EXTENDED KALMAN FILTER IN BLIND SEPARATION OF NONSTATIONARY SIGNALS

Slavica Todorovic-Zarkula[1], Branimir Todorovic[2], Miomir Stankovic[2], Vlastimir Pavlovic[3]

[1]Technical School, Nis, Serbia & Montenegro
phone: +381 18 529-960, email: bssmtod@eunet.yu
[2]Faculty of Occupational Safety, University of Nis, Nis, Serbia & Montenegro
phone: +381 18 529 703, fax: +381 18 , email: {todor,mstan}@znrfak.znrfak.ni.ac.yu
[3]Faculty of Electronics, University of Nis, Nis, Serbia & Montenegro
phone: +381 18 529 206, email: vpavlovic@elfak.ni.ac.yu

ABSTRACT

In this paper we deal with on-line algorithms for blind source separation using second order statistics. We briefly describe separating algorithms based on stochastic gradient, and natural gradient optimization method, and propose an on-line separating algorithm based on the application of the extended Kalman filter. Performance evaluation of the proposed algorithm has shown that extended Kalman filter applied as an optimization algorithm can improve convergence speed and estimation accuracy of the separating algorithm.

1. INTRODUCTION

Blind separation of sources refers to the problem of recovering source signals from their mixtures using only the observed mixtures. The separation is called blind, because it assumes very weak assumptions on the source signals and mixing process. The key assumption is the statistical independence of source signals, and a goal is to obtain estimated output signals that are as independent as possible.

In the last few years, blind source separation has received considerable attention, due to its potential applications in various areas, such as telecommunications, speech processing, geophysics, biomedical and image processing. Since 1985, when blind source separation was initially proposed by Jutten, Herault and Ans, various approaches have been proposed to this problem [1,2,4,7,9]. In this paper, we consider on-line algorithms for blind separation of non-stationary signals using second-order statistics. It has been shown that, using the additional assumption on non-stationarity of sources, blind source separation can be achieved using only second-order statistics (SOS) [10, 11]. We are interested in second-order non-stationarity in the sense that source variances vary with time. We base our algorithm on diagonalization of the output correlation matrix in order to achieve decorrelation of the estimated output signals. In order to blindly separate source signals from the observed mixtures, we apply a self-organizing neural network with lateral connections, which uses the observed mixtures as inputs, and provides the estimated source signals as outputs. Throughout the learning process, the network weights are adapted in a direction that reduces correlation between the network outputs. As an optimization algorithm that minimizes cross-correlations between output signals, we propose an on-line algorithm derived from the extended Kalman filter (EKF). In order to evaluate our algorithm, we compare its performances in on-line learning with those of the separating algorithms based on stochastic and natural gradient. In numerical experiments with artificial mixtures of real-world speech signals, the EKF based algorithm outperformed both algorithms in convergence speed and estimation accuracy.

The paper is organized as follows. In Section 2, we formulate the problem of blind source separation. In Section 3, we briefly describe stochastic gradient (SG) and natural gradient (NG) based methods for blind separation of non-stationary sources. In section 4, we propose an EKF based separating algorithm. We define a simple contrast function, and apply EKF as an optimization algorithm in order to estimate unknown parameters of a demixing model and recover non-stationary sources. Section 5 contains results obtained in separation of speech signals from their artificial mixtures. In Section 6, we give the concluding remarks.

2. PROBLEM FORMULATION

Let $\mathbf{s} = [s_1 s_2 \ldots s_N]^T$ represent N zero-mean random source signals whose exact probability distributions are unknown. Suppose that M sensors $(M \geq N)$, receive instantaneous linear mixtures $\mathbf{x} = [x_1 x_2 \ldots x_M]^T$ of source signals. The mixtures are considered instantaneous if the differences of signal time arrivals between sensors can be neglected, so the propagation from sources to sensors can be represented by:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{1}$$

where $\mathbf{A}$ is the unknown $M \times N$ mixing matrix, and $\mathbf{x}$ is the vector of the observed mixtures. In a demixing system, source signals have to be recovered using the observed mixtures as inputs. As a result, we get an N-dimensional random vector $\mathbf{y}$ of separated components:

$$\mathbf{y} = \mathbf{B}\mathbf{x} = \mathbf{B}\mathbf{A}\mathbf{s} = \mathbf{G}\mathbf{s} \,. \tag{2}$$

