BLIND SEPARATION OF NONSTATIONARY SIGNALS USING NEURAL NETWORKS

Slavica Zarkula, Vlastimir Pavlović, *Faculty of Electronic Engineering, University of Niš*
Branimir Todorović, *Faculty of Occupational Safety, University of Niš*

Abstract - *The aim of this work is to examine the performances of the linear self-organized neural network for blind separation of nonstationary signals. The network performs signal separation from their linear combination in the absence of any prior information about the signal properties, except the fact that the source signals are statistically independent of each other. In order to achieve faster convergence, the basic learning algorithm is modified using the additional adaptive momentum term. To illustrate the performances of the proposed method for blind separation, the network is applied to the nonstationary seismic signals.*

I INTRODUCTION

The problem of how to separate source signals from their composite linear or nonlinear transformation appears in a number of fields which involve analysis and processing of the observed signals, such as speech, biomedical signals, etc. If the unknown source signals have to be recovered from their observed composite signals without any prior information about the properties of source signals and transformation coefficients, it is a question of blind separation. Although it seems that this problem can not be solved in general, several methods based on some prior assumptions on the statistical properties of the source signals have been developed. Conventional methods, applicable to the non-Gaussian signals, which offer the most general approach to the problem of blind separation, are based on the evaluation of higher-order statistics of the observed signals in order to find out the inverse of transformation and thus directly yield to the component signals. A neural approach, described in [1], enables both the linear and nonlinear decorrelation using the higher-order cumulants of the observed signals. The method proposed in [2], which is also a neural one, acquires the recovery of nonstationary source signals using only the second-order moments of the composite signals without regarding the type of statistical distribution of the source signals. In the particular case of periodic signals, the decomposition of a composite signal into its periodic components can be achieved by using the Singular Value Ratio Spectrum [3].

The objective of this paper is to examine the performances of neural network for blind separation of non-stationary signals originally proposed in [2]. The network performs a separation of the observed composite signals into their constituents using the fact that the source signals are statistically independent of each other. The network is self-organized in the sense that it changes its connection weights according to the anti-Hebbian rule in a direction that reduces the correlation between its outputs. In order to achieve faster convergence, the basic learning algorithm, proposed in [2], is modified by employing the additional adaptive momentum term. The performances of the proposed modified algorithm are demonstrated on the basis of real seismic data applied to the network inputs. In this case, the signal separation has been achieved in about four epochs through the actual sequences.

II SEPARATION NEURAL NETWORK

In this Section, we briefly describe the linear self-organized neural network with lateral connections for blind separation of nonstationary signals. If the number of source signals is equal to the number of observed composite signals, i.e. sensors, the network performs a decomposition of composite signals, represented by a linear combination of source signals, into their constituents.

In matrix notation, the observed composite signals $s_i(t), i=1,...,N$, considered in this work, are given by:

$$s(t) = Ax(t) \tag{1}$$

where $s(t)=[s_1(t), ..., s_N(t)]^T$ is a vector of composite signals and $A=[a_{ij}]$ is the affine transformation matrix, which is assumed to be nonsingular; $x(t)=[x_1(t), ..., x_N(t)]^T$, $x_j(t), j=1...N$ are reffered to as statistically independent source signals with zero means, which implies for the covariance matrix $R(t)$ of $x(t)$ to be a diagonal one:

$$R(t) = diag\{r_1(t),...,r_N(t)\} = diag\{\langle x_1^2(t)\rangle,...,\langle x_N^2(t)\rangle\}$$

It has been shown that, if $r_i(t)/r_j(t)$ change in time, it is possible to recover source signals from the observed signals in the absence of any prior knowledge about the properties of A and $x_j(t)$, except the fact that the source signals $x_j(t)$ are uncorrelated with each other [2]. However, the recovery of source signals has an ambiguity which exists in the nonzero scale parameters, d_i, since $d_1x_{p_1},...,d_Nx_{p_N}$, where $\{p_1,...p_N\}$ is any arbitrary permutation of $\{1,...N\}$, may also be considered source signals. Due to this ambiguity, blind separation is defined as any process which yields to the following type signals:

$$\overline{x}(t) = DPx(t) \tag{2}$$

D is a diagonal matrix $D=diag\{d_1,...,d_N\}$, and P is a permutation matrix. The proposed network for blind separation receives sensor signals as inputs and provides output signals $y_i(t), i=1,\dots,N$ as estimates of the source signals $x_i(t)$. In matrix notation, the dynamics of each output unit is given by the first-order linear differential equation:

$$\tau \frac{d\mathbf{y}(t)}{dt} + \mathbf{y}(t) = \mathbf{s}(t) - \mathbf{C}\mathbf{y}(t) \qquad (3)$$

