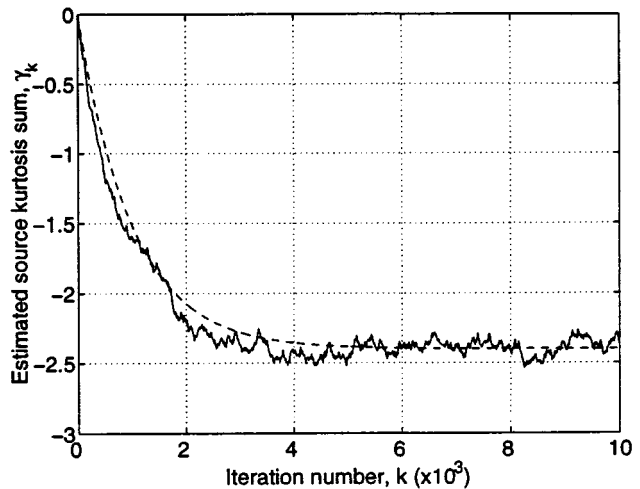


Fig. 5. Evolution of the source kurtosis sum adaptive estimate for a mixture of two uniformly distributed signals $\mu = 10^{-3}$. Solid line: Estimator (9). Dashed line: ODE approximation (15).

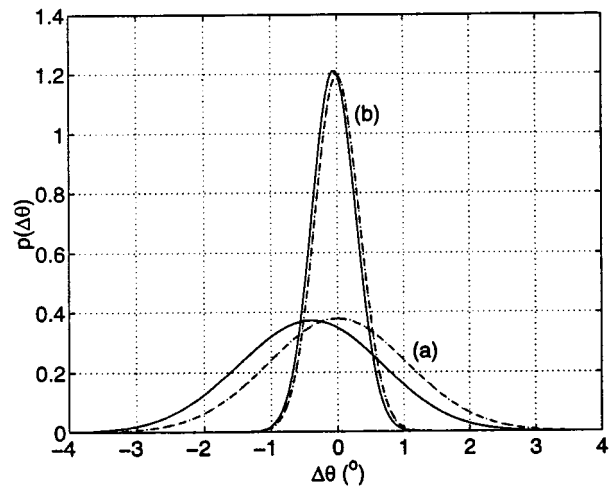


Fig. 6. adEML-error pdf for a mixture of two independent uniform distributions $\theta = 30^\circ$. Solid lines: Empirical pdf estimates, kernel method with function width w . Dotted lines: Analytic asymptotic pdf (20). Dashed lines: Gaussian approximation (23). (a) $\mu = 10^{-3}$, last $5 \cdot 10^3$ iterations, $w = 0.909^\circ$. (b) $\mu = 10^{-4}$, last $5 \cdot 10^4$ iterations, $w = 0.303^\circ$.

the kernel method, this time with $w = 0.303^\circ$. This estimate appears in the solid line of Fig. 6(b). Formulae (20) and (23) produce the analytic curves displayed by the dotted and dashed line, respectively, in the same figure. Now, the curves are even closer than before and, as expected, the variance is reduced relative to the previous case, roughly in the same proportion as σ_ξ^2 decreases. This corroborates the asymptotic-variance expression (25).

In order to demonstrate that all these asymptotic results also hold for long-tailed and asymmetric source distributions, analogous experiments are carried out with a Laplacian (long-tailed symmetric) and an exponential (long-tailed asymmetric) distribution for the sources. The results appear in Fig. 7 for $\theta = 30^\circ$, $R = 1.771 \cdot 10^4$ (calculated from a 10^5 -sample i.i.d. source

TABLE I
ESTIMATED-ANGLE BIAS \pm STANDARD
DEVIATION FOR $\theta = 30^\circ$ AND ADAPTION COEFFICIENTS $\lambda = \mu = 10^{-3}$,
EVALUATED FROM THE LAST $5 \cdot 10^4$ OUT OF $5.5 \cdot 10^4$ ITERATIONS.
ALL VALUES ARE EXPRESSED IN DEGREES.

