Shifeng OU, Xianyun WANG, Ying GAO

Yantai University

Adaptive Improved RLS Algorithm for Blind Source Separation

***Abstract**. Based on an adaptive combination of two RLS-type algorithms with different forgetting factor, an effective scheme is proposed to improve the performance of the RLS-type algorithm for blind source separation. A mixing parameter for adjusting the proportion of the two RLS algorithms is introduced in an attempt to put together the best properties of them, and its adaptive rule is obtained by means of a natural gradient criterion. Experimental results demonstrate the good performance of the proposed approach in different kinds of environments.*

***Streszczenie.** W artykule przedstawiono nową, efektywniejszą strukturę algorytmu RLS do ślepej separacji sygnałów, bazującą na adaptacyjnej kombinacji dwóch takich algorytmów z różnymi współczynnikami ważenia. W celu uzyskania jak najlepszego wykorzystania ich własności, zastosowano parametr, który pozwala na ich dostrojenie. Wyniki eksperymentalne potwierdzają skuteczność działania. (**Ulepszony algorytm adaptacyjny RLS do ślepej separacji sygnałów**).*

Keywords: Blind source separation, Recursive least squares, Nonlinear principle component analysis
Słowa kluczowe: ślepa separacja sygnałów, rekurencyjna metoda NK, RLS, analiza głównych składowych.

Introduction

Blind source separation (BSS) aims to estimate the original source signals from their mixtures captured by a number of sensors without knowing any prior information about the mixing channels. The problem of BSS has drawn lost of attention in various fields, such as speech signal processing, wireless communication, biomedical and image signal processing [1]. Over the past two decades, a variety of algorithms have been proposed for BSS, such as fast independent component analysis (FastICA), natural gradient, nonlinear principle component analysis (PCA) algorithm and so on [2-5]. Most of these approaches can be categorized into either batch or adaptive techniques. The adaptive BSS algorithms has drawn more applications as these algorithms have particularly practical advantages compared with the batch algorithms. However, most of the adaptive algorithms belong to the LMS-type algorithms, and they often converges slowly. One solution to solve this problem is allowing the step size of the adaptive algorithm can be adjusted in a proper way [6,7], but the step size adaptation technique often results in introducing some hyper parameters which are not easy to establish. The other effective solution is to adopt RLS formulation. It is well known that the RLS algorithm offers superior convergence rate as compared to the LMS algorithm.

Our proposal to improve the performance of RLS BSS algorithm is to use a scheme by combining two RLS algorithms with different forgetting factor adaptively. The proposed combination is carried out in an attempt to put together the best properties of the two RLS algorithms, and the mixing parameter that controls the combination is adapted by means of a natural gradient algorithm. Since the combination is adaptive itself, the overall scheme is equivalent to a variable forgetting factor procedure in which the forgetting factor is optimally selected at every time point.

RLS-type BSS

Suppose n unknown, statistically independent source signals, with at most one is Gaussian distributed, contained within s(t)$\in\Re^n$, are mixed by an unknown mixing matrix A $\in\Re^{m\times n}$, therefore, m mixed signals x(t)$\in\Re^m$ are observed x(t)=As(t)+d(t), where t is the time index, and d(t)$\in\Re^m$ represents the additive noise which is ignored in this paper. To simplify the problem, we assume that the number of sources matches the number of mixtures i.e. m=n. The main problem of BSS is then to estimate the original source s(t) only using the observed mixtures x(t). This is equivalent to estimate the mixing matrix A and achieves the original sources by finding a separating matrix B, which performs the inverse operation as subsequently used in the separation model. Then, the output signal vector can be obtained by y(t)= Bx(t), where y(t) represent the estimated source signals. Whitening is often considered as a necessary condition for BSS. After whitening, the BSS tasks usually become somewhat easier and less ill-conditioned. Suppose the whitening matrix V is obtained in batch mode or online algorithm, and let the whitened signals be v(t)= Vx(t) such that E{v(t)vT(t)}=I, then an orthogonal matrix W is learned to achieve the separation of mixed signals. Hence, the total separating matrix can be written by B=VW. There exist various possible criteria to determine the orthogonal matrix W. The focus of this paper is the nonlinear PCA criterion which has been theoretically proved to perform the BSS employing the orthogonality constraint as follows

$$J_M[\boldsymbol{W}(t)] = E\left\{\left\|\boldsymbol{v}(t) - \boldsymbol{W}^T(t)\boldsymbol{g}\left[\boldsymbol{W}(t)\boldsymbol{v}(t)\right]\right\|^2\right\} \tag{1}$$

