

BLIND EXTRACTION OF SOURCE SIGNALS WITH SPECIFIED STOCHASTIC FEATURES

Ruck THAWONMAS *and Andrzej CICHOCKI

Laboratory for Artificial Brain Systems
FRP, The Institute of Chemical and Physical Research (RIKEN)
2-1 Hirosawa, Wako-shi, Saitama 351-01, Japan
{ruck, cia}@zoo.riken.go.jp

ABSTRACT

We present a neural-network approach which allows sequential extraction of source signals from a linear mixture of multiple sources in the order determined by absolute values of normalized kurtosis. To achieve this, we develop a non-linear Hebbian learning rule for extraction of a single signal. We discuss several techniques which enable extraction of signals not randomly but in the desired order. To prevent the same signals from being extracted several times, a robust deflation technique is used which eliminates from the mixture the already extracted signals. Extensive computer simulations confirm the validity and high performance of our method.

1. INTRODUCTION

Separation of original source signals from a linear mixture of them is a problem of interest in many applications such as biomedical signal processing (ECG or EEG), speech recognition (cocktail party problem), image enhancement and telecommunication [1,3-10].

Most neural solutions to this problem in the literature [1,3,5,7,9] take an approach which attempts to separate all of the original source signals simultaneously. This approach employs an assumption that the number of sources is known and usually equal to the number of sensors. In practice, however, the number of active sources is not known *a priori*. Moreover, in some applications we may have a mixture of hundreds or thousands of signals and we want to extract from this mixture only the most "interesting" or significant signals with specified stochastic properties. Such signals usually bring the most useful information [2].

Our objective in this paper is to extract at a time the source signal which is most deviated from Gaussian signals. From practical point of view, this objective is crucial. With a deflation procedure discussed in [4] which eliminates from the mixture the already extracted signals, we can extract significant source signals step-by-step. In addition, we can choose to terminate the extraction as soon as a newly extracted signal is closed to Gaussian or the amplitude of every signal in the new mixture is closed zero. As a result, the *a priori* knowledge of the number of sources is not needed.

To extract a single source signal, methods for blind equalization or deconvolution problems [11] can be used, as done

in [4,6,8,10]. Namely, extraction of an independent source signal can be achieved by maximizing (and/or minimizing) the fourth order cumulants subject to certain constraints. To prevent duplicate extraction, an adaption of the orthogonal Schur eigenvalue deflation technique was used in [6]. This technique is, however, not suited for on-line, real-time applications due to its rather high complexity. In [8], the hierarchical orthogonalization technique was used. However, it is rather difficult to choose proper values for the coefficients corresponding to the orthogonalizing feedback terms, unless *a priori* knowledge of the kurtosis of sources signals is known.

In the rest of this paper, we present on-line extraction and deflation learning algorithms in Section 2. To ensure extraction of source signals in the desired order, we discuss several techniques in Section 3. We show exemplary simulation results in Section 4, and summarize the paper by conclusions and open problems in Section 5.

2. ON-LINE EXTRACTION AND DEFLATION LEARNING ALGORITHMS

The observed sensor signals at discrete time t can be expressed through the following linear model

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad t = 0, 1, 2, \dots \quad (1)$$

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 signals, and $\mathbf{A}$ is an $n \times m$ unknown mixing matrix. Now let us consider a single processing unit $y_1 = \mathbf{w}_1^T \mathbf{x} = \sum_{j=1}^n w_{1j} x_j$. The unit successfully extracts a source signal if $\mathbf{w}_1(t) = \mathbf{w}_1$ satisfy the relation $\mathbf{w}_1^T \mathbf{A} = \mathbf{e}_k$, where $\mathbf{e}_k$ denotes the k -th column of the $n \times n$ identity matrix $\mathbf{I}$.