METHOD \rightarrow SOURCE DISTRIBUTION \downarrow	adEML	EASI	AROT
Uniform-uniform	0.054 \pm 0.591	0.033 \pm 1.002	0.156 \pm 0.762
Uniform-Gaussian	-0.180 \pm 3.037	-0.100 \pm 3.433	-0.045 \pm 3.324
Laplacian-exponential	-0.138 \pm 1.575	No Convergence	-0.540 \pm 1.690
Exponential-Rayleigh	-0.267 \pm 2.722	No Convergence	-0.376 \pm 3.068

TABLE II
ESTIMATED-ANGLE BIAS \pm STANDARD DEVIATION FOR $\theta = 30^\circ$ AND
ADAPTION COEFFICIENTS $\lambda = \mu = 10^{-4}$, EVALUATED FROM THE LAST
 $5 \cdot 10^4$ OUT OF 10^5 ITERATIONS. ALL VALUES ARE EXPRESSED IN DEGREES.

METHOD \rightarrow SOURCE DISTRIBUTION \downarrow	adEML	EASI	AROT
Uniform-uniform	0.159 \pm 0.151	0.157 \pm 0.364	0.189 \pm 0.179
Uniform-Gaussian	-0.819 \pm 0.766	-0.880 \pm 0.909	-0.842 \pm 0.573
Laplacian-exponential	0.466 \pm 0.357	0.457 \pm 1.887	0.574 \pm 0.411
Exponential-Rayleigh	0.634 \pm 0.779	0.090 \pm 2.715	0.759 \pm 0.909

realization; population value $1.902 \cdot 10^4$, $\kappa^+ = 9$ (from the population sks), and the same general conditions as above. The two cases are a) $\mu = 10^{-3}$, $\sigma_\xi = 2.976$, $\sigma_\xi/\kappa^+ = 0.331$, and $w = 4.444^\circ$; and b) $\mu = 10^{-4}$, $\sigma_\xi = 0.941$, $\sigma_\xi/\kappa^+ = 0.105$, and $w = 1.379^\circ$. The resemblance between the curves is also outstanding.

Remark that in all the previous examples, the ratio $(\sigma_\xi/\kappa^+)/w$ is approximately constant so that the pdf estimation can be considered as reasonably "fair" in all cases.

B. Performance Comparison With Other Adaptive Algorithms

Stationary Parameters: The adEML estimator is to be compared with two other adaptive procedures: the EASI [3]² and the AROT [5] methods. The prewhitening stage in the latter procedure is removed so that only the angle estimation from the whitened observations is considered. In order to disclose the remaining orthogonal rotation, the AROT method requires to estimate the fourth-order cross-cumulants of the whitened sensor outputs, which can be done from their moments (just as suggested in [5]). To this end, the adaption coefficient λ is used to estimate the second-order moments, whereas the adaption coefficient μ is employed to estimate the fourth-order moments. For the part of the EASI method, when cubic nonlinearities are chosen, the convergence condition is such that the sum of kurtosis of every pair of sources are negative [3]. Accordingly, if the sks is positive, the choice of negative cubic nonlinearities, in principle, guarantees the convergence. Table I displays the results, in the form "bias \pm standard deviation," obtained for the angle estimated by those three methods for several source-pair distributions. The true angle is $\theta = 30^\circ$, and, inspired by EASI, equal values are chosen for the second and fourth-order adaption coefficients $\lambda = \mu = 10^{-3}$. The results were calculated

²Original MATLAB code was downloaded from <http://www-sig.enst.fr/~cardoso/guidesepsou.html>.

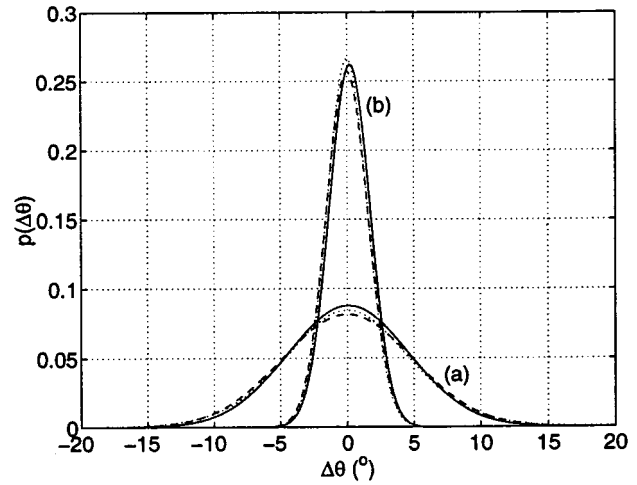


Fig. 7. adEML-error pdf for a mixture of a Laplacian and an exponential distribution $\theta = 30^\circ$. Solid lines: Empirical pdf estimates, kernel method with function width w . Dotted lines: Analytic asymptotic pdf (20). Dashed lines: Gaussian approximation (23). (a) $\mu = 10^{-3}$, last $5 \cdot 10^3$ iterations, $w = 4.444^\circ$. (b) $\mu = 10^{-4}$, last $5 \cdot 10^4$ iterations, $w = 1.379^\circ$.