where $\boldsymbol{g}(\boldsymbol{y})$=[$g(y_1)$, $\ldots g(y_n)]^T$ denote the nonlinear function. By replacing the mean-square error in (1) with the exponentially weighted sum of error squares and approximating the unknown vector $\boldsymbol{g}[\boldsymbol{W}(t)\boldsymbol{v}(t)]$ by $z(t)$=$\boldsymbol{g}[\boldsymbol{W}(t-1)\boldsymbol{v}(t)]$, the least-squares type criterion can be gotten

$$J_L[\boldsymbol{W}(t)] = \sum_{i=1}^{t}\beta^{t-i}\left\|\boldsymbol{v}(i) - \boldsymbol{W}^T(t)\boldsymbol{z}(i)\right\|^2 \tag{2}$$

where β is the forgetting factor which is kept in interval (0, 1]. The gradient of $J_L[\boldsymbol{W}(t)]$ with respect to $\boldsymbol{W}(t)$ is

$$\nabla J_L[\boldsymbol{W}(t)] = \sum_{i=1}^{t}\beta^{t-i}\left[-\boldsymbol{z}(i)\boldsymbol{v}^T(i) + \boldsymbol{z}(i)\boldsymbol{z}^T(i)\boldsymbol{W}(t)\right] \tag{3}$$

Based on natural gradient learning and approximating the vector $\boldsymbol{W}(t)\boldsymbol{v}(i)$ by $\boldsymbol{y}(i)$=$\boldsymbol{W}(i-1)\boldsymbol{v}(i)$, we can obtain the optimal weighting matrix at time t

$$\boldsymbol{W}_{opt}(t) = \left[\sum_{i=1}^{t}\beta^{t-i}\boldsymbol{y}(i)\boldsymbol{z}^T(i)\right]^{-1}\left[\sum_{i=1}^{t}\beta^{t-i}\boldsymbol{z}(i)\boldsymbol{v}^T(i)\right] = \boldsymbol{R}^{-1}(t)\boldsymbol{C}(t) \tag{4}$$

where $\boldsymbol{R}(t)$= $\sum_{i=1}^{t}\beta^{t-i}\boldsymbol{y}(i)\boldsymbol{z}^T(i)$ and $\boldsymbol{C}(t)$= $\sum_{i=1}^{t}\beta^{t-i}\boldsymbol{z}(i)\boldsymbol{v}^T(i)$. By applying the matrix inversion lemma, the matrix $\boldsymbol{P}(t)$=$\boldsymbol{R}^{-1}(t)$ can be recursively computed, ant it yields an adaptive RLS-type algorithm for BSS with the summary as follows

$$
(5) \quad \begin{aligned}
&\boldsymbol{y}(t)=\boldsymbol{W}(t-1)\boldsymbol{v}(t),\ \boldsymbol{z}(t)=\boldsymbol{g}[\boldsymbol{y}(t)]\\
&\boldsymbol{Q}(t)=\frac{\boldsymbol{P}(t-1)}{\beta+\boldsymbol{z}^T(t)\boldsymbol{P}(t-1)\boldsymbol{y}(t)}\\
&\boldsymbol{P}(t)=\frac{1}{\beta}\left[\boldsymbol{P}(t-1)-\boldsymbol{Q}(t)\boldsymbol{y}(t)\boldsymbol{z}^T(t)\boldsymbol{P}(t-1)\right]\\
&\boldsymbol{W}(t)=\boldsymbol{W}(t-1)\\
&\quad+\left[\boldsymbol{P}(t)\boldsymbol{z}(t)\boldsymbol{v}^T(t)-\boldsymbol{Q}(t)\boldsymbol{y}(t)\boldsymbol{z}^T(t)\boldsymbol{W}(t-1)\right]
\end{aligned}
$$

where the forgetting factor β is close to 1, it controls the convergence rate of the RLS algorithm but also determines the final misadjustment.

