Proc. the 1996 International Symposium on Nonlinear Theory and its Applications (NOLTA'96)
Kochi, JAPAN, pp. 377-380, 1996.

Blind Signal Extraction Using Self-Adaptive Non-linear Hebbian Learning Rule

Andrzej CICHOCKI†, Shun-ichi AMARI‡ and Ruck THAWONMAS†*

†Lab. for Artificial Brain Systems and ‡Lab. for Information Representation
Frontier Research Program, The Institute of Chemical and Physical Research (RIKEN)
2-1 Hirosawa, Wako-shi, Saitama 351-01 JAPAN

Abstract

A simple neural network with local, biologically plausible, non-linear Hebbian learning rule is developed to perform sequential extraction of independent signals from a linear mixture of them. Instead of separating all signals simultaneously, we rather extract them sequentially using a new deflation technique. It is shown that with suitable designed nonlinear functions and applying self-normalization principle, we are able to extract source signals with (specified) predefined order (e.g., extracting signals according to the maximum absolute value of normalized kurtosis). Computer simulation experiments confirm the validity of the proposed algorithm and demonstrate that it is useful extension of existing neural network solutions to blind separation problem as extraction of particular source signals with predefined stochastic characteristic is possible.

I. Introduction

In blind separation problem, the task is to recover original signals from their linear mixture without knowing the mixing coefficients. Such kind of problem occurs in many areas of science and engineering, e.g., in biomedical signal processing (ECG or EEG), speech recognition (cocktail party problem), image enhancement and in telecommunication [1,3-6,8-11].

Most of the known efficient algorithms separate signals simultaneously under an assumption that the number of sources is known [1,3,4,9,11]. In many practical situations, the number of active sources is unknown and changed in time. Moreover, we are very interested not in separation of all signals but rather in extraction of some signals with specific stochastic properties, e.g., a signal with maximum positive and/or minimum negative kurtosis.

For a non-neural, but adaptive approach, Delfosse and Loubaton proposed in [5] an algorithm to extract one of the source signals and a deflation procedure which, when used together with the extraction algorithm, successfully allows the extraction of the other sources. It was analytically proved that convergence to a correct solution can be ensured. However, the developed procedures are rather complicated and not suitable for on-line, real-time applications.

In this paper, we consider a neural network approach which enables us to successfully extract primary sources. Very recently, Hyvärinen and Oja [8] as well as Malouche and Macchi [10] developed simple learning algorithms for sequential extraction of source signals. However, all these algorithms do not allow extraction of signals with specified order according to their stochastic properties. The main objective of this paper is to develop a new learning rule which enables us to extract signals sequentially according to decreasing absolute values of their kurtosis. The second objective is to propose a simple and efficient deflation (projection) procedure which enables us to apply the same local learning rule for each processing unit (neuron).

II. A Single Neuron for Signal Extraction

Let us assume that we observe the following sensor signals at discrete time t ($t = 0, 1, 2, \dots$).

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

where $\mathbf{x}(t)$ is an $n \times 1$ sensor vector, $\mathbf{s}(t)$ is an $m \times 1$ unknown source vector, whose elements are assumed to be independent and zero-mean, and $\mathbf{A}$ is an $n \times m$ unknown mixing matrix. Moreover, it is assumed that the number of sources is unknown but satisfies the relation $m \leq n$.

In order to successfully extract desired signals, especially for ill-conditioned problems, we may preprocess the mixed signals $\mathbf{x}(t)$ by de-correlating them.

*The third author is supported by the Special Postdoctoral Researchers Program of RIKEN.

This preprocessing procedure called sphering or pre-whitening is a linear transformation $\mathbf{x}_1(t) = \mathbf{V}\mathbf{x}(t)$, where the matrix $\mathbf{V} \in \mathbf{R}^{n\times n}$ can be updated using a simple local learning rule [1,4] (see Fig. 1).

$$\mathbf{V}(t+1) = \mathbf{V}(t) + \eta(t)(\mathbf{I}_n - \mathbf{x}_1(t)\mathbf{x}_1^T(t)) \qquad (2)$$

or a global learning rule

$$\mathbf{V}(t+1) = \mathbf{V}(t) + \eta(t)(\mathbf{I}_n - \mathbf{x}_1(t)\mathbf{x}_1^T(t))\mathbf{V}(t). \qquad (3)$$

Note that the systems reach a steady (equilibrium) state iff $\mathbf{R}_{x_1x_1} = E\big[\mathbf{x}_1(t)\mathbf{x}_1^T(t)\big] = \mathbf{I}_n$.

