

Robust learning algorithm for blind separation of signals

A. Cichocki, R. Unbehauen and E. Rummert

Indexing terms: Neural nets, Learning systems, Signal processing

The authors present a novel, efficient, self-normalising, unsupervised adaptive learning algorithm for the on-line (real-time) separation of statistically independent unknown source signals from a linear mixture of them. In contrast to the known algorithms the new algorithm allows the separation (or extraction) of 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.

Introduction: In a variety of engineering and scientific applications observations are made of the outputs of an unknown multiple-input multiple-output linear system from which it is of interest to recover the input (source) signals. The above problem is designated as a blind separation of sources or as waveform-preserving blind estimation of multiple independent sources because the problem is to estimate the multiple source signals from an array of sensors without knowledge of the transmission channels [1-10]. The first neural networks for on-line blind separation of a linear superposition of sources has been developed by Herault and Jutten [1, 2]. The basic problem of blind separation of sources can be formulated as follows. For a given set of time variable sensor signals $x_i(t)$ described by

$$x(t) = As(t) \quad (1)$$

(where $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ is an unknown nonsingular mixing matrix, $x(t) = [x_1(t), x_2(t), \dots, x_n(t)]^T$ is a vector of measured (observed) sensor signals and $s(t) = [s_1(t), s_2(t), \dots, s_n(t)]^T$ is an unknown vector of mutually independent zero-mean signals (primary sources)) it is necessary to find or estimate a new output vector $y(t) = [y_1(t), y_2(t), \dots, y_n(t)]^T$ which is proportional to the input vector $s(t)$. Such a problem has an inherited indeterminacy [4]. This indeterminacy is characterised 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 [9].

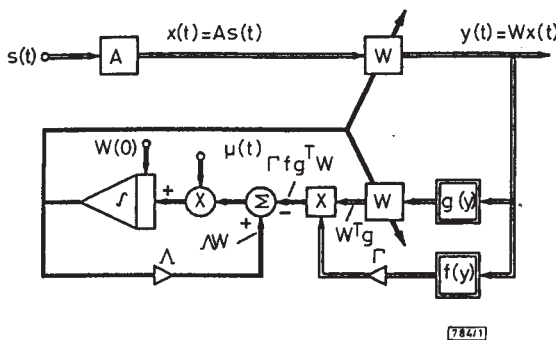


Fig. 1 Functional block diagram illustrating the implementation of new learning algorithm

New results: Self-normalising adaptive learning algorithm: In our approach we have used a single layer, feedforward neural network consisting of n neurons described by (Fig. 1)

$$y(t) = Wx(t) \quad (2)$$

where $W = [w_{ij}] \in \mathbb{R}^{n \times n}$ is the matrix of synaptic weights and $y(t) = [y_1(t), y_2(t), \dots, y_n(t)]^T$ is the vector of 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. 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. Owing to space limitation we omit here the detailed derivation of the new algorithm which can be written in the compact matrix form

$$\frac{dW(t)}{dt} = \mu(t) [\Lambda - \Gamma f[y(t)]g^T[y(t)]] W(t) \quad (3)$$

with any nonsingular $W(0) \neq 0$, where

$$\mu(t) = \text{diag}\{\mu_1(t), \mu_2(t), \dots, \mu_n(t)\}, \text{ typically } \mu_i(t) = \mu(t) > 0 \forall i$$

$\Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$, Λ is the positive definite ^{diagonal} matrix, typically $\Lambda = \Gamma = I$,

and

$$f[y(t)] = [f_1[y_1(t)], f_2[y_2(t)], \dots, f_n[y_n(t)]]^T$$

$$g^T[y(t)] = [g_1[y_1(t)], g_2[y_2(t)], \dots, g_n[y_n(t)]]$$

are vectors of odd activation functions. The main justification for using the nonlinear activation functions $f_i(y_i)$ and $g_i(y_i)$ 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, the use of nonlinear odd activation functions $f_i(y_i)$ and $g_i(y_i)$ is not novel in our proposal. However, our proposal differs from the known neural network models and associated learning algorithms in several respects. First of all, the proposed learning algorithm ensures a self-normalisation of the amplitude of the output signals. In other words, the algorithm has an inherent self-adaptive gain control mechanism which allows source signals with large amplitude ratios to be separated. Secondly, the mixing matrix can be ill-conditioned (near singular).

Such a neural network can be relatively easily realised in currently available VLSI technologies either as analogue or as digital circuits. Of course, the proposed learning algorithm can be simulated and realised on a standard digital computer if off-line solutions are acceptable. In this case the continuous-time algorithm (eqn. 3) can be converted to an iterative time-discrete process (system of difference equations). Note that the neural network does not require the computation of the inversion of matrices at every iteration step.

Experimental results: Extensive simulations on a computer for a large variety of difficult separation problems have been carried out to check the validity and performance of the proposed neural network and the associated adaptive learning algorithm. Very good results have been obtained.

