Electronics Letters , Vol. 33, No. 1, pp. 64-65, 1997.

Sequential Blind Signal Extraction in Order Specified by Stochastic Properties

Andrzej CICHOCKI†, Ruck THAWONMAS†*

and

Shun-ichi AMARI‡

†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

{cia, ruck, amari}@zoo.riken.go.jp

Indexing terms: Signal Processing, Neural Networks, Learning Systems

Abstract: A new neural-network adaptive algorithm is proposed for performing extraction of independent source signals from a linear mixture of them. Using a suitable non-linear hebbian learning rule and a new deflation technique, the developed neural network is able to extract the source signals (sub-Gaussian and/or super-Gaussian) one-by-one with specified order according to their stochastic properties, namely, in decreasing order of absolute normalized kurtosis. The validity and performance of the algorithm are confirmed through extensive computer simulations.

Introduction: In many areas of science and engineering, we need to recover original signals from a linear mixture of them, received at the sensors, when

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

the mixing coefficients are not known. Such a problem is called blind separation of sources [1-7].

This problem can be formulated as follows: let the sensor signals at discrete time t ($t = 0, 1, 2, \dots$) be described by

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

where $\mathbf{x}(t)$ is an $n \times 1$ sensor vector, $\mathbf{s}(t)$ is an $m \times 1$ unknown source vector having independent and zero-mean elements, and $\mathbf{A}$ is an $n \times m$ unknown mixing matrix.

Most of the efficient algorithms in the literature perform separation of the source signals simultaneously under an assumption that the number of sources is known and equal to the number of sensors [1-4]. On the contrary, in this paper we employ a more pragmatic assumption that the number of sources is unknown but satisfies the relation $m \leq n$. A feasible solution under the employed assumption is to take a deflation approach which extracts the original sources sequentially [5, 6]. The adaptive algorithms for the blind signal extraction using deflation approach were proposed for the first time by Delfosse and Loubaton in [5]. These algorithms, however, are not suited for on-line, real-time applications due to their complexity. Recently, neural-network algorithms were proposed by Hyvärinen and Oja [6] as well as Malouche and Macchi [7]. These algorithms, however, do not ensure extraction of source signals with specified order according to their stochastic properties. In this paper, we develop a neural network algorithm which is able to extract signals in decreasing order according to absolute values of their normalized kurtosis. Such kind of extraction could be essential in many applications, especially, when the number of sources is large and only some of them are interesting or when useful signals are buried in Gaussian noises. The proposed algorithm consists of two parts: signal extraction and deflation parts, which are described in the followings.

Signal extraction: To cope with ill-conditioned cases, we first de-correlate the sensor signals $\mathbf{x}(t)$ by a linear transformation known as sphering or pre-whitening, i.e., $\mathbf{x}_1(t) = \mathbf{V}\mathbf{x}(t)$ such that $\mathbf{R}_{x_1x_1} = E[\mathbf{x}_1(t)\mathbf{x}_1^T(t)] = \mathbf{I}_n$.

In [5-7], it was shown that blind extraction of independent source signals can be achieved by maximizing (and/or minimizing) the fourth order cumulants $\kappa_4(y_1)$ subject to $E(y_1^2) = m_2 = 1$ and/or $\|\mathbf{w}_1\| = 1$, where $y_1 \stackrel{\text{def}}{=} \mathbf{w}_1^T\mathbf{x}_1 = \sum_{j=1}^n w_{1j}(t)x_{1j}(t)$.

Here, to represent the stochastic properties of the source signals, we use the normalized kurtosis $\bar{\kappa}_4(y_1) = \kappa_4(y_1)/m_2^2 = E[y_1^4]/E^2[y_1^2] - 3$. We formulate the above criterion as

$$\text{minimize} \quad \mathcal{J}_1(\mathbf{w}_1) = -\frac{1}{4}|\bar{\kappa}_4(y_1)| \tag{2}$$

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

Applying the standard gradient descent approach to (2), we derive a new learning rule

$$\bar{\mathbf{w}}_1(t+1) = \mathbf{w}_1(t) + \eta_1(t)sgn(\bar{\kappa}_4(t))f[y_1(t)]\mathbf{x}_1(t), \tag{3}$$

$$\mathbf{w}_1(t+1) = \bar{\mathbf{w}}_1(t+1)/\left\|\bar{\mathbf{w}}_1(t+1)\right\|, \tag{4}$$

where the nonlinear function $f[y_1(t)]$ is derived by

$$f[y_1(t)] = m_2^2\frac{\partial\bar{\kappa}_4(y_1)}{\partial y_1} \cong y_1^3(t) - \frac{m_4(t)}{m_2(t)}y_1(t), \tag{5}$$

and the following on-line estimations are performed ($p = 2,\ 4$)

$$m_p(t+1) = (1-\eta(t))m_p(t) + \eta(t)y_1^p(t), \text{and} \tag{6}$$

$$\bar{\kappa}_4(t+1) = \frac{m_4(t+1)}{m_2^2(t+1)} - 3. \tag{7}$$

It should be noted that the activation function $f[y_1(t)]$ (in general, $f[y_k(t)]$) is not fixed but changed during the learning process depending on the value of estimated moments $m_4(t)$ and $m_2(t)$.