where the matrix $\mathbf{C} = [c_{ij}]$ denotes the mutual lateral connections between the output units, which are adapted according to the anti-Hebbian rule. If we assume that the time constant τ is sufficiently small, the equation (3) can be replaced by:

$$\mathbf{y}(t) = (\mathbf{I} + \mathbf{C})^{-1}\mathbf{s}(t) \qquad (4)$$

The aim is to determine matrix C so that the outputs $y_i(t)$ be the estimates of the source signals $x_i(t)$. According to the equations (2) and (4), C takes the general form:

$$\mathbf{C} = \mathbf{A}\mathbf{P}^T\mathbf{D}^{-1} - \mathbf{I} \qquad (5)$$

With respect to the additional constraint $c_{ii} = 0$, which leads to $\mathbf{D} = diag\{\mathbf{A}\mathbf{P}^T\}$, C becomes:

$$\mathbf{C} = \mathbf{A}\mathbf{P}^T\left[diag\{\mathbf{A}\mathbf{P}^T\}\right]^{-1} - \mathbf{I} \qquad (6)$$

Matrix C, obviously, can not be calculated directly from (6), since the transformation matrix A is unknown. However, it has been proved [2] that the result given by (6) is equivalent to the following statements:

(I) in the output covariance matrix, the terms $\langle y_i(t) y_j(t) \rangle$, $i,j = 1,\dots,N; i \neq j$, where $\langle * \rangle$ denotes the ensemble average of $*$, are all zeros at any time instant t;

(II) the non-negative scalar function given by:

$$Q(\mathbf{C}, \mathbf{R}(t)) = \frac{1}{2}\left\{\sum_{i=1}^{N}\log\langle y_i^2(t)\rangle - \log\left|\langle \mathbf{y}(t)\mathbf{y}(t)^T\rangle\right|\right\} \qquad (7)$$

where $\mathbf{R}(t) = diag\{r_1(t),\dots,r_N(t)\}$ is the covariance matrix of the network outputs $\mathbf{y}(t)$, takes zero at any time instant t.

This means that signal separation can be realized through the learning process by determining C so that $Q(\mathbf{C}, \mathbf{R}(t))$ has the minimum every time:

$$T\frac{dc_{ij}}{dt} = -\frac{\partial Q(\mathbf{C}, \mathbf{R}(t))}{\partial c_{ij}}, \quad i,j = 1,\dots N; i \neq j \qquad (8)$$

According to the equation (8), $Q(\mathbf{C}, \mathbf{R}(t))$ attains its minimum with a steepest descent search. T is the time constant that controlles the learning speed. If T is chosen to be large, i.e. C varies very slowly, (8) may be approximated by the following equation,

$$T\left\langle\frac{d\mathbf{C}}{dt}\right\rangle = \left(\mathbf{I} + \mathbf{C}^T\right)^{-1}\left\{\left(diag\langle \mathbf{y}(t)\mathbf{y}(t)^T\rangle\right)^{-1}\langle \mathbf{y}(t)\mathbf{y}(t)^T\rangle - \mathbf{I}\right\} \qquad (9)$$

which describes the way of adapting the connection weights in the sense of the statistical average. The implementation of (9) requires for $diag\langle \mathbf{y}(t)\mathbf{y}(t)^T\rangle$, or $\langle y_i^2(t)\rangle$, to be evaluated in real time.The basic learning algorithm, developed from (9), uses the moving average $\phi_i(t)$ in order to estimate $y_i^2(t)$ in real time, and is given by the equations [2]:

$$T'\frac{d\phi_i(t)}{dt} + \phi_i(t) = y_i^2(t), \; i = 1,\dots,N \qquad (10)$$

$$T\frac{d\mathbf{C}}{dt} = \left(\mathbf{I} + \mathbf{C}^T\right)^{-1}\left\{\left(diag\,\Phi(t)\right)^{-1}\mathbf{y}(t)\mathbf{y}(t)^T - \mathbf{I}\right\} \qquad (11)$$

where $\Phi(t) = diag\{\phi_1(t),\dots,\phi_N(t)\}$ and $\phi_i(t)$ denotes the moving average of $y_i^2(t)$ in time interval T', in which $r_i(t)$ are assumed to be constant.

