Self-Adaptive Source Separation, Part I: Convergence Analysis of a Direct Linear Network Controled by the Hérault-Jutten Algorithm

Odile Macchi, *Fellow, IEEE*, and Eric Moreau, *Member, IEEE*

***Abstract*—It is known that self-adaptive separation of a linear mixture of non-Gaussian independent sources can be achieved with a feedback linear neural network that is adapted by the Hérault-Jutten algorithm. Yet, realizability of the feedback requires implementation constraints. In this paper, an equivalent direct (without feedback) network is considered that is free of these constraints while the self-adaptive rule is kept unchanged. The separating states are shown to be equilibrium points. Their stability status is studied in the case of two sources. Then, we show that the algorithm is convergent in the "quasi"-quadratic mean sense toward a separating state for a small enough step-size.**

I. Introduction

In the source separation problem, several linear mixtures of unknown, random, zero-mean, and statistically independent signals called sources are observed [1]–[13]. The sources must be recovered without knowing the mixture parameters. This is a "self-learning" or an "unsupervised" inverse problem. The inverse system has to be learned with the sole knowledge of the observed mixtures. This problem is often qualified blind[1]. It has many applications in diverse fields of engineering and applied sciences, like communications, array processing, airport surveillance (localizing and recognizing the planes), etc. ...

The key property that makes separation possible is mutual independence of the sources. In the independent component analysis technique [1], by a linear transform, a new vector is looked for with independent components that will hopefully correspond to the sources.

The two different ways to deal with this problem are the block approach and the adaptive one. In the block approach, certain statistics of order 3 or more are calculated for the observed mixture vector. Then, identification of the mixture parameters follows. For instance, the contribution [2] is based on the eigenvalue decomposition of a fourth-order cumulant matrix of the observations. In [1], a contrast criterion based on rth-order cumulants $(r \geq 3)$ is maximized, corresponding to zero cross-cumulants.

In this paper, like in [10]–[14], we adopt the adaptive approach that solves the problem in real time. The first successful method was a linear adaptive feedback structure that has been investigated independently in two fields of application. For instance, in the case of two sources and two observations x_1, x_2, it computes

$$\begin{aligned} y_1 &= x_1 + f_{12} y_2 \\ y_2 &= x_2 + f_{21} y_1. \end{aligned} \tag{1}$$

On the one hand, this structure has been successfully used under the denomination of "bootstrapped algorithm" in satellite and radio communication [12], [13] using second-order moments in the adaptation laws: minimum output power and/or zero correlation together with output nonlinearities called signal discriminators. On the other hand, this feedback structure has been formulated in a neural network context [5], [6], e.g., for speech enhancement in noise [7]. In the latter approach, the structure is controlled by an unsupervised local Hebbian learning rule. A nonlinear odd function is applied to the outputs to produce independence and not only decorrelation.

This paper uses the original updating rule of [5] and [6], even though the structure is changed. Indeed, the joint system (1) raises a problem of realizability. Moreover, it can yield instability. To cope with this problem, it is assumed in [5] and [12] that the sources are random processes in continuous time. Then, oversampling makes the feedback realizable at the price of an increased sampling rate, which is much beyond Shannon's requirement. Therefore, implementation complexity is increased.

Thus, it is better to consider a direct structure as in [10] and [11]. For two sources, it reads

$$\begin{aligned} y_1 &= x_1 + d_{12} x_2 \\ y_2 &= x_2 + d_{21} x_1. \end{aligned} \tag{2}$$

The present Part I is devoted to this direct structure. Now, one can design a mixed realizable direct-feedback structure as in [10], which will benefit from the advantages of both the direct and the feedback structures. It is the purpose of Part II of this contribution to compare the direct, feedback, and mixed structures.

Manuscript received June 15, 1994; revised October 22, 1996. The associate editor coordinating the review of this paper and approving it for publication was Dr. Dirk Slock.

O. Macchi is with the Laboratoire des Signaux et Systèmes (LSS), CNRS-Supelec-University of Paris 11, Gif sur Yvette, France.

E. Moreau was with the Laboratoire des Signaux et Systèmes (LSS), CNRS-Supelec-University of Paris 11, Gif sur Yvette, France. He is now with the Groupe d'étude des Signaux et Systémes (MS-GESSY), ISITV, Toulon, France.

Publisher Item Identifier S 1053-587X(97)02586-5.

[1] Many authors use the adjective "blind" despite its pejorative connotation. On the other hand, the positive adjective "self-learning" emphasizes the power of this approach.

After stating the problem in Section II, we investigate the separating states in Section III. Then, we consider the particular updating rule imagined by Hérault and Jutten for the feedback structure and apply it to the direct structure. Its equilibrium states with their associated stability are studied for the general case of m sources with a particular emphasis on the case of two sources (Section IV). Section V is devoted to the study of the (average) deterministic algorithm, whereas Section VI shows that the stochastic algorithm is convergent in the "quasi"-quadratic mean sense toward the desired system which separates the sources.

II. Statement of the Problem

We observe the n outputs $x_i, i = 1, \cdots, n$ of an unknown multidimensional linear system called "mixture," which is driven by unobserved realizations of m random, zero-mean, and statistically independent input sources $a_j, j = 1, \cdots, m$. In matrix and vector notations, this model reads

$$\boldsymbol{x} = \boldsymbol{Ga} \tag{3}$$

where

$\boldsymbol{x}$ $(n, 1)$ vector of observations,
$\boldsymbol{a}$ $(m, 1)$ vector of sources
$\boldsymbol{G}$ (n, m) mixture matrix.

It is a noise-free, time-free model.

To recover a vector $\boldsymbol{y}$ close to the source vector $\boldsymbol{a}$ *knowing the vector* $\boldsymbol{x}$ *only*, one should estimate some inverse of $\boldsymbol{G}$, which is denoted $\boldsymbol{H}$ in the sequel. The corresponding estimate of $\boldsymbol{a}$ is

$$\boldsymbol{y} = \boldsymbol{Hx} \tag{4}$$

where $\boldsymbol{H}$ is an (m, n) matrix. Clearly, this problem has no solution when $m > n$. When $m < n$, a subvector $\boldsymbol{x}_s$ can be extracted from $\boldsymbol{x}$ with m specific coordinates of which the other will be linear combinations. Then separation is done with the help of the subvector $\boldsymbol{x}_s$. It will thus be assumed in the sequel that $n = m$.