A possible loss (contrast) function can be formulated as [11]:

$$\mathcal{J}_1(\mathbf{w}_1) = -\frac{1}{4} |\kappa_4(y_1)| \quad (2)$$

subject to the constraint that $\mathbf{R}_{\mathbf{x}\mathbf{x}} = \mathbf{I}$ (i.e., the mixing signals are prewhitened or decorrelated), where $\kappa_4(y_1)$ is

normalized kurtosis defined as $\kappa_4(y_1) = \frac{E[y_1^4]}{E^2[y_1^2]} - 3$. Min-

imization of such a loss function according to stochastic gradient descent (SGD) approach leads to a simple learning rule:

$$\mathbf{w}_1(t+1) = \mathbf{w}_1(t) + \mu_1(t) f(y_1(t)) \mathbf{x}_1(t), \quad (3)$$

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

where $\mu_1(t) > 0$ is a learning rate, $\mathbf{x}_1(t)$ is a prewhitened version of the mixing signals, $f(y_1(t)) = \text{sign}[\kappa_4(y_1(t))]\left\{y_1(t) - \frac{m_2(y_1(t))}{m_4(y_1(t))}y_1^3(t)\right\} \frac{m_4(y_1(t))}{m_2^3(y_1(t))}$ with $m_p(y_k(t)) \stackrel{\text{def}}{=} E[|y_k(t)|^p]$ and $p = 2, 4$, and $y = \mathbf{w}_1^T \mathbf{x}_1$. The high order moments m_2, m_4 and the sign of kurtosis κ_4 can be estimated on-line using the following averaging formula:

$$m_p(y_k(t)) = (1 - \eta)m_p(y_k(t-1)) + \eta y_k^p(t), \quad (4)$$

where η is a small positive constant. In the special case when statistics of signals are known, we could use fixed nonlinearity with $\alpha(t) = \frac{m_2(y_1(t))}{m_4(y_1(t))} = \text{constant}$.

After a successful extraction of the first source signal $y_1(t) \sim s_j(t)$ ($j \in \{1, \dots, n\}$), we can apply a deflation procedure which removes previously extracted signals from the mixture. This means that we are looking for such an on-line linear transformation given by

$$\mathbf{x}_{k+1}(t) = \mathbf{x}_k(t) - \tilde{\mathbf{w}}_k(t)y_k(t) \quad k = 1, 2, \dots \quad (5)$$

which ensures minimization of the generalized energy (loss) function

$$\tilde{J}_k(\tilde{\mathbf{w}}_k) = \frac{1}{2} \|\mathbf{x}_{k+1}\|^2, \quad (6)$$

where

$y_k = \mathbf{w}_k^T \mathbf{x}_k$, $\mathbf{x}_k = [x_{k1}, x_{k2}, \dots, x_{kn}]^T$, $\mathbf{w}_k(t+1) = \mathbf{w}_k(t) + \Delta \mathbf{w}_k(t)$, $\Delta \mathbf{w}_k(t) \stackrel{\text{def}}{=} \mu_k(t)f(y_k(t))\mathbf{x}_k(t)$, and $f_k(y_k(t)) = \text{sign}[\kappa_4(y_k(t))]\left\{y_k(t) - \frac{m_2(y_k(t))}{m_4(y_k(t))}y_k^3(t)\right\} \frac{m_4(y_k(t))}{m_2^3(y_k(t))}$. The last term m_4/m_2^3 can be absorbed by learning rate $\mu_k(t)$ since it is always positive. Minimization of the above defined loss function leads to a simple learning rule:

$$\tilde{\mathbf{w}}_k(t+1) = \tilde{\mathbf{w}}_k(t) + \tilde{\mu}_k(t)y_k(t)\mathbf{x}_{k+1}(t) \quad k = 1, 2, \dots \quad (7)$$

The procedure can be continued until all of the estimated source signals are recovered, i.e., until the normalized kurtosis of the newly extracted signal $\kappa_4(y_k(t))$ or the amplitude of each deflated signal $\mathbf{x}_{k+1,i}$ are below given thresholds. This means that it is not necessary to know the number of source signals in advance.