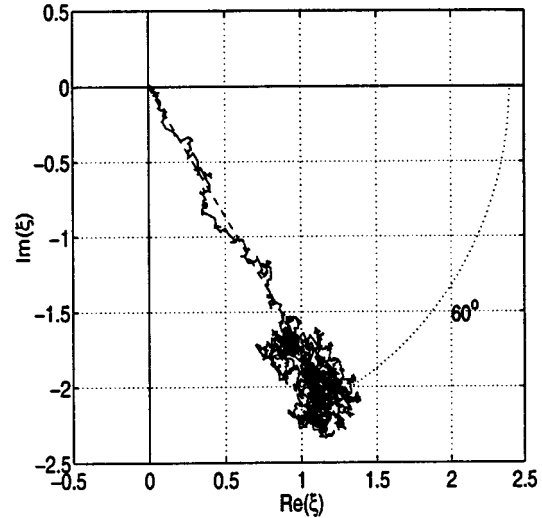


Fig. 8. Example of the adEML centroid trajectory for a mixture of two uniform distributions $\mu = 2 \cdot 10^{-3}$; first $5 \cdot 10^3$ iterations. Solid line: Adaptive estimator (8). Dashed line: ODE solution (14).

from the last $5 \cdot 10^4$ out of a total of $5.5 \cdot 10^4$ performed iterations. The first two rows correspond to source pairs with negative kurtosis sum (uniform-uniform: -2.4 , uniform-Gaussian: -1.2), whereas in the last two the source kurtosis sum is positive (Laplacian-exponential: 9 , exponential-Rayleigh: ≈ 6.25). The same mixture realizations are fed into all three methods. Note that EASI does not converge for the last two source pairs (which, probably, could have been prevented with the normalized version of the algorithm [3], although for the sake of a meaningful comparison, only non-normalized implementations of the methods were considered). For the standard deviation, it is found that $\sigma_{\hat{\theta}_{\text{adEML}}} < \sigma_{\hat{\theta}_{\text{AROT}}} < \sigma_{\hat{\theta}_{\text{EASI}}}$. Table II shows the outcome of a similar experiment but, this time, with $\lambda = \mu = 10^{-4}$ and taking the last $5 \cdot 10^4$ estimates from a total of 10^5 iterations.

Again, the above relationship among the respective standard deviations seems to hold.³

Regarding the convergence speed, it was observed that for the same adaption coefficient, adEML converges considerably faster than EASI and AROT. This is due to the fact that in order to estimate the missing angle, the magnitude of relevance is not the exact centroid location itself but only the centroid orientation. This orientation is estimated with great accuracy in just a few iterations when the centroid is initialized at the origin of the complex plane ($\xi_0 = 0$). This fast convergence feature is illustrated in Fig. 8, which plots the centroid trajectory for a mixture of two uniform sources, with a true angle of 30° (providing a centroid orientation of $\angle \xi = 4 \times 30^\circ + 180^\circ = 300^\circ$; see [21]) and step size $\mu = 2 \cdot 10^{-3}$. Further experiments (see Fig. 9) confirm that the convergence speed is practically independent of the particular choice of μ . In short, the adEML estimator exhibits a high convergence speed that is extremely robust with respect to the adaption coefficient.

Tracking Nonstationarities: A time-varying rotation angle is applied to two uniformly distributed sources. Specifically, the true angle used is mathematically described by

$$\theta_k(^{\circ}) = \begin{cases} -30 + 60 \sin(2\pi k/4000), & 0 < k \leq 1000 \\ -30 + 30(k - 1000)/1000, & 1000 < k \leq 2000 \\ 40 - 60(k - 2000)/1000, & 2000 < k \leq 3000 \\ -20, & k > 3000 \end{cases} \quad (27)$$

