Blind Separation of Non-stationary Signals Using Extended Kalman Filter

Slavica Todorovic-Zarkula[1*], Branimir Todorovic[2], Claudio Moraga[3]

[1] Bul. Cara Konstantina 80-84, 18000 Nis, Serbia and Montenegro
Email: bssmtod@eunet.yu
[2]Faculty of Occupational Safety, University of Nis, 18000 Nis, Serbia and Montenegro
Email: bssmtod@eunet.yu
[3]Department of Computer Science, University of Dortmund, Germany
Email: moraga@cs.uni-dortmund.de

Abstract. The problem of blind source separation consists in retrieving unknown source signals from their mixtures. In the estimation of sources, most of the proposed algorithms use stochastic or natural gradient to optimize some contrast function obtained from the independence property of sources. In this paper, we deal with non-stationary source signals, and propose a new approach, based on the second order statistics, to blindly separate them. Our approach is based on the assumption that the cross correlations of the output signals should be equal to zero. The cross correlations are parameterized by the weights of a self organized neural network, which implements a demixing model. The dynamics of the neural network weights and observations of the cross correlations are represented using the state space model. Extended Kalman filter (EKF) is applied to estimate the unknown weights. As an on-line second order optimization algorithm, EKF showed superior convergence properties compared to the stochastic gradient descent based separating algorithm.

1 Introduction

Blind source separation refers to the problem of recovering source signals from their mixtures using the observed mixtures only. The separation is called "blind", because nothing is assumed known about statistical distributions of sources, and the mixing process. The lack of a priori knowledge of the mixing system is compensated by a statistically strong assumption on the independence of sources, and by assumptions on the mixing model (linear or nonlinear, instantaneous or convolutive). The problem of blind separation appears in many areas of signal processing where an array of sensors picks up mixtures of source signals. Examples include speech separation (cocktail party problem), processing of multi-sensor biomedical data (EEG, ECG), seismic signal processing, arrays of radar and sonar recordings, image restoration, etc.

In the last few years, the problem of blind source separation has received considerable attention. From 1984, when blind source separation was initially proposed by

[*] The work of S. Todorović-Zarkula and B. Todorović was supported by a Scholarship of the German Academic Exchange Service (DAAD) under the Stability Pact for South East Europe.

Jutten and Herault, various approaches have been proposed [1]. These include independent component analysis - ICA [2], information maximization [3], the natural gradient approach [4], etc. Most of the approaches use the independence property either directly, through optimization of criteria based on the Kullback-Leibler divergence, or indirectly, through minimization of criteria based on the cumulants [5, 6]. Source signals are usually estimated iteratively using learning algorithms derived from heuristic considerations, or from optimization of some contrast function of the output. Contrast-based approaches are *direct* in the sense that they try to estimate directly source vector by searching for demixing system whose output is as close as possible to the original sources. *Indirect* approaches, which can be based on cumulant matching, aim at identifying the mixing matrix in a first stage, and the sources are estimated in a second stage [6]. Most of the approaches are based on the prior assumption that the transformation matrix is of the full rank, which implies that the number of mixtures is equal or greater than the number of sources. However, the under-determined case, i.e. the case when the number of sources is greater than the number of mixtures, has also been examined [7].

Having in mind the independence property of sources, the task of blind separation is to recover independence of the estimated output signals. By definition, random variables are said to be independent, if their mutual probability density function (pdf) is the product of their marginal pdf's. Since the independence of sources implies that cumulants of all orders should be equal to zero, the problem is obviously related to the higher-order statistics. It has been shown that the fourth-order statistics (moments and cumulants) are enough to achieve independence, and they are used in many algorithms [1 5]. Application of higher-order statistics is limited to non-Gaussian signals, because for Gaussian signals, cumulants of order higher than two vanish.