3. TECHNIQUES FOR AVOIDING LOCAL MINIMA

It can be shown that the loss function in (2) has no spurious local minima [11]. In other words, each local minimum corresponds to a successful extraction of a single source. However, only global minima correspond to an extraction of the signal with maximum value of normalized kurtosis. Unfortunately, SGD does not guarantee to achieve global (optimal) minima. In order to avoid being stuck in local minima, we could apply several techniques.

The first novel technique is to apply the nonlinear function $f(y) = \text{sign}[\kappa_4(y)]\{y - \alpha(t)y^3\}$ with $\alpha(t)$ gradually changing from zero to the value determined by $\alpha = m_2(y)/m_4(y)$ during the learning process. However, this technique is not suitable for highly nonstationary signals.

Alternative techniques for avoiding local minima and improving performance of learning are to add to inputs, outputs and/or synaptic weights a Gaussian uncorrelated noise signals, i.e.,

(a) additive noise for outputs

$$f(\hat{y}_k(t)) = f[y_k(t) + \nu_k(t)], \quad (8)$$

(b) additive noise for inputs

$$\hat{x}_{kj}(t) = x_{kj}(t) + \nu_{kj}(t), \quad (9)$$

(c) additive noise for synaptic weights

$$\hat{w}_{kj}(t+1) = w_{kj}(t) + \Delta w_{kj} + \nu_{kj}(t). \quad (10)$$

The basic idea is to change the shape of loss functions by incorporating auxiliary noise. Of course, this noise should be gradually decreased to zero in analogy to the concept of the simulated annealing algorithm [5,12].

The approach is illustrated by a simple 2-D example with only two sources (s_1 and s_4 in Fig. 2.a) mixed by mixing matrix $\mathbf{A} = [1.0, 0.5; 0.5, 1.0]$. Fig. 1 shows the shape of the loss function for different levels of noise added to the output where $\nu_1 = 0, n_1, 3n_1$, and $5n_1$ in (a), (b), (c), and (d), respectively, and n_1 is a Gaussian noise with mean 0.0 and variance 1.0.

Combinations of these techniques are also possible. We are currently investigating the effects of these techniques from theoretical view points.

4. COMPUTER SIMULATIONS

We confirmed the validity and performance of our methods using extensive computer simulations for a variety of problems. Below, due to limit of space, we only present an illustrative example of typical results in Fig. 2. In this example, we added the following noise $\nu_k(t)$ to the nonlinear function $f(y_k(t)) + \nu_k(t)$ with $\nu_k(t) = 10n_k \exp(-0.005t)$ for $t \leq 1000$; 0 otherwise, where n_k is a Gaussian noise with mean 0.0 and variance 1.0. The learning rates for extraction μ_k and deflation $\tilde{\mu}_k$ were initialized to 0.01 and 0.001, respectively, which were then adaptively changed according to the adaptive learning rate scheme in [4]. Execution of a next processing unit is delayed for 5000 time steps after initiating execution of the previous unit.

Fig. 2 shows the results of extraction of four unknown signals (Fig. 2.a) from a mixture of them received at four sensors (Fig. 2.b) with the following normalized kurtosis: $\kappa_4(s_1) = -1.5$, $\kappa_4(s_2) = 0.4$, $\kappa_4(s_3) = -1.49$, and $\kappa_4(s_4) = -2$. The prewhitened signals, obtained by a standard algorithm, are shown in Fig. 2.c. The extracted outputs are shown in Fig. 2.d and the deflated signals are given in Fig. 2.e. Visual comparison of Figs. 2.a and 2.d confirms that the source signals were successfully extracted, and, in addition, in decreasing order of absolute values of normalized kurtosis. The fact that the number of source signals was four, which was not known to the system, was confirmed by having the very small amplitude of every element in $\mathbf{x}_5$.