Finally, let $Q_1, \cdots, Q_m$ designate the respective powers of $a_1, \cdots, a_m$. The normalized (unit power) sources $a'_i = a_i/\sqrt{Q_i}$ and the modified matrix $\boldsymbol{G}'$ with entries $g'_{ij} = g_{ij}\sqrt{Q_j}$ yield the same observed vector $\boldsymbol{x} = \boldsymbol{G}'\boldsymbol{a}' = \boldsymbol{Ga}$. Therefore, it can be assumed in the sequel that all the sources have unit power.

For this problem, the key assumption is joint independence of the sources, which is a property that is much stronger than decorrelation for non-Gaussian sources. However, the requirement of independent components for $\boldsymbol{y}$ is not sufficient, on its own, to ensure that $\boldsymbol{y} = \boldsymbol{a}$. Let $\boldsymbol{P}$ be any arbitrary permutation matrix (with one and only one nonzero element—which is a 1—in each row and each column), and let $\boldsymbol{\Lambda} = \text{Diag}(\lambda_1, \cdots, \lambda_m)$ be any diagonal matrix. The vector

$$\boldsymbol{y} = \boldsymbol{\Lambda Pa} \tag{5}$$

has components

$$y_i = \lambda_i a_{p(i)}, \qquad i = 1, \cdots, m \tag{6}$$

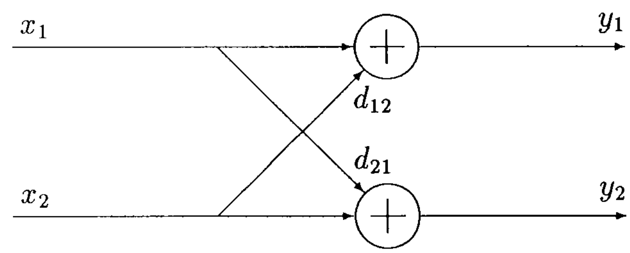

Fig. 1. Direct architecture for the case of two sources.

where $p(i)$ is the label of the column where the 1 of the ith row stands. Therefore, $\boldsymbol{y}$ has m independent components proportional to the sources a_j, provided none of the λ_i is zero, i.e., the matrix $\boldsymbol{\Lambda}$ is invertible. Conversely, it has been shown in [1] for non-Gaussian sources that the linear mixture $\boldsymbol{y} = \boldsymbol{Sa}$ cannot have independent components unless $\boldsymbol{S} = \boldsymbol{\Lambda P}$. Therefore, the matrix $\boldsymbol{H}$ in (4) generates such a vector $\boldsymbol{y}$ if and only if (iff)

$$\boldsymbol{HG} = \boldsymbol{\Lambda P}. \tag{7}$$

Such a matrix $\boldsymbol{H}$ will be called "separating." Applied to the mixture $\boldsymbol{x}$, the matrix $\boldsymbol{H}$ restores the m individual sources a_i up to certain specific nonzero gains (the coefficients λ_i) and up to a certain reordering (the permutation $\mathcal{P} = (p(1), \cdots, p(m))$). With the help of the entries k_{ij} of the inverse mixture matrix $\boldsymbol{K} = \boldsymbol{G}^{-1}$, it is easily seen that the gains λ_i are

$$\lambda_i = 1/k_{p(i)i}, \quad \forall i. \tag{8}$$

Since $\boldsymbol{\Lambda}$ and $\boldsymbol{P}$ are invertible, (7) shows that the separation problem has no solution when $\boldsymbol{G}$ is not invertible. Thus, throughout this paper (Parts I and II), $\boldsymbol{G}$ *is assumed regular*.

Since $\boldsymbol{\Lambda}$ and $\boldsymbol{P}$ are arbitrary, this independence problem is an ill-posed problem with a number of unspecified (arbitrary) parameters. The three structures investigated in Parts I and II of this paper put constraints onto the m nonzero gains λ_i. For the direct structure, this is obtained by fixing at the value 1 the diagonal entries h_{ii} [cf. (2)]. As for the ordering of restored sources, there is only a finite number of possibilities.

Finally, it is assumed that the m sources a_i have zero mean and finite fourth-order moments:

$$\text{E}a_i = 0 \tag{9}$$

$$\text{E}a_i^2 = 1 \tag{10}$$

$$\text{E}a_i^4 \triangleq A_i^2 < \infty, \quad A_i > 0. \tag{11}$$

The sources can be characterized equivalently by their so-called "kurtosis":

$$\kappa_i \triangleq A_i^2 - 3 \tag{12}$$

which is very often used in the literature because they are null for Gaussian variables.

III. The Separating States

The direct architecture was presented in [12] and [10]. It is depicted in Fig. 1 for two sources $(m = 2)$.

It obeys the input/output equation

$$\boldsymbol{y} = (\boldsymbol{I} + \boldsymbol{D})\boldsymbol{x}, \qquad d_{ii} = 0 \quad \forall i \tag{13}$$

where $\boldsymbol{I}$ is the identity matrix. There are m linear cells with m inputs and one output each and one input coefficient constrained to 1. It is a single-layer feedforward network. It involves $m(m-1)$ multiplications, which is a low computational complexity.

The separating states are those matrices $\boldsymbol{D}$ for which (5) holds.

Case of Two Sources: For $m = 2$, it is clear that the changed quantities

$$\boldsymbol{a}' \triangleq (a_2, a_1) \quad \text{and} \quad \boldsymbol{G}' \triangleq \begin{pmatrix} g_{12} & g_{11} \\ g_{22} & g_{21} \end{pmatrix}$$

satisfy the relationship

$$\boldsymbol{G}\boldsymbol{a} = \boldsymbol{G}'\boldsymbol{a}'. \tag{14}$$

Hence, by relabeling the sources if required, it can be assumed that the entries of $\boldsymbol{G}$ satisfy

$$|g_{12}g_{21}| \leq |g_{11}g_{22}|. \tag{15}$$

For instance, this inequality holds when a_1 is dominant in x_1 and a_2 in x_2, i.e., when $|g_{12}| \leq |g_{11}|$ and $|g_{21}| \leq |g_{22}|$. Thanks to the invertibility of $(\boldsymbol{I} + \boldsymbol{D})$, one can write

$$(\boldsymbol{I} + \boldsymbol{D}) = \gamma(\boldsymbol{I} - \boldsymbol{D})^{-1}, \qquad \gamma = 1 - d_{12}d_{21}. \tag{16}$$

Hence, according to (13)

$$\begin{aligned} y_1 &= \gamma x_1 + d_{12}y_2 \\ y_2 &= \gamma x_2 + d_{21}y_1. \end{aligned} \tag{17}$$

Comparison of the systems (1) and (17) shows that the direct and feedback structures having the same vectors $D \triangleq (d_{12}, d_{21})^T$ yield the same outputs (resp. $\boldsymbol{y}_D$ and $\boldsymbol{y}_F$) up to the scale factor γ

$$\boldsymbol{y}_D = \gamma \boldsymbol{y}_F. \tag{18}$$

In this sense, the structures are equivalent. This will reduce the efforts by one half.

