A New On-Line Adaptive Learning Algorithm for Blind Separation of Source Signals

A. Cichocki*, R. Unbehauen, L. Moszczynski* and E. Rummert
* Warsaw University of Technology, Poland and
University Erlangen-Nürnberg, Germany
E-Mail: cia@iem.pw.edu.pl

ABSTRACT

In this paper a novel, very efficient, temporarily stable, self-normalizing, unsupervised adaptive learning algorithm for on-line (real-time) separation of statistically independent unknown source signals from a linear mixture of them is proposed. In contrast to the known algorithms the new algorithm allows to separate (or extract) extremely badly scaled signals (i.e. some or even all of the source and/or sensor signals can be very weak). Moreover, the mixing matrix can be very ill-conditioned (i.e. all sensors can be almost identical or they can have almost identical transmission channel parameters). The new algorithm can be considered as an essential improvement, a modification and/or extension of the well-known Herault-Jutten algorithm and/or as a generalization of nonlinear PCA (principal component analysis) adaptive algorithms.

I. INTRODUCTION

Blind separation of sources has a wide range of applications in signal processing, sensors, measurements, medical science, biology and physics [9] - [15]. The first neural networks for on-line blind separation of a linear superposition of sources has been developed by Herault and Jutten [1] - [3]. Further works on off-line blind separation of independent sources are based on cumulants (higher order moments) [6] - [10]. The above developments have been limited principally to signals that are linearly mixed. An extension to more complex cases (e.g. convolutive mixing with causal FIR filters and nonlinear mixing of signals) has been proposed more recently [10] - [15]. Moreover, several VLSI implementations of such algorithms have been developed [4], [5]. A number of useful algorithms for blind separation of independent sources have been already developed but most of them are either relatively complex and much computation is involved which prohibits the use of them for on-line (real-time) separation [6] - [10] and/or their performance seems to be rather limited [1], [2]; especially, they may behave rather poor if the signal-to-noise ratio is small and/or the source signals are badly scaled and/or the mixing matrices are ill-conditioned (near singular). Moreover, a serious problem often arises to ensure the system stability.

The main purpose of this paper is to present a new high-performance, on-line adaptive learning algorithm which avoids or at least alleviates the above mentioned drawbacks. The high performance of the proposed algorithm is illustrated and confirmed by extensive computer simulation experiments.

II. PROBLEM FORMULATION

The problem of blind separation of sources is formulated as follows. For a given set of time variable sensor signals $x_i(t)$ described by

$$x_i(t) = \sum_{j=1}^{n} a_{ij} s_j(t), \quad (i = 1, 2, \ldots, n) \tag{1a}$$

or in matrix form by (cf. Fig. 1)

$$\boldsymbol{x}(t) = \boldsymbol{A}\, \boldsymbol{s}(t), \tag{1b}$$

(where $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ is an unknown nonsingular mixing matrix, $x(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^T$ is a vector of measured (observed) sensor signals and $s(t) = [s_1(t), s_2(t), \ldots, s_n(t)]^T$ is an unknown vector of mutually independent signals (primary sources)) it is necessary to find or estimate a new output vector $y(t) = [y_1(t), y_2(t), \ldots, y_n(t)]^T$ which is proportional to the input vector $s(t)$. Such a problem has an inherited indeterminacy [8], [9]. This indeterminacy is characterized by the magnitude scaling and the order in which the estimated signals $y_i(t)$ are arranged with respect to the original sources $s_i(t)$. In many applications this indeterminacy is fully acceptable, because the most relevant information of the source signals is contained in the waveforms of the signals (e.g. speech, music) rather than in their magnitudes and orders in which they are arranged. In fact, in a large number of cases, the signals received by an array of sensors (e.g. microphones, antennas, transducers, etc.) are sums (linear mixtures) of primary, original source signals. These sources are usually time-variable and totally unknown as in the case of aerial processing of radar or sonar signals. Of course, the mixing process of input signals can be described by different mathematical models in different situations and different a priori informations about the signals are available. In this paper we will focus on the general problem, where no a priori information on the sources themselves is available and only the sensor (mixed) signals are observed or available. For simplicity of our further considerations we assume that all the signals are zero-mean although this condition can be easily alleviated [1], [19].