III MODIFICATION OF LEARNING ALGORITHM

In order to achieve faster convergence, we have applied the "heavy ball" method [4] as a modification of the steepest descent algorithm described in the previous section. This method employes the fact that the connection weights c_{ij}, by means of which the minimization is performed, possess their own "mass", i.e. inertion. This feature of the parameters being adapted may be expressed by means of an additional term in the equations describing the learning algorithm. Due to this, it is possible to fasten the convergence of the learning algorithm and to enable the escape from the local minima during training. The modified learning algorithm, developed using the Euler approximation of the equation (8), is given by:

$$c_{ij}(k+1) = c_{ij}(k) - \beta\frac{\partial Q(\mathbf{C}, \mathbf{R}(k))}{\partial c_{ij}} + \gamma_{ij}(k)\left(c_{ij}(k) - c_{ij}(k-1)\right) \qquad (12)$$

The term $\gamma_{ij}(k)$ represents the "mass" associated with any state of the connection weights during adaptation. The values of $\gamma_{ij}(k)$, which depend on the $\partial Q(\mathbf{C}, \mathbf{R}(k))/\partial c_{ij}$, are limited to $0 \le \gamma_{ij}(k) \le 1$. In order to enable the escape from the local minima, the term $\gamma_{ij}(k)$ needs to be defined in the following manner:

- if the value of $\frac{\partial Q(\mathbf{C}, \mathbf{R}(k))}{\partial c_{ij}}$ increases, $\gamma_{ij}(k) \to 0$;
- if $\frac{\partial Q(\mathbf{C}, \mathbf{R}(k))}{\partial c_{ij}} \to 0$, $\gamma_{ij}(k) \to 1$

In our experiment, we have applied the following definition of $\gamma_{ij}(k)$:

$$\gamma_{ij}(k) = \frac{1}{1 + \left|\frac{\partial Q(\mathbf{C}, \mathbf{R}(k))}{\partial c_{ij}}\right|} \qquad (13)$$

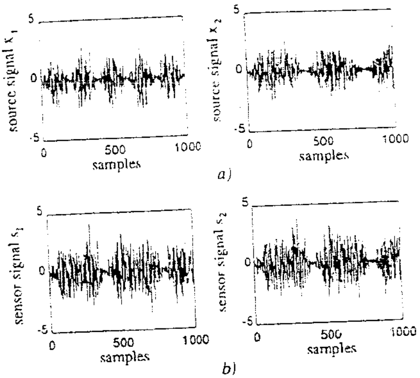

Fig. 1. a) Source signals b) Sensor signals

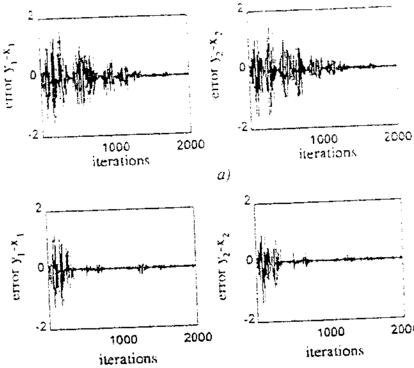

Fig2. Errors obtained by applying: a) gradient method b) "Heavy ball" method

To demonstrate the properties of the proposed "heavy ball" method, we have applied it to decompose the linear combination of two source signals given by:

$$x_1(k) = 2\sin(k\pi / 200) \cdot n(k) \qquad (14)$$

$$x_2(k) = 2\sin(k\pi / 400) \cdot n'(k) \qquad (15)$$

$n(k)$ and $n'(k)$ are Gaussian signals with zero mean and unity variance. The coefficients of the linear transformation were given by the matrix A:

$$A = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix} \qquad (16)$$

and the parameters of the learning dynamics are chosen to be $\alpha = 0.8$ and $\beta = 0.001$. Source signals given by the equations (14) and (15), and sensor signals, which represent their linear combinations, are shown in figures 1.a) and 1.b) respectively.

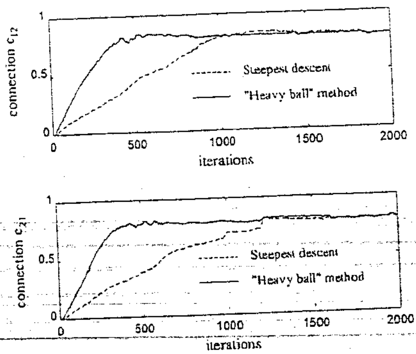

Fig. 3. Adaptation of the connection weights

Comparison between the performances of the original steepest descent algorith and the proposed "heavy ball" algorithm applied to the signals at hand is given in Fig. 2 and 3. Figures 2.a) and 2.b) are graphical representations of the differences between the original source signals, $x_i(k)$, and their estimates obtained at the network outputs, $y_i(k)$. Figure 3. shows the adaptation of the connection weights $c_{ij}(k)$ through the learning process, in which the weights tend towards the corresponding values a_{ij} of the transformation matrix A (16). It is obvious that the proposed modified learning algorithm provides faster convergence compared to those of the original algorithm proposed in [2]. The learning speed of both algorithms does not depend on the magnitudes of source signals, but on the $r_i(t)/r_j(t)$, and is faster for stronger fluctuations of $r_i(t)/r_j(t)$.