Now, there are two possible configurations.

First Configuration: When

$$g_{11}g_{22}g_{12}g_{21} \neq 0 \tag{19}$$

there are two separating matrices corresponding to the permutation matrices $\boldsymbol{I}$ and

$$\boldsymbol{I}' \triangleq \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The vector D of the first separating state (associated with $\boldsymbol{P} = \boldsymbol{I}$) is

$$D^1 = -\left(\frac{g_{12}}{g_{22}}, \frac{g_{21}}{g_{11}}\right)^T \tag{20}$$

with the outputs

$$w_1 = \frac{D_g}{g_{22}}a_1; \quad w_2 = \frac{D_g}{g_{11}}a_2 \tag{21}$$

where

$$D_g \triangleq g_{11}g_{22} - g_{12}g_{21} (\neq 0) \tag{22}$$

is the determinant of $\boldsymbol{G}$. We call D^1 the "natural" separating state because it restores a_1 and a_2 in the natural order. Channel 1, where a_1 is dominant, generates a_1 and similarly for a_2.

The second solution (associated with $\boldsymbol{P} = \boldsymbol{I}'$) is the "reversing" separating state:

$$D^2 = -\left(\frac{g_{11}}{g_{21}}, \frac{g_{22}}{g_{12}}\right)^T \tag{23}$$

with the outputs

$$w_1' = -\frac{D_g}{g_{21}}a_2; \quad w_2' = -\frac{D_g}{g_{12}}a_1. \tag{24}$$

Second Configuration: When

$$g_{11}g_{22}g_{12}g_{21} = 0 \tag{25}$$

we must specify which one(s) among the coefficients g_{ij} is (are) zero. According to (15), $g_{11} = 0$ or $g_{22} = 0$ mean that the matrix $\boldsymbol{G}$ has a row or a column that is null. Thus, $\boldsymbol{G}$ is not invertible; this is a case that must be excluded. Consequently, (25) cannot hold unless

$$g_{11}g_{22} \neq 0. \tag{26}$$

When $g_{12} = 0$, $\boldsymbol{G}$ is a lower triangular matrix. There is one and only one separating state that is D^1. It corresponds to the output

$$y_1 = x_1, \quad y_2 = -\frac{g_{21}}{g_{11}}x_1 + x_2. \tag{27}$$

When $g_{21} = 0$, $\boldsymbol{G}$ is an upper triangular matrix, and the (unique) solution is similar.

IV. Adaptation Law

We return to the general case of m sources and use the vector

$$D \triangleq (d_{12}, \cdots, d_{1m}, d_{21}, d_{23}, \cdots, d_{2m}, d_{m1}, \cdots, d_{m(m-1)})^T \tag{28}$$

to characterize the matrix $\boldsymbol{D}$ in (13). Adaptation is usually performed in a discrete recursive way, i.e., D is incremented along

$$D(n) = D(n-1) - \mu Z(n), \qquad \mu > 0 \tag{29}$$

where μ is a positive step-size. We use the following adaptation law [10] for the element z_{ij} of Z:

$$z_{ij} = y_i^3 y_j, \qquad i \neq j. \tag{30}$$

This is the mere transposition to the direct structure of the updating rule used in [5] and [6] for the feedback structure for which it has turned out successfully. It is built up solely with the system outputs. Therefore, it makes sense to use the same increment with the two different structures. The rest of this Part I will bring full justification for this adaptation choice. In

(30), Z is evaluated taking the previous value $D = D(n-1)$ to calculate y_i and y_j. This yields a stochastic algorithm.

We will denote $\overline{Z}$ the expectation of the increment. It is conditioned on the state of D:

$$\overline{z}_{ij}(D) \triangleq \mathrm{E}_{/D} y_i^3 y_j, \qquad i \neq j \tag{31}$$

and the result depends on D. The deterministic algorithm associated with (29) and (30) ["deterministic Hérault Jutten" (DHJ)] is

$$D(n) = D(n-1) - \mu \overline{Z}(D(n-1)). \tag{32}$$

Consider a separating state D corresponding to certain matrices $\boldsymbol{\Lambda}$ and $\boldsymbol{P}$. Clearly, thanks to independence of the a_i

$$\overline{z}_{ij}(D) = \lambda_i^3 \lambda_j \mathrm{E} a_{p(i)}^3 \mathrm{E} a_{p(j)}, \qquad i \neq j. \tag{33}$$

Since the a_i are zero-mean

$$\overline{Z}(D) = 0. \tag{34}$$

It means that *separating states are equilibria of the deterministic algorithm (32)*. However, the converse is not true: Some equilibrium points of (32) might not be separating states, as will be discussed later.

The next issue is stability of these equilibria. A first approach is the so-called "ordinary differential equation" (ODE) technique [16], [17], which replaces the recurrence (32) by a differential equation according to

$$\frac{dD(t)}{dt} = -\overline{Z}(D(t)). \tag{35}$$

This system is locally stable near an equilibrium point E iff the corresponding tangent linear system

$$\frac{dD(t)}{dt} = -\boldsymbol{J}(E)(D(t) - E) \tag{36}$$

is stable, where $\boldsymbol{J}(\cdot)$ is the matrix of partial derivatives with entries

$$J_{kk'}(D) = \frac{\partial \overline{Z}_k(D)}{\partial d_{k'}}, \qquad k, k' = 1, \cdots, m(m-1) \tag{37}$$

and k is the rank of an entry d_{ij} in the definition (28) of D. According to (31)

$$J_{kk'} = 3\mathrm{E} y_i^2 y_j \frac{\partial y_i}{\partial d_{i'j'}} + \mathrm{E} y_i^3 \frac{\partial y_j}{\partial d_{i'j'}}. \tag{38}$$

Moreover, it follows from (13) that

$$\frac{\partial y_l}{\partial d_{i'j'}} = \delta_{li'} x_{j'} \qquad (i' \neq j') \tag{39}$$

where δ_{li} is the Kronecker symbol. Accordingly

$$J_{kk'} = \begin{cases} 0 & \text{if } i' \neq i, i' \neq j, \\ 3\mathrm{E} y_i^2 y_j x_{j'} & \text{if } i' = i, \\ \mathrm{E} y_i^3 x_{j'} & \text{if } i' = j. \end{cases} \tag{40}$$