Most of the known procedures for blind signal extraction of independent source signals are based on maximization (and/or minimization) of higher order cumulants (with order $r > 2$). The general criteria for extraction of a single source signal $y_1(t) \sim s_{i^*}(t)$ could be formulated as follows:

$$\text{minimize} \quad \mathcal{J}_1(\mathbf{w}_1) = -|\kappa_r(y_1)|^p \quad (p = 1 \text{ or } 2) \qquad (4)$$

subject to one of the following constraints: $E(y_1^2) = m_2 = 1$ or $\|\mathbf{w}_1\| = 1$ or $w_{11} = 1\ \forall t$, where $y_1 \stackrel{\text{def}}{=} \mathbf{w}_1^T\mathbf{x}_1 = \sum_{j=1}^n w_{1j}(t)x_{1j}(t)$ and $\kappa_r(y_1)$ is a rth order cumulant.

In a special case, where some a priori knowledge is available, e.g., all source signals are sub-Gaussian or super Gaussian, the above optimization criteria could be simplified [8,10]. Unfortunately, such criteria do not guarantee extraction of a signal with the maximum absolute value of normalized kurtosis. In order to extract signals with decreasing order of the absolute values of normalized kurtosis $\bar{\kappa}_4(y_1) = E\big[y_1^4\big]/E^2\big[y_1^2\big] - 3$, we formulate the problem as:

$$\text{minimize} \quad \mathcal{J}_1(\mathbf{w}_1) = -\frac{1}{4}|\bar{\kappa}_4(y_1)| \qquad (5)$$

subject to the constraint $\|\mathbf{w}_1\| = 1$.

Applying the gradient descent procedure, we obtain the new learning rule:

$$\bar{\mathbf{w}}_1(t+1) = \mathbf{w}_1(t) + \eta_1(t) sgn(\bar{\kappa}_4(t)) f\big[y_1(t)\big]\mathbf{x}_1(t), \qquad (6)$$

$$\mathbf{w}_1(t+1) = \bar{\mathbf{w}}_1(t+1)/\|\bar{\mathbf{w}}_1(t+1)\|, \qquad (7)$$

where the nonlinear function $f\big[y_1(t)\big]$ is estimated as follows

$$f\big[y_1(t)\big] = m_2^2\frac{\partial\bar{\kappa}_4(y_1)}{\partial y_1} = m_3(t) - \frac{m_4(t)}{m_2(t)}y_1(t), \qquad (8)$$

with (p = 2, 3, 4)

$$m_p(t+1) = (1-\eta(t))m_p(t) + \eta(t)y_1^p(t), \text{and} \qquad (9)$$

$$\bar{\kappa}_4(t+1) = \frac{m_4(t+1)}{m_2^2(t+1)} - 3. \qquad (10)$$

In a special case, let us assume in (8) that $\frac{m_4}{m_2} = c_1$, and replace the moment by instantaneous values. Thereby, we obtain a rough approximation of $f(y_1)$, i.e., $f(y_1) = y_1^3 - c_1y_1$ for $c_1 > 0$.

Remark 1: The above learning algorithm could be easily extended by maximizing the loss function (generalized normalized kurtosis measure) $\mathcal{J}_\beta^\alpha(\mathbf{W}_1) = \alpha^{-1}E|y_1|^\alpha/(E|y_1|^\beta)^{\alpha/\beta}$, which leads to the generalized activation function $f(y_1) = |y_1|^{(\alpha-1)}sgn(y_1) - (E|y_1|^\alpha/E|y_1^\beta|)|y_1|^{\beta-1}sgn(y_1)$. For example, for $\alpha = 1$ and $\beta = 2$, we have $f(y_1) = sgn(y_1) - (E|y_1|/E|y_1|^2)y_1$ which is a slight modification of the Sato function; for $\alpha = 4$ and $\beta = 2$, we get $f(y_1) = y_1^3 - (E|y_1|^4/E|y_1|^2)y_1$ which is a modification of constant modulus or Godard function [7]. The optimal choice of parameters α and β depends on the distribution of source signals and the trade off between the asymptotic performance (accuracy) and tracking ability.

