

An Iterative Inversion Approach to Blind Source Separation

Sergio Cruces-Alvarez, *Associate Member, IEEE*, Andrzej Cichocki, *Member, IEEE*, and Luis Castedo-Ribas, *Associate Member, IEEE*

Abstract—In this paper we present an iterative inversion (II) approach to blind source separation (BSS). It consists of a quasi-Newton method for the resolution of an estimating equation obtained from the implicit inversion of a robust estimate of the mixing system. The resulting learning rule includes several existing algorithms for BSS as particular cases giving them a novel and unified interpretation. It also provides a justification of the Cardoso and Laheld step size normalization [12]. The II method is first presented for instantaneous mixtures and then extended to the problem of blind separation of convolutive mixtures. Finally, we derive the necessary and sufficient asymptotic stability conditions for both the instantaneous and convolutive methods to converge.

Index Terms—Adaptive signal processing, blind source separation (BSS), equivariant learning algorithms, higher order statistics, independent component analysis (ICA), multichannel blind deconvolution (MBD).

I. INTRODUCTION

BLIND source separation (BSS) is the problem of estimating unobserved signals (sources) from observations (sensor signals) of linear mixtures of them. It is an important problem in signal processing because it can be solved using a minimum amount of prior information [19], namely that the mixing system is invertible and that the sources are non-Gaussian and mutually independent. It is thus useful in a large number of extremely diverse applications such as array processing, multiuser communications, voice restoration and biomedical engineering [37], [38], [29]. In addition, BSS plays an important role in neural network theory because separation is carried out with a linear system that can be interpreted as the synaptic weights of a single layer neural network. Thus, it is recognized that adaptive algorithms for BSS adequately work as unsupervised learning rules for neural networks.

The BSS separation problem for noiseless situations is typically formulated as follows. Let us assume that an

array of sensors provides a vector of N observed signals $\mathbf{x}[k] = [x_1[k], x_2[k], \dots, x_N[k]]^T$ that are linear mixtures of N unobserved random processes $s_i[k]$, $i = 1, 2, \dots, N$ termed sources. The exact probability density function (pdf) of the sources is unknown. We will only assume that they are real-valued (although we will later relax this assumption), zero-mean, non-Gaussian distributed and mutually independent. Additionally, we will assume that they are temporally independent and identically distributed (i.i.d.) and spatially independent. When considering convolutive mixtures, the observations are related to the sources as follows:

$$\mathbf{x}[k] = \mathbf{H}[k] * \mathbf{s}[k] = \sum_{m=-\infty}^{\infty} \mathbf{H}[m] \mathbf{s}[k-m] \quad (1)$$

where $\mathbf{s}[k] = [s_1[k], s_2[k], \dots, s_N[k]]^T$ is the vector of sources and $\mathbf{H}[k]$ is the sequence of $N \times N$ impulse response matrices corresponding to the mixing system. To recover the sources, the observations are processed by a linear multiple input–multiple output (MIMO) system with memory to produce the outputs

$$\mathbf{y}[k] = \mathbf{W}[k] * \mathbf{x}[k] = \sum_{m=-l_w}^{l_w} \mathbf{W}[m] \mathbf{s}[k-m] \quad (2)$$

where $\mathbf{W}[k]$ is the sequence of $N \times N$ finite impulse response matrices corresponding to the separating system of length $L_w = 2l_w + 1$. We will denote by $\mathbf{G}[k] = \mathbf{W}[k] * \mathbf{H}[k]$ to the matrix impulse response of the overall mixing (convolutive) and separating (deconvolutive) system. The mixing and demixing models are considerably simplified when instantaneous mixtures are considered. In this case, $\mathbf{H}[k] = \mathbf{H}[0] \delta[k]$ and $\mathbf{W}[k] = \mathbf{W}[0] \delta[k]$, and equations (1) and (2) reduce to¹

$$\mathbf{x} = \mathbf{H} \mathbf{s} \quad (3)$$

$$\mathbf{y} = \mathbf{W} \mathbf{x}. \quad (4)$$

The aim in BSS is to estimate the demixing and deconvolutive matrices $\mathbf{W}[k]$ such that each output of the separating system retrieves an original source signal [37], [11].

Let us start by considering the BSS of instantaneous mixtures. Since the pioneering work of Jutten and Herault [23], a lot of efficient and robust adaptive algorithms have been proposed and their properties investigated. Several authors [13], [39], [6] have interpreted this problem as the estimation of the separation matrix \mathbf{H}^{-1} from the observed data \mathbf{x} . In order to perform this es-

¹Indexes with respect to k are omitted in the instantaneous case to simplify notation.

Manuscript received February 11, 1999; revised November 29, 1999 and June 14, 2000. Part of this work is the result of a stay at the Brain Science Institute (RIKEN), supported by the Frontier Research Program of Japan. This work was also supported by CICYT (Grant TIC96-0500-C10-08) and FEDER funds (Grant 1FD97-0082).

S. Cruces-Alvarez is with the Signal Processing Group, Escuela de Ingenieros, University of Seville, 41092 Seville, Spain (e-mail: sergio@viento.us.es).

A. Cichocki is with the Laboratory for Open Information Systems, Brain Science Institute, RIKEN, Japan (e-mail: cia@brain.riken.go.jp); he is currently on leave from the Warsaw University of Technology, Warsaw, Poland.

L. Castedo-Ribas is with the Department of Electronic and Systems, University of La Coruña, Spain (e-mail: luis@des.fi.udc.es).

Publisher Item Identifier S 1045-9227(00)09854-4.

timization, different approaches have been proposed: minimizing the mutual information of the outputs (MI criterion [19]), maximizing the likelihood of the observations (ML criterion [34]) or maximizing the information transfer of a neural network (INFOMAX criterion [7]). The cost functions $\Psi(\mathbf{Y})$ associated to the previous criteria [13], [39] are obtained in terms of the Kullback–Leibler divergence $D(\cdot\|\cdot)$ between the joint probability density function (pdf) of the outputs, $p_{\mathbf{Y}}(\mathbf{y})$, and another distribution with factorized components, $p_{\mathbf{Z}}(\mathbf{z}) = \prod_{i=1}^N p_{Z_i}(z_i)$, i.e.,

$$\Psi(\mathbf{Y}) = -D\left(p_{\mathbf{Y}}(\mathbf{y}) \left\| \prod_{i=1}^N p_{Z_i}(z_i) \right.\right) \quad (5)$$

In the MI criterion $p_{Z_i}(z_i)$ is the marginal pdf of the outputs, $p_{Y_i}(y_i)$, whereas in the ML and INFOMAX criteria is the marginal pdf of the sources, $p_{S_i}(s_i)$. A necessary condition to obtain the estimate of the separation system \mathbf{W}_* is [6]

$$\left. \frac{\partial \Psi(\mathbf{Y})}{\partial \mathbf{W}} \right|_{\mathbf{W}=\mathbf{W}_*} = (\mathbf{R}_{\psi\mathbf{x}} - \mathbf{W}^{-T})|_{\mathbf{W}=\mathbf{W}_*} = 0 \quad (6)$$

where $\psi(\cdot) = [- (d \log p_{Z_1}(z_1)/dz_1), \dots, - (d \log p_{Z_N}(z_N)/dz_N)]^T$ is a component-wise nonlinear function and $\mathbf{R}_{\psi\mathbf{x}} = E[\psi(\mathbf{y})\mathbf{x}^T]$ is the correlation matrix between a nonlinear function of the outputs $\psi(\mathbf{y})$ and the input \mathbf{x} .

However, the exact knowledge of $\psi(\cdot)$ is not available in many situations and we should replace this desired function by an estimate $\hat{\psi}(\cdot)$. As a consequence, the contrast $\Psi(\cdot)$ should be replaced by the approximate contrast $\hat{\Psi}(\cdot)$. This is the case when the marginal pdf of the sources is approximated by the truncated Edgeworth [19] or Gram–Charlier [39] expansion in the MI criterion or when the *a priori* model for the pdf of the sources is not accurate [13] in the ML and INFOMAX criteria. Due to these mismatches, the maximization of the approximate contrast by means of gradient approaches such as the ordinary gradient or the natural gradient (NG) may fail to converge to the separation solution because the mismatch can make this solution not to be a maximum of $\hat{\Psi}(\cdot)$. Since the unknown function $\psi(\cdot)$ is infinity dimensional we face up a semiparametric estimation problem that is extremely difficult to solve. In order to circumvent this difficulty, Amari and Cardoso [4] suggest to use the method of estimating functions and estimating equations.

In this paper we will follow this latter approach and obtain a special class of estimating equations from the implicit inversion of an approximate estimate of the mixing system. We will show that under certain conditions the iterative inversion of the estimate leads to the blind separation solution. We will also propose two quasi-Newton algorithms that perform the iterative inversion of the estimate. These two algorithm provide us a unified perspective to BSS: different adaptive algorithms for BSS can be interpreted as particular cases of the resultant iterative inversion algorithms. This enables us to explain their convergence behavior and derive a specific normalized step size that yields to high convergence speed and numerical stability. Finally, we will present a straightforward and simple extension of the algorithms to the convolutive case.

This paper is structured as follows. Section II presents the new guiding principle to perform BSS that will be termed the iterative inversion (II) method. Section III presents some results referred to the local stability of the resulting II algorithms. Section IV analyzes the step size normalization to achieve a fast convergence. Section V shows how several existing algorithms for BSS can be interpreted as II algorithms. Section VI extends the method and the stability results to the convolutive mixtures case. Section VII presents the results of several computer simulations and, finally, Section VIII is devoted to the conclusions.

II. THE ITERATIVE INVERSION APPROACH FOR INSTANTANEOUS MIXTURES

Let us consider again the necessary condition for BSS given by (6). This condition can be rewritten in the form of the following estimating equation:

$$\mathbf{W}_* = \mathbf{R}_{\hat{\psi}}^{-1}(\mathbf{W}_*). \quad (7)$$

It is interesting to see from this equation that when the model is correct ($\hat{\psi}(\cdot) = \psi(\cdot)$) the nonlinear correlation function $\mathbf{R}_{\psi\psi}(\mathbf{W}_*)$ can be interpreted as an estimate of the mixing system, \mathbf{H} . Note also that when the model is not correct ($\hat{\psi} \neq \psi$) the implicit equation (7) still gives a valid criteria for BSS because it aims the diagonalization of a nonlinear correlation matrix, i.e., $\mathbf{R}_{\psi\psi}(\mathbf{W}_*) = \mathbf{I}$. This only occurs when the outputs are mutually independent and, therefore, when the sources are recovered.

Taking into account that statistical independence yields source separation, we suggest the following estimating equation to solve the BSS problem

$$\mathbf{W}_* = \hat{\mathbf{H}}^{-1}(\mathbf{W}_*). \quad (8)$$

This equation can be interpreted as a generalization of (7) and represents the implicit inversion of the mixture system estimate $\hat{\mathbf{H}}(\mathbf{W})$. Let us define, now, two functions of the outputs that act component-wise on their elements $\mathbf{f}(\mathbf{y})$ and $\mathbf{g}(\mathbf{y})$, and the following implicit function $\mathbf{F}(\mathbf{y}) = \mathbf{W}^{-1}\mathbf{f}(\mathbf{y})$. We also suggest that the estimate of the mixing system be given by the nonlinear correlation matrix

$$\hat{\mathbf{H}}(\mathbf{W}) = \mathbf{R}_{Fg}(\mathbf{W}) = E[\mathbf{F}(\mathbf{y})\mathbf{g}^T(\mathbf{y})]. \quad (9)$$

Thus, the estimating equation (9) reduces to

$$\mathbf{W}_* \hat{\mathbf{H}}(\mathbf{W}_*) = \mathbf{R}_{fg}(\mathbf{W}_*) = \mathbf{I}. \quad (10)$$

It is apparent from this equation that (9) is a valid estimator of the mixing system because the diagonalization of \mathbf{R}_{fg} is only attained at separation.