In this paper, we consider blind separation of non-stationary signals using second order-statistics. Non-stationarity means that the signal power changes with time. In [8, 9], it has been shown that, using the additional assumption on the non-stationarity of sources, blind source separation can be achieved using only second-order statistics. We base our algorithm on diagonalization of the output correlation matrix in order to achieve decorrelation of the output signals. As an optimization algorithm, that minimizes cross-correlations between output signals, we propose the extended Kalman filter. The rest of the paper is organized as follows. In Section II, we define the problem of blind source separation. In Section III, we describe a self-organized neural network for blind separation and we give the stochastic gradient descent algorithm for weight estimation. In section IV, based on the state space model of the weight dynamics and observations of cross corelations, we derive the extended Kalman filter for weight estimation. Section V contains the simulation results. Concluding remarks are given in Section VI.

2 Problem Formulation

Let $\mathbf{s} = [s_1 s_2 ... s_N]^{\mathrm{T}}$ represent N zero-mean random source signals whose exact probability distributions are unknown. Suppose that L sensors receive linear mixtures

$\mathbf{x} = [x_1\, x_2 ... x_L]^{\mathrm{T}}$ of source signals. If we ignore delays in signal propagation, this can be expressed in the matrix form:

$$\mathbf{x} = \mathbf{As}\,, \tag{1}$$

where **A** is the unknown $L \times N$ linear combination matrix that is not necessarily of the full rank, and 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 result we get generally an *M*-dimensional ($M \le N$) random vector of separated components:

$$\mathbf{y} = \mathbf{Bx} = \mathbf{BAs} = \mathbf{Cs}\,, \tag{2}$$

where **B** is an $M \times L$ matrix, and **C** is an $M \times N$ matrix. Since it is of interest to obtain single source signals and/or complementary groups of signals as separated components, the matrix **C** has to represent a partition matrix [10]. In the partition matrix element c_{ij} equals 1 if a source signal s_j belongs to a signal subset S_i, and otherwise $c_{ij} = 0$. For example, a particular source signal s_1 can be separated out if the first row of the matrix **C** is $[1, 0, 0, \ldots, 0]$. Ideally, if **C** is an identity matrix, the set of sources is completely separable. Therefore, the problem is to obtain, if possible, a matrix **B** such that each row of **C** contains only one nonzero element. It should be noted that the problem has inherent indeterminacy – the matrix **A** is not identifiable from the observed signals even if it should be possible to extract all source signals, because their ordering remains unknown. The magnitudes of the source signals are also not recoverable, because a scalar multiple of s_j , ks_j , can not be distinguished from multiplication of the *j*-th column of **A** by the same scalar k. Therefore, we can obtain at best $\mathbf{y} = \mathbf{DPs}$, where P is a permutation matrix, and D is a nonsingular diagonal scaling matrix [11]. This means that only permuted and rescaled source signals can be recovered from mixture signals. In most cases, such solution is satisfactory. In our further consideration we assume for simplicity that $\mathbf{DP} = \mathbf{I}$.

In the following, we assume that the sources are nonstationary mutually independent random signals, the mixtures are linear and instantaneous, and the number of observed mixtures is equal to the number of sources, and number of separated components, i.e. *L*=*N*=*M*.

3 Neural Network for Blind Source Separation

In order to separate non-stationary source signals from their linear mixtures, we have applied a linear self-organized neural network with lateral connections originally described in [9].

The network receives sensor signals $\mathbf{x}_t$ as inputs and provides estimates of the source signals $\mathbf{y}_t$ as outputs (Fig. 1). In matrix notation, the dynamics of each output unit is given by the first-order linear differential equation:

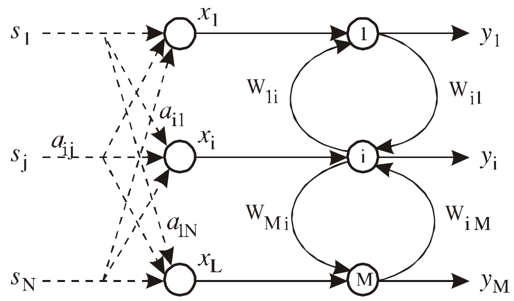

Fig. 1. Self-organized neural network with lateral connections for blind source separation

