Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Neurocomputing 70 (2006) 229–240
NEUROCOMPUTING
www.elsevier.com/locate/neucom

A modified infomax algorithm for blind signal separation

Hyung-Min Park[a,*], Sang-Hoon Oh[b], Soo-Young Lee[a]

[a]Brain Science Research Center and Department of Biosystems, Korea Advanced Institute of Science and Technology, Daejeon, 305-701, Republic of Korea
[b]Department of Information Communication Engineering, Mokwon University, Daejeon, 302-729, Republic of Korea

Received 19 April 2005; received in revised form 17 March 2006; accepted 17 March 2006
Communicated by T. Heskes
Available online 29 June 2006

Abstract

We present a new algorithm to perform blind signal separation (BSS), which takes a trade-off between the ordinary gradient infomax algorithm and the natural gradient infomax algorithm. Analyzing the algorithm, we show that desired equilibrium points are locally stable by choosing appropriate score functions and step sizes. The algorithm provides better performance than the ordinary gradient algorithm, and it is free from approximation error and the small-step-size restriction of the natural gradient algorithm. In simulations on convolved mixtures, the algorithm provides much better performance than the other algorithms while requiring less computation.

Keywords: Independent component analysis; Blind signal separation; Entropy maximization; Gradient learning

1. Introduction

The blind signal separation (BSS) problem is to find a transform that recovers source signals from their mixtures without knowing how the sources are mixed [13,19]. Although the term 'blind' means that no prior information is available, many BSS algorithms rely on statistical independence of source signals [5,8]. Only with this statistical-independent assumption, BSS shows good performance in many applications and it has received extensive attention in signal and speech processing, machine learning, and neuroscience communities.

Although many researchers have proposed algorithms to perform BSS, a large number of these are batch-type with prewhitened signals of instantaneous mixtures. In many practical applications, however, all mixing data are not given in advance, and outputs have to be immediately provided for each input sample. In addition, batch-type algorithms cannot be used for non-stationary environments. Furthermore, convolved mixtures of natural signals which have correlation among time samples are often addressed. For such practical applications, it is necessary for BSS algorithms to have separation capability of convolved mixtures with on-line adaptation even without prewhitening. Unfortunately, the majority of algorithms cannot handle these applications because they have been developed to separate instantaneous mixtures or whitened signals with batch-type processing [6,15,16,25,30].

As an approach to BSS without these difficulties, an ordinary gradient algorithm for entropy maximization is notable for its simple and biologically plausible formulation [4,29]. However, the parameter space is not orthogonal in the Riemannian manifold, which is usually encountered in practical problems. In this case, the ordinary gradient does not indicate the most efficient direction for a desired solution, thereby causing a slow convergence. As a much more efficient strategy, Amari et al. proposed the natural gradient, which can consider the relationship between the Riemannian manifold and the Euclidean manifold [1–3]. In addition, Cardoso and Laheld independently proposed the same, which they termed the relative gradient, and proved that the gradient has the 'equivariance property' [7].

The ordinary gradient algorithm has a slow convergence property in many practical problems and involves matrix inversion which is computationally intensive. On the other hand, the natural gradient algorithm is quite efficient and does not involve the matrix inversion. However, it still requires

*Corresponding author. Tel.: +82 42 869 5351; fax: +82 42 869 8490.
E-mail address: hmpark@kaist.ac.kr (H.-M. Park).

0925-2312/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2006.03.009

additional computation such as convolution for convolved mixtures and matrix multiplication for instantaneous mixtures. Moreover, the natural gradient algorithm has a serious problem in dealing with convolved mixtures. The exact form of the natural gradient algorithm for BSS of convolved mixtures involves non-causal terms and requires very intensive computation. To remove the non-causal terms and reduce the computational complexity, it is necessary to approximate the algorithm on the assumption that the updating amounts of filter coefficients are very small [3,8]. To fulfill the assumption, the step size should be very small, which results in slow convergence. In addition, the approximation may induce errors in updating adaptive filter coefficients.

In an attempt to obtain better performance than the ordinary gradient algorithm and overcome the disadvantages of the natural gradient algorithm, we present a new modification of the algorithms. In the modification, the algorithm provides a compromise between the ordinary gradient algorithm and the natural gradient algorithm. The algorithm maintains spatial and temporal independence, and requires less computation than the other algorithms. Simulation results demonstrate the efficiency of the proposed algorithm. For theoretical support, local stability on desired solutions of the algorithm is proven.