A. First Iterative Inversion Algorithm

Unfortunately, the implicit inversion (8) cannot be carried out directly since we do not have access to the estimate $\hat{\mathbf{H}}(\mathbf{W})$ at the separating solution \mathbf{W}_* . Instead, we consider an special class of estimates denoted as *robust estimates*. Let \mathbf{W}^{-1} be the actual estimate of the mixing system. We will define *robust estimates* of the mixing system as those functions $\hat{\mathbf{H}}(\mathbf{W})$ that always give a good reference to the update of \mathbf{W}^{-1} in the direction of the

mixing system (\mathbf{H}) and that at the separation solution coincide exactly with it.

An interesting property of the *robust estimates* is that the implicit inversion (8) can be solved by means of an iterative inversion procedure similar to that used by Bussgang methods for blind deconvolution [8]. Here iterative inversion means the iterative correction of the current estimate of the mixing system \mathbf{W}^{-1} in the direction of the robust estimate $\hat{\mathbf{H}}(\mathbf{W})$ (see Fig. 1), i.e.,

$$\mathbf{W}^{-1(n+1)} = \mathbf{W}^{-1(n)} + \eta \left(\hat{\mathbf{H}}^{(n)} - \mathbf{W}^{-1(n+1)} \right). \quad (11)$$

Defining $\mu = \eta/(1 + \eta)$ we can rewrite the above implicit iteration in an explicit form as

$$\mathbf{W}^{-1(n+1)} = (1 - \mu)\mathbf{W}^{-1(n)} + \mu\hat{\mathbf{H}}^{(n)}. \quad (12)$$

Note from (12) that $\mathbf{W}^{-1(n+1)}$ can be interpreted as an estimate of $\hat{\mathbf{H}}^{(n)}$ with an exponential window. Moreover, it is also possible to obtain this iteration from the application of a quasi-Newton method to find the zeros of the estimating function $\mathcal{F}(\mathbf{W}^{-1}) = \hat{\mathbf{H}}(\mathbf{W}) - \mathbf{W}^{-1}$ (see Appendix I). Rewriting (12) in terms of the separation system we obtain

$$\mathbf{W}^{(n+1)} = \left(\mathbf{I} + \mu \left(\mathbf{W}^{(n)} \hat{\mathbf{H}}^{(n)} - \mathbf{I} \right) \right)^{-1} \mathbf{W}^{(n)} \quad (13)$$

where $\hat{\mathbf{H}}^{(n)} = \mathbf{R}_{fg}(\mathbf{W}^{(n)})$. We can avoid the calculation of the implicit function $\mathbf{F}(\mathbf{y})$ if we take into account that $\mathbf{W}^{(n)} \hat{\mathbf{H}}^{(n)} = \mathbf{R}_{fg}^{(n)}$ and rewrite (13) as

$$\mathbf{W}^{(n+1)} = \left(\mathbf{I} + \mu \left(\mathbf{R}_{fg}^{(n)} - \mathbf{I} \right) \right)^{-1} \mathbf{W}^{(n)}. \quad (14)$$

This recursion will be termed *first iterative inversion* (II1) algorithm. Note that when the estimate $\hat{\mathbf{H}}^{(n)}$ is accurate this algorithm is of the quasi-Newton type and, therefore, exhibits a very fast convergence.

Let us consider now the selection of the step-size. It is desirable for derivative-based optimization methods to restrict the algorithm to work in a domain where the function $\mathcal{F}(\mathbf{W}^{-1})$ and its derivative are continuous. Assuming that $\hat{\mathbf{H}}^{(n)}$ is a sufficiently smooth function, the only discontinuities occur at the points where the separation matrix \mathbf{W} is singular. Therefore, constraining the algorithm to work in a closed region that avoids the singularity of the separating matrix, \mathbf{W} , discontinuities on the estimating function will not occur. A sufficient condition to ensure that discontinuities are not crossed when updating $\mathbf{W}^{(n)}$ to $\mathbf{W}^{(n+1)}$ is that the step-size $\mu^{(n)}$ satisfies

$$\mu^{(n)} < \frac{1}{\left\| \mathbf{R}_{fg}^{(n)} - \mathbf{I} \right\|_p} \quad (15)$$

where $\|\cdot\|_p$ denotes the p-norm of a matrix. This way we prevent matrix $\mathbf{I} + \mu(\mathbf{R}_{fg}^{(n)} - \mathbf{I})$ from being singular and guarantee its invertibility.

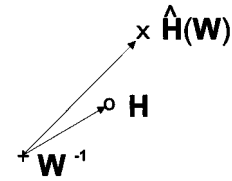


Fig. 1. Direction of update given by the robust estimate of the mixing system compared with the optimal update.

B. Second Iterative Inversion Algorithm

When condition (15) is satisfied we can express the inverse in (14) as an infinite expansion [22] and make the following approximation:

$$\begin{aligned} \left(\mathbf{I} + \mu^{(n)} \left(\mathbf{R}_{fg}^{(n)} - \mathbf{I} \right) \right)^{-1} &= \sum_{i=0}^{\infty} \left(-\mu^{(n)} \left(\mathbf{R}_{fg}^{(n)} - \mathbf{I} \right) \right)^i \\ &\approx \mathbf{I} - \mu^{(n)} \left(\mathbf{R}_{fg}^{(n)} - \mathbf{I} \right). \end{aligned} \quad (16)$$

Finally, substituting (16) into (14) we arrive at the *second iterative inversion* (II2) algorithm

$$\mathbf{W}^{(n+1)} = \mathbf{W}^{(n)} - \mu^{(n)} \left(\mathbf{W}^{(n)} \hat{\mathbf{H}}^{(n)} - \mathbf{I} \right) \mathbf{W}^{(n)} \quad (17)$$

that can also be interpreted as an iterative version of a classical quasi-Newton algorithm for the inversion of $\hat{\mathbf{H}}^{(n)}$ (see [32], [28] for more details). It is straightforward to show that the II1 and II2 algorithms exhibit the equivariant property [12] in the absence of noise.

The most crucial question now is to determine what pairs of functions $\mathbf{f}(\mathbf{y})$ and $\mathbf{g}(\mathbf{y})$ provide robust estimates of the mixing system. In the next section we perform an analysis of the local stability of the II algorithms and obtain an answer to this question: the functions that verify the asymptotic stability conditions for the algorithms to converge provide local robust estimates.

III. STABILITY ANALYSIS

In this section we perform a local stability analysis of the II1 and II2 algorithms. We will obtain the necessary and sufficient conditions that the nonlinearities must satisfy in order to ensure that the separating solutions are stable points.

Let us adopt the following notation $f_i = [\mathbf{f}(\mathbf{s})]_i$, $f'_i = [\partial \mathbf{f}(\mathbf{s}) / \partial \mathbf{s}]_{ii}$, $g_i = [\mathbf{g}(\mathbf{s})]_i$, $g'_i = [\partial \mathbf{g}(\mathbf{s}) / \partial \mathbf{s}]_{ii}$. We will assume that the sources \mathbf{s} are properly scaled and the functions $\mathbf{f}(\mathbf{s})$ and $\mathbf{g}(\mathbf{s})$ adequately chosen in order that the conditions for the separation solution be a stationary point, i.e.,

$$E[f_i g_i] = 1 \quad \text{and} \quad E[f_i g_j] = 0 \quad (18)$$

are satisfied for all $i, j |_{i \neq j} = 1, \dots, N$.

Theorem 1: Assuming that

- 1) \mathbf{s} is a vector of real and zero mean sources.
- 2) The functions $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ are component-wise and differentiable in the domain of support.
- 3) The step-size μ is small enough.

the II2 algorithm with a constant step-size

$$\mathbf{W}^{(n+1)} = \mathbf{W}^{(n)} - \mu \left(\mathbf{R}_{fg}^{(n)} - \mathbf{I} \right) \mathbf{W}^{(n)} \quad (19)$$

contains an asymptotically stable point at the separation solution if and only if the following conditions are satisfied:

$$E[f'_i s_i g_i] + E[f_i s_i g'_i] > 0 \quad (20)$$

$$E[f'_i]E[s_j g_j] + E[f'_j]E[s_i g_i] > 0 \quad (21)$$

$$E[f'_i]E[f'_j]E[s_i g_i]E[s_j g_j] > E[g'_i]E[g'_j]E[s_i f_i]E[s_j f_j] \quad (22)$$

for all $i, j |_{i \neq j} = 1, \dots, N$.

This theorem is proved in Appendix III.

The previous conditions are a generalization and extension of the following stability conditions derived by Amari *et al.* in [5].

Corollary 1: If the second function is linear, i.e., $\mathbf{g}(\mathbf{y}) = \mathbf{y}$ and assuming that the sources are properly scaled so that

$$E[f_i s_i] = 1 \quad (23)$$

for all $i = 1, \dots, N$, the asymptotic stability conditions reduce to

$$E[f'_i s_i^2] + 1 > 0 \quad (24)$$

$$E[f'_i] > 0 \quad (25)$$

$$E[f'_i]E[f'_j]E[s_i^2]E[s_j^2] > 1 \quad (26)$$

for all $i, j |_{i \neq j} = 1, \dots, N$.

Proof: Note that when $\mathbf{g}(\mathbf{y}) = \mathbf{y}$ the constraint $E[f_i g_j] = 0, \forall i \neq j$ is always true at the correct equilibrium point since, due to the independence and zero mean assumptions for the sources, $E[f_i g_j] = E[f_i]E[g_j] = E[f_i]E[s_j] = 0 \forall i \neq j$. The conditions (24) and (26) are straightforwardly obtained from (20) and (22) taking into account (23) and the fact that $g_i = s_i$ and $g'_i = 1$ for all $i = 1, \dots, N$. On the other hand, condition (21) is rewritten as

$$E[f'_i]E[s_j^2] + E[f'_j]E[s_i^2] > 0. \quad (27)$$

Taking into account (26), it is easy to see that we need that $\text{sign}(E[f'_i]) = \text{sign}(E[f'_j]) \forall i, j$ and, therefore, condition (27) simplifies to (25). ■

However, from Theorem 1 we can see that a different result is obtained when the first function is linear.

Corollary 2: If the first function is linear, i.e., $\mathbf{f}(\mathbf{y}) = \mathbf{y}$, and assuming that the sources are properly scaled so that

$$E[s_i g_i] = 1 \quad (28)$$

for all $i = 1, \dots, N$, the asymptotic stability conditions reduce to

$$1 + E[s_i^2 g'_i] > 0 \quad (29)$$

$$E[g'_i]E[g'_j]E[s_i^2]E[s_j^2] < 1 \quad (30)$$

for all $i, j |_{i \neq j} = 1, \dots, N$.

Proof: This corollary is proved by substituting $E[s_i g_i] = 1$ and $E[f'_i] = 1$ in (20)–(22). ■

An interesting consequence of this new corollary is that those functions $\mathbf{g}(\mathbf{y})$ whose derivative is continuous and almost null

on the whole support domain of the sources guarantee the stability of the algorithm. As an example of application, if we assume that the sources never take zero value² and choose the nonlinearity as the signum function $g(y_i) = \text{sign}(y_i)$ we guarantee, in the vicinity of the separation, that $E[s_i^2 g'_i]$ and $E[g'_i]$ are both null and, therefore, the asymptotic stability conditions (29) and (30) are always true. Note also that in this case we will have that \mathbf{R}_{Fg} is a robust estimate for the mixing system. Since $\mathbf{F}(\mathbf{y}) = \mathbf{x}$ and $\mathbf{g}(\mathbf{y})$ is the signum function, \mathbf{R}_{Fg} will be constant in a wide domain around the separation solution leading to a quadratic convergence of the II1 and II2 algorithms.

When the sources have even pdf and the nonlinearities are strictly monotonically increasing odd functions, (20) and (21) are always satisfied. These assumptions, however, do not guarantee that condition (22) is true. In this case, the following corollary shows how to stabilize the algorithm.

Corollary 3: If

$$E[f'_i]E[f'_j]E[s_i g_i]E[s_j g_j] < E[g'_i]E[g'_j]E[s_i f_i]E[s_j f_j] \quad (31)$$

an interchange of $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ ensures the asymptotical stability of the II2 algorithm.