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

where 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. Assuming the time constant τ is sufficiently small, the equation (3) can be replaced by:

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

In [9], signal separation is realized through the learning process by determining **W** so that the non-negative scalar function given by:

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

takes minimum every time. In (5), $\mathbf{R}_{x,t}$ is the diagonal correlation matrix of $\mathbf{x}_t$.

$$T \frac{dw_{ij}}{dt} = - \frac{\partial Q(\mathbf{W}, \mathbf{R}_{x,t})}{\partial w_{ij}} , i, j = 1, ..., N; i \neq j \tag{6}$$

T is the time constant that controls the learning speed. The learning algorithm is obtained using the Euler approximation of the equation (6), and is given by [9]:

$$w_{ij,k} = w_{ij,k} + \beta \frac{y_{i,k} y_{j,k}}{\phi_{i,k}} \tag{7}$$

$$\phi_{i,k} = \alpha \phi_{i,k} + (1 - \alpha) y_{i,k}^2 \tag{8}$$

The learning algorithm, obtained using stochastic gradient descent, uses moving average $\phi_{i,k}$ in order to estimate $< y_{i,k}^2 >$ in real time. The neural network, applied as a demixing model, is self-organized in a sense that it changes its connection weights in a direction that reduces the correlation between its outputs.

4 Parameter Estimation Using the Extended Kalman Filter

Our approach to non-stationary blind signal separation is based on the assumption that cross-correlations of the output signals should be equal to zero. Let us consider the vector r_k, formed of the cross-correlations of the output signals, i.e. the non-diagonal elements of the output correlation matrix. Taking into account that the vector of cross-correlations r_k of output signals y_k at time step is parameterized by the unknown mixing weights, the problem can be formulated as minimization of the instantaneous cost function:

$$J(w_k) = r_k(w_k)^{\mathrm{T}} r_k(w_k) \tag{9}$$

where w_k is the vector of the neural network weights.

By forcing the observation of the cross-correlations r_k to be equal to zero, and using Gauss Markov first order random process as weight dynamics, we can form the following state space model:

$$w_k = w_{k-1} + d_{k-1}, \quad d_{k-1} \sim N(0, Q_{k-1}) \tag{10}$$

$$z_k = -r_k(w_k) + v_k, \quad v_k \sim N(0, R_k)\,. \tag{11}$$

Note that observations z_k are equal to zero at every time step *k*. The process noise d_{k-1} and the observation noise v_k are assumed mutually independent, white and Gaussian.

Estimation of the unknown parameters can be put in the Bayesian filtering framework using the state space formulation of the problem. The posterior filtering probability density function (pdf) $p(w_k / z_{1:k})$ of the parameters w_k given the observations $z_{1:k} = \{z_i, i = 1,2,...,k\}$ represents the complete solution of the estimation problem. Assuming the initial pdf of parameters $p(w_0)$ is known, the posterior pdf $p(w_k / z_{1:k})$ can be obtained recursively using the Bayes rule and the state space formulation (10)-(11):

$$p(w_k / z_{1:k}) = \frac{1}{C} p(z_k / w_k) p(w_k / z_{1:k-1}), \tag{12}$$

where $p(w_k / z_{1:k-1})$ represents the prior pdf of parameters, given by:

$$p(w_k / z_{1:k-1}) = \int p(w_k / w_{k-1}) p(w_{k-1} / z_{1:k-1}) \mathrm{d}w_{k-1}, \tag{13}$$

and *C* is the constant (independent of w_k):

$$C = \int p(z_k / w_k) p(w_k / z_{1:k-1}) \mathrm{d}w_k\,. \tag{14}$$

If we approximate the posterior pdf of the parameters at time step $k-1$ with the Gaussian: $p(w_{k-1} / z_{1:k-1}) = N(\hat{w}_{k-1}, P_{k-1})$, we have:

$$p(w_k / z_{1:k-1}) = N(\hat{w}_{k-1}, P_{k-1} + Q_{k-1}) \quad \textbf{(15)}$$

$$p(z_k / w_k) = N(-r_k(w_k), R_k)\,. \quad \textbf{(16)}$$

The estimate of the parameters $\hat{w}_k$ at time step k can be obtained by maximizing posterior pdf:

$$\begin{aligned}\hat{w}_k &= \arg\max_{w_k} p(w_k / z_{1:k}) \\ &= \arg\max_{w_k} \{p(z_k / w_k) p(w_k / z_{1:k-1})\}\end{aligned} \quad \textbf{(17)}$$

or, equivalently, by minimizing the cost function defined as the negative logarithm of the $p(w_k / z_{1:k})$:

$$J(w_k) = -\log p(z_k / w_k) - \log p(w_k / z_{1:k-1})\,. \quad \textbf{(18)}$$

Taking into account that the observations z_k are equal to zero and neglecting unnecessary constants, (18) becomes:

$$\begin{aligned}J(w_k) = r_k(w_k)^{\mathrm{T}} R_k^{-1} r_k(w_k) + \\ (w_k - \hat{w}_{k-1})(P_{k-1} + Q_{k-1})^{-1}(w_k - \hat{w}_{k-1})\end{aligned} \quad \textbf{(19)}$$

After linearization of $r_k(w_k)$ around the $\hat{w}_{k-1}$, the maximum a posteriory estimate is obtained by setting the gradient of the cost function (9) to zero $\nabla_{w_k} J(w_k) = 0$,

$$\hat{w}_k = \hat{w}_{k-1} + K_k r_k(\hat{w}_{k-1}) \quad \textbf{(20)}$$

where K_k represents the Kalman gain:

$$K_k = (P_{k-1} + Q_{k-1}) H_k^T (R_k + H_k (P_{k-1} + Q_{k-1}) H_k^T)^{-1}, \quad \textbf{(21)}$$

and

$$H_k = \partial r_k(w_k) / \partial w_k \big|_{w_k = \hat{w}_{k-1}}\,. \quad \textbf{(22)}$$

The posterior pdf of parameters is approximated with the Gaussian: $p(w_k / z_{1:k}) = N(\hat{w}_k, P_k)$, where the estimate $\hat{w}_k$ is given by (20) and the covariance P_k is:

$$P_k = (I - K_k H_k)(P_{k-1} + Q_{k-1})\,. \quad \textbf{(23)}$$

Recursions (20) and (23) represent the basic equations of the extended Kalman filter for the problem defined by the state space model (10)-(11).

5 Simulation Results

In our examples, we consider the separation of the following sources:

$$\begin{aligned} s_{1,k} &= \sin(\pi\, k/400)\cdot n_{1,k} \quad n_{1,k} \sim N(0,1) \\ s_{2,k} &= \sin(\pi\, k/100)\cdot n_{2k} \quad n_{2,k} \sim N(0,1) \end{aligned} \qquad \textbf{(24)}$$

where $s_{1,k}$ and $s_{2,k}$ are non-stationary sources that have to be estimated, and $n_{1,k}$ and $n_{2,k}$ are zero-mean Gaussian noises, with unity variances.

Example 1. In the first example, the mixing matrix is given by:

$$\mathbf{A} = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

and mixtures **x** are obtained according to $\mathbf{x} = \mathbf{As}$. As a demixing system, we have applied the self-organized neural network (Fig. 1). For the estimation of the network weights, we have applied our EKF based separation algorithm. The convergence rate of the proposed algorithm is controlled by the process noise variance $Q_{k,}$ and is higher for the higher values of Q_k. In the given example, Q_k is set to the constant value, $Q_k = 10^{-8}$. The observation noise covariance R_k is equal to unity.

In order to demonstrate performances of our algorithm, we have compared the results with those obtained by stochastic gradient algorithm proposed in [9]. Fig. 2 shows the adaptation of the network weights through iterations. The EKF based algorithm converged to the true values in approximately 200 time steps, while the stochastic gradient (SG) based algorithm converged in more than 1000 iterations. Fig. 3 and Fig. 4 show error evolution through iterations for both algorithms.