so that θ_k exhibits two abrupt changes: one at $k = 1000$ and another one at $k = 2000$. This function is represented by the steady solid line of Fig. 10. The oscillating solid line in that plot shows the angle estimated by the adEML from the set of resultant whitened observations for $\mu = 2 \cdot 10^{-2}$ averaged over 100 independent source realizations. The angle obtained by EASI (without the prewhitening part in its serial updating rule) with $\lambda = 5.5 \cdot 10^{-3}$ and averaged over the same 100 Monte Carlo (MC) runs is displayed by the dashed line of the same figure. The adaption coefficients of both methods were selected so that they produced the same estimated-angle variance in steady state in a bid to compare their convergence speed under the same accuracy conditions. It can be observed in Fig. 10 that adEML is faster than EASI in tracking the time-varying parameter and recovering from its abrupt oscillations. Nevertheless, the adEML high-speed convergence characteristic remarked in the previous paragraph could still be further exploited by means of a suitable abrupt-change detection scheme in order to yield a faster more efficient performance in the presence of such rapid variations. Essentially, the centroid should be taken to the origin of the complex plane as soon as the abrupt change is detected. Some abrupt-change detectors are studied in [1].

Performance of the Full Separation System: As explained in Section II, angle estimator (10) can be turned into a complete separation system just by combining it with an adaptive prewhitening stage. In the following simulations, the serial updating rule (2) is used to obtain adaptively the whitening matrix and generate the corresponding samples of the whitened

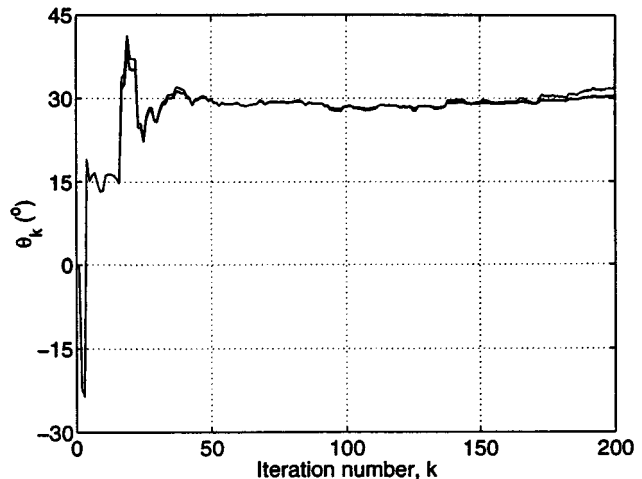


Fig. 9. adEML angle evolution for adaption coefficients $\mu = \{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}\}$ (superimposed) and a fixed mixture realization of uniformly distributed sources with true angle $\theta = 30^\circ$. Convergence is invariably reached in all cases at around the 50th sample.

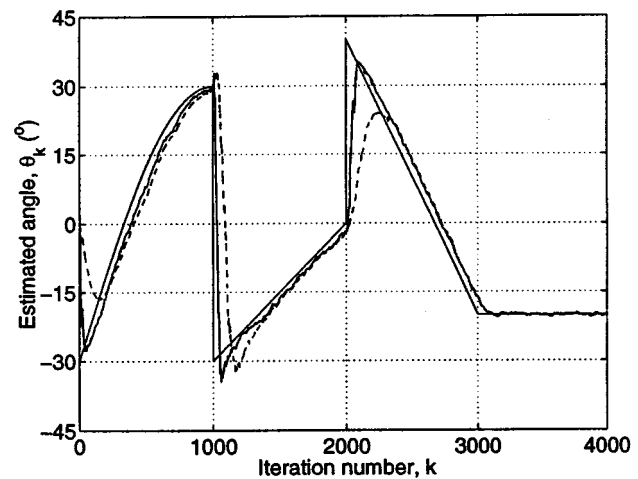


Fig. 10. Nonstationary rotation angle. Two uniform sources. Steady solid line: True angle θ . Oscillating solid line: Angle estimated by adEML $\mu = 2 \cdot 10^{-2}$. Dashed line: Angle estimated by EASI $\lambda = 5.5 \cdot 10^{-3}$. Results are averaged over 100 independent Monte Carlo iterations.

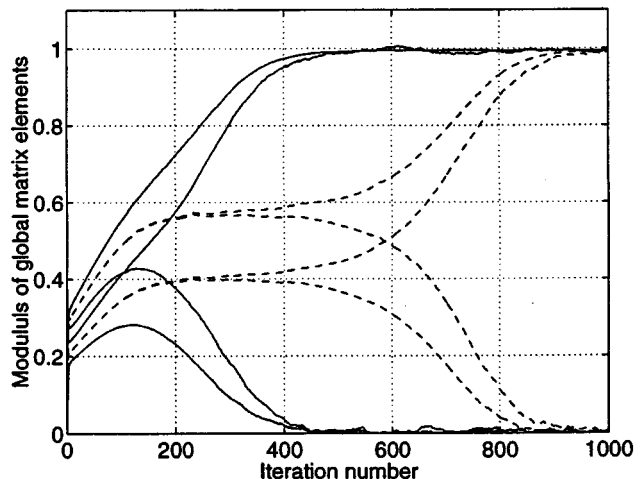


Fig. 11. Evolution of the global system averaged over 100 independent signal-realizations in the separation of a uniform and a binary source from two sensors by the adEML method $\lambda = \mu = 10^{-2}$ fixed mixing matrix with condition number n_M . Solid lines: $n_M = 10$. Dashed lines: $n_M = 10^3$.