III. NEW RESULTS – SELF NORMALIZING ADAPTIVE LEARNING ALGORITHM

In our approach we have used a single layer, feed-forward neural network consisting of n neurons described by

$$y_i(t) = \sum_{j=1}^{n} w_{ij}\, x_j(t), \quad (i = 1, 2, \ldots, n) \tag{2a}$$

or in matrix form as (cf. Fig. 1)

$$y(t) = W\, x(t), \tag{2b}$$

where $W = [w_{ij}] \in \mathbb{R}^{n \times n}$ is the matrix of the synaptic weights, $x(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^T$ is the vector of the observed (measured) sensor signals and $y(t) = [y_1(t), y_2(t), \ldots, y_n(t)]^T$ is the vector of the output signals which after the learning phase (i.e. after the synaptic weights w_{ij} achieve an equilibrium point) must be mutually statistically independent under the assumption that the input sources are also independent. In order to separate the input sources $s_i(t)$ the neural network must adapt on-line and without a supervisor (teacher) which is a rather challenging and very difficult type of learning. In other words, the key problem is to develop or derive a suitable unsupervised learning algorithm which allows the on-line adjustment (update) of the synaptic weights $w_{ij}(t)$. Unfortunately, the derivation of such an unsupervised adaptive learning algorithm is a rather difficult, tedious and computationally complex task. Due to limit of space and clarity we omit here a detailed derivation of the new algorithm which can take the basic form

$$\boxed{\begin{aligned} &\frac{dw_{ij}(t)}{dt} = \mu_i(t)\left[\lambda_i w_{ij}(t) - f_i[y_i(t)] \sum_{k=1}^{n} w_{kj}(t)\, g_k[y_k(t)]\right] \\ &(i, j = 1, 2, \ldots, n) \quad \text{with} \quad w_{ij} \neq 0 \end{aligned}} \tag{3}$$

where $\mu_i(t) > 0$ is the learning rate (typically exponentially decreasing to zero, i.e. $\mu_i(t) = \mu_0 / t^\gamma$, $\mu_0 > 0$, $\frac{1}{2} < \gamma < 1$), λ_i, are positive scaling factors (typically $\lambda_i = 1 \ \forall i$), w_{ij} are the synaptic weights, $y_i(t)$ are the outputs of the neurons described by (2a,b) and $f_i(y_i)$, $g_i(y_i)$ are nonlinear, odd and different activation functions, typically $f_i(y_i) = f(y_i)$ and $g_i(y_i) = g(y_i) \ \forall i$, e.g. $f(y) = y^3$ and $g(y) = \alpha \tanh(\beta y)$, $\alpha, \beta > 0$ (typically $\alpha = 3$, $\beta = 10$) or $f(y) = y^2 \operatorname{sign} y$ and $g(y) = \operatorname{sign}(y)$, or $g(y) = y$.

The learning algorithm (3) can be written in a compact, generalized matrix form as

$$\boxed{\frac{\mathrm{d}\boldsymbol{W}(t)}{\mathrm{d}t} = \mu(t)\left[\Lambda - f[\boldsymbol{y}(t)]\boldsymbol{g}^{T}[\boldsymbol{y}(t)]\right]\boldsymbol{W}(t)} \tag{4}$$

with any nonsingular $\boldsymbol{W}(0) \neq \mathbf{0}$, where

$$\Lambda = \mathrm{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_n\},$$

Γ is a symmetrical positive definite matrix (typically $\Lambda = I$)

$$f[\boldsymbol{y}(t)] = [f_1[y_1(t)],\ f_2[y_2(t)], \ldots, f_n[y_n(t)]]^T,$$

$$\boldsymbol{g}^T[\boldsymbol{y}(t)] = [g_1[y_1(t)],\ g_2[y_2(t)], \ldots, g_n[y_n(t)]]^T.$$

The learning matrix $\mu(t)$ is an $n \times n$ matrix with all elements positive, modified according to the following local self-adaptation rule:

$$\boxed{\mu_{ij}(t) = h_1(t) * \left| h_2(t) * \frac{\mathrm{d}w_{ij}}{\mathrm{d}t} \right|,} \tag{5}$$