Example 2. In the second example, we used the same definitions of the sources and EKF parameters, but the mixing matrix **A** is chosen to be non-symmetrical:

$$\mathbf{A} = \begin{bmatrix} 1 & 0.5 \\ 0.9 & 1 \end{bmatrix}$$

As shown in Figures 6. and 7., the EKF based algorithm converged in less than 200 iterations, and the SG based algorithm failed to converge in less that 2000 iterations (in this case, SG algorithm converged in approximately 3500 iterations).

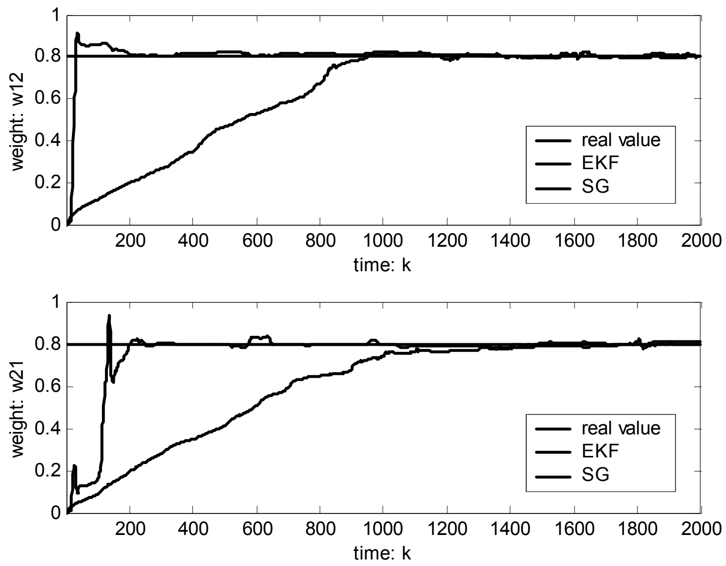

Fig. 2. Adaptation of the neural network weights through iterations (Example 1)

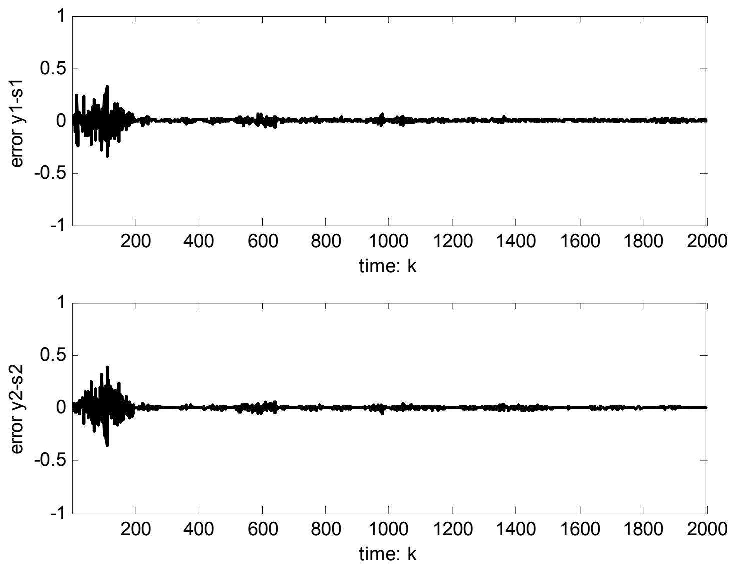

Fig. 3. Estimation errors obtained using separating algorithm based on the extended Kalman filter (Example 1)