In (2), $\mathbf{B}$ is an $N \times M$ matrix, and $\mathbf{G}$ is an $N \times N$ global system matrix. Since it is of interest to obtain separated components that represent possibly scaled and permuted versions of sources, the matrix **G** has to represent a generalized permutation matrix. Therefore, the problem is to obtain, if possible, a matrix **B** such that each row and each column of **G** contains only one nonzero element. Due to the lack of prior information, the ordering and scaling of the recovered sources remain undetermined, so we can obtain at best $\mathbf{y}=\mathbf{DPs}$, where **P** is a permutation matrix, and **D** is a nonsingular diagonal scaling matrix. Due to the scaling ambiguity, in the estimation of sources some parameters have to be arbitrarily fixed. In most of the approaches, this is accomplished by fixing diagonal elements of a demixing matrix to unity, or by constraining output signal variance to be equal to unity.

3. SEPARATING ALGORITHMS BASED ON SOS

Blind source separation using the additional assumption on non-stationarity of sources was initially proposed by Matsuoka et al. [11]. It was shown that non-stationary signals can be separated from their mixtures using SOS if signal variances change with time, and fluctuate independently of each other during the observation [11]. In order to separate sources from their instantaneous mixtures, a linear self-organizing neural network with lateral connections (Fig. 1) was applied as a demixing model [11].

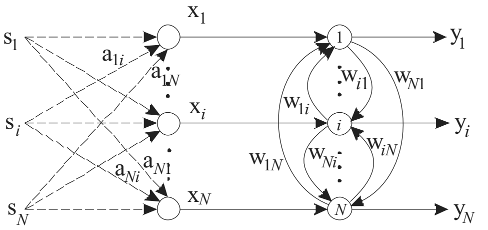


Figure 1: Self-organizing linear neural network with lateral connection for blind source separation

The dynamics of each output unit is given by the first-order linear differential equation:

$$\tau\frac{d\mathbf{y}_t}{dt}+\mathbf{y}_t=\mathbf{x}_t-\mathbf{W}\mathbf{y}_t \tag{3}$$

where the matrix $\mathbf{W}=[w_{ij}]$ denotes the mutual lateral connections between the output units. The output units have no self-connections, and therefore w_{ii}=0. In the steady state, the equation (3) becomes:

$$\mathbf{y}_t=(\mathbf{I}+\mathbf{W})^{-1}\mathbf{x}_t \tag{4}$$

Using the self-organized neural network (Fig. 1) as a demixing model, Matsuoka et al. [11] have derived an on-line stochastic gradient based algorithm for blind separation of non-stationary sources. The algorithm was obtained by minimization of the following contrast function [11]:

$$Q\left(\mathbf{W},\mathbf{R}_{y,t}\right)=\frac{1}{2}\left\{\sum_i \log < y_{i,t}^2 > -\log\left|< \mathbf{y}_t\mathbf{y}_t^{\mathrm{T}} >\right|\right\} \tag{5}$$

where $\mathbf{R}_{y,t}$ is the output correlation matrix, and $<\cdot>$ denotes expectation. It should be noted that in the case of zero-mean signals, correlation matrix is equal to covariance matrix. In discrete-time *k*, the SG based separating algorithm is given by the following equations for adaptation of the network weights $w_{ij,k}$, $i,j=1,\ldots,N$ [11]:

$$w_{ij,k}=w_{ij,k}+\beta\frac{y_{i,k}y_{j,k}}{\phi_{i,k}} \tag{6a}$$

$$\phi_{i,k}=\alpha\phi_{i,k}+(1-\alpha)y_{i,k}^2 \tag{6b}$$

In (6a), the learning rate β is assumed to be a very small positive constant, and the constant α in (6b) is a forgetting factor $0<\alpha<1$. The learning algorithm (6a)-(6b) uses moving average $\phi_{i,k}$ in order to estimate $< y_{i,k}^2 >$ in real time. In practice, the expected values are not available, and time-averaged, or instantaneous values can be used instead of them.