where $h_1(t)$ and $h_2(t)$ are impulse responses of suitable low-pass filters and the symbol $*$ denotes the convolution operator. The rule (5) enables us to automatically adjust the learning rate individually for each synaptic weight. The main justification of using the nonlinear activation functions $f(y)$ and $g(y)$ is that they introduce higher-order statistics into the computation similarly to the known Herault-Jutten algorithm [1], [2]. As can be experimentally verified our algorithm is very robust and it can operate properly for a large class of nonlinear odd functions, e.g. for those proposed by Herault and Jutten. In fact, using nonlinear odd activation functions $f(y)$ and $g(y)$ is nothing new in our proposal. However, our proposal differs from known neural network models and associated learning algorithms in several respects. First of all, in contrast to the basic Herault-Jutten model, which is recurrent, our neural network is feedforward and it is always temporally stable independent of the initial conditions. Moreover, our neural network does not require the computation of the inversion or pseudo-inversion of matrices at every iteration step. Secondly, our learning algorithm ensures a self-normalization of the amplitude of the output signals (to a value approximately equal to 1 for $\lambda_i = \gamma_i = 1$ and $f_i(y_i) = y_i^3, g(y_i) = y_i\ \ \forall i$). In other words, the algorithm has an inherent self-adaptive gain control mechanism (due to the adaptive synaptic weights w_{ii}) which allows to separate the source signals with an extremely wide range of amplitudes.

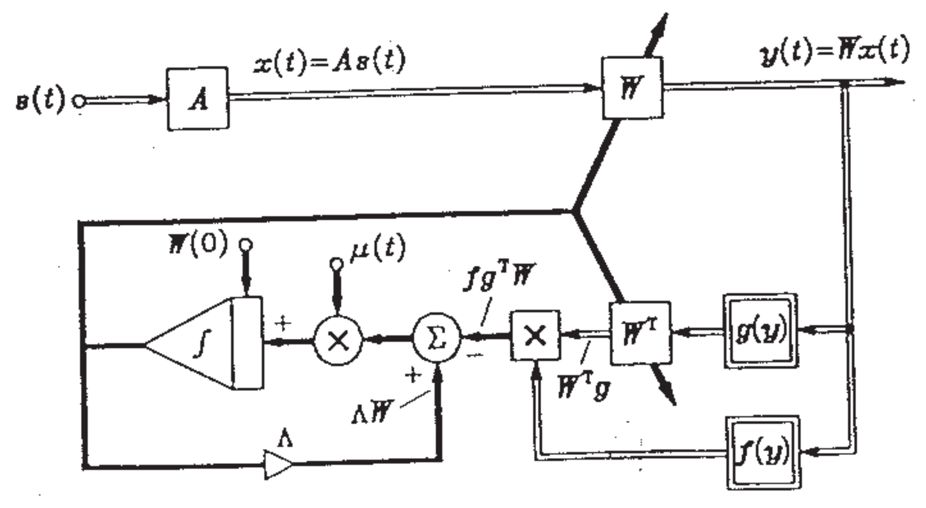

Fig. 1 Functional block diagram illustrating the implementation of the new learning algorithm

It is interesting to note that in the special case of linear activation functions, i.e. $f(y) = g(y) = y$ the proposed algorithm performs a kind of subspace PCA (principal component analysis) which provides only decorrelation of the output signals [16], [17]. However, in the general case, for which the activation functions $f(y)$ and $g(y)$ are suitable nonlinear functions the algorithms (4) and (5) perform a so-called INdependent Component Analysis (INCA). The INCA concept, first introduced in the open literature by Herault and Jutten [2] is in many cases a much more powerful or meaningful decomposition or transform than the PCA transform, since INCA provides mutual independence of the output signals, i.e. full separability of the original source signals that form a mixture of input signals whereas the PCA ensures only decorrelation of the output signals but they may be still dependent, i.e. they are almost always some linear combination of the source signals [1], [2], [16], [17].

The neural network (separation network) of Fig. 1 may be seen as an adaptive system which receives the sensor signals $\mathbf{x}(t)$ at the inputs and provides the separated (mutually independent) signals $\mathbf{y}(t)$ at the output. Note that since the algorithm (4), (5) is adaptive by nature, it can be used in situations where the mixing parameters a_{ij} and/or the shape of the waveforms are slowly varying in time.

IV. EXPERIMENTAL RESULTS

In order to check the validity and performance of the proposed neural network and the associated adaptive learning algorithm, it has been extensively simulated on a computer for a large variety of difficult separation problems. Very good results have been obtained. Due to limited space we shall present in this paper only one illustrative Example.