REFERENCES

- [1] S. Amari, A. Cichocki, and H.H. Yang, "A new learning algorithm for blind signal separation," *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*, Dec 1996.
- [4] A. Cichocki, S. Amari and R. Thawonmas, "Blind signal extraction using self adaptive non-linear hebbian learning rule," *NOLTA '96*, October 1996, pp. 377-380.
- [5] A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*, John Wiley, Chichester, 1993.
- [6] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: A deflation approach," *Signal Processing*, Vol. 45, 1995, pp. 59-83.
- [7] M. Girolami and C. Fyfe, "Blind separation of sources using exploratory projection pursuit networks," *Int. Conf. on Eng. Applications of Neural Networks*, 1996, pp. 249-252.
- [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] O. Shalvi and E. Weinstein "Universal method for blind deconvolution," *Blind Deconvolution*, S. Haykin, Ed., pp. 121-180, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1994.
- [12] S. Kirkpatrick, C. Gelatt, and M. Vecchi "Optimization by simulated annealing," *Science* 220, 1983, pp. 671-680.

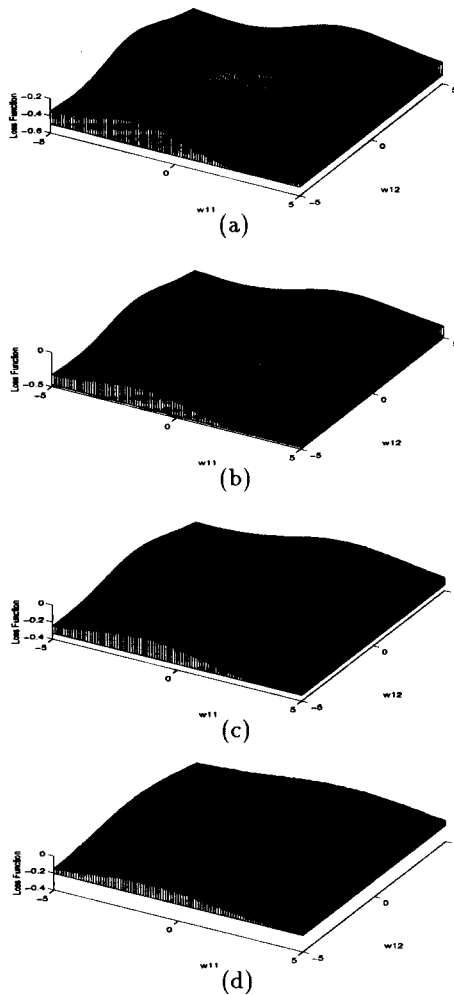
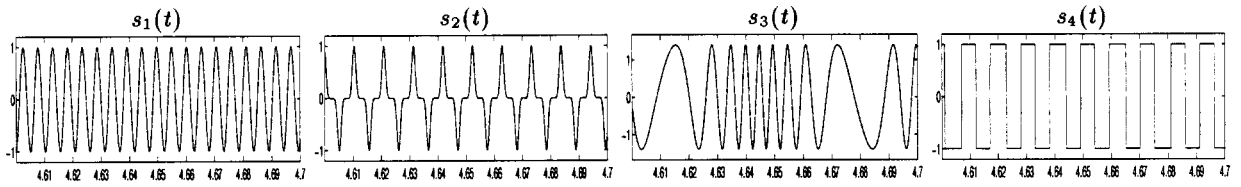


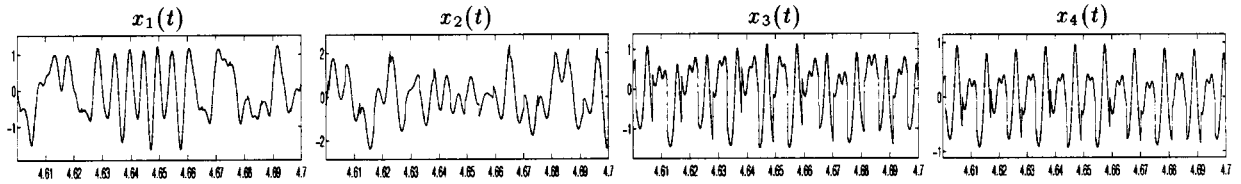
Figure 1. Loss functions without noise injection (a) and with different levels of Gaussian noise added to the output (b) - (d).

