Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Signal Processing 84 (2004) 801 – 804

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

Fast communication

A novel adaptive learning rate sequential blind source separation algorithm

M.G. Jafari[a,*], J.A. Chambers[b], D.P. Mandic[c]

[a] *Centre for Digital Signal Processing Research, King's College London, London, UK*
[b] *Cardiff School of Engineering, Cardiff University, Cardiff, UK*
[c] *Communications and Signal Processing Research Group, Imperial College, London, UK*

Received 19 May 2003; received in revised form 29 September 2003

Abstract

A new member of the family of natural gradient algorithms for on-line blind separation of independent sources is proposed. The method is based upon an adaptive step-size which varies in sympathy with the dynamics of the input signals and properties of the de-mixing matrix, and is robust to the perturbations in the initial value of the learning rate parameter. As a result, the convergence speed is significantly improved, especially in non-stationary mixing environments. Simulations support the expected improvement in convergence speed of the approach.

Keywords: Blind source separation; Natural gradient algorithm; Adaptive step-size

1. Introduction

In many practical applications, such as communications and biomedical signal processing, it is necessary to estimate the unknown source signals from a set of observed measurements. When m unobservable source signals contained within $\mathbf{s}(k) \in \mathbb{R}^m$ are mixed by an unknown, full column rank matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$, n mixture signals are generated according to $\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k)$, where $\mathbf{x}(k) \in \mathbb{R}^n$ represents the vector of observed signals, and k denotes the discrete time index. The goal of blind source separation (BSS) is then to recover the original sources given only the observed mixtures, using the separating model $\mathbf{y}(k) = \mathbf{W}(k)\mathbf{x}(k)$, where $\mathbf{y}(k)$ is an estimate of $\mathbf{s}(k)$ to within the well-known permutation and scaling ambiguities, and $\mathbf{W}(k) \in \mathbb{R}^{m\times n}$ is the separating matrix. The crucial assumption at the core of conventional BSS is that the original sources are mutually statistically independent. In addition, it is assumed that the sources have unit variance, and in this paper, for convenience of presentation, we assume that there are as many sources as mixtures, that is $m = n$. The blind separation of sources effectively involves the application of blind channel identification and signal estimation techniques, since the separation procedure requires the estimation, explicit or otherwise, of an unmixing matrix, which is subsequently used to evaluate the separating model, leading to the recovery of the sources. The natural gradient algorithm [1] is a BSS method that updates the separating matrix coefficients according to

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu[\mathbf{I} - \mathbf{f}(\mathbf{y}(k))\mathbf{y}^{\mathrm{T}}(k)]\mathbf{W}(k), \quad (1)$$

* Corresponding author.
E-mail address: maria.jafari@kcl.ac.uk (M.G. Jafari).

0165-1684/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2003.12.012

where $\mathbf{f}(\mathbf{y}(k))$ is an odd non-linear function which acts upon the elements of the output vector $\mathbf{y}(k)$, and μ is a positive step-size parameter.

2. Adaptive step-size algorithm

The function of the step-size parameter in the stochastic gradient descent setting is to control the magnitudes of the updates of the estimated parameters at each iteration. Its choice is crucial to the performance of the algorithm, and generally the use of a fixed step-size parameter can lead to slow convergence speed and poor tracking performance. An alternative approach is to use an adaptive step-size, whose value is adjusted according to the time-varying dynamics of the signals and the separating matrix [5–7]. In [3], the learning rate changes according to a non-linear function of the mean values of the gradient components:

$$\hat{\mathbf{g}}(k) = (1-\beta_2)\hat{\mathbf{g}}(k-1) + \beta_2\tilde{\mathbf{g}}(k),$$

$$\eta(k) = (1-\beta_1)\eta(k-1) + \beta_1\gamma\phi(\|\hat{\mathbf{g}}(k)\|), \tag{2}$$