The Improved RLS Algorithm

To improve the performance of RLS-type BSS algorithm, the fundamental idea of our proposed adaptive combination is that two RLS algorithms with different forgetting factor are adopted separately, and the output signals of both algorithms are combined by a mixing parameter in such a way that the advantages of both components are retained: the rapid convergence from the fast algorithm (small β) and the reduced misadjustment from the slow algorithm (large β). The output vector of the proposed scheme using such an adaptive combination of two RLS algorithms is

$$
(6) \quad \boldsymbol{y}(t)=\boldsymbol{W}(t)\boldsymbol{v}(t)=\lambda(t)\boldsymbol{y}_1(t)+\left[1-\lambda(t)\right]\boldsymbol{y}_2(t)
$$

where $\boldsymbol{y}_1(t)$ and $\boldsymbol{y}_2(t)$ are the respective outputs of either RLS algorithm, i.e., $\boldsymbol{y}_l(t)=\boldsymbol{W}_l(t\text{-}1)\boldsymbol{v}(t)$, $l=1,2$, $\lambda(t)$ is the mixing para meter which is kept in interval [0, 1], and $\boldsymbol{W}(t)$ is the overall s eparating matrix that can be thought of as

$$
(7) \quad \boldsymbol{W}(t)=\lambda(t)\boldsymbol{W}_1(t-1)+\left[1-\lambda(t)\right]\boldsymbol{W}_2(t-1)
$$

The idea is that if $\lambda(t)$ is assigned appropriate values at each iteration, then the combined separating matrix $\boldsymbol{W}(t)$ in (7) would extract the best properties of the individual separating matrix $\boldsymbol{W}_1(t\text{-}1)$ and $\boldsymbol{W}_2(t\text{-}1)$. Both matrix operate completely decoupled from each other using the RLS adaptation rule in (5) with different forgetting factor

$$
(8) \quad \mathrm{RLS}_i \begin{cases}
\boldsymbol{y}_i(t)=\boldsymbol{W}_i(t-1)\boldsymbol{v}_i(t),\ \boldsymbol{z}_i(t)=\boldsymbol{g}[\boldsymbol{y}_i(t)]\\
\boldsymbol{Q}_i(t)=\dfrac{\boldsymbol{P}_i(t-1)}{\beta_i+\boldsymbol{z}_i^T(t)\boldsymbol{P}_i(t-1)\boldsymbol{y}_i(t)}\\
\boldsymbol{P}_i(t)=\dfrac{1}{\beta_i}\left[\boldsymbol{P}_i(t-1)-\boldsymbol{Q}_i(t)\boldsymbol{y}_i(t)\boldsymbol{z}_i^T(t)\boldsymbol{P}_i(t-1)\right]\\
\boldsymbol{W}_i(t)=\boldsymbol{W}_i(t-1)\\
\quad+\boldsymbol{P}_i(t)\boldsymbol{z}_i(t)\boldsymbol{v}_i^T(t)-\boldsymbol{Q}_i(t)\boldsymbol{y}_i(t)\boldsymbol{z}_i^T(t)\boldsymbol{W}_i(t-1)
\end{cases}
$$

where $i=1,2$, and $\beta_1<\beta_2$, which means the RLS_1 separating system is faster but results in more misadjustment than the RLS_2 system. The key point of the proposed algorithm is to control the mixing parameter according to the performance of the two components. However, instead of modifying $\lambda(t)$ directly, we adopt a variable parameter $\rho(t)$ that defines via a sigmoidal function. The sigmoidal function is

$$
(9) \quad \lambda(t)=\mathrm{sgm}\left[\rho(t)\right]=\left[1+e^{-\rho(t)}\right]^{-1}
$$

and the update equation for $\rho(t)$ is given by

$$
(10) \quad \rho(t+1)=\rho(t)-\mu_\rho\nabla_\rho J_M(t)\big|_{\rho=\rho(t)}
$$

where μ_ρ is a constant, and $J_M(t)$ is the cost function in (1) from which the natural gradient based nonlinear PCA algorithm is derived. To proceed, we use an inner product of matrices defined as $\langle\boldsymbol{C},\boldsymbol{D}\rangle=\mathrm{tr}\left(\boldsymbol{C}^T\boldsymbol{D}\right)$, where $\langle\cdot\rangle$ denotes the inner product, $\mathrm{tr}(\cdot)$ is the trace operator, and $\boldsymbol{C}$, $\boldsymbol{D}\in\mathbb{R}^{m\times n}$. Therefore, the gradient term on the right hand side of (10) can be evaluated as

$$
(11) \quad \begin{aligned}
\nabla_\rho J_M(t)\big|_{\rho=\rho(t)}&=\left\langle\frac{\partial J(t)}{\partial\boldsymbol{W}(t)},\frac{\partial\boldsymbol{W}(t)}{\partial\lambda(t)}\frac{\partial\lambda(t)}{\partial\rho(t)}\right\rangle\\
&=\mathrm{tr}\left[\left(\frac{\partial J(t)}{\partial\boldsymbol{W}(t)}\right)^T\frac{\partial\boldsymbol{W}(t)}{\partial\lambda(t)}\frac{\partial\lambda(t)}{\partial\rho(t)}\right]
\end{aligned}
$$