Remark 2: The loss function could have many local minima. In order to avoid local minima and to increase convergence speed, we introduce auxiliary additive noise to the nonlinearity function, i.e., $f(\tilde{y}_1(t)) = f\big[y_1(t)+\nu_1(t)\big]$, where $\nu_1(t)$ is a noise gradually decreasing to zero.

III. On-Line Deflation Learning Rule

After successful extraction of the first independent signal $y_1(t) = s_{i^*}(t)$ (or more generally, after extraction of $k = 1, 2, \ldots, n$ sources, in which case the subscript 1 becomes k), we could apply a deflation procedure which eliminates from the mixture the already extracted signals. In other words, we employ a linear transformation (see Fig. 1)

$$\mathbf{x}_{k+1}(t) \stackrel{\text{def}}{=} \mathbf{x}_k(t) - \widetilde{\mathbf{w}}_k(t)y_k(t), \qquad (11)$$

which ensures minimization of the loss (energy) function

$$\widetilde{\mathcal{J}}_k(\widetilde{\mathbf{w}}_k) = \rho(\mathbf{x}_{k+1}) = \sum\nolimits_{j=1}^n \rho(x_{k+1,j}), \qquad (12)$$

where $y_k = \mathbf{w}_k^T\mathbf{x}_k$, $\mathbf{x}_k = \big[x_{k,1}, x_{k,2}, \ldots, x_{k,n}\big]^T$, $\rho(\mathbf{x}_k)$ is a loss function, e.g., $\rho(\mathbf{x}_k) = \frac{1}{2}\|\mathbf{x}_k\|^2$, and $\mathbf{w}_k(t+1) = \bar{\mathbf{w}}_k(t+1)/\|\bar{\mathbf{w}}_k(t+1)\|$, with $\bar{\mathbf{w}}_k(t+1) = \mathbf{w}_k(t) + \eta_k(t)sgn(\bar{\kappa}_4(t))f\big[y_k(t)\big]\mathbf{x}_k(t)$.

Minimization of the above defined loss function (12) by using the standard gradient descent method leads to the following simple learning rule:

$$\widetilde{\mathbf{w}}_k(t+1) = \widetilde{\mathbf{w}}_k(t) + \tilde{\eta}_k(t)y_k(t)g(\mathbf{x}_{k+1}(t)), \qquad (13)$$

where $g(\mathbf{x}_{k+1}) = \big[g(x_{k+1,1}), \ldots, g(x_{k+1,n})\big]^T$ and $g(x_{k+1,i}) \stackrel{\text{def}}{=} \frac{\partial\rho(\mathbf{x}_{k+1})}{\partial x_{k+1,i}}$, e.g., $g(x_{k+1,i}) = x_{k+1,i}$ for $\rho(x_{k+1,i}) = \frac{1}{2}x_{k+1,i}^2$.

IV. Learning of Learning Rate

Till now, we have assumed that the learning rates $\eta(t), \eta_k(t)$, and $\tilde{\eta}_k(t)$ are fixed positive constants or exponentially decreasing to zero. However, in non-stationary environment, the learning rate should be adjusted in a self-adaptive way. We have developed a simple learning algorithm for the learning rate [4]:

$$\eta_k(t) = \eta_k(t-1) + \alpha \eta_k(t-1)\left[\beta \left\|\mathbf{r}_k(t)\right\| - \eta_k(t-1)\right] \quad (14)$$

$$\mathbf{r}_k(t) = \mathbf{r}_k(t-1) + \rho\big(\mathbf{g}_k(t) - \mathbf{r}_k(t-1)\big), \quad (15)$$

where $\alpha > 0, \beta > 0,\ 0 < \rho < 1$ are appropriate positive constants, and $\mathbf{g}_k(t) \stackrel{\text{def}}{=} \frac{\partial \mathcal{J}_k(\mathbf{w}_k(t))}{\partial \mathbf{w}_k} = sgn\left[\bar{\kappa}_4(t)\right] f\left[y_k(t)\right] \mathbf{x}_k(t)$.