where $0<\beta_1<1$, $0<\beta_2<1$, and $\gamma>0$ are fixed coefficients, $\tilde{\mathbf{g}}(k) = -(\mathbf{I}-\mathbf{f}(\mathbf{y}(k))\mathbf{y}^{\mathrm{T}}(k))\mathbf{W}(k)$, is the gradient at time k, and $\phi(\|\hat{\mathbf{g}}(k)\|)$ is a non-linear function defined in [3]. However, the algorithm in (2) has the disadvantage of requiring the selection of three additional parameters used in the update, and its performance is sensitive to their values. Thus, in this paper we propose a gradient adaptive step-size algorithm, which updates the learning rate according to

$$\mu(k) = \mu(k-1) - \rho\nabla_\mu J(k)|_{\mu=\mu(k-1)}, \tag{3}$$

where ρ is a small constant, and $J(k)$ is the natural gradient algorithms (NGA) cost function. Methods exist for the exact calculation of this quantity [4], but they result in greatly increased computational complexity and are sensitive to small variations in the parameters which are used in the update equation. Here, we propose a simple, yet effective update of $\mu(k)$ by employing a gradient-based adaptation. In the derivation of the algorithm, following the approach in [2], we introduce an inner product of matrices, defined as

$$\langle \mathbf{C},\mathbf{D}\rangle = \mathrm{tr}(\mathbf{C}^{\mathrm{T}}\mathbf{D}) \tag{4}$$

to evaluate the gradient term on the right-hand side of (3). In Eq. (4) $\langle\cdot,\cdot\rangle$ denotes the inner product, $\mathrm{tr}(\cdot)$ is the trace operator, and $\mathbf{C},\mathbf{D}\in\mathbb{R}^{m\times n}$. Notice that, due to the complexities of matrix differential calculus

$$\nabla_\mu J(k)|_{\mu=\mu(k-1)} \neq \frac{\partial J(k)}{\partial \mathbf{W}(k)} \times \frac{\partial \mathbf{W}(k)}{\partial \mu(k-1)}. \tag{5}$$

Hence, to evaluate the gradient term on the right-hand side of (3), we employ the inner product as defined in (4), leading to

$$\begin{aligned}\nabla_\mu J(k)|_{\mu=\mu(k-1)} &= \left\langle \frac{\partial J(k)}{\partial \mathbf{W}(k)}, \left(\frac{\partial \mathbf{W}(k)}{\partial \mu(k-1)}\right)\right\rangle \\ &= \mathrm{tr}\left(\frac{\partial J(k)}{\partial \mathbf{W}(k)} \times \left(\frac{\partial \mathbf{W}(k)}{\partial \mu(k-1)}\right)^{\mathrm{T}}\right),\end{aligned} \tag{6}$$

where

$$\frac{\partial J(k)}{\partial \mathbf{W}(k)} = -[\mathbf{I}-\mathbf{f}(\mathbf{y}(k))\mathbf{y}^{\mathrm{T}}(k)]\mathbf{W}(k), \tag{7}$$

which is the instantaneous estimate of the natural gradient of the cost function $J(k)$ [1]. Notice that from (1), the separating matrix at time k is given by

$$\begin{aligned}\mathbf{W}(k) = \mathbf{W}(k-1) + \mu(k-1)[\mathbf{I}-\mathbf{f} \\ \times(\mathbf{y}(k-1))\mathbf{y}^{\mathrm{T}}(k-1)]\mathbf{W}(k-1).\end{aligned} \tag{8}$$

Following the approach from [7], and considering the direct path only, from (8) we have

$$\begin{aligned}\frac{\partial \mathbf{W}(k)}{\partial \mu(k-1)} = [\mathbf{I}-\mathbf{f}(\mathbf{y}(k-1)) \\ \times \mathbf{y}^{\mathrm{T}}(k-1)]\mathbf{W}(k-1).\end{aligned} \tag{9}$$

Substituting (7) and (9) into (6) leads to

$$\begin{aligned}\frac{\partial J(k)}{\partial \mu(k-1)} &= -\mathrm{tr}([\mathbf{I}-\mathbf{f}(\mathbf{y}(k))\mathbf{y}^{\mathrm{T}}(k)]\mathbf{W}(k) \\ &\quad \times \mathbf{W}^{\mathrm{T}}(k-1)[\mathbf{I}-\mathbf{y}(k-1)\mathbf{f}^{\mathrm{T}}(\mathbf{y}(k-1))]).\end{aligned} \tag{10}$$