From (1), we can get

$$
(12) \quad \frac{\partial J_M(t)}{\partial\boldsymbol{W}(t)}=-\left\{\boldsymbol{g}[\boldsymbol{y}(t)]\boldsymbol{v}^T(t)-\boldsymbol{y}(t)\boldsymbol{g}^T[\boldsymbol{y}(t)]\boldsymbol{W}(t)\right\}
$$

According to (7) and (9), we can obtain

$$
(13) \quad \frac{\partial\boldsymbol{W}(t)}{\partial\lambda(t)}=\boldsymbol{W}_1(t-1)-\boldsymbol{W}_2(t-1)\,,\ \frac{\partial\lambda(t)}{\partial\rho(t)}=\lambda(t)\left[1-\lambda(t)\right].
$$

Denoting $\boldsymbol{\Gamma}(t)=\left\{\boldsymbol{g}[\boldsymbol{y}(t)]\boldsymbol{v}^T(t)-\boldsymbol{y}(t)\boldsymbol{g}^T[\boldsymbol{y}(t)]\boldsymbol{W}(t)\right\}$ and $\boldsymbol{H}(t)=\boldsymbol{W}_1(t-1)-\boldsymbol{W}_2(t-1)$. Finally, we can obtain

$$
(14) \quad \nabla_\rho J_M(t)\big|_{\rho=\rho(t)}=-\mathrm{tr}\left[\boldsymbol{\Gamma}^T(t)\boldsymbol{H}(t)\right]\lambda(t)\left[1-\lambda(t)\right]
$$

and thus the adaptive update equation for $\rho(t)$ with the form of (10) can be written as

$$
(15) \quad \rho(t+1)=\rho(t)+\mu_\rho\,\mathrm{tr}\left[\boldsymbol{\Gamma}^T(t)\mathbf{H}(t)\right]\lambda(t)\left[1-\lambda(t)\right]
$$

Here, the parameter $\rho(t)$ should be restricted to the interval $[-\rho^+,\rho^+]$, which limits the range of $\lambda(t)$ to $[1-\lambda^+,\lambda^+]$, where $\lambda^+=\mathrm{sgm}(\rho^+)$ is a constant close to 1. A good choice for ρ^+ is 4 that constrains $\lambda(t)$ to [0.02, 0.98].

This proposed scheme has a very simple interpretation: in situations where a fast convergence speed would be desirable, the fast RLS algorithm will outperform the slow one, making $\lambda(t)$ approaches towards 1 and $\boldsymbol{W}(t)\approx\boldsymbol{W}_1(t)$. However, in stationary intervals, it is the slow algorithm which operates better, making $\lambda(t)$ get close to 0 and $\boldsymbol{W}(t)\approx\boldsymbol{W}_2(t)$. It is also possible to further improve the performance of the our combination algorithm by using the good convergence properties of the fast RLS algorithm to speed up the convergence rate of the slow one. We can do this by step-by-step transferring a part of weight matrix $\boldsymbol{W}_1(t)$ to $\boldsymbol{W}_2(t)$, and this transfer procedure is only applied if the fast algorithm is significantly outperforming the slow one.

Simulations

In this section, the performance of the proposed improved RLS algorithm was compared to that of the nonlinear PCA algorithm and the RLS for BSS. Firstly, three zero mean Sub-Gaussian sources were mixed by a 3×3 mixing matrix, and the nonlinear function $g(\cdot)=\tanh(\cdot)$ is used. The value of step size for the onlinear PCA was initialized at 0.02, and the forgetting factor parameters of the proposed new algorithm were set to β_1=0.9, β_2=0.996 and the constant μ_ρ is 8. The value of forgetting factor for the RLS algorithm were set at 0.9 and 0.996, respectively. The performance index (PI) as a function of the system matrix ***G=WVA,*** was used to evaluate the performance of these BSS algorithms, which ideally attains its minimum value, zero, when separation is achieved. Fig. 1 plots the PI values obtained from the simulations of each BSS algorithm averaged over 50 Monte Carlo trials. Fig. 2 plots the evolution of $\lambda(k)$ in our new algorithm, The estimated mixing parameter $\lambda(k)$ can adaptively adjust its value according to the evolution of algorithm's performance. The evolution of this parameter matches the requirements of the adaptive combination.