With the help of (13) and (40), the matrix $\boldsymbol{J}$ can thus be evaluated. For instance, with $m = 2$

$$\boldsymbol{J}(D^1) = \frac{D_g^3}{g_{11}^3 g_{22}^3} \begin{pmatrix} 3g_{11}^2 g_{22}^2 & g_{11}^4 A_1^2 \\ g_{22}^4 A_2^2 & 3g_{11}^2 g_{22}^2 \end{pmatrix},$$
$$\boldsymbol{J}(D^2) = \frac{-D_g^3}{g_{12}^3 g_{21}^3} \begin{pmatrix} 3g_{12}^2 g_{21}^2 & g_{12}^4 A_2^2 \\ g_{21}^4 A_1^2 & 3g_{12}^2 g_{21}^2 \end{pmatrix}. \tag{41}$$

Then, it is possible to test the local stability of the ODE system (35) at the equilibrium states $E = D$. The procedure is to calculate the roots of the characteristic polynomial associated with the matrix $\boldsymbol{J}(E)$ and to check whether or not all of them have a positive real part. This procedure is indeed feasible but tedious.

A simpler (but partial) answer is to test the real part of the sum of eigenvalues. It is well known that this sum is the trace of $\boldsymbol{J}$. Hence, we have a necessary condition for stability of the equilibrium state E:

$$T(E) \triangleq \tfrac{1}{3}\, \mathrm{Trace}(\boldsymbol{J}(E)) > 0. \tag{42}$$

It follows from (40) that

$$T(E) = \sum_{\substack{i,j=1 \\ i \neq j}}^{m} \mathrm{E} y_i^2 y_j x_j. \tag{43}$$

Moreover, if this equilibrium state E is indeed a separating state D associated with a specific permutation $\mathcal{P} = (p(1), \cdots, p(m))$, it follows that

$$T(D) = \sum_{\substack{i,j=1 \\ i \neq j}}^{m} \lambda_i^2 \lambda_j \mathrm{E} a_{p(i)}^2 a_{p(j)} x_j, \tag{44}$$

$$T(D) = \sum_{\substack{i,j=1 \\ i \neq j}}^{m} \frac{g_{jp(j)}}{k_{p(i)i}^2 k_{p(j)j}} \tag{45}$$

where the last equality follows from independence of the a_i and from (8)–(10). Based on (45), it is possible to test the necessary (but insufficient) stability condition (42) at each separating equilibrium D.

The rest of this paper is devoted to the case of $m = 2$.

V. The Deterministic Algorithm

Stability and convergence of the deterministic algorithm (31) and (32) toward a separating state is a prerequisiste for the stochastic algorithm (29) and (30) itself to be stable and convergent. Therefore, we first investigate the deterministic algorithm.

A. Equilibrium States

Under the fourth-order moment condition

$$A_1 A_2 - 3 \neq 0 \tag{46}$$

the equilibrium states of the DHJ algorithm associated to the feedback structure are investigated in [18] and [19] and recalled in Part II of this paper. The equivalence (18) between the direct and feedback structures having the same vector D can be used to avoid solving the system (34). Therefore, the results below hold:

First Configuration: [i.e., $g_{11}g_{22}g_{12}g_{21} \neq 0$ according to (19)]. There are two separating equilibria, namely, $E^1 = D^1, E^2 = D^2$ [see (20) and (23)] plus, in general,[2] two nonseparating equilibria E^3 and E^4 (for which the outputs y_1 and y_2 remain mixtures of the sources a_1 and a_2).

[2] when a certain equality relates the fourth-order moments A_1^2 and A_2^2 to one of the two lines of the matrix $\boldsymbol{G}$, there is only one nonseparating equilibrium.

Second Configuration: [i.e., $g_{12}g_{21} = 0, g_{11}g_{22} \neq 0$ according to (25) and (26)]. There is one separating equilibrium $E^1 = D^1$ plus, in general, the two nonseparating equilibria E^3 and E^4.

B. Local Stability in D^1 and D^2

Consider the stability in the natural separating state D^1 first. For any vector D, we define the deviation from optimality

$$V = (v_1, v_2)^T \triangleq D - D^1 = \left(d_{12} + \frac{g_{12}}{g_{22}}, d_{21} + \frac{g_{21}}{g_{11}}\right)^T. \tag{47}$$

By the definition (13) of the direct structure, the output corresponding to a fixed (nonoptimal) vector D is

$$y_1 = w_1 + x_2 v_1; \quad y_2 = w_2 + x_1 v_2 \tag{48}$$

where w_1 and w_2 are given in (21). As a result, we get the following expansion for the algorithm increment $Z(D)$ versus increasing powers of the deviation V:

$$Z(D) = Z^{(1)} + \boldsymbol{A}V + \boldsymbol{B}V^2 + \boldsymbol{F}V^3 + \boldsymbol{K}V^4 \tag{49}$$

where

$$V^2 \triangleq (v_1^2, v_1 v_2, v_2^2)^T,$$
$$V^3 \triangleq (v_1^3, v_1^2 v_2, v_1 v_2^2, v_2^3)^T \quad \text{and}$$
$$V^4 \triangleq (v_1^3 v_2, v_1 v_2^3)^T$$

are of order at least two versus V, and where

$$Z^{(1)} \triangleq \begin{pmatrix} w_1^3 w_2 \\ w_1 w_2^3 \end{pmatrix}, \quad \boldsymbol{A} \triangleq \begin{pmatrix} 3w_1^2 w_2 x_2 & w_1^3 x_1 \\ w_2^3 x_2 & 3w_1 w_2^2 x_1 \end{pmatrix} \tag{50}$$

while

$$\boldsymbol{B} \triangleq \begin{pmatrix} 3w_1 w_2 x_2^2 & 3w_1^2 x_1 x_2 & 0 \\ 0 & 3w_2^2 x_1 x_2 & 3w_1 w_2 x_1^2 \end{pmatrix}, \tag{51}$$

$$\boldsymbol{F} \triangleq \begin{pmatrix} w_2 x_2^3 & 3w_1 x_1 x_2^2 & 0 & 0 \\ 0 & 0 & 3w_2 x_1^2 x_2 & w_1 x_1^3 \end{pmatrix} \tag{52}$$

$$\boldsymbol{K} \triangleq \begin{pmatrix} x_1 x_2^3 & 0 \\ 0 & x_1^3 x_2 \end{pmatrix}. \tag{53}$$