Proof: If the pdf of the sources is even and the nonlinearities strictly odd, $E[\mathbf{f}(\mathbf{s})] = \mathbf{0}$ and $E[\mathbf{g}(\mathbf{s})] = \mathbf{0}$. On the other hand, when the nonlinearities are strictly monotonically increasing odd functions, $E[f'_i s_i g_i] > 0$ and $E[f_i s_i g'_i] > 0 \forall i$, and condition (20) is true. For the same reasons, $E[f'_i] > 0$ and $E[s_i g_i] > 0 \forall i$ and (21) is also true. Finally, condition (22), when the equality does not hold, is not as critical since it can be forced by just permuting the order of the functions $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$. ■

This property has been previously observed in [36] for the Herault–Jutten algorithm and later in [5] for the special case of $\mathbf{g}(\mathbf{y}) = \mathbf{y}$. Our results, however, consider the general case of two nonlinear functions.

Finally, the next corollary extends the previous results to the II1 algorithm.

Corollary 4: The II1 algorithm

$$\mathbf{W}^{-1(n+1)} = \mathbf{W}^{-1(n)} \left(\mathbf{I} + \mu \left(\mathbf{R}_{fg}^{(n)} - \mathbf{I} \right) \right) \quad (32)$$

has the same asymptotical stability conditions given by (20)–(22) as the II2 algorithm.

Proof: In order to extend the previous results to the II1 algorithm we will show that the local behavior of both algorithms is equivalent. Let $\mathbf{W}_1^{(n+1)}$ and $\mathbf{W}_2^{(n+1)}$ be the result of applying the II1 and II2 algorithms, respectively, to the same initial value $\mathbf{W}^{(n)}$. It is apparent that if both algorithms were identical, the product $\mathbf{W}_2^{(n+1)} \mathbf{W}_1^{-1(n+1)}$ would be the identity matrix. This will not be the case since, due to the approximation in (16), algorithms II1 and II2 are different. However, it is straightforward to show that the following equality holds true:

$$\left\| \mathbf{W}_2^{(n+1)} \mathbf{W}_1^{-1(n+1)} - \mathbf{I} \right\|_p = \mu^2 \left\| \mathbf{R}_{fg}^{(n)} - \mathbf{I} \right\|_p^2. \quad (33)$$

²This hypothesis is valid for several communications signals where the symbols do not lie on the principal axis of the constellation.

This result means that both algorithms perform the same when μ is small or $\mathbf{R}_{fg}^{(n)}$ is very close to the identity matrix, i.e., the algorithms are close to convergence.

Moreover, a general bound in the difference of behavior can be obtained. Since we have hypothesized that $\mu^{(n)} < 1/\|\mathbf{R}_{fg}^{(n)} - \mathbf{I}\|_p$ then

$$\left\| \mathbf{W}_2^{(n+1)} \mathbf{W}_1^{-1(n+1)} - \mathbf{I} \right\|_p < 1$$

is an upper bound of the difference between both algorithms. ■

IV. PRACTICAL CONSIDERATIONS AND ALGORITHMS NORMALIZATION

The nonlinear autocorrelation matrix $\mathbf{R}_{fg}^{(n)}$ used by the II algorithms is a statistical average that is not known in practical applications but can be estimated from the incoming data. There are two possibilities: on-line adaptations in which the correlation matrix $\mathbf{R}_{fg}^{(n)}$ is replaced by its single sample estimate, i.e.,

$$\mathbf{R}_{fg}^{(n)} \approx \mathbf{f}(\mathbf{y}[n])\mathbf{g}^T(\mathbf{y}[n]) \quad (34)$$

and batch adaptations where, if we assume that the stationarity and ergodicity properties hold on a block of L observations, we can replace the statistical average by the temporal average

$$\mathbf{R}_{fg}^{(n)} \approx \frac{1}{L} \sum_{k=0}^{L-1} \mathbf{f}(\mathbf{y}[k])\mathbf{g}^T(\mathbf{y}[k]). \quad (35)$$

Another major practical issue is the selection of the algorithm step-size in order to ensure both numerical stability and high convergence speed. Whenever $\hat{\mathbf{H}}^{(n)} = \mathbf{R}_{Fg}^{(n)}$ is an approximately constant estimate of the mixing system, the III algorithm exhibits the fastest convergence whereas the II2 algorithm also converges superlinearly. This fact explains why the III algorithm performs better than the II2 algorithm for robust estimates. In an adaptive filtering context, however, the direct implementation of the III algorithm is more complex than that of the II2 algorithm as a consequence of the matrix inversion. Notwithstanding, we will show in the sequel that we can indirectly implement the III algorithm by means of an iteration of the II2 type together with a proper step-size normalization.

Let us start considering the on-line version of the III algorithm with the one sample estimates

$$\mathbf{W}^{(n+1)} = (\mathbf{I} + \mu (\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y}) - \mathbf{I}))^{-1} \mathbf{W}^{(n)}. \quad (36)$$

Let us rewrite again (36) in terms of $\eta = \mu/(1 - \mu)$ and apply the Sherman–Morrison inversion formula [35] to obtain

$$\mathbf{W}^{(n+1)} = (1 + \eta) \left(\mathbf{I} - \eta \frac{\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y})}{1 + \eta\mathbf{g}^T(\mathbf{y})\mathbf{f}(\mathbf{y})} \right) \mathbf{W}^{(n)}.$$

When implementing on-line algorithms it is a common practice to choose $\eta \ll 1$ since the single sample estimates of $\mathbf{R}_{fg}^{(n)}$ introduce a lot of misadjustment noise into the algorithm. As a consequence, we can neglect the η^2 and higher order terms and arrive at

$$\mathbf{W}^{(n+1)} = \mathbf{W}^{(n)} - \left(\eta \frac{\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y}) - \mathbf{I}}{1 + \eta\mathbf{g}^T(\mathbf{y})\mathbf{f}(\mathbf{y})} \right) \mathbf{W}^{(n)}. \quad (37)$$

To ensure the algorithm works in a continuous region we must satisfy condition (15). Accordingly we should modify the step-size and use

$$\mu^{(n)} = \frac{\eta}{1 + \eta|\mathbf{g}^T(\mathbf{y})\mathbf{f}(\mathbf{y})|} \quad (38)$$

where $0 < \eta < 1$. Thus, the on-line stochastic normalized implementation of the III algorithm is

$$\mathbf{W}^{(n+1)} = \mathbf{W}^{(n)} - \left(\eta \frac{\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y}) - \mathbf{I}}{1 + \eta|\mathbf{g}^T(\mathbf{y})\mathbf{f}(\mathbf{y})|} \right) \mathbf{W}^{(n)}. \quad (39)$$

It is interesting to note that this recursion can be interpreted as the II2 algorithm with the variable step-size normalization $\mu^{(n)} = \eta/(1 + \eta|\mathbf{g}^T(\mathbf{y})\mathbf{f}(\mathbf{y})|)$. This is the normalization suggested by Cardoso *et al.* in [12] for the EASI algorithm. However, we obtained it from the implementation of the III algorithm and not from the II2 algorithm as it would appear at a first glance. Since the convergence behavior of the III algorithm is better than that of the II2 algorithm when accurate estimates for the mixing system are used we obtain a first theoretical argument that justifies the practical efficiency and fast convergence that it is achieved with this normalization.

For the batch adaptation we propose the following step size which is a generalization of (38):

$$\mu^{(n)} = \frac{\eta}{1 + \eta \left\| \mathbf{R}_{fg}^{(n)} \right\|_p} \quad (40)$$

where $0 < \eta < 1$. Note that when $p = 2$ and $\mathbf{R}_{fg}^{(n)} \approx \mathbf{f}(\mathbf{y}[n])\mathbf{g}^T(\mathbf{y}[n])$ (40) reduces to (38). Taking into account that $\left\| \mathbf{R}_{fg}^{(n)} - \mathbf{I} \right\|_p \leq 1 + \left\| \mathbf{R}_{fg}^{(n)} \right\|_p$ it is easy to show that the batch step size (40) always verifies condition (15) for any given p -norm. In practical implementations of the batch algorithm we will typically choose η close to unity since we are interested in the algorithm to converge as fast as possible. The selection of the norm in (40) depends on practical considerations. In Appendix II we demonstrate that the 2-norm provides the fastest convergence for a fixed η . Other norms such as the 1-norm or the ∞ -norm are easier to evaluate and therefore lead to more robust step-sizes. Nevertheless, we can replace the 2-norm by coarse estimates, such as $\|\mathbf{R}_{fg}^{(n)}\|_2 \approx \max\{|\text{diag}\{\mathbf{R}_{fg}^{(n)}\}|\}$, whose accuracy improves as we approach to convergence.

V. RELATION WITH EXISTING ALGORITHMS

In this section we show how several existing algorithms for BSS can be derived as particular cases of the on-line implementation of the III algorithm given by (39). This enables us to interpret them as quasi-Newton algorithms that iteratively invert a nonlinear autocorrelation matrix.

A. Generalized Learning Rule

The algorithm (39) is the generalized learning rule derived by Cichocki *et al.* [16], [17] from an heuristic principle. Here, we have presented an alternative formal derivation and proposed the normalized step-size (38) to considerably improve its convergence behavior.

B. Decorrelation Algorithm

Choosing $\mathbf{f}(\mathbf{y}) = \mathbf{g}(\mathbf{y}) = \mathbf{y}$, the adaptation rule (39) reduces to the normalized decorrelation algorithm proposed by Almeida [2] (see also [21]).

$$\mathbf{W}^{(n+1)} = \mathbf{W}^{(n)} - \eta \frac{\mathbf{y}\mathbf{y}^T - \mathbf{I}}{1 + \eta\mathbf{y}^T\mathbf{y}} \mathbf{W}^{(n)}. \quad (41)$$

C. Natural Gradient or EASI Algorithm

Selecting $\mathbf{g}(\mathbf{y}) = \mathbf{y}$ and $\mathbf{f}(\mathbf{y})$ as an arbitrary nonlinear function, the III algorithm (39) converts into

$$\mathbf{W}^{(n+1)} = \mathbf{W}^{(n)} - \eta \frac{\mathbf{f}(\mathbf{y})\mathbf{y}^T - \mathbf{I}}{1 + \eta|\mathbf{y}^T\mathbf{f}(\mathbf{y})|} \mathbf{W}^{(n)} \quad (42)$$

which is a normalized version of the natural gradient or EASI algorithm independently developed by Amari *et al.* [6] and Cardoso and Laheld [12]. When this algorithm does not converge, Amari *et al.* [5] have suggested that the nonlinearities be interchanged [i.e., $\mathbf{f}(\mathbf{y}) = \mathbf{y}$ and $\mathbf{g}(\mathbf{y})$]. This interchange leads to an algorithm that is not of the natural gradient type because it does not follow the descent lines of a cost function. Note, however, that it can be interpreted as an iterative inversion algorithm.

It has been proved by Amari and Cardoso [4] that the most asymptotically efficient form of algorithm (39) is obtained when one of the two functions, $\mathbf{f}(\mathbf{y})$ or $\mathbf{g}(\mathbf{y})$, is linear. But, unfortunately, this result only characterizes the local behavior so it does not exclude the possibility of using two nonlinearities and obtain better global convergence properties. In fact, in this article we will present some exemplary cases where algorithms with two nonlinearities exhibit faster convergence. Another argument in favor of using two nonlinearities is the robustness to the noise present in the mixture. It has been shown in [18] that two nonlinearities usually lead to a lower bias in the estimation.