Using natural gradient approach to the optimization of contrast function (5), Choi et al. [6] have derived an efficient equivariant algorithms for both fully connected recurrent and feedforward network. We consider here the learning algorithm derived for the fully connected feedforward network. The output of the network is given by:

$$\mathbf{y}_k=\mathbf{W}_k\mathbf{x}_k \tag{7}$$

The natural gradient separating algorithm for the feedforward neural network is given by [6]:

$$\begin{aligned}\Delta\mathbf{W}_k&=\eta(I-\mathbf{\Lambda}^T{}_k\,\mathbf{y}_k\mathbf{y}_k{}^T)\mathbf{W}_k=\\&=\eta\mathbf{\Lambda}^{-1}{}_k(\mathbf{\Lambda}_k-\mathbf{y}_k\mathbf{y}_k{}^T)\mathbf{W}_k\end{aligned} \tag{8}$$

where $\mathbf{\Lambda}_k$ is a diagonal matrix whose elements $\lambda_{i,k}$ are the variances of the output signals $y_{i,k}$ that can be estimated by the moving average (6b).

4. EXTENDED KALMAN FILTER IN BLIND SEPARATION

In order to apply the Kalman filter to the blind source separation, we have formulated the problem of blind separation as minimization of the instantaneous contrast function [13]:

$$J(\mathbf{w}_k)=\tfrac{1}{2}\mathbf{r}_k(\mathbf{w}_k)^T\mathbf{r}_k(\mathbf{w}_k)\,. \tag{9}$$

where $\mathbf{r}_k$ is the vector formed of the non-diagonal elements of the output correlation matrix, i.e. the cross-correlations $< y_i(\mathbf{w}_k)y_j(\mathbf{w}_k) >$ of the network outputs $\mathbf{y}_k$ at time step *k*, parameterized by the unknown mixing weights $\mathbf{w}_k$. As a demixing model, we have applied a neural network with lateral connections (Fig. 1). The network outputs, which represent the recovered source signals, are calculated at every time step according to:

$$y_{i,k}=x_{i,k}-\sum_{j=1,\,j\neq i}^{N}w_{ij,k}y_{j,k}\,,\quad i,j=1,\ldots N\,. \tag{10}$$

The state space model of the parameter dynamics is given in the observed-error form [13]:

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \mathbf{d}_{k-1}, \quad \mathbf{d}_{k-1} \sim N(0, \mathbf{Q}_{k-1}) \qquad (11a)$$

$$\mathbf{z}_k = -\mathbf{r}_k(\mathbf{w}_k) + \mathbf{v}_k, \quad \mathbf{v}_k \sim N(0, \mathbf{R}_k). \qquad (11b)$$

Note that the observations $\mathbf{z}_k$ of cross-correlations $\mathbf{r}_k(\mathbf{w}_k)$ are equal to zero at every time step *k*. The process noise $\mathbf{d}_{k-1}$ and the observation noise $\mathbf{v}_k$ are assumed mutually independent, white and Gaussian and with variances equal to $\mathbf{Q}_{k-1}$ and $\mathbf{R}_k$, respectively. The parameters can be estimated by maximizing a posteriory probability density function:

$$p(\mathbf{w}_k / \mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k / \mathbf{w}_k) p(\mathbf{w}_k / \mathbf{z}_{1:k-1})}{\int p(\mathbf{z}_k / \mathbf{w}_k) p(\mathbf{w}_k / \mathbf{z}_{1:k-1})\, d\mathbf{w}_k} \qquad (12)$$

where $\mathbf{z}_{1:k} = \{\mathbf{z}_i, i = 1,...,k\}$, or, equivalently by minimizing the negative logarithm of (11):

$$J(\mathbf{w}_k) = -\log(p(\mathbf{z}_k / \mathbf{w}_k)) - \log(p(\mathbf{w}_k / \mathbf{z}_{1:k-1})) \quad (13).$$