³Bear in mind, however, that the performance of AROT's estimator deteriorates as the true parameter θ approaches zero [22].

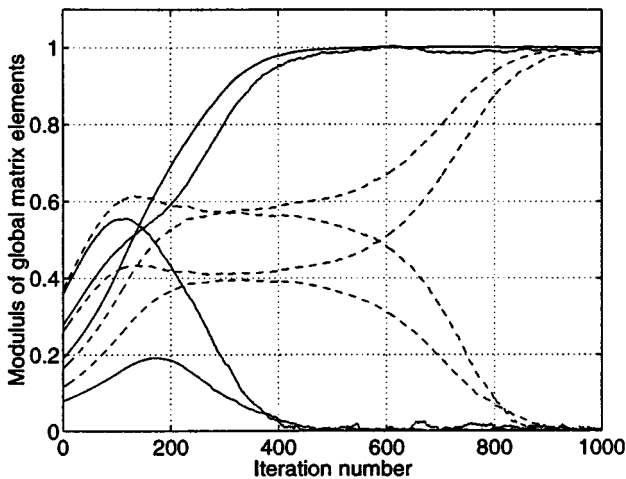


Fig. 12. Global system trajectories averaged over 100 independent signal-realizations in the separation of a uniform and a binary source from two observations by the EASI method, cubic nonlinearities, $\lambda = 10^{-2}$, fixed mixing matrix with condition number n_M . Solid lines: $n_M = 10$. Dashed lines: $n_M = 10^3$.

observations from which the adEML estimator (10) can then be employed in order to find the missing rotation. The same prewhitening strategy is used for AROT. The incorporation of prewhitening allows us, in turn, to pay attention to the impact of the mixing-matrix conditioning on the algorithms' convergence.

a) Two signals: A two-sensor mixture is made up of a uniformly distributed source, a binary source, and a fixed regular mixing matrix with condition number $n_M = 10^4$. We choose the step sizes $\lambda = \mu = 10^{-2}$. The trajectories of the modulus of the elements of the global mixing-unmixing system (which relates true and estimated sources; see Fig. 1) by the adEML averaged over 100 independent source-realizations are plotted in the solid lines of Fig. 11. Two of the elements converge to 1, whereas the other two do so to 0, thus achieving a successful source separation. The same mixture realizations are processed by EASI (this time including the prewhitening stage as well) with positive cubic nonlinearities and identical adaption coefficient. The selected nonlinearities fulfill the stability condition of this method for the mentioned kinds of sources. The averaged global matrix elements obtained by EASI are represented in Fig. 12 (solid lines). The results obtained by AROT under exactly the same conditions appear in Fig. 13 (solid lines). All three procedures converge nearly as fast, with minor differences in their oscillations once convergence is reached (at about the 500th iteration). The adEML and EASI results are quite similar, whereas AROT seems to exhibit more difficulties in the transient period until the moments are estimated with enough accuracy. The inclusion of the prewhitening stage accounts for the decrease in speed of the adEML method relative to the results of the preceding sections. A theoretical study of the influence of prewhitening on the method's performance is required to confirm this hypothesis. However, we lack space here for such thorough analysis.

On the other hand, the impact of the mixing-matrix conditioning is evidenced by the dashed lines of Figs. 11–13 obtained