D. Generalized EASI Algorithm

The generalized EASI algorithm [24] is an extension of the EASI algorithm [12] that considers two nonlinearities instead of a single one. The generalized EASI algorithm is known to perform adequately in practice [24] but has not been interpreted from an information theoretic perspective yet. In the sequel, we will derive this algorithm from the III recursion following a procedure similar to that described in [12]. Let us start decomposing the separating system in two stages, i.e., $\mathbf{W} = \mathbf{W}_b\mathbf{W}_a$. The first matrix \mathbf{W}_a will be selected to decorrelate the input, i.e., to diagonalize the symmetric matrix \mathbf{R}_{yy} . This can be done with the decorrelation algorithm (41). On the other hand, the second matrix \mathbf{W}_b will be selected to diagonalize the matrix \mathbf{R}_{fg} . This objective can be achieved using the generalized learning rule (39). However, in order to integrate both adaptations into a single recursion for the overall separating system, \mathbf{W} , we need that they be orthogonal at a first order in η [12]. This can be accomplished if we replace $\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y}) - \mathbf{I}$ by its projection onto the space of skew-symmetric matrices. Thus, the resulting algorithm is

$$\mathbf{W}_b^{(n+1)} = \left(\mathbf{I} - \frac{\eta}{2} \frac{\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y}) - \mathbf{g}(\mathbf{y})\mathbf{f}^T(\mathbf{y})}{1 + \eta|\mathbf{g}^T(\mathbf{y})\mathbf{f}(\mathbf{y})|} \right) \mathbf{W}_b^{(n)}. \quad (43)$$

Combining the decorrelation adaptation (41) and (43) so that $\mathbf{W}^{(n+1)} = \mathbf{W}_b^{(n+1)}\mathbf{W}_a^{(n+1)}$ and neglecting the high-order terms in η we arrive at

$$\mathbf{W}^{(n+1)} = \left(\mathbf{I} - \eta \frac{\mathbf{y}\mathbf{y}^T - \mathbf{I}}{1 + \eta\mathbf{y}^T\mathbf{y}} - \frac{\eta}{2} \frac{\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y}) - \mathbf{g}(\mathbf{y})\mathbf{f}^T(\mathbf{y})}{1 + \eta|\mathbf{g}^T(\mathbf{y})\mathbf{f}(\mathbf{y})|} \right) \cdot \mathbf{W}^{(n)}$$

which is the normalized version of the generalized EASI algorithm.

E. Herault–Jutten algorithm

It is also possible to obtain the classical Herault–Jutten (HJ) algorithm from the III algorithm. This is done under some simple hypothesis for which the HJ algorithm is known to converge properly. We will consider only two sources and that the mixing system is soft in the sense that there is a direct component from the sources to the observations that is stronger than that of the interfering signals. This is the case when each sensor is placed close to a different source.

Let us define the diagonal operator \mathbf{D}_M that keeps the diagonal part of a matrix \mathbf{M} . Similarly, we define the off-diagonal operator \mathbf{O}_M that keeps the off-diagonal terms. Let us assume also that at a given iteration the inverse of the separating system can be decomposed as $\mathbf{W}^{-1(n)} = (\mathbf{I} + \mathbf{C}^{(n)})$ where $\mathbf{C}^{(n)}$ has zero elements at its diagonal.

Consider $\mathbf{W}^{-1(n+1)}$ as the result of the iteration given by the III algorithm. If we scale this matrix to preserve across the iterations the identity at the main diagonal we find the composite iteration

$$\mathbf{C}^{(n+1)} = \mathbf{O}_{\mathbf{W}^{-1(n+1)}} \mathbf{D}_{\mathbf{W}^{-1(n+1)}}^{-1} \quad (44)$$

where the diagonal and off-diagonal parts of $\mathbf{W}^{-1(n+1)}$ are

$$\begin{aligned} \mathbf{D}_{\mathbf{W}^{-1(n+1)}} &= \mathbf{I} + \mu \left(\mathbf{D}_{\mathbf{R}_{fg}^{(n)}} + \mathbf{D}_{(\mathbf{C}^{(n)}\mathbf{O}_{\mathbf{R}_{fg}^{(n)}})} - \mathbf{I} \right) \quad (45) \\ \mathbf{O}_{\mathbf{W}^{-1(n+1)}} &= \mathbf{C}^{(n)} + \mu \left(\mathbf{C}^{(n)} \left(\mathbf{D}_{\mathbf{R}_{fg}^{(n)}} - \mathbf{I} \right) \right. \\ &\quad \left. + \mathbf{O}_{\mathbf{R}_{fg}^{(n)}} + \mathbf{O}_{(\mathbf{C}^{(n)}\mathbf{O}_{\mathbf{R}_{fg}^{(n)}})} \right). \quad (46) \end{aligned}$$

However, for a sufficiently small step-size, (i.e., $\mu \ll 1$), it is possible to expand $\mathbf{D}_{\mathbf{W}^{-1(n+1)}}^{-1}$ in powers of μ and neglect the higher order terms $o(\mu) \approx 0$ to arrive at the following result:

$$\begin{aligned} \mathbf{C}^{(n+1)} &= \mathbf{C}^{(n)} + \mu \mathbf{O}_{\mathbf{R}_{fg}^{(n)}} \\ &\quad + \mu \left(\mathbf{O} \left(\mathbf{C}^{(n)}\mathbf{O}_{\mathbf{R}_{fg}^{(n)}} \right) - \mathbf{C}^{(n)}\mathbf{D}_{(\mathbf{C}^{(n)}\mathbf{O}_{\mathbf{R}_{fg}^{(n)}})} \right). \quad (47) \end{aligned}$$

Now we can realize that in the two sources case ($N = 2$) the term $\mathbf{O}_{(\mathbf{C}^{(n)}\mathbf{O}_{\mathbf{R}_{fg}^{(n)}})}$ vanishes because $\mathbf{C}^{(n)}\mathbf{O}_{\mathbf{R}_{fg}^{(n)}}$ is diagonal. Therefore, the iteration can be rewritten in scalar form as

$$c_{ij}^{(n+1)} = c_{ij}^{(n)} + \mu \left(1 - c_{ij}^{(n)} c_{ji}^{(n)} \right) \left[\mathbf{R}_{fg}^{(n)} \right]_{ij} \quad (48)$$

for $i, j |_{i \neq j} = 1, 2$. When assuming soft mixtures the following approximation is valid $1 - c_{ij}^{(n)} c_{ji}^{(n)} \approx 1$. Finally, considering the instantaneous stochastic approximation for the correlation matrix we obtain the Herault–Jutten algorithm

$$c_{ij}^{(n+1)} = c_{ij}^{(n)} + \mu f(y_i[n])g(y_j[n]) \quad (49)$$

for $i, j |_{i \neq j} = 1, 2$.

The previous analysis is reinforced by the fact that the local stability conditions of the II1 and HJ algorithms for the off-diagonal terms are equivalent. Indeed, under the given hypothesis, the local stability conditions for the convergence of the HJ algorithm are [15]

$$E[f'_1]E[s_2g_2] + E[f'_2]E[s_1g_1] > 0 \quad (50)$$

$$E[f'_1]E[f'_2]E[s_1g_1]E[s_2g_2] > E[g'_1]E[g'_2]E[s_1f_1]E[s_2f_2]. \quad (51)$$

But these are also the necessary and sufficient asymptotic stability conditions (21) and (22) for the convergence of the off-diagonal terms of \mathbf{W}^{-1} in the II1 algorithm. Note that the additional condition (20) does not apply here because is associated with the convergence of the diagonal terms of \mathbf{W}^{-1} which are always constant in the HJ algorithm.

F. Nonlinear PCA Algorithm

The relationship between the nonlinear PCA algorithm (developed by Oja and Karhunen in [31], [24]) and other approaches, such as those derived from the maximum likelihood criterion, has been first explored in [25]. Here we will show that when the mixing system \mathbf{H} is orthogonal it is also possible to relate the II1 algorithm with the nonlinear PCA algorithm. Note that the orthogonality of the mixing system is not a limitation since is also assumed in the derivation of the nonlinear PCA algorithm. Let us consider the following estimate for the mixing system:

$$\hat{\mathbf{H}}(\mathbf{W}) = \mathbf{W}^{-1} (\mathbf{I} + \mathbf{R}_{ff} - \mathbf{R}_{fs}\mathbf{H}^T\mathbf{W}^{-1}). \quad (52)$$

A necessary condition for (52) to be a robust estimate is that it should exactly attain the value of the mixing system at separation, i.e., $\hat{\mathbf{H}}(\mathbf{W}_0) = \mathbf{H}$. Since at separation \mathbf{R}_{ff} and \mathbf{R}_{fs} are diagonal, $\mathbf{H}^T\mathbf{W}_0^{-1} = \mathbf{H}^T\mathbf{H} = \mathbf{I}$ which is equivalent to the orthogonality of the mixing system. By substituting the stochastic form of this estimate in the II2 algorithm we obtain

$$\mathbf{W}^{(n+1)} = \mathbf{W}^{(n)} + \mu \mathbf{f}(\mathbf{y}) \left(\mathbf{x}^T - \mathbf{f}^T(\mathbf{y})\mathbf{W}^{(n)} \right) \quad (53)$$

which is the nonlinear PCA algorithm. Also, a new form of the nonlinear PCA adaptation with a fast convergence is obtained if we substitute the estimate (52) in the II1 algorithm, i.e.,

$$\mathbf{W}^{(n+1)} = \left(\mathbf{I} + \mu \mathbf{f}(\mathbf{y}) \left(\mathbf{f}^T(\mathbf{y}) - \mathbf{x}^T\mathbf{W}^{-1(n)} \right) \right)^{-1} \mathbf{W}^{(n)}. \quad (54)$$

Note that the stability conditions of the previous section do not apply to (53) and (54) since the estimate given in (52) is not exactly of the form of a nonlinear correlation matrix.

VI. EXTENSION TO CONVOLUTIVE MIXTURES

In this section we extend the previous ideas to the case of convolutive mixtures of temporally i.i.d. and spatially independent sources. The approach consists in exploiting the i.i.d. assumption of the sources in order to asymptotically approach the convolutive model by an instantaneous model of a much larger dimension. We will solve the asymptotically equivalent instantaneous problem and exploiting the special structure of the involved matrices we return to the convolutive notation to rewrite the resulting adaptation.

Since we assume that the sequences of source samples are temporally i.i.d. we can consider that the samples obtained from the sources at different lags ($\mathbf{s}[k]$, $k = 0, \dots, L_c - 1$) can be rearranged in the following $NL_c \times 1$ vector of sources:

$$\bar{\mathbf{s}} = [\mathbf{s}_{L_c-1}^T, \dots, \mathbf{s}_0^T]^T \quad (55)$$

where $\mathbf{s}_k = [s_1[k], s_2[k], \dots, s_N[k]]^T$, $k = 0, \dots, L_c - 1$. Note that all the components in this vector are mutually independent. Analogously, we can rearrange the observations and the outputs as

$$\bar{\mathbf{x}} = [\mathbf{x}_{L_c-1}^T, \dots, \mathbf{x}_0^T]^T \quad (56)$$

$$\bar{\mathbf{y}} = [\mathbf{y}_{L_c-1}^T, \dots, \mathbf{y}_0^T]^T \quad (57)$$

where $\mathbf{x}_k = [x_1[k], x_2[k], \dots, x_N[k]]^T$, $k = 0, \dots, L_c - 1$ and $\mathbf{y}_k = [y_1[k], y_2[k], \dots, y_N[k]]^T$, $k = 0, \dots, L_c - 1$.

Let L_g be the significant length of the impulse response $\mathbf{G}[k] = \mathbf{W}[k] * \mathbf{H}[k]$. In order that L_g be finite we consider that $\mathbf{H}[k]$ is a stable mixing systems with roots distant from the unit circle so as to guarantee a sufficient decaying impulse response. On the other hand, we have considered in (2) that $\mathbf{W}[k]$ is a finite impulse response (FIR) system of length L_w and symmetrically centered around the origin.