After taking into account the assumptions on the process noise and observation noise, and after neglecting the constant terms, the cost function (13) becomes:

$$J(\mathbf{w}_k) = \tfrac{1}{2}\mathbf{r}_k(\mathbf{w}_k)^{\mathbf{T}} \mathbf{R}_k^{-1} \mathbf{r}_k(\mathbf{w}_k) + \tfrac{1}{2}(\mathbf{w}_k - \hat{\mathbf{w}}_{k-1})^{\mathbf{T}} (\mathbf{P}_{k-1} + \mathbf{Q}_{k-1})^{-1} (\mathbf{w}_k - \hat{\mathbf{w}}_{k-1})$$

The estimate of the network weights $\hat{\mathbf{w}}_k$ and its associated covariance $\mathbf{P}_k$ at time step *k*, are given by:

$$\hat{\mathbf{w}}_k = \hat{\mathbf{w}}_{k-1} + \mathbf{K}_k \mathbf{r}_k(\hat{\mathbf{w}}_{k-1}) \qquad (14a)$$

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \cdot (\mathbf{P}_{k-1} + \mathbf{Q}_{k-1}), \qquad (14b)$$

where $\mathbf{K}_k$ is the Kalman gain:

$$\mathbf{K}_k = (\mathbf{P}_{k-1} + \mathbf{Q}_{k-1}) \mathbf{H}_k^T (\mathbf{R}_k + \mathbf{H}_k (\mathbf{P}_{k-1} + \mathbf{Q}_{k-1}) \mathbf{H}_k^T)^{-1} \qquad (15)$$

and:

$$\mathbf{H}_k = \partial \mathbf{r}_k(\mathbf{w}_k) \big/ \partial w_k \Big|_{w_k = \hat{w}_{k-1}}. \qquad (16)$$

Recursions (14a) and (14b) represent the basic equations of the extended Kalman filter for the problem defined by the state space model (11).

5. SIMULATION RESULTS

In the problem of blind source separation, it is important to obtain an estimator whose properties do not depend on the mixing matrix, but only on the source signals [3]. Uniformity, or equivariance property of the estimator is desirable in on-line learning, because it avoids necessity to tune parameters of the algorithm in order to obtain optimal results and uniform performances in the estimation of sources for various mixing matrices. In order to evaluate performances of our EKF-based algorithm in blind source separation, we have compared it with the stochastic gradient and natural gradient separating algorithms given in Section 3. The task was to separate two non-stationary sources from their mixtures. The mixtures were obtained artificially, using a set of mixing matrices with diagonal elements set to unity, and non-diagonal elements (a_{12}, a_{21}) were taken from the interval [0.1,0.9] with step 0.1. In this framework, we can measure the performance of the algorithm in terms of the performance index PI defined by [5]:

$$PI = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left[\left(\sum_{k=1}^{n} \frac{|g_{ik}|}{\max_j |g_{ij}|} - 1 \right) + \left(\sum_{k=1}^{n} \frac{|g_{ki}|}{\max_j |g_{ji}|} - 1 \right) \right]$$

where g_{ij} is the (i, j)-th element of the global system matrix $\mathbf{G} = (\mathbf{I} + \mathbf{W})^{-1}\mathbf{A}$. The performance index indicates how far the global system matrix $\mathbf{G}$ is from a generalized permutation matrix. When perfect signal separation is achieved, the performance index is zero.