Finally, substituting back into (3), the resulting adaptive step-size algorithm is given by

$$\begin{aligned}\mu(k) = \mu(k-1) + \rho\,\mathrm{tr}([\mathbf{I}-\mathbf{f}(\mathbf{y}(k))\mathbf{y}^{\mathrm{T}}(k)]\mathbf{W}(k) \\ \times \mathbf{W}^{\mathrm{T}}(k-1)[\mathbf{I}-\mathbf{y}(k-1)\mathbf{f}^{\mathrm{T}}(\mathbf{y}(k-1))]).\end{aligned} \tag{11}$$

Thus, the proposed algorithm effectively introduces memory in the step-size update, as well as preserving information regarding the direction of descent, so that it is capable of reacting in sympathy with the changes in the mixing environment.

3. Simulations

Two sub-Gaussian sources were mixed by a real stationary channel, and zero mean, independent white Gaussian noise was added such that the signal-to-noise ratio was 20 dB. The mixtures were separated using conventional NGA with fixed step-size $\mu = 5 \times 10^{-4}$, Cichocki's method in (2), with $\eta(0) = 5 \times 10^{-4}$, $\beta_1 = \beta_2 = 0.01$, and $\gamma = 4 \times 10^{-4}$, and the proposed algorithm with adaptive learning rate (11), where $\mu(0) = 5 \times 10^{-4}$, and $\rho = 10^{-7}$. Notice that the user needs only to select a single parameter ρ to control the dynamics of $\mu(k)$. The results were evaluated using the performance index (PI), as conventionally employed to assess BSS algorithms [3]. Fig. 1 shows the PI resulting from the application of the three methods, and averaged over 100 Monte Carlo trials, and illustrates that the convergence speed of NGA is considerably faster when the proposed adaptive step-size algorithm is employed. In particular, NGA was found to require less than 800 samples to converge to a PI of 0.01 when the proposed algorithm was employed, while it required over 1800 samples with Cichocki's method, and did not achieve this value within 4000 samples when the fixed step-size parameter was used. Fig. 2 depicts the average performance indices for the three methods when the sources were mixed by a time-varying mixing channel, whose elements varied according to independent first-order Gauss–Markov models, and changed abruptly after 2000 samples. Additive white Gaussian noise was also present, resulting in a signal-to-noise ratio of 20 dB. The convergence curves in Fig. 2 shows that the average performance of the NGA algorithm improved considerably when separation was carried out using the proposed adaptive step-size method. This was especially evident during initial convergence, and following the abrupt change in the mixing channel, where the use of a fixed learning rate resulted in slow convergence speed, while the variable step-size approach presented here ensured that the algorithm reacted quickly to the changes in the mixing channel. Fig. 3 shows the logarithm of the average PI obtained after 2000 iterations (PI_c), for various values of the fixed step-size μ, and with $\eta(0) = \mu(0) = \mu$ in (2) and (11). The proposed NGA-type algorithm was found to consistently outperform conventional NGA with a fixed step-size, and the adaptive learning rate method in (2), over the set of values considered.

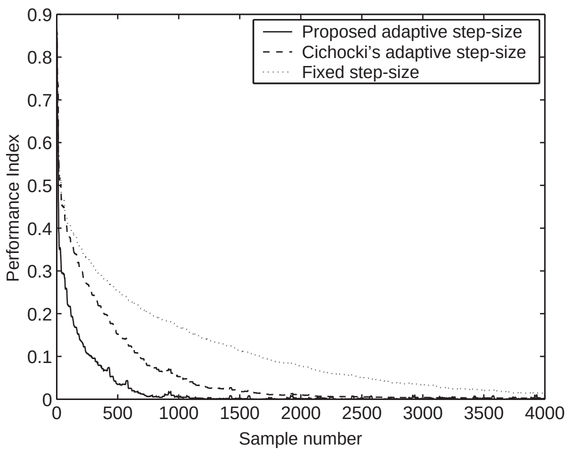

Fig. 1. Average performance indices obtained for NGA with fixed step-size and adaptive learning rates in (2) and (11).