Let us define the set of lags $\mathcal{N}_0 = \{k : k \in [-L_g, -1] \cup [L_c, L_c + L_g - 1]\}$ and assume that the sources and the observations satisfy the following conditions $\mathcal{B} = \{\mathbf{s}_k = \mathbf{x}_k = \mathbf{0} \forall k \in \mathcal{N}_0\}$. Under these conditions, the convolutive model described by (1) and (2) reduces to the truncated model

$$\mathbf{x}[k] = \sum_{m=0}^{L_c-1} \mathbf{H}[k-m]\mathbf{s}[m] \quad (58)$$

and

$$\mathbf{y}[k] = \sum_{m=0}^{L_c-1} \mathbf{W}[k-m]\mathbf{x}[m]. \quad (59)$$

These equations can be rewritten in a Sylvester matrix representation [27] as

$$\bar{\mathbf{x}} = \bar{\mathbf{H}}\bar{\mathbf{s}} \quad (60)$$

$$\bar{\mathbf{y}} = \bar{\mathbf{W}}\bar{\mathbf{x}} \quad (61)$$

where

$$\bar{\mathbf{H}} = \begin{bmatrix} [\bar{\mathbf{H}}]_{11} & [\bar{\mathbf{H}}]_{12} & \dots & [\bar{\mathbf{H}}]_{1L_c} \\ [\bar{\mathbf{H}}]_{21} & [\bar{\mathbf{H}}]_{22} & \dots & [\bar{\mathbf{H}}]_{2L_c} \\ \vdots & \vdots & \ddots & \vdots \\ [\bar{\mathbf{H}}]_{L_c 1} & [\bar{\mathbf{H}}]_{L_c 2} & \dots & [\bar{\mathbf{H}}]_{L_c L_c} \end{bmatrix}$$

$$\bar{\mathbf{W}} = \begin{bmatrix} [\bar{\mathbf{W}}]_{11} & [\bar{\mathbf{W}}]_{12} & \dots & [\bar{\mathbf{W}}]_{1L_c} \\ [\bar{\mathbf{W}}]_{21} & [\bar{\mathbf{W}}]_{22} & \dots & [\bar{\mathbf{W}}]_{2L_c} \\ \vdots & \vdots & \ddots & \vdots \\ [\bar{\mathbf{W}}]_{L_c 1} & [\bar{\mathbf{W}}]_{L_c 2} & \dots & [\bar{\mathbf{W}}]_{L_c L_c} \end{bmatrix}. \quad (62)$$

These matrices are square block-Toeplitz of dimension $NL_c \times NL_c$, i.e., each block satisfies the following:

$$[\bar{\mathbf{H}}]_{ij} = \mathbf{H}[j - i] \quad (63)$$

$$[\bar{\mathbf{W}}]_{ij} = \mathbf{W}[j - i]. \quad (64)$$

Having in mind (60) and (61), the separation of a convolutive mixture of N sources can be formulated as the separation of an instantaneous mixture of $N \times L_c$ sources under the boundary condition given by \mathcal{B} . This problem can be solved using the algorithms developed in Section II by substituting the true contrast $\Psi(\bar{\mathbf{Y}})$ presented in (5) by the conditional contrast

$$\Psi(\bar{\mathbf{Y}}|\mathcal{B}) = -D \left(p_{\bar{\mathbf{Y}}}(\bar{\mathbf{y}}|\mathcal{B}) \left\| \prod_{i=1}^N p_{\bar{z}_i}(\bar{z}_i|\mathcal{B}) \right\| \right) \quad (65)$$

where $p_{\bar{\mathbf{Y}}}(\bar{\mathbf{y}}|\mathcal{B})$ and $p_{\bar{z}_i}(\bar{z}_i|\mathcal{B})$ denote the conditional pdf of the outputs and sources over \mathcal{B} .

As long as the set of boundary conditions refer to a finite set of samples and the involved mixing and separation filter do not infinitely propagate their effect (i.e., that the significant length of the overall system L_g is also finite), border effects become less important as the dimension of the model (NL_c) increases, and the conditional pdf approaches asymptotically to the true pdf [10]. Therefore

$$\Psi(\bar{\mathbf{Y}}|\mathcal{B}) \xrightarrow{L_c \gg L_g} \Psi(\bar{\mathbf{Y}}). \quad (66)$$

In particular, we can always assume $L_c \gg L_g$ in order that the approximation becomes true.

At this point, we can follow the same analysis of Section II to obtain the iterative inversion algorithms as iterative techniques that find the solution of the estimating equation (8). Thus, the II1 algorithm yields

$$\bar{\mathbf{W}}^{-1(n+1)} = \bar{\mathbf{W}}^{-1(n)} \left(\mathbf{I} + \mu^{(n)} \left(\mathbf{R}_{\bar{f}\bar{g}}^{(n)} - \mathbf{I} \right) \right) \quad (67)$$

where $\mathbf{R}_{\bar{f}\bar{g}} = E[\bar{\mathbf{f}}(\bar{\mathbf{y}})\bar{\mathbf{g}}^T(\bar{\mathbf{y}})]$. In addition, if $\mu^{(n)} < \|\mathbf{R}_{\bar{f}\bar{g}}^{(n)} - \mathbf{I}\|_p^{-1}$ we arrive at the II2 algorithm

$$\bar{\mathbf{W}}^{(n+1)} = \bar{\mathbf{W}}^{(n)} - \mu^{(n)} \left(\mathbf{R}_{\bar{f}\bar{g}}^{(n)} - \mathbf{I} \right) \bar{\mathbf{W}}^{(n)}. \quad (68)$$

Note that the matrix $\mathbf{R}_{\bar{f}\bar{g}}^{(n)}$ is asymptotically block Toeplitz (i.e., $[\mathbf{R}_{\bar{f}\bar{g}}^{(n)}]_{ij} = \mathbf{R}_{\bar{f}\bar{g}}^{(n)}[j - i]$) and has the same structure and dimensions as $\bar{\mathbf{W}}$. Finally, using this property and the convolution operator $*$ we can rewrite the convolutive II algorithms (67) and (68) as

$$\mathbf{W}^{-1(n+1)}[k] = \mathbf{W}^{-1(n)}[k] * \left(\mathbf{I}\delta[k] + \mu^{(n)} \left(\mathbf{R}_{\bar{f}\bar{g}}^{(n)}[k] - \mathbf{I}\delta[k] \right) \right) \quad (69)$$

and

$$\mathbf{W}^{(n+1)}[k] = \mathbf{W}^{(n)}[k] - \mu^{(n)} \left(\mathbf{R}_{\bar{f}\bar{g}}^{(n)}[k] - \mathbf{I}\delta[k] \right) * \mathbf{W}^{(n)}[k] \quad (70)$$

where

$$\mu^{(n)} = \frac{\eta}{1 + \eta \sum_k \left\| \mathbf{R}_{\bar{f}\bar{g}}^{(n)}[k] \right\|_p} \quad (71)$$

with $p = 1$ or $p = \infty$ and $\eta < 1$.

A. Stability Analysis

The local stability analysis for the convolutive algorithms parallels that presented for the instantaneous case. It is possible to extend all the stability results obtained for the instantaneous case by means of the following theorem.

Theorem 2: For temporally i.i.d. and spatially independent sources and under the same assumptions of Theorem 1, the II1 (69) and II2 (70) convolutive algorithms with a constant step-size asymptotically exhibits a local stable point at the separation solution if and only if the conditions (20)–(22) are satisfied for all $i, j|_{i \neq j} = 1, \dots, N$.

Proof: This theorem is straightforward to prove since in the neighborhood of the separation $L_g \approx 0$ and we have shown that the convolutive problem of temporally i.i.d. sources can be asymptotically formulated as an instantaneous separation problem and, therefore, the convolutive algorithms will asymptotically share the same set of stability conditions of the instantaneous case. ■

B. Convolutive Relations

In this section we will show how the convolutive II algorithms provide a unified perspective of some existing and also novel algorithms for BSS of convolutive mixtures.

When the function $\mathbf{g}(\cdot)$ is linear, the algorithm (70) reduces to the multichannel blind deconvolution algorithm derived by Amari *et al.* [3] from an information theoretic perspective using the natural gradient [6]. On the other hand, our approach can be used to justify the nongradient algorithms obtained when $\mathbf{g}(\cdot)$ is nonlinear and also provides a step size $\mu^{(n)}$ that exhibits a fast convergence. Moreover, note that Theorem 2 is the convolutive

TABLE I
FILTER COEFFICIENTS OF THE MIXING SYSTEM $\mathbf{H}[k]$ FOR THE FIRST EXAMPLE

<i>Filter</i> \ <i>Lag</i>	$\delta[k+3]$	$\delta[k+2]$	$\delta[k+1]$	$\delta[k]$	$\delta[k-1]$	$\delta[k-2]$	$\delta[k-3]$
$H_{11}[k]$	0.1	0.2	0.6	1	0	0	0
$H_{12}[k]$	0.1	-0.3	0.5	-0.6	0.5	-0.3	0.1
$H_{13}[k]$	0.1	-0.3	0.5	-0.6	0.5	-0.3	0.1
$H_{21}[k]$	-0.1	0.2	-0.4	-0.5	-0.4	0.2	-0.1
$H_{22}[k]$	0	0	0	1	0.6	0.2	0.1
$H_{23}[k]$	-0.1	0.2	-0.4	-0.5	-0.4	0.2	-0.1
$H_{31}[k]$	0.1	0.2	0.4	0.5	0.4	0.2	0.1
$H_{32}[k]$	0.1	0.2	0.4	0.5	0.4	0.2	0.1
$H_{33}[k]$	0	0	0	1	0.6	0.2	0.1

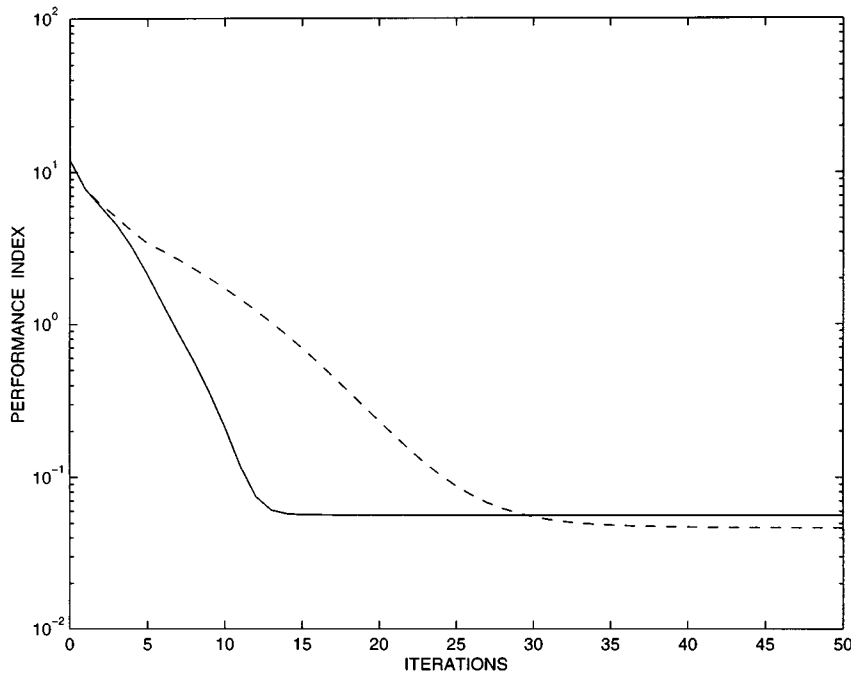


Fig. 2. Performance index versus iterations of the batch II2 convolutive algorithm for the first example. Comparison between two different nonlinearities: case a) $\mathbf{f}(y) = |y|^2 y$ and $\mathbf{g}(y) = \text{sign}(\mathcal{R}e\{y\}) + j \text{sign}(\mathcal{I}m\{y\})$ is the continuous line, case b) $\mathbf{f}(y) = |y|^2 y$ and $\mathbf{g}(y) = y$ is the dashed line.

extension of the stability conditions presented by Amari *et al.* [4] for the instantaneous case.

Using a similar procedure to that shown in Section V-D, it is straightforward to extend the family of equivariant (EASI) algorithms to the convolutive case as

$$\mathbf{W}^{(n+1)}[k] = \left(\mathbf{I}\delta[k] - \eta \frac{\mathbf{R}_{yy}^{(n)}[k] - \mathbf{I}\delta[k]}{1 + \eta \sum_k \left\| \mathbf{R}_{yy}^{(n)}[k] \right\|_p} \right)$$

$$- \frac{\eta \mathbf{R}_{fg}^{(n)}[k] - \mathbf{R}_{fg}^{T(n)}[-k]}{2 \left(1 + \eta \sum_k \left\| \mathbf{R}_{fg}^{(n)}[k] \right\|_p \right)} * \mathbf{W}^{(n)}[k] \quad (72)$$

for $p = 1$ or $p = \infty$ and $\eta < 1$.