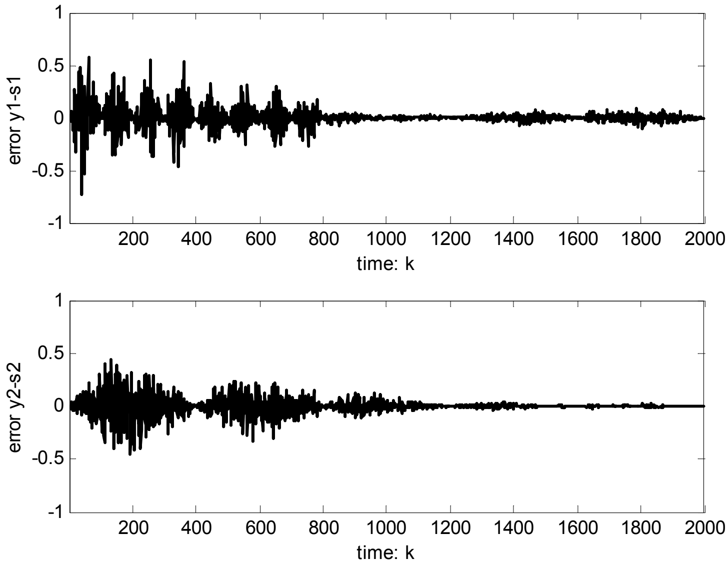

Fig. 4. Estimation errors obtained using separating algorithm based on the stochastic gradient approach (Example 1)

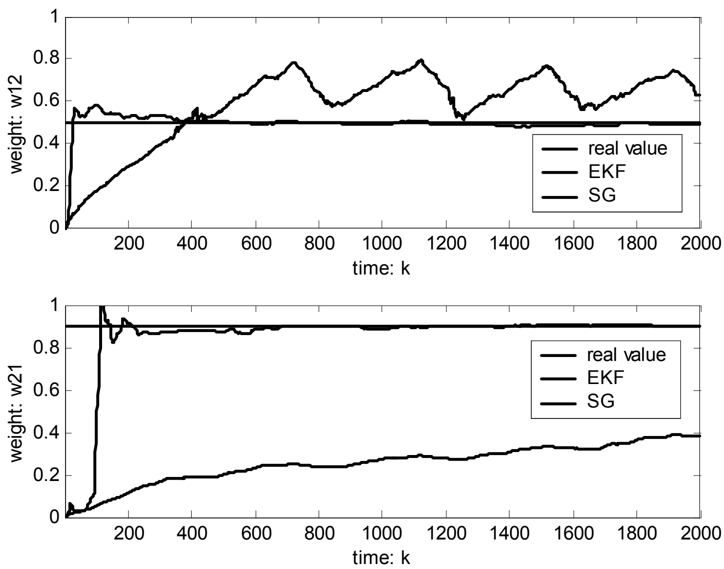

Fig. 5. Adaptation of the neural network weights through iterations (Example 2)

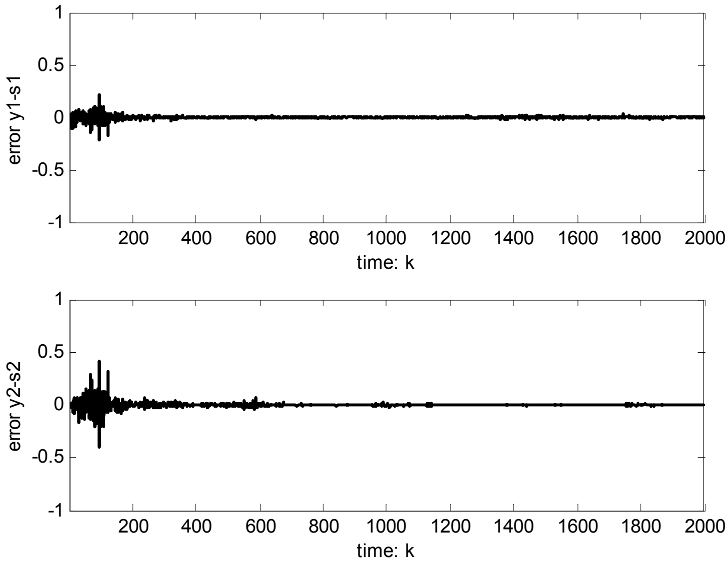

Fig. 6. Estimation errors obtained using separating algorithm based on the extended Kalman filter (Example 2)