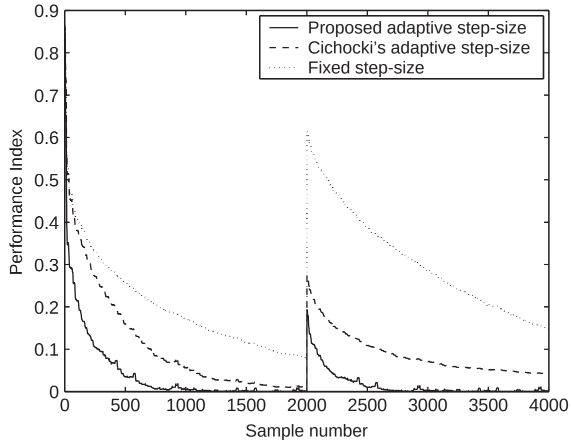

Fig. 2. Behaviour of NGA with the fixed and adaptive learning rates (2) and (11), when the mixing channel is non-stationary.

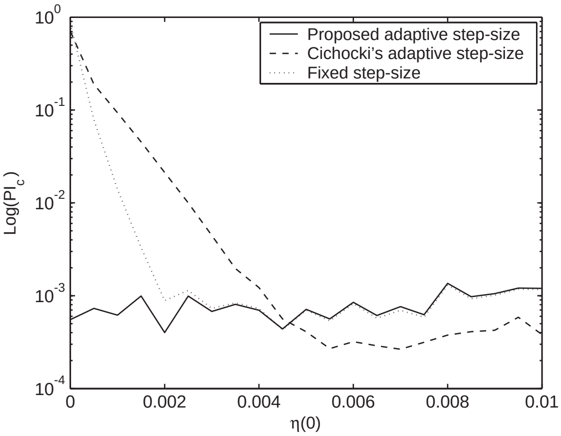

Fig. 3. Performance of the average PI after 2000 samples when the fixed step-size (or initial step-size) is varied.

4. Conclusions

A novel natural gradient-based adaptive step-size algorithm for the blind separation of sources has been proposed. The algorithm has been shown to be robust to the perturbations in the initial value of the learning rate parameter, and is of the same order of complexity as the standard algorithm. By varying the learning rate in response to changes in the dynamics of the estimated parameters, this technique improves the performance of the conventional natural gradient algorithm, and is especially well suited to the separation of sources mixed by a time-varying environments. In particular, simulation results have shown that improved convergence rate is achieved when the sources are extracted by the proposed approach.

References

[1] S. Amari, A. Cichocki, Adaptive blind signal processing–neural network approaches, Proc. IEEE 86 (1998) 2026–2048.

[2] A. Cichocki, S. Amari, Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications, Wiley, New York, 2002.

[3] A. Cichocki, B. Orsier, A. Back, S. Amari, On-line adaptive algorithms in non-stationary environments using a modified conjugate gradient approach, in: Proceedings of the IEEE Workshop on Neural Networks for Signal Processing, Amelia Island Plantation, FL, 1997, pp. 316–325.

[4] S.C. Douglas, A. Cichocki, Adaptive step size techniques for decorrelation and blind source separation, Proceedings of the Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, 1998, Vol. 2, pp. 1191–1195.

[5] M.G. Jafari, J.A. Chambers, Normalised natural gradient algorithm for the separation of cyclostationary sources, in: Proceedings of the IEEE International Conference On Acoustic, Speech, and Signal Processing, Hong Kong, 2003, Vol. 5, pp. 301–304.

[6] D.P. Mandic, A. Cichocki, An online algorithm for blind extraction of sources with different dynamical structures, in: Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation, Nara, Japan, 2003, pp. 645–651.

[7] V.J. Mathews, Z. Xie, A stochastic gradient adaptive filter with gradient adaptive step size, IEEE Trans. Signal Process. 41 (1993) 2075–2087.