Under the assumption of two sources and convolutive soft mixture (i.e., $\delta[k] - c_{ij}^{(n)}[k] * c_{ji}^{(n)}[k] \approx \delta[k]$ for all $i \neq j$) we can obtain the Herault–Jutten convolutive algorithm [30] starting from the convolutive version of the III1 algorithm. Following the same steps as in Section V-E we find that the III1 algorithm with

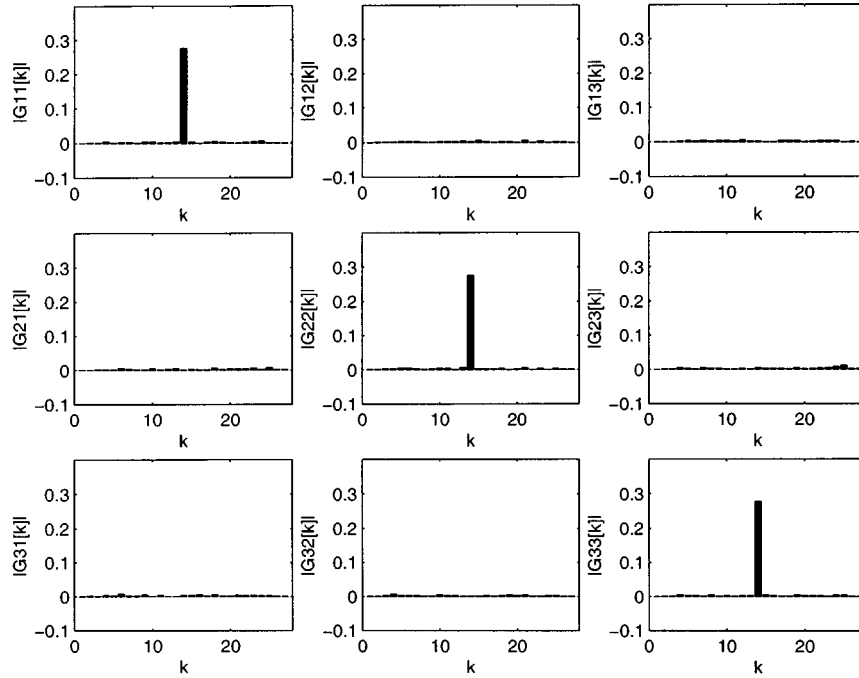


Fig. 3. Coefficients of the overall transfer system $\mathbf{G}[k]$ after convergence in the first computer experiment.

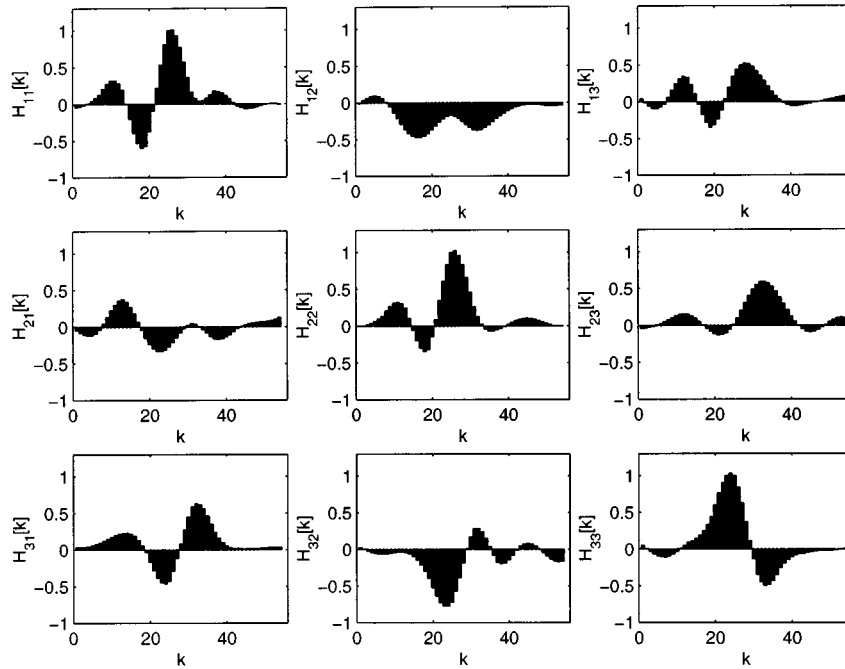


Fig. 4. Coefficients of the mixing system $\mathbf{H}[k]$ for the second computer experiment.

a sufficiently small step-size can be approximated by the following recursion:

$$c_{ij}^{(n+1)}[k] = c_{ij}^{(n)}[k] + \mu \left(\delta[k] - c_{ij}^{(n)}[k] * c_{ji}^{(n)}[k] \right) * \left[\mathbf{R}_{fg}^{(n)}[k] \right]_{ij} \quad (73)$$

for $i, j |_{i \neq j} = 1, 2; \forall k$. Nevertheless, when applying the soft mixture assumption, this result can be approximated by the convolutive Herault–Jutten algorithm [30]

$$c_{ij}^{(n+1)}[k] = c_{ij}^{(n)}[k] + \mu \left[\mathbf{R}_{fg}^{(n)}[k] \right]_{ij} \quad (74)$$

for $i, j |_{i \neq j} = 1, 2; \forall k$.

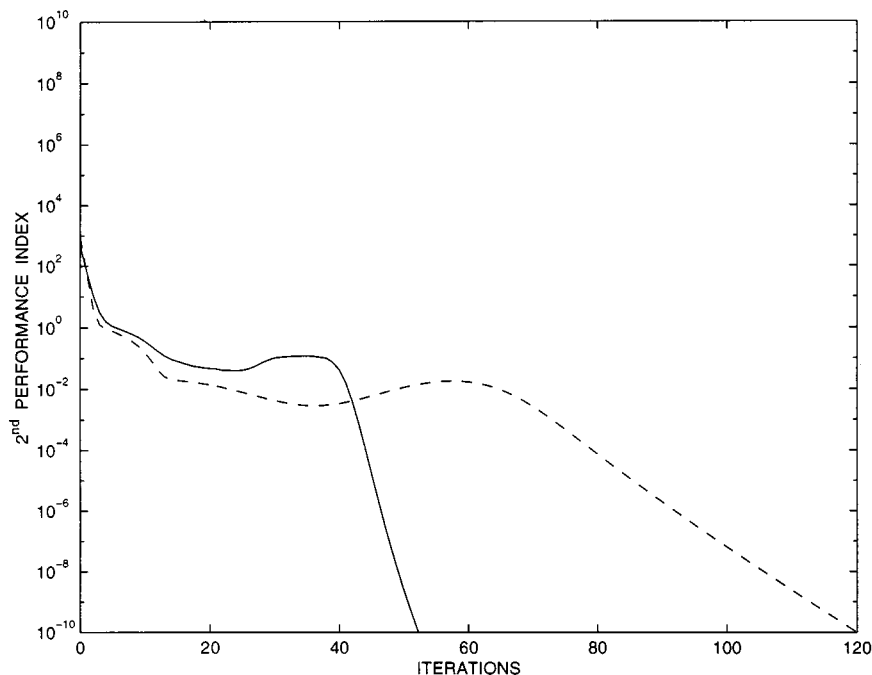


Fig. 5. Performance index versus iterations for the second computer experiment. Comparison between two different nonlinearities: case a) $\mathbf{f}(y) = |y|^2y$ and $\mathbf{g}(y) = \text{sign}(\mathcal{R}e\{y\}) + j \text{sign}(\mathcal{I}m\{y\})$ is the continuous line, case b) $\mathbf{f}(y) = |y|^2y$ and $\mathbf{g}(y) = y$ is the dashed line.

When the mixing system satisfies the following orthogonality condition $\mathbf{H}^{T(n)}[k] * \mathbf{H}^{(n)}[k] = \mathbf{I}\delta[k]$ we can use the convolutive extension of the mixing system estimate (52) to obtain a novel convolutive form of the nonlinear PCA algorithm

$$\mathbf{W}^{(n+1)}[k] = \mathbf{W}^{(n)}[k] + \mu \left(\mathbf{R}_{fx}^{(n)}[k] - \mathbf{R}_{ff}^{(n)}[k] * \mathbf{W}^{(n)}[k] \right). \quad (75)$$

Finally, we wish to point out that Iterative Inversion convolutive algorithms are also well posed in the degenerate case where there is an unique source and the separation is performed only on the temporal domain. This case is usually termed *blind deconvolution* and the derivation presented in Section VI is still valid since it does not make any assumption on the spatial dimensions of the mixing and separating systems.

VII. COMPUTER SIMULATIONS

Computer simulations were carried out to illustrate the performance of the II algorithms. We considered three i.i.d. 16-QAM complex sources. The II algorithms can be extended to the complex-valued signals case just replacing the transpose operator by an Hermitian operator. In a first computer experiment we considered a convolutive mixture obtained with the mixing system impulse response

$$\mathbf{H}[k] = \begin{pmatrix} H_{11}[k] & H_{12}[k] & H_{13}[k] \\ H_{21}[k] & H_{22}[k] & H_{23}[k] \\ H_{31}[k] & H_{32}[k] & H_{33}[k] \end{pmatrix} \quad (76)$$

where the filter coefficients are given in Table I. The separating system for this first experiment is a three-input–three-output system with nine FIR filters having 21 taps each. To adapt the separating system we used the II2 convolutive algorithm given by (70) and (71) over a data window of 5000 samples. In the

implementation of the algorithm the cross-correlation sequence $\mathbf{R}_{fg}[k]$ was truncated and only its values inside the window $k \in [-30, 30]$ were preserved.

We defined the performance index of the separation system in terms of the overall response $\mathbf{G}[k]$ as

$$P_{\text{index}} = \sum_{i=1}^N \left(\sum_{j=1}^N \sum_k \frac{|G_{ij}[k]|^2}{\max_{t,m} \{|G_{it}[m]|^2\}} - 1 \right) + \sum_{j=1}^N \left(\sum_{i=1}^N \sum_k \frac{|G_{ij}[k]|^2}{\max_{t,m} \{|G_{tj}[m]|^2\}} - 1 \right). \quad (77)$$

We set the step size $\eta = 0.9$, we chose the infinite norm ($p = \infty$) and we considered two choices for the nonlinearities: case a) in which $\mathbf{f}(y) = |y|^2y$ and $\mathbf{g}(y) = \text{sign}(\mathcal{R}e\{y\}) + j \text{sign}(\mathcal{I}m\{y\})$ and case b) where $\mathbf{f}(y) = |y|^2y$ and $\mathbf{g}(y) = y$. Fig. 2 plots the evolution of the performance index versus the iterations for the two cases. We can see that a better asymptotic performance is achieved in case b) whereas case a) exhibits a much faster convergence since the algorithm achieves its final state in less than 15 iterations. This result is consistent with the conclusions obtained by Amari and Cardoso in [4] where they show that, regarding to the local asymptotic behavior, the best estimating functions $(\mathbf{f}(\mathbf{y})\mathbf{g}^T(\mathbf{y}) - \mathbf{I})$ are obtained when one of the functions $\mathbf{f}(\cdot)$ or $\mathbf{g}(\cdot)$ is linear. Nevertheless, the simulations indicate that other estimating functions with two nonlinearities may have preferred properties such as faster convergence.

Fig. 3 shows the overall transfer system after convergence for the case a). Since the overall system has converged to a scaled version of the identity matrix at zero lag and zero in the rest (i.e., $\mathbf{G}[k] \approx \gamma \mathbf{I}\delta[k]$) the spatio-temporal separation of the sources has been performed.