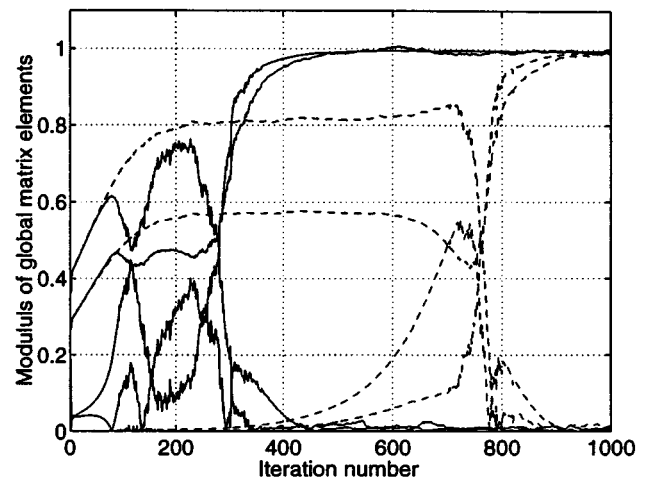


Fig. 13. Global matrix coefficients averaged over 100 independent signal-realizations in the separation of a uniform and a binary source from two measurements by the AROT method, $\lambda = \mu = 10^{-2}$, fixed mixing matrix with condition number n_M . Solid lines: $n_M = 10$. Dashed lines: $n_M = 10^3$.

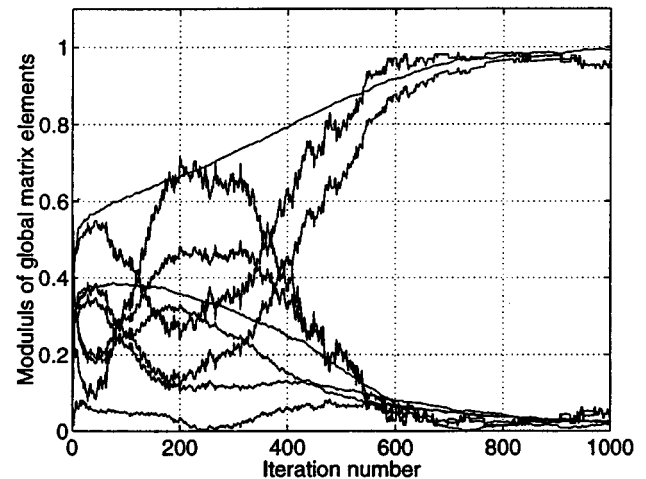


Fig. 14. Modulus of global system coefficients averaged over 100 mixture realizations. Three-source (binary-uniform-sinusoid) three-sensor scenario. adEML method, $\lambda = \mu = 4.5 \cdot 10^{-3}$, fixed mixing matrix with condition number $n_M = 10$.

for $n_M = 10^3$. Convergence is delayed as the mixing-matrix conditioning worsens.

b) More than two signals: In this scenario, the extension described in Section IV-B and Fig. 3 is used for the adEML method. On the part of the AROT method, only a single sweep is performed over the signal pairs, following the guidelines of [5]. A fixed regular 3×3 mixing matrix is chosen, with $n_M = 10$, which generates a set of three mixtures when applied on a set of three independent sources:

- binary sequence;
- uniformly distributed process;
- sinusoid with random frequency and initial phase.

The adaption coefficients are now selected as $\lambda = \mu = 4.5 \cdot 10^{-3}$ for all three methods. Figs. 14 and 15 display the trajectories of the global system entries obtained by adEML and EASI, respectively, averaged over 100 independent MC realizations. AROT's results are not shown since the procedure did not

⁴The condition number n_M is defined as the ratio of the largest to the smallest singular value of M .

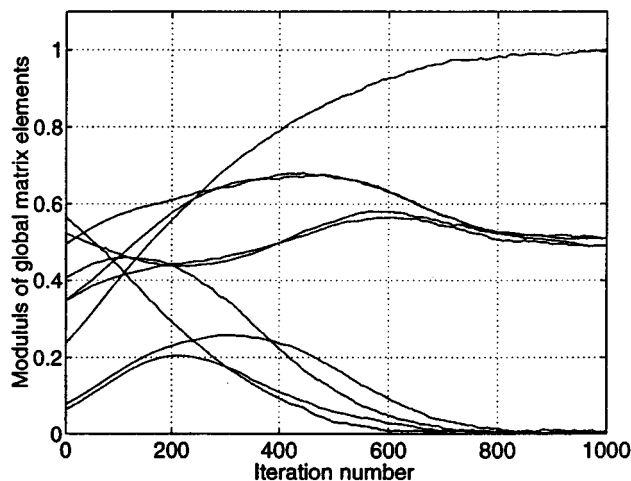


Fig. 15. Modulus of global system coefficients averaged over 100 mixture realizations. Three-source (binary-uniform-sinusoid) three-sensor scenario. EASI method, cubic nonlinearities, $\lambda = 4.5 \cdot 10^{-3}$, fixed mixing matrix with condition number $n_M = 10$.

converge on 20% of the MC runs (perhaps this negative outcome could have been prevented with some sort of normalization in the algorithm; this could be a subject of further investigations). Looking at Fig. 14, we can see that the proposed extension works fine, which is a conclusion supported by additional experiments. EASI's results are smoother than adEML's, but the former method consistently converged to a nonseparating solution, in which one of the sources was extracted but the other two remained mixed at the separator output. The influence of the mixing-matrix conditioning on the convergence was observed to be analogous to that in the two-signal case.