Alternative approach is to use individual learning rates η_{ij} for each synaptic weight [4]:

$$w_{kj}(t+1) = w_{kj}(t) + \eta_{kj}(t) g_{kj}(t), \quad (16)$$

where

$$\eta_{kj}(t) = \eta_{kj}(t-1) + \alpha \eta_{kj}(t-1)\left[\beta |r_{kj}(t)| - \eta_{kj}(t-1)\right], \quad (17)$$

$$r_{kj}(t) = r_{kj}(t) + \rho\big(g_{kj}(t) - r_{kj}(t-1)\big), \text{and} \quad (18)$$

$$g_{kj}(t) \stackrel{\text{def}}{=} \frac{\partial \mathcal{J}_k}{\partial w_{kj}} = sgn\left[\bar{\kappa}_4(t)\right] f\left[y_k(t)\right] x_{k,j}(t). \quad (19)$$

V. Simulation Results

To check validity of the proposed extraction and deflation methods, we performed extensive computer simulations. Table 1 describes an illustrative example with three sources and indicates the kurtosis of each source. In our simulations, we used mixing matrices **A** whose elements were randomly selected in the range [-1, 1]. We initialized all the weights such that they had the elements randomly lying between -0.1 and 0.1.

Figure 2 shows exemplary simulation results after convergence. It can be seen that the source signals have been successfully extracted in decreasing order of the absolute value of kurtosis.

VI. Conclusions

In practice, sequential extraction of source signals from their linear mixture is more crucial than simultaneous separation of them. This paper described simple, but efficient methods based on neural networks for doing so. In addition, it is more interesting to extract the source signals in such a way that the order of extraction is guaranteed to be in decreasing (or increasing) order of negative (or positive) kurtosis. Namely, a signal which is most different from the Gaussian signal is extracted first. In our simulations, the proposed methods perform this task well. The proposed methods can be extended to signal blind deconvolution.

Source	Kurtosis
$s_1(t) = 10^{-2} sign(cos(550t) - 5sin(99t))$	-2
$s_2(t) = 0.5sin^7(100t)$	0.4
$s_3(t) = cos(400t + 10sin(90t))$	-1.4

Table 1: Characteristic of source signals for the illustrative example.

References

[1] S. Amari, A. Cichocki, and H.H. Yang, "A new learning algorithm for blind signal separator", *NIPS 95*, MIT Press, Vol. 8, 1996, pp. 757-763.

[2] R. Baddeley, "Searching for filters with interesting output distributions: an uninteresting direction to explore", *Network: Comput. Syst.*, Vol. 7, 1996, pp. 409-421.

[3] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation", *To appear in IEEE Trans. on Signal Processing*, 1996.

[4] A. Cichocki, S. Amari, M. Adachi, and W. Kasprzak, "Self-adaptive neural networks for blind separation of sources", *ISCAS 96*, May 1996, Vol. II, pp. 157-160.

[5] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: A deflation approach", *Signal Processing*, Vol. 45, 1995, pp. 59-83.

[6] M. Girolami and C. Fyfe, "Blind separation of sources using exploratory projection pursuit networks", *Int. Conf. on Eng. Applic.*, London, 1996, pp. 249-252.

[7] S. Haykin (Ed.), Blind Deconvolution, Prentice Hall, New Jersey, 1994.

[8] A. Hyvärinen and E. Oja, "A neuron that learns to separate one signal from a mixture of independent sources", *ICNN 96*, June 1996, Vol. I, pp. 62-67.

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

[10] Z. Malouche and O. Macchi, "Extended anti-hebbian adaptation for unsupervised source extraction", *ICASSP 96*, May 1996, pp. 1664-1667.

[11] E. Oja and J. Karhunen, "Signal separation by nonlinear hebbian learning", *ICNN 95*, Dec 1995, pp. 83-97.

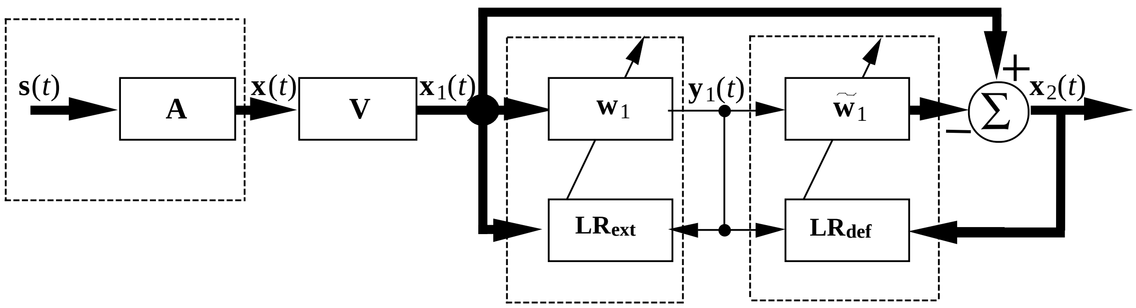

Figure 1: Functional block diagram of the 1st extraction single neuron and the 1st deflation network. $\mathbf{LR}_{\mathrm{ext}}$ and $\mathbf{LR}_{\mathrm{def}}$ stand for the extraction learning rule and the deflation learning rule, respectively.