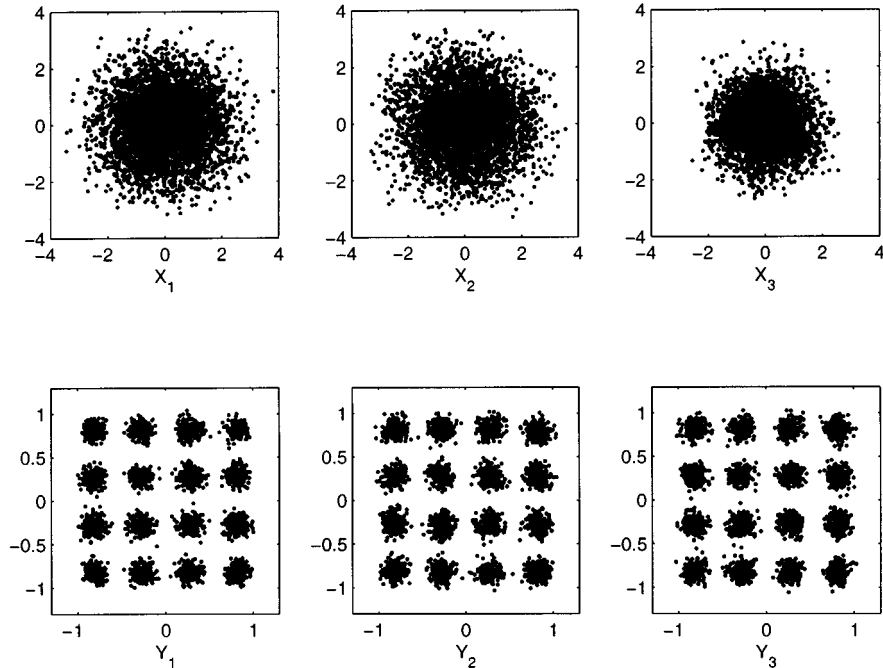


Fig. 6. Constellation of the observations in the first row and of the recovered sources in the second row, both for the second computer experiment.

In the second computer experiment we considered the same kind of sources but a different 3×3 mixing system with the impulse response sequence of 55 taps length plotted in Fig. 4. We considered the same parameters of the previous experiment and a data window of 10 000 samples. Due to the equivariant property of the II2 convolutive algorithm we have implemented it for this second experiment in the way suggested by Cardoso in [14]: instead of evaluating the separating system, the outputs of one iteration are directly computed from the previous outputs using the following recursion:

$$\mathbf{y}^{(n+1)}[k] = \mathbf{y}^{(n)}[k] - \mu^{(n)} \left(\mathbf{R}_{fg}^{(n)}[k] - \mathbf{I}\delta[k] \right) * \mathbf{y}^{(n)}[k]. \quad (78)$$

The advantage of this approach is that it greatly reduces the number of operations. Since the separating system $\mathbf{W}[k]$ will not be computed we will use a different performance index given by $\rho(n) = \|\mathbf{R}_{fg}^{(n)}[k] - \mathbf{I}\delta[k]\|_F^2$ where $\|\cdot\|_F$ denotes de Frobenius norm of a matrix. As in the previous experiment the convergence of the algorithm with two nonlinearities is faster that with only one nonlinearity as can be seen from Fig. 5. Fig. 6 shows the observations of the three mixed sources and the resulting outputs after convergence for the case a). It is clearly seen from these constellations that the separation has been performed.

VIII. CONCLUSION

In this paper we present an unified view of several blind source separation algorithms from a novel perspective. This unifying approach is termed II because it consists of a quasi-Newton iterative procedure to invert a nonlinear correlation matrix of the outputs. We determine the variable step-sizes that convert conventional algorithms for BSS in algorithms of the Newton type that exhibit a faster convergence. Although initially presented for instantaneous mixtures, the method is also extended to consider the more realistic case of convolutive

mixtures. Finally, we obtain the necessary and sufficient conditions that ensure local convergence of the instantaneous and convolutive II algorithms and generalize the convergence conditions obtained by Amari *et al.* in [5].

APPENDIX I QUASI-NEWTON RECURSION

We can find the zeros of the following matrix functional:

$$\mathcal{F}(\mathbf{W}^{-1}) = \hat{\mathbf{H}}(\mathbf{W}) - \mathbf{W}^{-1} \quad (79)$$

with the following quasi-Newton recursion:

$$\mathbf{W}^{-1(n+1)} = \mathbf{W}^{-1(n)} - \mu \mathcal{F} \left(\mathbf{W}^{-1(n)} \right) \mathbf{B}^{(n)} \quad (80)$$

where $\mathbf{B}^{(n)}$ is an approximation to the inverse of the derivative matrix $\partial \mathcal{F}(\mathbf{W}^{-1}) / \partial \mathbf{W}^{-1}$ that must satisfy certain conditions described in [26]. As can be seen from Fig. 1, the robust estimates of the mixing system $\hat{\mathbf{H}}^{(n)}$ verify that $\mathcal{F}(\mathbf{W}^{-1}) = \hat{\mathbf{H}}(\mathbf{W}) - \mathbf{W}^{-1}$ forms an acute angle with the direction that points to the solution $\mathbf{H} - \mathbf{W}^{-1}$. Therefore

$$\mathcal{F}(\mathbf{W}^{-1}) \propto \mathbf{H} - \mathbf{W}^{-1}. \quad (81)$$

As a consequence

$$\left(\frac{\partial \mathcal{F}(\mathbf{W}^{-1(n)})}{\mathbf{W}^{-1(n)}} \right)^{-1} \propto -\mathbf{I} = \mathbf{B}^{(n)} \quad (82)$$

and we arrive again at (12).

APPENDIX II STEP-SIZE NORM

In this appendix we demonstrate that the II algorithms using the step-size expression (40) achieves its faster convergence

when the 2-norm is used. Let λ be an eigenvalue of $\mathbf{R}_{fg}^{(n)}$ whose corresponding eigenvector is \mathbf{v} . Then

$$|\lambda| \|\mathbf{v}\| = \|\lambda \mathbf{v}\| = \|\mathbf{R}_{fg}^{(n)} \mathbf{v}\| \leq \|\mathbf{R}_{fg}^{(n)}\| \|\mathbf{v}\| \quad (83)$$

and $|\lambda| \leq \|\mathbf{R}_{fg}^{(n)}\|$ for any matrix norm. In our case, the matrix $\mathbf{R}_{fg}^{(n)}$ will be symmetric at convergence and thus the 2-norm will asymptotically tend to the spectral radio $\rho(\cdot)$ of the matrix

$$\|\mathbf{R}_{fg}^{(n)}\|_2 = \left[\rho \left(\mathbf{R}_{fg}^{T(n)} \mathbf{R}_{fg}^{(n)} \right) \right]^{1/2} \rightarrow \rho \left(\mathbf{R}_{fg}^{(n)} \right). \quad (84)$$

Since the proposed step-size for the II2 algorithm is selected according to (40) the fastest convergence for a given η is obtained when the 2-norm is used ($p = 2$).

APPENDIX III

STABILITY ANALYSIS OF THE II2 ALGORITHM

In this appendix we will prove that the conditions (20)–(22) ensure the asymptotic stability of the II2 algorithm under the assumptions of Theorem 1. Let us start multiplying both sides of (19) by the mixing matrix \mathbf{H} to express the II2 algorithm in terms of the global transfer matrix \mathbf{G} as

$$\mathbf{G}^{(n+1)} = \mathbf{G}^{(n)} - \mu \left(\mathbf{R}_{fg}^{(n)} - \mathbf{I} \right) \mathbf{G}^{(n)}. \quad (85)$$

Let $C(\mathbf{G}\mathbf{s}) = \mathbf{f}(\mathbf{G}\mathbf{s})\mathbf{g}^T(\mathbf{G}\mathbf{s}) - \mathbf{I}$. Condition (18) is necessary to allow the separation solution $\mathbf{G}_* = \mathbf{I}$ to be an equilibrium point of the algorithm, i.e., $E[C(\mathbf{G}_*\mathbf{s})] = \mathbf{0}$.

The next step is to study the asymptotic stability of the algorithm at this point. Toward this aim we will study the behavior of the ordinary differential equation (ODE) associated to the algorithm at the separation solution [9]. Let us define the mean field of the algorithm as $M(\mathbf{G}) = E[C(\mathbf{G}\mathbf{s})\mathbf{G}]$ and an arbitrary small perturbation ϵ of the global transfer matrix at the separation solution \mathbf{G}_* . Then

$$M(\mathbf{I} + \epsilon) = E[C(\mathbf{s} + \epsilon\mathbf{s})(\mathbf{I} + \epsilon)] = E[C(\mathbf{s} + \epsilon\mathbf{s})] + o(\epsilon). \quad (86)$$

In order to find the linear approximation to the mean field in terms of ϵ , we will replace both functions $\mathbf{f}(\mathbf{s} + \epsilon\mathbf{s})$ and $\mathbf{g}(\mathbf{s} + \epsilon\mathbf{s})$ by its first-order Taylor expansion at $\mathbf{G}_* = \mathbf{I}$

$$\begin{aligned} f(\mathbf{s} + \epsilon\mathbf{s}) &= \mathbf{f}(\mathbf{s}) + \mathbf{f}'(\mathbf{s})\epsilon\mathbf{s} + o(\epsilon) \\ g(\mathbf{s} + \epsilon\mathbf{s}) &= \mathbf{g}(\mathbf{s}) + \mathbf{g}'(\mathbf{s})\epsilon\mathbf{s} + o(\epsilon) \end{aligned} \quad (87)$$

where $\mathbf{f}'(\mathbf{s}) = \partial\mathbf{f}(\mathbf{s})/\partial\mathbf{s}$ and $\mathbf{g}'(\mathbf{s}) = \partial\mathbf{g}(\mathbf{s})/\partial\mathbf{s}$ are two diagonal matrices since $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ act component-wise. Substituting (87) in (86) and having in mind that $E[C(\mathbf{G}_*\mathbf{s})] = \mathbf{0}$ we can express the algorithm mean field as

$$M(\mathbf{I} + \epsilon) = E[\mathbf{f}'(\mathbf{s})\epsilon\mathbf{s}\mathbf{g}^T(\mathbf{s})] + E[\mathbf{f}(\mathbf{s})\mathbf{s}^T\epsilon^T(\mathbf{g}'(\mathbf{s}))^T] + o(\epsilon). \quad (88)$$

Defining $f_i = [\mathbf{f}(\mathbf{s})]_i$, $f'_i = [\mathbf{f}'(\mathbf{s})]_{ii}$, the terms of the mean field can be written as

$$M_{ij}(I + \epsilon) = \sum_k E[f'_i s_k g_j] \epsilon_{ik} + \sum_k E[f_i s_k g'_j] \epsilon_{jk} + o(\epsilon). \quad (89)$$

Taking into account the independence assumptions for the sources at the separation solution and using the zero mean assumption of the sources $E[s_i] = 0 \forall i$ we obtain

$$M_{ij}(I + \epsilon) = \begin{cases} (E[f'_i s_i g_i] + E[f_i s_i g'_i]) \epsilon_{ii} + o(\epsilon) & \text{if } i = j \\ E[f'_i] E[s_j g_j] \epsilon_{ij} + E[f'_i s_i] E[g_j] \epsilon_{ii} \\ \quad + E[f_i s_i] E[g'_j] \epsilon_{ji} + E[f_i] E[s_j g'_j] \\ \quad \cdot \epsilon_{jj} + o(\epsilon) & \text{if } i \neq j. \end{cases} \quad (90)$$

Next, let us rearrange the elements of the matrix $\mathbf{G} = [G_{ij}]$ in the following vector as:

$$\theta = [G_{11}, \dots, \underbrace{G_{kk}, \dots, G_{ij}, G_{ji}, \dots}_{k=2 \dots N, 1 \leq i < j \leq k \leq N}]^T. \quad (91)$$

Thus, the ODE associated to the algorithm (85) can be rewritten as

$$\frac{d\theta}{dt} = -\mu \psi \left(\theta^{(t)}, \mathbf{z}^{(t)} \right) \quad (92)$$

where $\mathbf{z}^{(t)}$ is the state vector. Let θ_* denote the stationary point of the algorithm that gives the separation solution. The mean vector field of the algorithm at this point is $\mathcal{M}(\theta_*, \mathbf{z}^{(t)}) = E[\psi(\theta_*, \mathbf{z}^{(t)})]$ and can be obtained from (90).