In addition, we calculate the signal-to-interference-ratio-improvement SIRI according to [6]:

$$SIRI_i = \frac{E\{(x_i - s_i)^2\}}{E\{(y_i - s_i)^2\}}$$

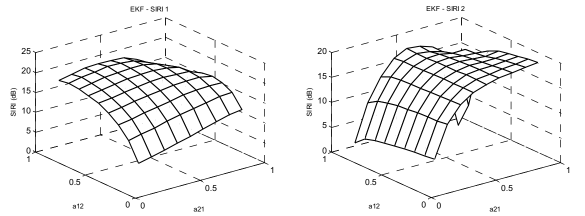
Before calculation of SIRI, the scaling and ordering ambiguities are resolved. In performance comparison, the source signals were two normalized speech recordings recorded in anechoic chamber at sampling rate 12kHz and each recording consisted of 60000 samples. Each signal sample was presented to the network input only once. In simulations, the learning rate β_{SG} and β_{NG} for the SG and NG-based algorithm were set to $\beta_{SG} = \beta_{NG} = 0.0001$, and forgetting factor were set to: $\alpha_{SG} = 0.8$, and $\alpha_{NG} = 0.8$. The observation noise covariance $\mathbf{R}_k$ for the EKF-based algorithm was fixed to identity matrix, $\mathbf{R}_k = \mathbf{I}$, and the process noise covariance $\mathbf{Q}_k$ was a diagonal matrix which entries were exponentially decayed in the range $(10^{-1} - 10^{-6})$ in order to achieve fast convergence at the beginning of the learning process, and to retain good tracking abilities. For both algorithms, initial values of w_{12} and w_{21}, as well as the initial correlations in the moving average (6b), were set to zeros. The initial value of estimation error covariance **P** in EKF-based algorithm was set to $\mathbf{P}_k = 0.01 \cdot \mathbf{I}$.

The obtained results are given in the following figures. Fig. 2. presents dependence of SIRI on the mixing coefficients (a_{12}, a_{21}) obtained using EKF, SG and NG based separating algorithms. Fig. 3 and Fig. 4 show the examples of time evolution of the performance index PI obtained for the mixing matrices given by:

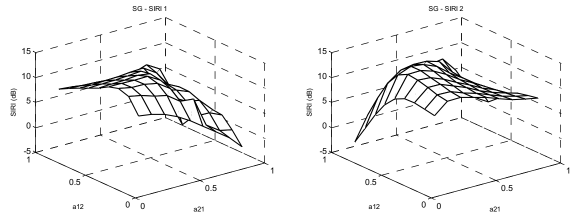
$$A_1 = \begin{bmatrix} 1 & 0.6 \\ 0.8 & 1 \end{bmatrix}, \text{ and } A_2 = \begin{bmatrix} 1 & 0.7 \\ 0.6 & 1 \end{bmatrix}.$$

We have observed the following:

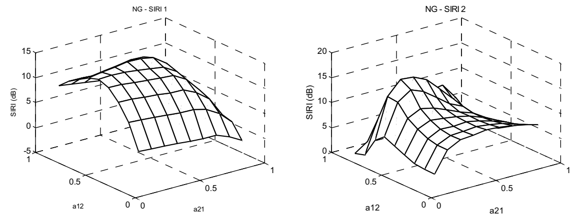
- For all considered algorithms, the values of PI and SIRI depend on the mixing coefficients, as well as on the source signals.
- Regarding convergence speed and estimation accuracy in on-line learning, the best results were obtained using EKF based algorithm.



a)



b)



c)

Figure 2: $SIRI_1$ and $SIRI_2$ versus mixing coefficients (a_{12}, a_{21}) obtained using: a) EKF based algorithm, b) SG based algorithm, c) NG based algorithm

- The results obtained by NG and SG based algorithms can be improved by several passes - epochs through the recordings, whereas for EKF based algorithm no improvements were observed in that case.
- For all considered algorithms, the best results were obtained for moderate mixtures; the quality of estimation decreased for badly-scaled sources, and for mixing matrix aproaching ill-conditioned matrix. For $a_{12} = a_{21} = 0.9$, EKF based algorithm did not converge within 60000 time steps and predefined parameter values.