Because D^1 is a separating state, we have

$$\mathrm{E}[Z^{(1)}] = 0. \tag{54}$$

Moreover, according to (21) and (41), it is easily seen that

$$\mathrm{E}[\boldsymbol{A}] = \boldsymbol{J}(D^1). \tag{55}$$

Consider the nonlinear mapping $D \to D'$ associated with the deterministic algorithm (32), and let us write it with the help of the corresponding deviation vectors V and V', i.e.

$$V' = \mathcal{G}(V). \tag{56}$$

It reads

$$V' = (\boldsymbol{I} - \mu \boldsymbol{J}(D^1))V - \mu(\mathrm{E}[\boldsymbol{B}]V^2 + \mathrm{E}[\boldsymbol{F}]V^3 + \mathrm{E}[\boldsymbol{K}]V^4). \tag{57}$$

For $\mu < \mu_o$, the matrix $\boldsymbol{I} - \mu\boldsymbol{J}(D^1)$ is invertible. Hence, for $|V|$ suitably upperbounded, $\mathcal{G}$ is a diffeomorphism. We call $\mathcal{U}$ the corresponding (open) neighborhood of D^1. Then, by virtue of the Hartman-Grossman theorem [22], the equilibrium point D^1 of the recursion (32) is asymptotically stable iff the associated tangent linear recurrence

$$V(n) = (\boldsymbol{I} - \mu\boldsymbol{J}(D^1))V(n-1) \tag{58}$$

is stable. In a similar way, when the reversing separating state D^2 exists (that is, when $g_{12}g_{21} \neq 0$), the linear system tangent in D^2 is written

$$V(n) = (\boldsymbol{I} - \mu\boldsymbol{J}(D^2))V(n-1) \tag{59}$$

where $V \triangleq D - D^2$. Hence, local stability in $D^i, i = 1, 2$ holds iff the eigenvalues of $\boldsymbol{K}_i \triangleq \boldsymbol{I} - \mu\boldsymbol{J}(D^i)$ have modulus less than 1. This is known to be equivalent to the set of inequalities

$$\mathrm{Det}(\boldsymbol{K}_i) < 1 \tag{60}$$

$$-1 - \mathrm{Det}(\boldsymbol{K}_i) < \mathrm{Trace}(\boldsymbol{K}_i) < 1 + \mathrm{Det}(\boldsymbol{K}_i). \tag{61}$$

The entries of $\boldsymbol{J}(D^i)$ are given by (41). It is not difficult to show that

$$\mathrm{Det}(\boldsymbol{K}_i) = 1 - 6\mu\beta_i + \mu^2\alpha\beta_i^2 \tag{62}$$

$$\mathrm{Trace}(\boldsymbol{K}_i) = 2 - 6\mu\beta_i \tag{63}$$

where

$$\beta_1 = D_g^3/(g_{11}g_{22}) \tag{64}$$

$$\beta_2 = -D_g^3/(g_{12}g_{21}) \tag{65}$$

$$\alpha = 9 - A_1^2 A_2^2. \tag{66}$$

The RHS of condition (61) cannot be true, neither in D^1 nor in D^2, unless α is positive:

$$A_1 A_2 < 3. \tag{67}$$

This is the condition for the existence of at least one stable separating state. It has only to do with the statistics of the sources a_1 and a_2. For instance, it is valid when both sources are sub-Gaussian, i.e., with negative kurtosis κ_i [cf. (12)]. However, (67) is more general. It can be valid with one sub-Gaussian and one super-Gaussian source. Note that one source can be Gaussian but not both of them. When (67) holds, we shall say that the two sources are "globally sub-Gaussian." Since the step-size μ is positive, inequality (60) is equivalent to

$$\mu\alpha\beta_i^2 < 6\beta_i. \tag{68}$$

Since α is positive, this inequality cannot hold unless β_i is positive. Now, it follows from (15) that

$$\beta_1 > 0 \tag{69}$$

$$-\beta_2 g_{11}g_{22}g_{12}g_{21} > 0. \tag{70}$$

As a result, for globally sub-Gaussian sources, the natural separating state D^1 is always locally stable. The reversing state D^2 is also locally stable iff

$$g_{11}g_{22}g_{12}g_{21} < 0. \tag{71}$$

Finally, (68) provides two upperbounds over μ for the respective stability of D^1 and D^2. These inequalities are refined by the LHS of (61) according to

$$0<\mu<\mu_1 \triangleq \frac{2g_{11}g_{22}}{D_g^3(3+A_1A_2)} \quad \text{for } D^1 \tag{72}$$

$$0<\mu<\mu_2 \triangleq \frac{-2g_{12}g_{21}}{D_g^3(3+A_1A_2)} \quad \text{for } D^2. \tag{73}$$

In summary, under conditions (67) and (72), for $D(0)$ inside a certain neighborhood $\mathcal{U}$ of D^1, the deterministic algorithm (32) converges to D^1. It is similar for D^2 under the additional assumption (71).

Three Remarks:

1) The stepsize upperbounds in (72) and (73) are approximated values because the problem has first been linearized.
2) These upperbounds increase sharply when D_g decreases, i.e., when the mixture matrix $\boldsymbol{G}$ is approaching irregularity. This result might appear paradoxical but will be better understood in Part II.
3) They are unaffected when the statistics of sources approach the limiting (forbidden) situation, where $\alpha \to 0$.

C. Instability of the Nonseparating Equilibria E^3 and E^4

This analysis is performed using the ODE. A straightforward (but rather tedious) calculation shows that when α is positive according to condition (67) (i.e., when there exists a stable separating equilibrium), then E^3 and E^4 are exponentially unstable. This result ensures that the recursive algorithm will not converge toward ill solutions of the separation problem. Finaly, note that the ODE (35) has exactly the same equilibrium states D^1, D^2, E^3, E^4 as the DHJ algorithm with the same status, except that there is, of course, no step-size condition such as (72) and (73).

We are now ready to investigate convergence of the random sequence of vectors $D(n)$ generated by the discrete algorithm [see (29) and (30)].

VI. "Quasi"-Convergence of the Stochastic Algorithm

The convergence properties of a stochastic adaptive algorithm are easy to handle only when the associated deterministic algorithm (or associated ODE) has only one stable equilibrium (say D^1). If, in addition, the ODE is globally stable, then the following general convergence result for stochastic algorithms has been proved in [16] and [17] under certain regularity and polynomial increase conditions for the increment Z, which are obviously fulfilled in the present case

Result 1 (General Algorithm): If the ODE is globally stable and has a unique stable equilibrium D^1, then

$$\forall \mu < \mu_o, \forall \epsilon, \lim_{n\to\infty} \sup P\{|D(n) - D^1| > \epsilon\} \le C(\mu) \tag{74}$$

where

$$C(\mu) \to 0, \quad \mu \to 0. \tag{75}$$

•

This result means that in steady-state, the probability that $D(n)$ departs (significantly) from D^1 is negligibly small if the step-size is small enough.