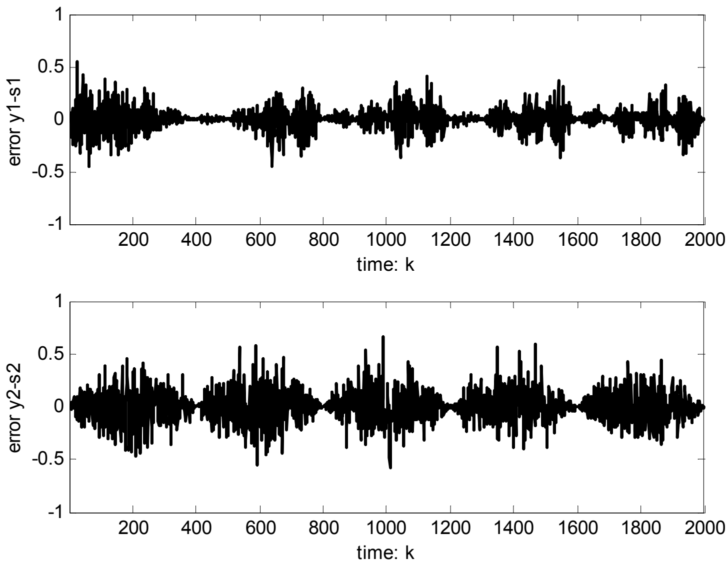

Fig. 7. Estimation errors obtained using separating algorithm based on the stochastic gradient approach (Example 2)

6 Conclusion

In this paper, we consider the blind separation of non-stationary sources using second-order statistics. We propose a state-space model of the mixing coefficient dynamics and apply extended Kalman filter to estimate the mixing model. The obtained EKF equations represent the on-line second-order optimization algorithm that minimizes cross-correlations of the estimated outputs signals. The proposed algorithm shows superior convergence properties compared to the stochastic gradient method.

References

1. Ali Mansour, Allan Kardec Barros, Noboru Ohnishi, Blind Separation of Sources: Methods, Assumptions and Applications, *IEICE Tranactions Fundamentals* E83-A(8) (2000)
2. Pierre Comon, Independent Component Analysis, a New Concept, *Signal Processing* 36: 287-314 (1994)
3. A.J.Bell, T.J.Sejnowski, An Information Maximization Approach to Blind Separation and Blind Deconvolution, *Neural Computation*, 7(6):1004-1034 (1995)
4. S.Amari, Natural Gradient Works Efficient in Learning, *Neural Computation*, 10:251-276, (1998)
5. Jean-Francois Cardoso, High-order Contrasts for Independent Component Analysis, *Neural Computation* 11:157-192 (1999)
6. Pierre Comon, From Source Separation to Blind Equalization – Contrast-based Approaches, In *Proceedings ICISP Conference*, Morocco (2001) pp.20-32
7. A.Jourjine, S.Rickard, O. Yilmaz, Blind Separation of Disjoint Orthogonal Signals: Demixing N Sources from 2 Mixtures, *Proceedings IEEE ICASSP Conference*, 2000
8. A.Mansour, The Blind Separation of Nonstationary Signals by only Using the Second Order Statistics, *Proceedings ISSPA'99 Conference*, Australia, (1999) pp. 235-238
9. K.Matsuoka, M.Ohya, M.Kawamoto, A Neural Net for Blind Separation of Nonstationary Signals, *Neural Networks*, 8(3):411-419, (1995)
10. Xi-Ren Cao, Ruey-Wen Liu, General Approach to Blind Source Separation, *IEEE Transactions on Signal Processing*, 44(3):562-571, (March 1996)
11. S.Amari, T.Chen, A.Cichocki, Stability Analysis of Learning Algorithms for Blind Source Separation, *Neural Networks*, 10(8):1345-1351, (1997)