It is well known that the asymptotic stability condition for the algorithm at θ_* is that all the eigenvalues of the derivative of the mean field at this point have positive real parts. Noting that $(\partial\mathcal{M}(\theta; \mathbf{z})/\partial\theta)|_{\theta=\theta_*} = (\partial\mathcal{M}(\theta; \mathbf{z})/\partial\epsilon)|_{\epsilon=\mathbf{0}}$ and using (90) it is found that the gradient of the mean field at separation takes a very simple form: it is a block-triangular matrix with diagonal block elements of size one and two. This fact, that has been observed before in several BSS algorithms (see for example [12], [31]), greatly facilitates the asymptotic stability analysis.

One of the conditions for asymptotic stability is extracted from the 1×1 diagonal blocks

$$\frac{\partial M_{ii}(I + \epsilon)}{\partial \epsilon_{ii}} = E[f'_i s_i g_i] + E[f_i s_i g'_i] > 0. \quad (93)$$

The 2×2 diagonal blocks are of the form

$$\begin{aligned} & \begin{pmatrix} \frac{\partial M_{ij}(I + \epsilon)}{\partial \epsilon_{ij}} & \frac{\partial M_{ij}(I + \epsilon)}{\partial \epsilon_{ji}} \\ \frac{\partial M_{ji}(I + \epsilon)}{\partial \epsilon_{ij}} & \frac{\partial M_{ji}(I + \epsilon)}{\partial \epsilon_{ji}} \end{pmatrix} \\ &= \begin{pmatrix} E[f'_i] E[s_j g_j] & E[g'_j] E[s_i f_i] \\ E[g'_i] E[s_j f_j] & E[f'_j] E[s_i g_i] \end{pmatrix} \\ &= \begin{pmatrix} a & b \\ c & d \end{pmatrix} \end{aligned} \quad (94)$$

for all $i \neq j$. The eigenvalues of this block matrices are the roots of the characteristic polynomial equation $P(\lambda) = \lambda^2 - (a+d)\lambda + (ad-bc)$. In order to check if the real part of the eigenvalues is positive, we can use the necessary and sufficient Hurwitz stability conditions [1], obtaining $a+d > 0$ and $ad-bc > 0$. Then, substituting a, b, c, d into these expressions we arrive at the remaining necessary and sufficient conditions for the asymptotic stability of the I12 algorithm

$$E[f'_i]E[s_j g_j] + E[f'_j]E[s_i g_i] > 0 \quad (95)$$

$$E[f'_i]E[f'_j]E[s_i g_i]E[s_j g_j] > E[g'_i]E[g'_j]E[s_i f_i]E[s_j f_j] \quad (96)$$

for all $i \neq j$.

REFERENCES

- [1] J. Ackerman, *Sampled Data Control Systems*. Berlin, Germany: Springer-Verlag, 1985.
- [2] L. B. Almeida and F. M. Silva, "Adaptive decorrelation," in *Artificial Neural Networks*. Amsterdam, The Netherlands: Elsevier, 1992, vol. 2, pp. 149–156.
- [3] S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. IEEE Int. Workshop Wireless Commun.*, Paris, France, Apr. 1997, pp. 101–104.
- [4] S. Amari and J. F. Cardoso, "Blind source separation—Semiparametric statistical approach," *IEEE Trans. Signal Processing*, vol. 45, pp. 2692–2697, Nov. 1997.
- [5] S.-I. Amari, T.-P. Chen, and A. Cichocki, "Stability analysis of learning algorithms for blind source separation," *Neural Networks*, vol. 10, no. 8, pp. 1345–1351, 1997.
- [6] S. Amari, "Natural gradient works efficiently in learning," *Neural Comput.*, vol. 10, no. 2, pp. 251–276, 1998.
- [7] A. J. Bell and T. J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, pp. 1129–1159, 1996.
- [8] S. Bellini, "Bussgang techniques for blind deconvolution and equalization," in *Blind Deconvolution*, S. Haykin, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1994.
- [9] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. Berlin, Germany: Springer-Verlag, 1990.
- [10] G. E. P. Box and G. M. Jenkins, "Time series analysis, forecasting and control," in *Time Series and Digital Processing*. New York: Holden Day, 1976, pp. 208–212.
- [11] X.-R. Cao and R.-W. Liu, "General approach to blind source separation," *IEEE Trans. Signal Processing*, vol. 44, pp. 562–571, Mar. 1996.
- [12] J. F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, pp. 3017–3030, Dec. 1996.
- [13] J. F. Cardoso, "Infomax and maximum likelihood for blind separation," *IEEE Signal Processing Lett.*, vol. 4, Apr. 1997.
- [14] —, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, pp. 2009–2025, 1998.
- [15] N. Charkani and Y. Deville, "Self-adaptive separation of convolutive mixed signals with a recursive structure—Part I: Stability analysis and optimization of asymptotic behavior," *Signal Processing*, vol. 73, pp. 225–254, 1999.
- [16] A. Cichocki, R. Unbehauen, and E. Rummert, "Robust learning algorithm for blind separation of signals," *Electron. Lett.*, vol. 30, no. 17, pp. 1386–1387, August 1994.
- [17] A. Cichocki and R. Unbehauen, "Robust neural networks with on-line learning for blind identification and blind separation of sources," *IEEE Trans. Circuits Syst. I*, vol. 43, pp. 894–906, Nov. 1996.
- [18] A. Cichocki, S. C. Douglas, and S. Amari, "Robust techniques for independent component analysis (ICA) with noisy data," *Neurocomput.*, vol. 22, pp. 113–129, 1998.
- [19] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 3, no. 36, pp. 287–314, 1994.
- [20] S. Cruces, "An unified view of blind source separation algorithms," Ph.D. dissertation, Signal Processing Dept., Univ. Vigo, Spain, 1999.
- [21] S. Douglas and A. Cichocki, "Neural networks for blind decorrelation of signals," *IEEE Trans. Signal Processing*, vol. 45, pp. 2829–2842, Nov. 1997.
- [22] G. H. Golub and C. F. Van-Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
- [23] C. Jutten and J. Herault, "Blind separation of sources—Part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1–10, 1991.
- [24] J. Karhunen, E. Oja, L. Wang, R. Vigario, and J. Koutsensalo, "A class of neural networks for independent component analysis," *IEEE Trans. Neural Networks*, vol. 8, pp. 486–503, May 1997.
- [25] J. Karhunen, P. Pajunen, and E. Oja, "The nonlinear PCA criterion in blind source separation: Relations with other approaches," *Neurocomputing*, vol. 22, pp. 5–20, Nov. 1998.
- [26] C. T. Kelley, "Iterative methods for linear and nonlinear equations," in *Frontiers in Applied Mathematics 16*. Philadelphia, PA: SIAM, 1995.
- [27] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [28] A. S. Householder, *The Theory of Matrices in Numerical Analysis*, 2nd ed. New York: Dover, 1975.
- [29] S. Makeig, A. Bell, T.-P. Jung, and T.J. Sejnowski et al., *Independent Component Analysis in Electro-Encephalographic Data*, M. Mozer et al., Eds. Cambridge, MA: MIT Press, 1996, vol. 8, pp. 145–151.
- [30] H. L. Nguyen-Thi and C. Jutten, "Blind source separation for convolutive mixtures," *Signal Processing*, vol. 45, no. 2, pp. 209–229, 1995.
- [31] E. Oja, "The nonlinear pca learning rule and signal separation-mathematical analysis," Lab. Comput. Inform. Sci., Helsinki Univ. Technol., Espoo, Finland, Tech. Rep. A26, Aug. 1995.
- [32] V. Pan and J. Reif, "Efficient parallel solution of linear systems," in *Proc. 17th Annu. ACM Symp. Theory Comput.*, May 1985, pp. 143–152.
- [33] B. A. Pearlmutter and L. C. Parra, "A context-sensitive generalization of ICA," in *Proc. Int. Conf. Neural Inform. Processing (ICONIP)*, Hong Kong, Sept. 1996, pp. 151–156.
- [34] D. T. Pham and P. Garat, "Blind separation of mixture of independent sources through a quasimaximum likelihood approach," *IEEE Trans. Signal Processing*, vol. 45, pp. 1712–1725, July 1997.
- [35] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes In C*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 1992.
- [36] E. Sorouchyari, "Blind separation of sources—Part III: Stability analysis," *Signal Processing*, vol. 24, pp. 21–29, 1991.
- [37] L. Tong, R.-W. Liu, V.-C. Soon, and Y.-F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 499–509, May 1991.
- [38] K. Torkkola, "Blind separation for audio signals—Are we there yet?," in *Proc. 1st ICA'99 Congr.*, Assois, France, Jan. 1999, pp. 239–244.
- [39] H. H. Yang and S. Amari, "Adaptive on-line learning algorithms for blind source separation—Maximum entropy and minimum mutual information," *Neural Comput.*, 1997.



Sergio A. Cruces-Alvarez (S'94–A'99) was born in Vigo, Spain, in 1970. He received the Telecommunication Engineer degree (with honors) in 1994 and the Ph.D. degree in 1999, both at the University of Vigo.

From 1994 to 1995, he was working as a Project Engineer for the Department of Signal Theory and Communications of this university. In 1995, he joined the Signal Theory and Communications group of the University of Seville, where he currently teaches undergraduate and graduate courses on digital signal processing and mathematical methods

for communication. In 1997 and 1999, he was invited to visit the Laboratory for Advanced Brain Signal Processing under the Frontier Research Program RIKEN (Japan). His research interests include statistical signal processing, information theoretic approaches, blind equalization, and filter stabilization techniques.

Andrzej Cichocki (M'96) was born in Poland. He received the M.Sc. (with honors), Ph.D., and Habilitate Doctorate (Dr.Sc.) degrees, all in electrical engineering and computer science, from Warsaw University of Technology, Warsaw, Poland, in 1972, 1975, and 1982, respectively.

Since 1972, he has been with the Institute of Theory of Electrical Engineering and Electrical Measurements at the Warsaw University of Technology, where he became a full Professor in 1991. He is the coauthor of two international books: *MOS Switched-Capacitor and Continuous-Time Integrated Circuits and Systems* (Springer-Verlag, 1989) and *Neural Networks for Optimization and Signal Processing* (New York: Wiley and Teubner-Verlag, 1993) and author or coauthor of more than 150 scientific papers. He spent a few years at the University of Erlangen-Nuernberg, Germany, as Alexander Humboldt Research Fellow and Guest Professor. From 1995 to 1996, he has been working as a Team Leader of the Laboratory for Artificial Brain Systems, at the Frontier Research Program RIKEN, Japan, in the Brain Information Processing Group. Currently, he is Head of the Laboratory for Advanced Brain Signal Processing in the Brain Science Institute, Riken, Wako-schi, Japan. He is Reviewer of several international journals, including *Signal Processing*, *Circuits and Systems*, *Biological Cybernetics*, *Electronics Letters*, *Neurocomputing*, and *Neural Computation*. He is also member of several international Scientific Committees. His current research interests include signal and image processing (especially blind signal/image processing), neural networks and their electronic implementations, learning theory and algorithms, independent and principal component analysis, optimization problems, circuits and systems theory and their applications, artificial intelligence.

Dr. Cichocki is a Reviewer and Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS since January 1998.



Luis Castedo-Ribas (S'92–A'95) was born in Santiago de Compostela, Spain, in 1966. He received the Ingeniero de Telecomunicación and Doctor Ingeniero de Telecomunicación degrees, both from Universidad Politécnica de Madrid (UPM), Spain, in 1990 and 1993, respectively.

From 1990 to 1994, he was with the Departamento de Señales, Sistemas y Radiocomunicación at the UPM where he worked in array processing applied to digital communications. In 1994, he joined the Facultad de Informática at Universidad de A Coruña, Spain, where he is currently an Associate Professor. His current research interests include blind adaptive filtering and signal processing methods for space and code diversity exploitation in communication systems.