6. CONCLUSION

In this paper, we have proposed an on-line algorithm for blind source separation based on the second order statistics and the extended Kalman filter. Numerical experiments have shown that the EKF based algorithm outperforms stochastic gradient and natural gradient separating algorithms in convergence speed and estimation accuracy.

REFERENCES

[1] Amari, S.-I., Cichocki, A., Yang, H. (2000) “Blind signal separation and extraction: Neural and information–theoretic approach”, in Haykin, S. (Ed). Unsupervised adaptive filtering, Vol. I, John Willey & Sons, Ltd.

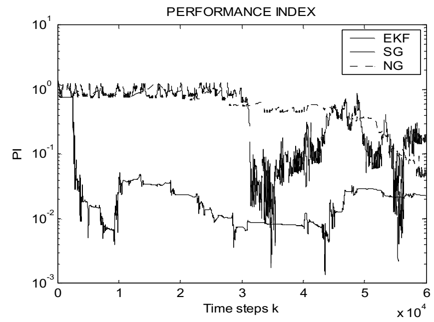


Figure 3 Time evolution of the performance index obtained for the mixing matrix A_1

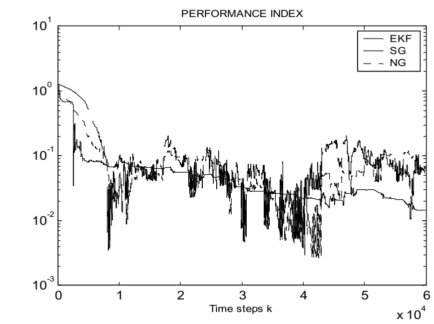


Figure 4: Time evolution of the performance index obtained for the mixing matrix A_2

[2] Bell,A.J., Sejnowski,T.J. (1995) “An information maximization approach to blind separation and blind deconvolution”, Neural computation, 7, no.6, pp. 1129-1159

[3] Cardoso, J.-F., Laheld, (1996) B.H., “Equivariant adaptive source separation”, IEEE Trans. Signal Processing, 44(12): 3017-3030

[4] Cardoso, J.-F. (1999) “High-order Contrasts for Independent Component Analysis”, Neural Computation 11, Massachusetts Institute of Technology, pp. 157-192

[5] Cichocki, A., Amari, S.-I., (2002), Adaptive blind signal and image processing: Learning algorithms and applications, John Willey & Sons, Ltd.

[6] Choi, S., Cichocki, A., Amari, S.-I., (2002) “Equivariant nonstationary source separation”, Neural Networks, 15, pp. 121-131

[7] Comon, P. (1994) “Independent component analysis, a new concept”, Signal Processing 36, pp. 287-314

[8] Haykin, S., Ed. (2001), Kalman filtering and neural networks, John Willey & Sons, Ltd.

[9] Mansour, A., Kardec Barros, A:, Ohnishi, N. (2000) “Blind separation of sources: Methods, sssumptions and spplications”, IEICE Trans. Fund., Vol.E83-A, No.8

[10] Mansour, A. (1999) “The blind separation of nonstationary signals by only using the second order statistics”, Proc. of the ISSPA’99, Australia, pp. 235-238

[11] Matsuoka, K., Ohya, M., Kawamoto, M. (1995) “A neural net for blind separation of nonstationary signals”, Neural networks, vol. 8, no. 3, pp. 411-419

[12] Todorović, B., Stanković, M, Todorović-Zarkula, S. (2000) “Structurally adaptive RBF network in non-stationary time series prediction”, In Proc. IEEE AS-SPCC, Lake Louise, Alberta, Canada, Oct. 1-4 pp. 224-229

[13] Todorovic-Zarkula, S., Todorovic, B., Moraga, C. (2003) “Blind Separation of Non-stationary Signals using Extended Kalman Filter”, In Proc. of 1st BCI, Greece, pp. 382-392