Example

Relatively, large uniformly distributed noise $s_5(t)$ with amplitude 1 volt is additively mixed with the following weak and very badly scaled signals:

$$s_1(t) = 10^{-5}\,\mathrm{sign}[\cos(155t)],$$

$$s_2(t) = 10^{-4}\,[\sin(800t)][\sin(60t)],$$

$$s_3(t) = 10^{-3}\,\sin[(300t) + 6\cos(60t)],$$

$$s_4(t) = 10^{-2}\,\sin 90t.$$

The mixing matrix $\mathbf{A}$ was randomly chosen as

$$\mathbf{A} = \begin{bmatrix} 0.79 & 0.19 & -0.22 & 0.12 & -0.48 \\ -0.92 & -0.90 & 0.27 & -0.93 & -0.69 \\ -0.45 & -0.91 & 1.00 & -0.78 & 0.58 \\ -0.88 & 0.99 & 0.19 & 0.10 & 0.29 \\ 0.84 & -0.33 & 0.82 & -0.26 & -0.51 \end{bmatrix}.$$

It was assumed that only the measured sensor signals are available. The learning rate $\mu(t)$ was chosen $\mu(t) = 200 = \mathrm{const}$ for $0 \leqq t \leqq t_0 = 0.25$ s (so-called "search" phase of learning) then exponentially decreasing to zero according to the relation $\mu(t) = 200\exp[-6(t - t_0)]$ for $t \geqq t_0 = 0.25$ s (the "converge" phase of the learning). The scaling parameters in algorithm (3), (4) was chosen as $\Lambda = I$ (i.e. $\lambda_i = 1$, $\forall i$) and the activation functions as $f(y_i) = y_i^2\,\mathrm{sign}(y_i)$, $g_i(y_i) = 3\tanh(10y_i)$, $\forall i$. Such a neural network (cf. Fig. 1) is able to extract (separate) the source signals in a time of a few hundred of milliseconds as illustrated in Fig. 2. Note that the output signals $y_i(t)$ $(i = 1, 2, \ldots, 5)$ are automatically normalized to approximately the same amplitude equal 1 volt. The Herault-Jutten algorithm has failed to separate such weak and badly scaled source signals.

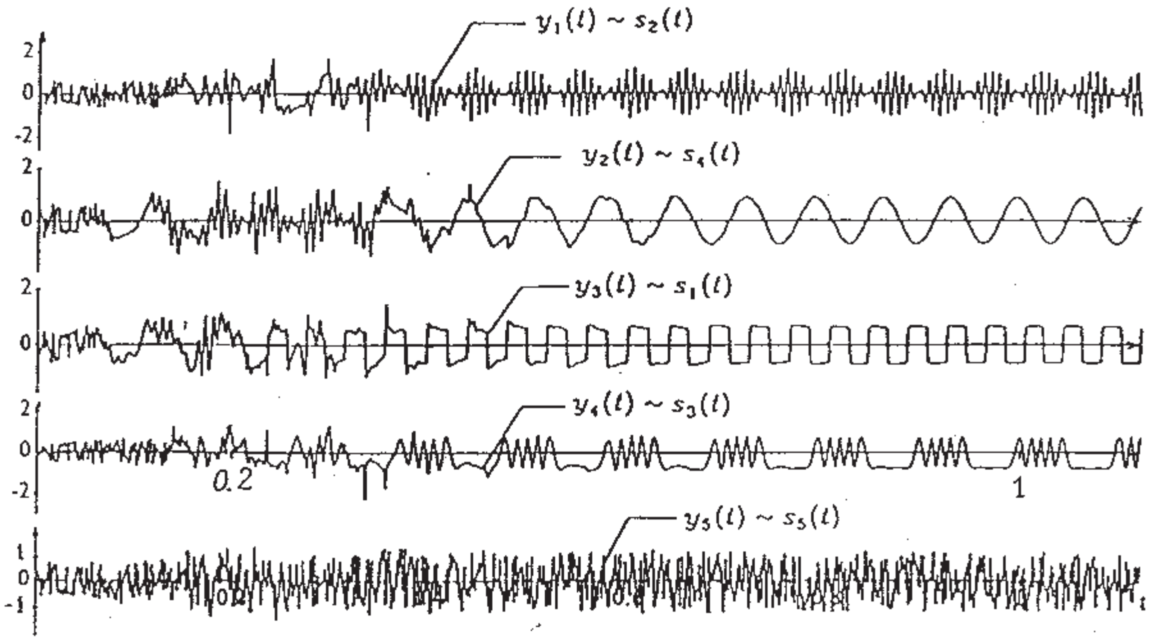

Fig. 2 Computer simulation results

V. CONCLUSIONS

A novel, efficient and robust adaptive unsupervising learning algorithm has been developed for blind separation of independent unknown source signals. The proposed algorithm is on-line because it does not necessarily require storage of input data. It can easily be implemented in VLSI technology using analog or digital techniques. Moreover, the neural network is temporarily stable independent of the initial conditions. The algorithm is especially suitable for blind separation of very badly scaled source or sensor signals and/or extremely ill-conditioned mixing matrices. Moreover, it is characterized by relatively good convergence properties. Extensive computer simulation experiments have fully confirmed the validity and high performance of the proposed learning algorithm.