To ensure extraction of a signal with the maximum absolute value of normalized kurtosis, we add auxiliary noise to the nonlinear function, i.e., $f(\tilde{y}_1(t)) = f[y_1(t) + \nu_1(t)]$, where $\nu_1(t)$ is a Gaussian noise gradually decreasing to zero. Introduction of Gaussian noises does not affect the loss function in (2) because the normalized kurtosis of Gaussian noises is zero. Finally, the above learning rule (3)-(7) could be easily generalized for extraction of next source signals, say, $y_2, \ldots, y_k$ in cooperation with the deflation procedure described below.

Deflation: Now let us suppose that y_k has been extracted, where the subscript k also indicates the total number of source signals being extracted so far. We exploit the knowledge of y_k to generate the new input vector $\mathbf{x}_{k+1}$ which will not include the already extracted signals $(y_1, \ldots, y_k)$. This can be easily done by the linear transformation

$$\mathbf{x}_{k+1}(t) \stackrel{\text{def}}{=} \mathbf{x}_k(t) - \tilde{\mathbf{w}}_k(t) y_k(t), \tag{8}$$

which minimizes of the loss function (generalized energy)

$$\widetilde{\mathcal{J}}_k(\tilde{\mathbf{w}}_k) = \rho(\mathbf{x}_{k+1}) = \sum_{j=1}^{n} \rho(x_{k+1,j}), \tag{9}$$

where $y_k = \mathbf{w}_k^T \mathbf{x}_k$, $\mathbf{x}_k = [x_{k,1}, x_{k,2}, \ldots, x_{k,n}]^T$, $\rho(\mathbf{x}_k)$ 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 $\bar{\mathbf{w}}_k(t+1) = \mathbf{w}_k(t) + \eta_k(t) sgn(\bar{\kappa}_4(t)) f[\tilde{y}_k(t)] \mathbf{x}_k(t)$.

We then derive the following rule by applying gradient descent to (9)

$$\tilde{\mathbf{w}}_k(t+1) = \tilde{\mathbf{w}}_k(t) + \tilde{\eta}_k(t) y_k(t) g(\mathbf{x}_{k+1}(t)), \tag{10}$$

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

Computer simulations: To check the validity and performance of the proposed algorithm, we performed extensive computer simulations for a variety of problems. Here, due to limit of space, we only present an illustrative example of typical results. In our simulations, we used mixing matrices $\mathbf{A}$ whose elements were randomly assigned in the range [-1, 1]. We initialized all the weights such that they had random elements lying between -0.1 and 0.1. We set the initial learning rates for extraction $\eta_k(t)$ and deflation $\tilde{\eta}_k(t)$ to 0.01 and 0.05, respectively. To these learning rates, we applied the learning of learning rate scheme as discussed in [4]. For separation, we added the following noise $\nu_k(t)$ to the nonlinear function $f(y_k(t))$.

$$\nu_k(t) = \begin{cases} 5n_k(t) & \text{for } t \leq 100, \\ 0 & \text{for } t > 100, \end{cases} \tag{11}$$

where $n_k(t)$ is a Gaussian noise with mean 0.0 and variance 1.0. Finally, we alternatively performed extraction and deflation in a quasi parallel manner with a delay of 3000 thousand time steps for each extracted signal y_k.

Fig. 1 shows the extraction of four signals, where $\bar{\kappa}_4(s_1) = -1.2, \bar{\kappa}_4(s_2) = -1.5, \bar{\kappa}_4(s_3) = 0.49$, and $\bar{\kappa}_4(s_4) = 0.04$. Because signals are extracted in decreasing order of absolute normalized kurtosis, we can terminate the algorithm when the newly extracted signal, say, y_k is closed to Gaussian, i.e., when $|\bar{\kappa}_4(y_k)| < \sigma$, where σ is a threshold of small value.

Conclusions: We proposed a new adaptive algorithm for blind signal extraction, consisting of separation and deflation parts, and showed its validity and high performance through computer simulations. The proposed algorithm does not require the knowledge of the number of sources in advance. In addition, with noise being added to the output of the separation neurons, the algorithm ensures extraction of source signals in decreasing order of their normalized kurtosis. These two features make the algorithm promising to be used in practise or applied to blind deconvolution.

References

[1] 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.

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

[3] 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.

[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] 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.

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

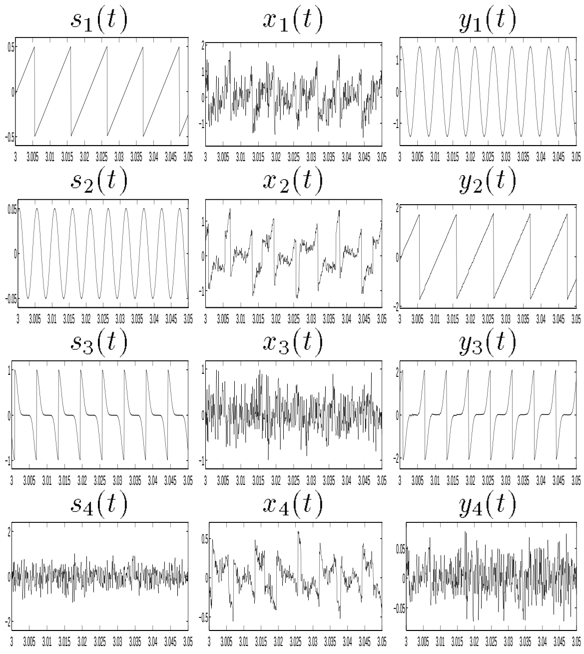

Figure 1: *A typical result of extraction of four sources (with 10 KHz sampling rate), where* s_k, x_k, *and* y_k *stand for the k-th source, mixed, and extracted signals, respectively.*