IV SEPARATION OF SEISMIC SIGNALS

To demonstrate the properties of the described neural network in practical applications, the modified learning algorithm is applied to seismic signals in order to separate the components generated by a number of origins.

Seismic signals, which represent the earth's response to excitations arising from natural phenomena or from man-made acoustic sources, are a complex of distinct components with various strengths and arrival times, which are generated and propagated independently of each other. The resulting composite signals, observed by seismic transducers, represent the linear combination of their constituents [5]. The waveform of the man-made seismic signals, considered in this experiment, depends primarily on the exciting force, i.e. on the actual class of origins of vibrations. Due to this fact, it is possible to determine the class of origin considering the detected waveform only. However, several origins, when acting simultaneously, result in a waveform which can not uniquely point out to the particular class of origins and it should be of interest to

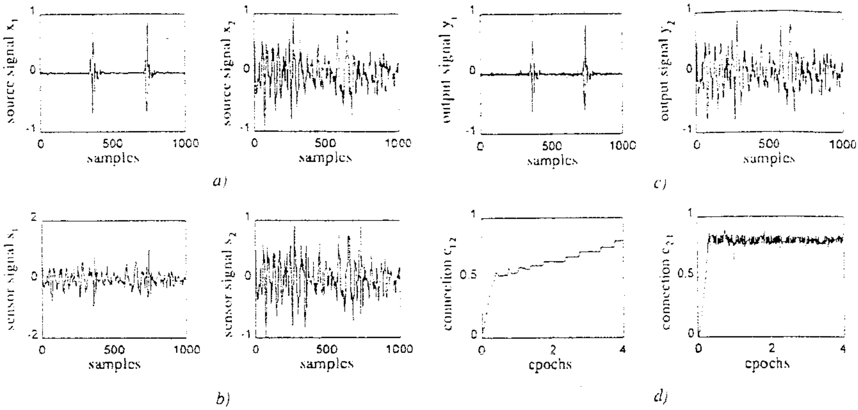

Fig. 4. a) Source signals b) Sensor signals c) Output signals d) Connection weights during training

separate it. In the example shown, we have applied real seismic data, given by a linear combination of two source signals, to the network inputs. The actual source signals, generated by a man and a car, are detected by geophones (Fig. 4.a). Sensor signals, which represent the linear combination of source signals, are obtained according to the equation (1), and the linear transformation matrix was given by (16).

The sequences, applied to the network input, (Fig. 4.b), consisted of 1000 samples per second. The network parameters are chosen to be $\alpha = 0.8$ and $\beta = 0.001$ and the initial values of c_{ij} and ϕ_{ij} are set to 0 and 1, respectively.

The figures above represent the results of our experiment. Output signals, obtained at the end of the learning process as the estimates of the source signals, are given in Figure 4.c). It has been shown that signal separation can be achieved in about four epochs through the considered sequences. Figure 4.d) represents the adaptation of the connection weights through the learning process.

V CONCLUSION

In this paper, we are concerned with the blind separation of nonstationary signals using the linear self-organized neural network with lateral connections. The proposed network performs a separation of composite signals into the source signals in the absence of any prior information about the statistical properties of the sources, except the fact that they are statistically independent of each other. The connection weights are iteratively adapted during the learning process in which the basic steepest descent algorithm minimizes the time-dependent cost function in order to obtain the estimates of the source signals as the network outputs. It has been shown that faster convergence can be achieved by modifying the steepest descent algorithm using the adaptive momentum term. The performances of the proposed modified algorithm are demonstrated on the example of real seismic data.

The described method for blind separation of nonstationary signals is valid only in those cases, in which the number of sources is equal to the number of sensors and the linear transformation matrix is nonsingular. In our further research, we will attempt to avoid these constraints by using the matrix generalized inverses. We will also consider the cases in which the coefficients of the linear transformation change in time.

REFERENCES

[1] G.Deco, W.Brauer, *Nonlinear Higher-Order Statistical Decorrelation By Volume-Conserving Neural Architectures*, Neural Networks, Vol. 8, No. 4, pp. 525-535, 1995

[2] K.Matsuoka, M.Ohya, M.Kawamoto, *A Neural Net for Blind Separation of Nonstationary Signals*, Neural Networks, Vol. 8, No. 3, pp. 411-419, 1995

[3] P.P.Kanjilat, S.Palit, *On Multiple Pattern Extraction Using Singular Value Decomposition*, IEEE Transactions on Signal Processing, Vol. 43, No. 6, June 1995

[4] I.I.Gurvich, G.N.Boganik, *Prospecting Seismology*, (in Russian), Nedra, Moskow, 1980

[5] S.Stojanov, *Methods and algorithms for optimization*, (in Bulgarian), Državno izdatelstvo Tehnika, Sofija, 1990