Unfortunately, even in the case where D^2 is unstable, we cannot apply Result 1 to the (stochastic) Hérault-Jutten algorithm because global stability of the ODE cannot be proved.[3]

When the ODE is not globally stable, the only convenient result is concerned with the decreasing step-size algorithm, i.e., $\mu = \mu_n$ is a function of the iteration step in the following way:

$$\sum_n \mu_n = \infty, \qquad \sum_n \mu_n^2 < \infty. \tag{76}$$

Then, the corresponding algorithm

$$D(n) = D(n-1) - \mu_n Z(n), \qquad \mu_n > 0 \tag{77}$$

has the ability to reach the optimal state D^1 (but not to track potential time variations $D^1(n)$) according to the following result (which is valid under the same regularity and polynomial increase conditions for Z as those in Result 1).

Result 2 (General Result): If $D(n)$ visits infinitely often some compact set $\mathcal{C}$ inside the ODE attraction basin of a stable equilibrium D^1, then $D(n)$ tends to D^1 almost surely, as n tends to infinity. •

In the stochastic approximation literature, this result is known as the "Kushner and Clark type of convergence" [17]. Result 2 is indeed applicable to the (stochastic) Hérault-Jutten algorithm under consideration here. However, it is but a weak result because of the infinitely often visiting hypothesis, which is essentially equivalent to the assumption that $D(n)$ is bounded. In most stochastic approximation problems, and particularly for this specific algorithm, the proof of boundedness seems a very difficult matter. For this reason, the approximate analysis that follows appears very valuable.

Result 3 (HJ Algorithm): For some fixed quantity η, the stochastic Hérault-Jutten algorithm satisfies

$$\mathrm{E}|V(n)|^2 \to \mu\eta, \qquad n \to \infty. \tag{78}$$

asymptotically for μ small and $D(0)$ close to D^1. •

A similar result holds with $D(0)$ close to D^2 when D^2 is a stable equilibrium of the (DHJ) algorithm. In the specific case where a_1 and a_2 have identical statistics, the proportionality coefficient is

$$\eta = \frac{g_{11}^4 + g_{22}^4}{g_{11}^5 g_{22}^5} \frac{3B^3 - A^6}{2(9 - A^4)} D_g^5 \tag{79}$$

where $B^3 \triangleq \mathrm{E}a_i^6$. This result is proved in Appendix A. It is of great practical value. Indeed, any preassigned level of accuracy can be reached for the steady-state power $\mathrm{E}|V(n)|^2$ of the deviation from optimality, provided μ is chosen small enough. This type of convergence can be called "quasi"-convergence in the mean square sense although it does not preclude the fact that with an extremely low probability, $D(n)$ can escape the vicinity of D^1. Such a result is familiar in the field of adaptivity because convergence of supervised adaptive LMS or RLS filters obeys a relationship similar to (78), as shown, for instance, in [20] and [21].

[3]The ODE does not appear to be globally stable. In fact, (35) is a 2×2 nonlinear differential system whose right-hand sides are rational fractions of fourth-order polynomials in the variables d_{12} and d_{21} and for which no Lyapounov function has been obtained.

Result 3 has been derived under the assumption that the sequence of pairs $\boldsymbol{a}(n) = (a_1(n), a_2(n))^T$ is i.d.d. This assumption is sufficient but far from necessary. We have used it because it makes the proof shorter.

In practice, our computer simulations have shown that (78) is valid when μ is smaller than $\mu_1/4$. This condition on μ is four times more restrictive than in the deterministic stability analysis.

VII. Conclusion

This paper is the first rigorous convergence analysis of the "Hérault-Jutten" stochastic algorithm for separating independent sources in a self-learning way. This algorithm was originally introduced for a feedback structure, but we have shown here that it can be applied to a direct structure. The separating states of this direct structure have been shown to be equilibria of the Hérault-Jutten algorithm. In the case of two sources, at least one of these equilibria is locally stable for the associated deterministic algorithm, provided the sources are globally sub-Gaussian, and the step-size is small enough. Finally, it has been proved that the stochastic algorithm itself is "quasi"-convergent in the mean square sense toward a separating state, subject to a suitable initialization in the attraction basin. This result requires a very small step-size μ.

All these theoretical results still have to be verified through computer simulations, and the achievements of this direct structure also have to be compared with those of the feedback one. This is the purpose of Part II of the present paper.

Appendix
"Quasi"-Mean Square Convergence of $D(n)$

When $D(0)$ is suitably close to D^1, the adaptation increment $Z(n)$ can be approximated by the first-order approximation of (49). Therefore, the deviation $V(n)$ from optimality obeys the approximate recursion

$$V(n) = (\boldsymbol{I} - \mu\boldsymbol{A}(n))V(n-1) - \mu Z^{(1)}(n). \tag{80}$$

Let us split $V(n)$ into two parts according to

$$V(n) = V^t(n) + V^f(n) \tag{81}$$

where

$$V^t(n) = (\boldsymbol{I} - \mu\boldsymbol{A}(n))V^t(n-1), \quad V^t(0) = V(0) \tag{82}$$

is called the transient deviation, while

$$V^f(n) = (\boldsymbol{I} - \mu\boldsymbol{A}(n))V^f(n-1) - \mu Z^{(1)}(n), \quad V^f(0) = 0 \tag{83}$$

is called the fluctuation deviation. Now, let

$$\begin{aligned} U_{n,j} &\triangleq (\boldsymbol{I} - \mu\boldsymbol{A}(n))\cdots(\boldsymbol{I} - \mu\boldsymbol{A}(j+1)), \qquad n > j,\\ U_{n,n} &\triangleq \boldsymbol{I} \end{aligned} \tag{84}$$

denote the transition matrix associated with the hereabove recursion. It is clear that the following relationships hold:

$$V^t(n) = U_{n,0}V(0) \tag{85}$$

$$V^f(n) = -\mu\sum_{j=1}^{n} U_{n,j}Z^{(1)}(j). \tag{86}$$

Although this assumption is not necessary, in order to reach a concise proof, we suppose that *the sequence $\boldsymbol{A}(\boldsymbol{n})$ is i.i.d.* It follows that all the factors in $U_{n,j}$ are independent. Moreover, $U_{n,j}$ and $Z^{(1)}(j)$ are independent. Finally, the pair $[\boldsymbol{A}(n), Z^{(1)}(n)]$ is independent of $V(n-1)$. Hence

$$\mathrm{E}V^t(n) = \prod_{j=1}^{n}(\boldsymbol{I} - \mu\mathrm{E}\boldsymbol{A}(j))V(0), \tag{87}$$

$$\mathrm{E}V^t(n) = (\boldsymbol{I} - \mu\boldsymbol{J}(D^1))^n V(0). \tag{88}$$