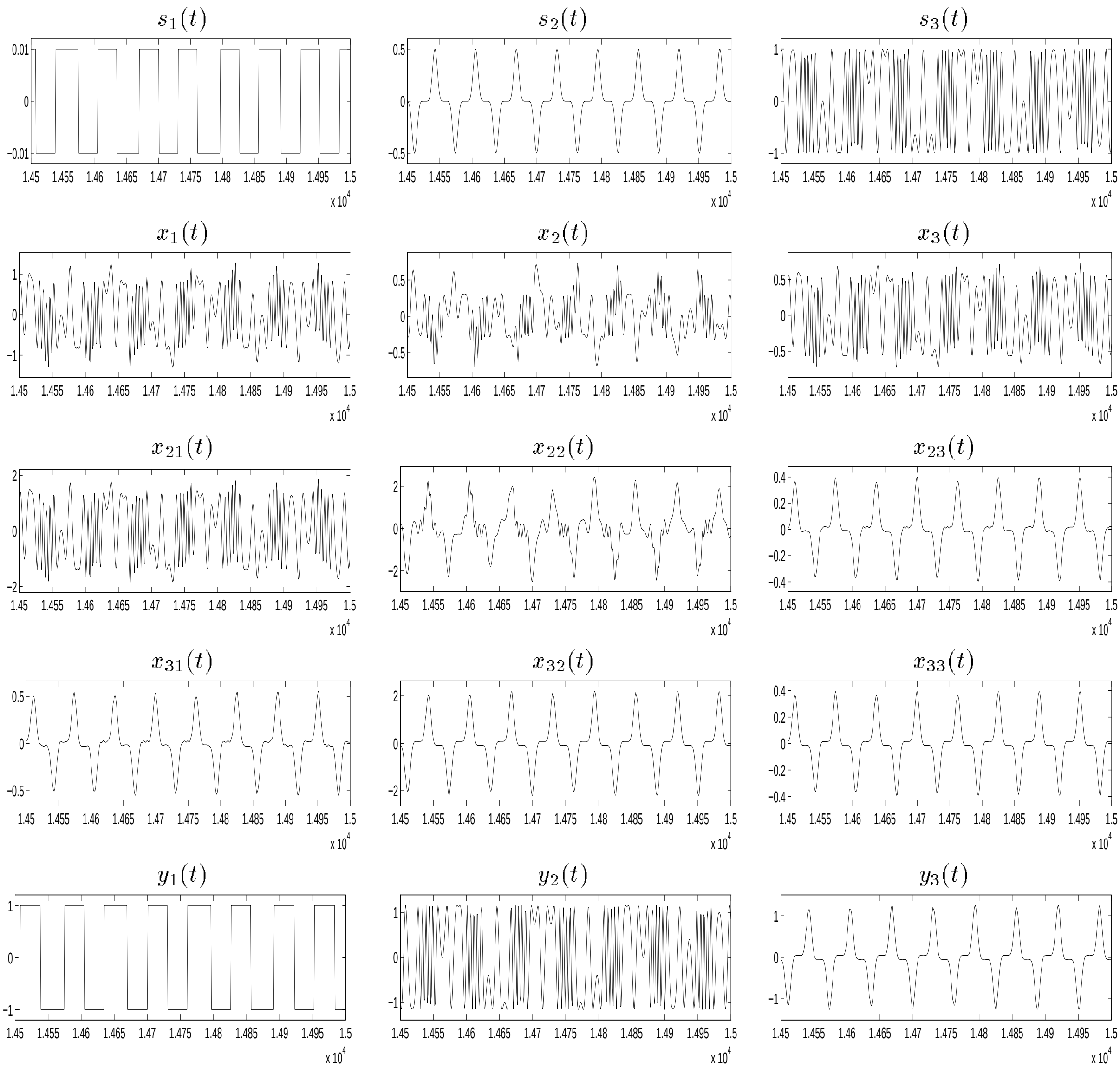

Figure 2: Results of sequential signal extraction of three sources.