5. CONCLUSIONS

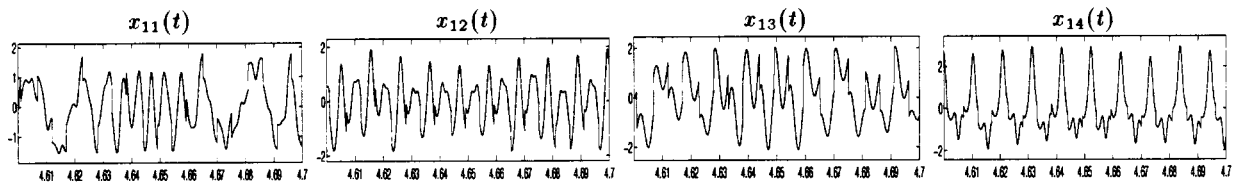
We have presented a neural-network approach for on-line blind signal extraction. Several techniques have been discussed which allow to avoid local minima and therefore enable to extract source signals with specified order, i.e., in decreasing order according to absolute values of their normalized kurtosis. Our approach has the following features: It uses a simple cost function (absolute value of normalized kurtosis) without any constraints. From this cost function, simple adaptive nonlinear functions are derived. These nonlinear functions change their shapes during the learning process. Moreover, the proposed algorithms are able to extract signals both sub-Gaussian and super-Gaussian. The developed learning algorithms are purely local and are biologically plausible; they could be considered as a generalization or extension of Hebbian/anti-Hebbian rules. The proposed methodology can be extended to multi-channel blind signal deconvolution or generalized to complex-valued signals.



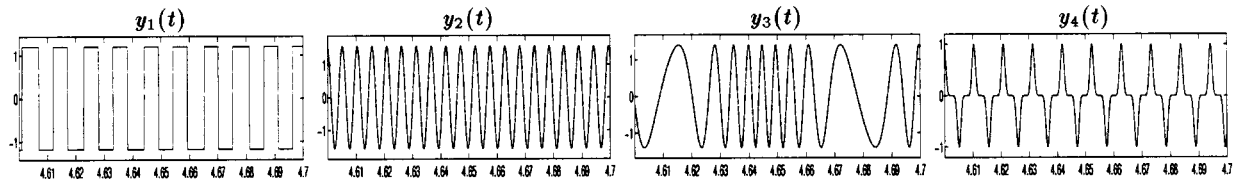
(a) original source signals



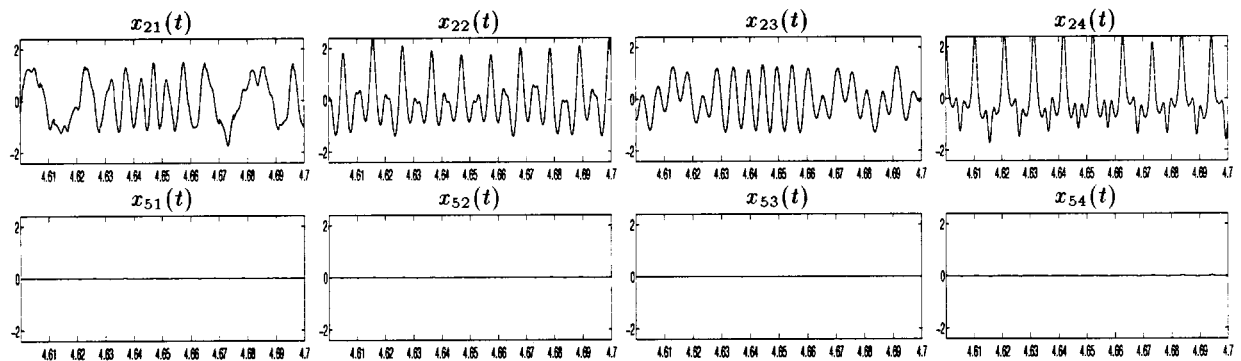
(b) mixed signals



(c) prewhitened signals



(d) extracted signals



(e) deflated signals after the 1-th and 4-th processing unit, respectively

Figure 2. A typical result of extraction of four sources mixed in four sensors (with 10 KHz sampling rate), where s_k , x_k , and y_k stand for the k -th source, mixed, and extracted signals, respectively.