References :

[1] J. Herault and C. Jutten, "Space or time adaptive signal processing by neural network models", in Proc. AIP Conf. Snowbird, UT, 1986, in the book Neural Networks for Computing, (Ed. J.S. Denker) American Institute of Physics, pp. 206-211, New York 1986.

[2] C. Jutten and J. Herault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture", Signal Processing, Vol. 24, pp. 1-20, 1991.

[3] P. Comon, C. Jutten and J. Herault, "Blind separation of sources, Part II: Problem statement", Signal Processing, Vol. 24, pp. 11-20, 1991.

[4] E.V. Vittoz and X. Arreguit, "CMOS integration of Herault-Jutten cells for separation of sources", Workshop of Analog VLSI and Neural Systems, Portland, Oregon, Kluwer Academic Press, MA, pp. 57-83, 1989.

[5] M.H. Cohen and A.G. Andreou, "Current-mode subthreshold MOS implementation of Herault-Jutten autoadaptive network", IEEE Journal Solid State Circuits, Vol. 27, pp. 714-727, May 1992.
[6] J.F. Cardoso, "Blind identification of independent components with higher order statistics", ICASSP, pp. 2109-2112, May 1989.
[7] P. Common, "Blind Identification in presence of noise", Proc. Signal Processing VI, pp. 835-838, 1992.
[8] 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 1993.
[9] L.Tong, R. Liu, V.C. Soon and Y.-F. Huang, "Indeterminacy and identifiability of blind identification", IEEE Trans. Circuits and Systems, Vol. 38, pp. 499-509, May 1991.
[10] J.L. Lacoume and P. Ruiz, "Separation of independent sources from correlated inputs", IEEE Trans. Signal Processing vol. 40, pp. 3074-3078, Dec. 1992.
[11] C. Jutten, H.L. Nguyen Thi, E. Dijkstra, E. Vittoz, J. Caelen, "Blind separation of sources: an algorithm for separation of convolutive mixtures", Int. Workshop on High Order Statistics, Chamrousse (France), pp. 273-276, July 1991.
[12] H.L. Nguyen Thi, C. Jutten, J. Caelen, "Speech enhancement: analysis and comparison of methods on various real situations", Proc. of Signal Processing VI Conf. 1992, Elsevier Science Publishers, pp. 303-306.
[13] J.C. Platt and F. Faggin, "Networks for the separation of sources that are superimposed and delayed", Advances in Neural Information Processing Systems 4, Morgan Kaufman Publishers, San Mateo, pp. 730-737, 1992.
[14] E. Weinstein, M. Feder, and A.V. Oppenheim, "Multi-channel separation by decorrelation, IEEE Trans. Speech and Audio Processing, Vol. 1, pp.405-413, 1993.
[15] D. Van Compernolle and S. Van Gerven, "Signal separation in a symmetric adaptive noise canceler by output decorrelation", In Proc. ICASSP92, San Francisco, Vol. IV, pp. 221-224, 1992.
[16] E. Oja, H. Ogawa and J. Wangviwattana, "Principal component analysis by homogeneous neural networks", Part I and II, IEICE Transactions on Information and Systems (Japan), Vol. E75-D, pp. 360-382, 1992.
[17] J. Karhunen and J. Joutsensalo, "Representation and separation of signals using PCA type learning", Neural Networks, Vol. 7, pp. 113-127, 1994.
[18] G. Burel, "Blind separation of sources: a nonlinear neural algorithm", Neural Networks, Vol. 5, pp. 937-947, 1992.
[19] A. Cichocki and R. Unbehauen, Neural Networks for Optimization and Signal Processing, Wiley-Teubner 1994, pp. 461-471.
[20] A. Cichocki and L. Moszczynski, "New learning algorithm for blind separation of sources", Electronics Letters, Vol. 28, pp. 1986-1987, 1992.