This quantity is exponentially vanishing under the three assumptions (15), (67), and (72), as proved in Section V-B. Moreover

$$\mathrm{E}V^f(n) = -\mu\sum_{j=1}^{n}(\boldsymbol{I} - \mu\boldsymbol{J}(D^1))^{n-j}\mathrm{E}Z^{(1)}(j). \tag{89}$$

Now, $Z^{(1)}(n)$ is zero-mean, and therefore

$$\mathrm{E}V^f(n) = 0. \tag{90}$$

It follows that in steady-state (i.e., for n large enough), we can assume that

$$\mathrm{E}V(n) = 0. \tag{91}$$

Now, consider second-order properties of $V(n)$. It follows from (80) that

$$\begin{aligned} V(n)V(n)^T = &(\boldsymbol{I} - \mu\boldsymbol{A}(n))V(n-1)V^T(n-1)\\ &\cdot(\boldsymbol{I} - \mu\boldsymbol{A}^T(n)) + \mu^2 Z^{(1)}(n)(Z^{(1)}(n))^T\\ &- \mu(V(n-1)(Z^{(1)}(n))^T\\ &+ Z^{(1)}(n)V^T(n-1))\\ &+ \mu^2(\boldsymbol{A}(n)V(n-1)(Z^{(1)}(n))^T\\ &+ Z^{(1)}(n)V^T(n-1)\boldsymbol{A}^T(n)). \end{aligned} \tag{92}$$

Denote

$$\boldsymbol{\Gamma}(n) \triangleq \mathrm{E}V(n)V^T(n) \tag{93}$$

$$B_i^3 \triangleq \mathrm{E}a_i^6, \qquad i = 1, 2, \quad B_i > 0 \tag{94}$$

and note that according to (50)

$$\begin{aligned} &\mathrm{E}Z^{(1)}(n)(Z^{(1)}(n))^T\\ &\quad= \frac{D_g^8}{g_{11}^6 g_{22}^6}\begin{pmatrix} B_1 g_{11}^3 & A_1^2 g_{11}^2 A_2^2 g_{22}^2\\ A_1^2 g_{11}^2 A_2^2 g_{22}^2 & B_2 g_{22}^3 \end{pmatrix} \triangleq \boldsymbol{L}. \end{aligned} \tag{95}$$

Thanks to the independence assumption and to (54), the expectation of (92) yields

$$\begin{aligned} \boldsymbol{\Gamma}(n) = &\boldsymbol{\Gamma}(n-1) - \mu\boldsymbol{J}(D^1)\boldsymbol{\Gamma}(n-1)\\ &- \mu\boldsymbol{\Gamma}(n-1)\boldsymbol{J}(D^1)^T + \mu^2\boldsymbol{L} \end{aligned} \tag{96}$$

where we have used the inequality

$$\mu \mathrm{E}[\boldsymbol{A}(n)\boldsymbol{\Gamma}(n-1)\boldsymbol{A}^T(n)] \ll \boldsymbol{J}(D^1)\boldsymbol{\Gamma}(n-1) + \boldsymbol{\Gamma}(n-1)\boldsymbol{J}(D^1)^T \quad (97)$$

which is valid for small values of μ, and where we have taken advantage of the result (91) to cancel the terms like $\mathrm{E}[\boldsymbol{A}(n)\mathrm{E}[V(n-1)](Z^{(1)}(n))^T]$. The recurrence (96) can be easily investigated. We note that $\boldsymbol{\Gamma}(n)$ is symmetrical by its very definition. It is equivalently represented by the 3-D vector

$$\gamma(n) \triangleq (\Gamma_{11}(n), \Gamma_{22}(n), \Gamma_{12}(n))^T \quad (98)$$

whose recurrence equation reads

$$\gamma(n) = (\boldsymbol{I} - \mu\boldsymbol{M})\gamma(n-1) + \mu^2 L \quad (99)$$

where

$$\boldsymbol{M} \triangleq \frac{D_g^3}{g_{11}^3 g_{22}^3}\begin{pmatrix} 6g_{11}^2 g_{22}^2 & 0 & 2g_{11}^4 A_1^2 \\ 0 & 6g_{11}^2 g_{22}^2 & 2g_{22}^4 A_2^2 \\ g_{22}^4 A_2^2 & g_{11}^4 A_1^2 & 6g_{11}^2 g_{22}^2 \end{pmatrix} \quad (100)$$

and

$$L = \frac{D_g^8}{g_{11}^6 g_{22}^6}(g_{11}^4 B_1^3, g_{22}^4 B_2^3, g_{11}^2 g_{22}^2 A_1^2 A_2^2)^T. \quad (101)$$

It is easily seen that the three eigenvalues of $\boldsymbol{M}$ are

$$\xi_\varepsilon = \frac{2D_g^3}{g_{11}g_{22}}\,(3 + \varepsilon A_1 A_2), \qquad \varepsilon = -1, 0, 1. \quad (102)$$

Because $A_1 A_2 \neq 0$, they are distinct. Moreover, they are positive thanks to condition (67). Thus, $\boldsymbol{M}$ can be diagonalized according to

$$\boldsymbol{M} = \boldsymbol{C}\ \mathrm{diag}(\xi_{-1}, \xi_0, \xi_1)\boldsymbol{C}^{-1} \quad (103)$$

where $\boldsymbol{C}$ is a regular matrix whose columns are eigenvectors, respectively, associated with the eigenvalues of $\boldsymbol{M}$. As a result, the sequence $\gamma(n)$ in (99) is convergent iff μ is suitably upperbounded along

$$\mu\xi_1 < 2 \quad (104)$$

i.e.

$$\mu < \frac{g_{11}g_{22}}{D_g^3(3 + A_1 A_2)}. \quad (105)$$

Under condition (105), the recurrence (99) is exponentially convergent toward

$$\gamma = \mu N, \qquad N \triangleq \boldsymbol{M}^{-1}L. \quad (106)$$

This result provides, in particular, the steady-state value of the mean-square deviation $\mathrm{E}|V(n)|^2$. It follows from (100), (101), and (106) that

$$\mathrm{E}|V(n)|^2 \to \mu\eta, \qquad \eta \triangleq (\eta_1 + \eta_2) \quad (107)$$

as stated by (78) in the text. Here, η_1 and η_2 are, respectively, the first and second components of the 3–D vector N in (106). Their calculation yields

$$\eta = \left(\frac{D_g}{g_{11}g_{22}}\right)^5 r \quad (108)$$

where r depends on $A_1, A_2, B_1, B_2, g_{11}$, and g_{22} in a straightforward manner. In the specific case where a_1 and a_2 have identical statistics ($A_1 = A_2 = A, B_1 = B_2 = B$), the numerical result is

$$r = (g_{11}^4 + g_{22}^4)\frac{3B^3 - A^6}{2(9 - A^4)} \quad (109)$$

which yields the result (79) in the text.