VIII. CONCLUSION

An adaptive estimator of the relevant separation parameter after prewhitening in the two-source two-sensor BSS scenario of real instantaneous linear mixtures has been derived. This adaptive estimator is implicitly based on the fourth-order statistics of the observed signals. Its convergence to a stable separation solution has been theoretically proven (in the i.i.d. case) for any source distribution combination, as long as the source kurtosis sum is not zero. This attribute is in sharp contrast with other adaptive methods (e.g., [3] and [11]), which are not guaranteed to converge to the desired solution. In addition, the simple algebraic formalism on which this adaptive method is based has allowed us to capitalize on the standard devices for the analysis of stochastic algorithms when studying the properties of the method. In particular, its asymptotic pdf has been determined, together with its Gaussian approximation and asymptotic variance, which depends exclusively on the source statistics. Simulations have shown the accuracy of the asymptotic results, as well as the fast convergence of the adaptive estimator. This high convergence speed has proven very robust to the adaption coefficient choice. By using a pairwise extension of computational cost comparable with other adaptive procedures, the estimator has also been made applicable in BSS contexts of more than two signals, which has been empirically supported. Experiments have also demonstrated the use of the estimator in a complete

separation scheme, revealing a performance up to the mark of other adaptive BSS methods.

Paths of further research include the extension of the procedure to complex mixtures and the study of alternative generalizations to the BSS scenario of more than two signals. The former issue begins to receive attention in [23]. The theoretical convergence of the pairwise scheme for the general BSS case remains an open question. The effects of noise, as well as of prewhitening, on the performance of the proposed adaptive estimator also deserve to be investigated.

REFERENCES

- [1] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, Berlin, Germany: Springer-Verlag, 1990.
- [2] J.-F. Cardoso *et al.*, "On the performance of orthogonal source separation algorithms," in *Proc. EUSIPCO*, M. Holt *et al.*, Eds., Edinburgh, U.K., Sept. 1994, pp. 776–779.
- [3] J.-F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, pp. 3017–3030, Dec. 1996.
- [4] B. W. Char, K. O. Geddes, G. H. Gonnet, B. L. Leong, M. B. Monagan, and S. M. Watt, *Maple V Library Reference Manual*. New York: Springer-Verlag, 1991.
- [5] P. Comon, "Separation of stochastic processes," in *Proc. Workshop Higher Order Spectral Anal.*, Vail, CO, June 1989, pp. 174–179.
- [6] P. Comon, "Independent component analysis, A new concept?," *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.
- [7] M. Gaeta and J.-L. Lacoume *et al.*, "Sources separation without a priori knowledge: The maximum likelihood solution," in *Proc. EUSIPCO*, L. Torres *et al.*, Eds., Barcelona, Spain, 1990, pp. 621–624.
- [8] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1989.
- [9] F. Harroy, J.-L. Lacoume, and M. A. Lagunas *et al.*, "A general adaptive algorithm for nongaussian source separation without any constraint," in *Proc. EUSIPCO*, M. Holt *et al.*, Eds., Edinburgh, U.K., Sept. 1994, pp. 1161–1164.
- [10] 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, Dec. 1996.
- [11] C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Process.*, vol. 24, no. 1, pp. 1–10, July 1991.
- [12] A. Mansour and C. Jutten, "Fourth-order criteria for blind sources separation," *IEEE Trans. Signal Processing*, vol. 43, pp. 2022–2025, Aug. 1995.
- [13] A. K. Nandi and V. Zarzoso, "Fourth-order cumulant based blind source separation," *IEEE Signal Processing Lett.*, vol. 3, pp. 312–314, Dec. 1996.
- [14] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*. New York: Wiley, 1980.
- [15] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, London, U.K.: Chapman & Hall, 1986.
- [16] E. Srouchyari, "Blind separation of sources, part III: Stability analysis," *Signal Process.*, vol. 24, no. 1, pp. 21–29, July 1991.
- [17] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [18] V. Zarzoso and A. K. Nandi, "The potential of decorrelation in blind separation of sources based on cumulants," in *Proc. ECSAP*, A. Procházka, Ed., Prague, Czech Republic, June 24–27, 1997, pp. 293–296.
- [19] V. Zarzoso, A. K. Nandi, and E. Bacharakis, "Maternal and foetal ECG separation using blind source separation methods," *IMA J. Math. Appl. Med. Biol.*, vol. 14, pp. 207–225, 1997.
- [20] V. Zarzoso and A. K. Nandi, "Generalization of a maximum-likelihood approach to blind source separation," in *Proc. EUSIPCO*, vol. IV, Rhodes, Greece, Sept. 8–11, 1998, pp. 2069–2072.
- [21] —, "Blind separation of independent sources for virtually any source probability density function," *IEEE Trans. Signal Processing*, vol. 47, no. 9, pp. 2419–2432, Sept. 1999.
- [22] —, "Analysis and development of closed-form estimators for blind separation of sources—Part I: Real mixtures," submitted for publication.
- [23] —, "Analysis and development of closed-form estimators for blind separation of sources—Part II: Complex mixtures," submitted for publication.