Example: Relatively large uniformly distributed noise source $s_2(t)$ with amplitude 1 V is additively mixed with the following weak and badly scaled signals:

$$s_1(t) = 10^{-5} \text{sgn}[\cos(155t)]$$

$$s_2(t) = 10^{-4} [\sin(350t)][\sin(60t)]$$

$$s_3(t) = 10^{-3} \sin[(300t) + 6 \cos(57t)]$$

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

The mixing matrix $A \in \mathbb{R}^{5 \times 5}$ was chosen as the Hilbert matrix with condition number 4.7×10^5 . It was assumed that only the measured sensor signals were available. The learning rate $\mu(t)$ was chosen to be $\mu(t) = 200 = \text{const.}$ for $0 \leq t \leq t_0 = 0.25s$ (the so called 'search' phase of learning) which then exponentially decreases to zero according to the relation $\mu(t) = 200 \exp[-6(t-t_0)]$ for $t \geq t_0 = 0.25s$ (the 'converge' phase of the learning). The scaling parameters in the algorithm (eqn. 3) were chosen to be $\Gamma = \Lambda = I$ and the activation functions $f_i(y_i) = y_i^2 \text{sgn}(y_i)$, $g_i(y_i) = 3 \tanh(10y_i)$, $\forall i$. Such a neural network is able to extract (separate) the source signals in a few hundred milliseconds, as illustrated in Fig. 2. Note that the output signals $y_i(t)$ ($i = 1, 2, \dots, 5$) are automatically normalised to approximately the same amplitude, equal to 1 V. The other known on-line algorithms [1-3, 8-10] have failed to separate such weak and badly scaled source signals mixed by using the ill-conditioned matrix.

Conclusions: A novel, efficient and robust adaptive unsupervised learning algorithm has been developed for the blind separation of independent unknown source signals. The algorithm is especially suitable for the blind separation of badly scaled source or sensor

signals and/or ill-conditioned mixing matrices. Moreover, it has relatively good convergence properties. Extensive computer simulation experiments have fully confirmed the validity and high performance of the proposed learning algorithm.

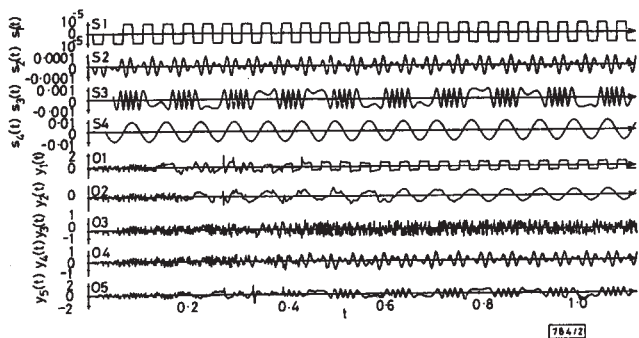


Fig. 2 Computer simulation results for the example

Acknowledgment: This research is partially supported by DFG, Germany.

© IEE 1994

7 July 1994

Electronics Letters Online No: 19940956

A. Cichocki, R. Unbehauen and E. Rummert (*Lehrstuhl für Allgemeine und Theoretische Elektrotechnik, University Erlangen-Nürnberg, Cauerstrasse 7, D-91058 Erlangen, Germany*)

References

- JUTTEN, C., and HERAULT, J.: 'Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture', *Signal Process.*, 1991, **24**, pp. 1-20
- COMON, P., JUTTEN, C., and HERAULT, J.: 'Blind separation of sources, Part II: Problem statement', *Signal Process.*, 1991, **24**, pp. 11-20
- LING, X.-T., HUANG, Y.F., and LIU, R.: 'A neural network for blind separation'. ISCAS-94, 1994, (London), pp. 69-72
- TONG, L., LIU, R., SOON, V.C., and HUANG, Y.-F.: 'Indeterminacy and identifiability of blind identification', *IEEE Trans.*, May 1991, **CAS-38**, pp. 499-509
- LACOUME, J.L., and RUIZ, P.: 'Separation of independent sources from correlated inputs', *IEEE Trans.*, 1992, **SP-40**, pp. 3074-3078
- JUTTEN, C., NGUYEN THI, H.L., DIJKSTRA, E., VITTOZ, E., and CAELEN, J.: 'Blind separation of sources: an algorithm for separation of convolutive mixtures'. Int. Workshop on High Order Statistics, July 1991, (Chamrousse, France), pp. 273-276
- PLATT, J.C., and FAGGIN, F.: 'Networks for the separation of sources that are superimposed and delayed', in: 'Advances in neural information processing systems 4' (Morgan Kaufman Publishers, San Mateo, 1992), pp. 730-737
- KARHUNEN, J., and JOUTSENSALO, J.: 'Representation and separation of signals using PCA type learning', *Neural Networks*, 1994, **7**, pp. 113-127
- CICHOCKI, A., and UNBEHAUEN, R.: 'Neural network for optimisation and signal processing' (Teubner-Wiley, 1993), pp. 461-471
- CICHOCKI, A., and MOSZCZYNSKI, L.: 'New learning algorithm for blind separation of sources', *Electron. Lett.*, 1992, **28**, pp. 1986-1987