ACKNOWLEDGMENT

The authors are endebted to the anonymous reviewer who has pointed out some inaccurracy in the formulation of the convergence result.

REFERENCES

[1] P. Comon, "Independent component analysis, A new concept?," *Signal Processing*, vol. 36, no. 3, pp 287–314, Apr. 1994.
[2] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *Proc. Inst. Elec. Eng.*, pt. F, vol. 140, pp 362–370, Dec. 1993.
[3] J. L. Lacoume and P. Ruiz, "Separation of independent sources from correlated inputs," *IEEE Trans. Signal Processing*, vol. 40, pp. 3074–3078, Dec. 1992.
[4] L. Tong, Y. Inouye and R. Liu, "Waveform-preserving blind estimation of multiple independent sources," *IEEE Trans. Signal Processing*, vol. 41, pp 2461–2470, July 1992.
[5] J. Hérault et B. Ans, "Réseaux de neurones á synapses modifiables: Décodage de messages sensoriels composites par apprentissage non supervisé et permanent," *C.R. Acad. Sc. Paris*, t.299, série III, no. 13, pp. 525–528, 1984 (in French).
[6] C. Jutten and J. Hérault, "Blind separation of sources, Part I: An adaptative algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1–10, 1991.
[7] L. H. N. Thi, C. Jutten, and J. Caelen, "Speech enhancement: Analysis and comparison of methods on various real situations," in *Proc. EUSIPCO'92, Euro. Signal Processing Conf.*, Brussels, Belgium, Aug. 1992, pp. 303–306.
[8] N. Delfosse and P. Loubaton, "Adaptive separation of independent sources: A deflation approach," in *Proc. ICASSP'94*, Adelaide, Australia, Apr. 1994, vol. IV, pp. 41–44.
[9] P. Loubaton and P. Regalia, "Blind deconvolution of multivariate signals: a deflation approach," in *Proc. ICC'93, IEEE Int. Conf. Commun.*, Geneva, Switzerland, May 1993, vol. 2, pp. 1160–1164.
[10] E. Moreau and O. Macchi, "Two novel architectures for the self-adaptive separation of signals," in *Proc. ICC'93, IEEE Int. Conf. Commun.*, Geneva, Switzerland, May 1993, vol. 2, pp. 1154–1159.
[11] A. Dinç and Y. Bar-Ness, "Convergence and performance comparison of the three different structures of bootstrap blind adaptive algorithm for multisignal co-channel separation," in *Proc. MILCOM'92*, pp. 913–918, 1992.
[12] Y. Bar-Ness, J. W. Carlin, and M. L. Steinberger, "Bootstrapping adaptive cross-pol canceller for satellite communication," in *Proc. ICC'82 Int. Conf. Commun.*, Philadelphia, PA, June 1982.
[13] J. W. Carlin, Y. Bar-Ness, S. Gross, M. L. Steinberger, and W. E. Studdiford, "An IF cross-pol canceller for microwave radio systems," *IEEE J. Select. Areas Commun.*, vol. SAC-5, pp. 502–514, Apr. 1987.
[14] B. Laheld and J. F. Cardoso, "Adaptive source separation without prewhitening," in *Proc. EUSIPCO'94*, Edinburgh, Scotland, Sept. 1994, pp. 183–186.
[15] A. Swami, G. Giannakis, and S. Shamsunder, "Multichannel ARMA processes," *IEEE Trans. Signal Processing*, vol. 42, pp. 898–913, Apr. 1994.
[16] A. Benveniste, M. Metivier, and P. Priouret, "Adaptive algorithms and stochastic approximations," in *Applications of Mathematics*. New York: Springer-Verlag, 1990, vol. 22.
[17] H. J. Kushner and D. S. Clark, "Stochastic approximation for constrained and unconstrained systems," in *Applied Mathematics and Science Series*. New York: Springer, 1978, vol. 26.
[18] E. Sorouchyari, "Blind separation of sources, Part III: Stability analysis," *Signal Processing*, vol. 24, pp. 21–29, 1991.
[19] J. C. Fort, "Stabilité de l'algorithme de séparation de sources de Jutten et Hérault," *Traitement du Signal*, vol. 8, pp. 35–42, 1991 (in French).
[20] S. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall, 2nd ed., 1991.
[21] O. Macchi, *Adaptive Processing: The Least Mean Squares Approach with Applications in Transmission*. New York: Wiley, 1994.
[22] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields*. New York: Springer-Verlag, 1983, ch. 1.

Odile Macchi (M'75–SM'84–F'89) was born in 1943 in Aurillac, France. She graduated from the École Normale Supérieure, Paris, France. She received the Master's degree in physics in 1964 and mathematics in 1965 and received the teaching degree "Agrégation" from Paris University in 1966. She received the degree of Doctor of Science in 1972.

Presently, she is a Director of Research of the French National Center for Scientific Research at the Laboratoire des Signaux et Systémes of the Ecole Supérieure d'Électricité in the field of adaptive communications, signal, and image digital processing.

Dr. Macchi was awarded three French prizes in the field of electricity and was nominated Distinguished Lecturer for 1994–1995 by the IEEE Signal Processing Society. In 1994, she was elected member of the French "Académie des Sciences."

Eric Moreau (S'93–M'96) was born in Lille, France, in 1964. He graduated from the Ecole Nationale Supérieure des Arts et Métiers (ENSAM), Paris, in 1989 and received the agrégation de physique degree from the Ecole Normale Supérieure de Cachan, France, in 1990. He received the M.Sc. degree in 1991 and the Ph.D. degree in 1995, both in the field of signal processing and from the University of Paris-Sud.

Since 1995, he has been with the Institut des Sciences de l'Ingénieur de Toulon et du Var (ISITV) as an Assistant Professor and with the Groupement d'Etude des Signaux et Systémes (MS-GESSY). His main research interests are in statistical signal processing and multidimensionnal signal processing using high-order statistics.