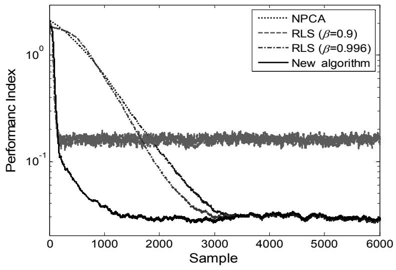

Fig.1. PI achieved by four algorithms in stationary environment

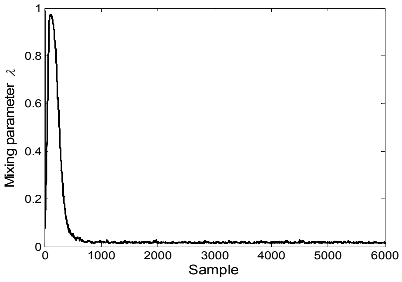

Fig.2. Evolution of $\lambda(t)$ in stationary environment

In the second experiment, the mixing matrix was changed abruptly at sample number 4000 and 7000, respectively, which means the performance of these adaptive BSS algorithms is compared in a non-stationary environment. The parameter values of all the algorithms were initialized the same as in the first stationary environment. The convergence performance of the PI measurements and the evolution of $\lambda(t)$ averaged over 50 Monte Carlo trials were showed in Fig. 3 and Fig. 4, which again confirmed the advantage of the proposed algorithm while following the abrupt change in the mixing channel.

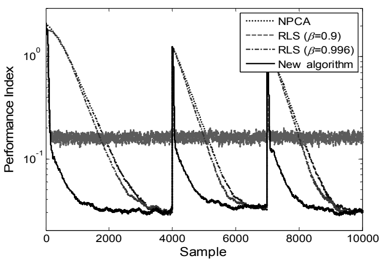

Fig.3. PI achieved by four algorithms in non-stationary environment

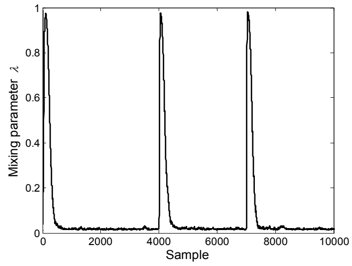

Fig.4. Evolution of $\lambda(t)$ in non-stationary environment

Conclusions

This paper proposed to use an adaptive combining structure solving the inherent compromise problem between convergence speed and misadjustment in the RLS-type BSS algorithm. Compared to the regular BSS algorithms, the proposed scheme was composed of two adaptive separating systems, and a mixing parameter was designed to control the proportion of the two components. Since the combination of the two RLS algorithms can extract the best properties of them, the fast convergence speed and small steady-state mean square error can be obtained. Simulation results shown the improved performance of the proposed approach.

Acknowledgments

This work was supported in part by NSFC under Grant Nos. 61005021, 61201457 and A Project of Shandong Province Higher Educational Science and Technology Program under contract J12LN27.

REFERENCES

[1] A. Cichocki and S. Amari. *Adaptive blind signal and image processing: learning algorithms and application*, New York: Wiley, (2002) 1-41
[2] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, *IEEE Transactions on Neural Networks*, 10(1999), No. 3, 626-634
[3] Y. Tang and J. P. Li, Normalized natural gradient in independent component analysis, *Signal Processing*, 90(2010), No. 9, 2773-2777
[4] S. C. Douglas, Self-stabilized gradient algortihms for blind source separation with orthogonality constraints, *IEEE Transactions on Neural Networks*, 11(2000), No. 6, 1490-1497
[5] J. M. Ye, H. H. Jin, S. T. Lou and K. J. You, An optimized EASI algorithm, *Signal Processing*, 89(2009), No. 3, 333–338
[6] L. X. Yuan, W. W. Wang and J. A. Chambers, Variable step-size sign natural gradient algorithm for sequential blind source separation, *IEEE Signal Processing Letters*, 12(2005) , No. 8, 589-592

***Authors**: Shifeng Ou, associate prof., Institute of Science and Technology for Opto-electronic Information, Yantai University, Yantai, China, E-mail: ousfeng@126.com*
Xianyun Wang, Institute of Science and Technology for Opto-electronic Information, Yantai University, Yantai, China, E-mail: cailiaoytu@126.com
Ying Gao (corresponding author), Institute of Science and Technology for Opto-electronic Information, Yantai University, Yantai, China, E-mail: claragaoying@126.com