Vicente Zarzoso (S'97) was born in Valencia, Spain, on September 12, 1973. He attended the Universidad Politécnica de Valencia for the first four years of his degree, and the University of Strathclyde, Glasgow, U.K., on an Erasmus exchange programme for the final year of his degree. In July 1996, he received the M.Eng. degree with the highest distinction (Premio Extraordinario de Terminación de Estudios) in Telecommunications Engineering from the Universidad Politécnica de Valencia. He was awarded a scholarship by the University of Strathclyde to

study in the Signal Processing Division of the Department of Electronic and Electrical Engineering toward his Ph.D. degree, the first year of which was also partly funded by a grant from the Defence Evaluation and Research Agency (DERA) of the U.K.

Since March 1999, he has been with the Signal Processing and Communications Group of the Department of Electrical Engineering and Electronics, The University of Liverpool, Liverpool, U.K., continuing his doctoral studies. His research interests include blind signal separation, higher order statistics, and statistical signal and array processing and their application to biomedical problems and communications.



Asoke K. Nandi (SM'96) received the Ph.D. degree from Trinity College, the University of Cambridge, Cambridge, U.K., in 1979. He held a scholarship at Trinity College during his doctoral studies at the Cavendish Laboratory, Cambridge.

Since then, he held several research positions: associateship with Rutherford Appleton Laboratory, Oxfordshire, U.K., and in the European Organization for Nuclear Research (CERN), Geneva, Switzerland, as well as Advanced Fellowship with the Department of Physics, Queen Mary College, London, U.K., and in the Department of Nuclear Physics, University of Oxford, Oxford, U.K. In 1987, he joined the Imperial College, London, U.K., as the Solartron Lecturer in the Signal Processing Section of the Electrical Engineering Department. In 1991, he joined the Signal Processing Division of the Electronic and Electrical Engineering Department, University of Strathclyde, Glasgow, U.K., as a Senior Lecturer. Subsequently, he was appointed a Reader in 1995 and a Professor in 1998. In March 1999, he moved to the University of Liverpool, Liverpool, U.K., to take up the appointment to the David Jardine Chair of Electrical Engineering within the Department of Electrical Engineering and Electronics. In 1983 he was part of the UAI team at CERN that discovered the three fundamental particles known as W^+ , W^- , and Z^0 , 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 the Head of the Signal Processing and Communications Research Group, which includes a number of tenured academics and a strong group of doctoral and post-doctoral researchers 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 machine condition monitoring, signal modeling, system identification, communication signal processing, time delay estimation, biomedical signals, underwater sonar, application of artificial neural networks, ultrasonics, blind source separation, and blind deconvolution. He has authored or coauthored over 200 technical publications including two books: *Automatic Modulation Recognition of Communication Signals* and *Blind Estimation Using Higher-Order Statistics*, and over 100 journal papers.

Professor Nandi was awarded the Mountbatten Premium, Division Award of the Electronics and Communications Division, of the Institution of Electrical Engineers of the U.K. in 1998 and the Water Arbitration Prize of the Institution of Mechanical Engineers of the U.K. in 1999. He is a Fellow of the Cambridge Philosophical Society, the Institution of Electrical Engineers, the Institute of Mathematics and its Applications, and the Institute of Physics. He is a Member of the British Computer Society and the European Association for Signal Processing.