2. Conventional algorithms for BSS

The goal of BSS is to separate source signals from linear mixtures of unknown independent source signals [13,19,20]. Let us consider a set of unknown sources, $\mathbf{s}(n) = [s_1(n), s_2(n), \ldots, s_M(n)]^{\mathrm{T}}$, such that the components $\{s_i(n),\ i = 1, 2, \ldots, M\}$ are zero-mean and mutually independent. Assume that a set of observations, $\mathbf{x}(n) = [x_1(n), x_2(n), \ldots, x_M(n)]^{\mathrm{T}}$, is obtained as a linear combination of the unknown sources. Then, the observations $\mathbf{x}(n)$ can be expressed as

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n), \tag{2.1}$$

where $\mathbf{A}$ is an unknown full rank mixing matrix. The task is to find an unmixing matrix $\mathbf{W}(n)$ such that estimated signals $\mathbf{u}(n)$ are the original sources up to permutation and scaling, where

$$\mathbf{u}(n) = \mathbf{W}(n)\mathbf{x}(n). \tag{2.2}$$

Bell and Sejnowski proposed training the unmixing matrix $\mathbf{W}(n)$ by maximizing the entropy of $\mathbf{y} = g(\mathbf{u})$, where g is a nonlinear function approximating the cumulative density function (cdf) of the sources [4]. The ordinary gradient for maximizing the entropy leads to the following learning rule called as the infomax algorithm:

$$\Delta\mathbf{W}(n) \propto [\mathbf{W}^{\mathrm{T}}(n)]^{-1} - \varphi(\mathbf{u}(n))\mathbf{x}^{\mathrm{T}}(n),$$

$$\varphi(\mathbf{u}(n)) = \left[-\frac{\partial p_1(u_1(n))/\partial u_1(n)}{p_1(u_1(n))}, \ldots, -\frac{\partial p_M(u_M(n))/\partial u_M(n)}{p_M(u_M(n))}\right]^{\mathrm{T}}, \tag{2.3}$$

where $\varphi(\cdot)$ is called a score function and $p_i(u_i(n))$ denotes the probability density function (pdf) of $u_i(n)$.

A much more efficient way to learn the unmixing matrix is to follow the natural gradient [2,7,9]. For instantaneous mixtures, the natural gradient rescales the ordinary gradient by post-multiplying it with $\mathbf{W}^{\mathrm{T}}(n)\mathbf{W}(n)$, giving

$$\Delta\mathbf{W}(n) \propto [\mathbf{I} - \varphi(\mathbf{u}(n))\mathbf{u}^{\mathrm{T}}(n)]\mathbf{W}(n). \tag{2.4}$$

It is known that the natural gradient finds the most efficient direction for updating the unmixing matrix when the parameter space belongs to the Riemannian manifold. Moreover, the gradient has the equivariance property such that its convergence property is independent of the mixing characteristics [7]. Because the natural gradient algorithm does not involve computationally intensive matrix inversion, it requires less computation than the ordinary gradient algorithm.

3. A modified infomax algorithm

Let us consider a ‘modified’ infomax algorithm as follows:

$$\Delta\mathbf{W}(n) \propto \mathbf{I} - \varphi(\mathbf{u}(n))\mathbf{u}^{\mathrm{T}}(n). \tag{3.1}$$

Comparing Eq. (3.1) with Eqs. (2.3) and (2.4), we can easily see that the algorithm takes a compromise between the ordinary gradient algorithm and the natural gradient algorithm.

Here, the dynamic property of the algorithm is investigated with a cost function $J(\mathbf{W})$, which derives the conventional infomax algorithms

$$J(\mathbf{W}) = -\log|\det(\mathbf{W})| - \sum_{i=1}^{M} \log(p_i(u_i)). \tag{3.2}$$

In an attempt to check if the cost function is a Lyapunov function, which rigorously proves the convergence of the corresponding algorithm, we derive

$$\frac{\mathrm{d}J(\mathbf{W})}{\mathrm{d}n} = \sum_{i=1}^{M}\sum_{j=1}^{M} \frac{\partial J}{\partial w_{ij}} \frac{\mathrm{d}w_{ij}}{\mathrm{d}n}. \tag{3.3}$$

Since $\partial J/\partial\mathbf{W} = -\mathbf{W}^{-\mathrm{T}} + \varphi(\mathbf{u})\mathbf{x}^{\mathrm{T}}$, the modified infomax algorithm can be represented as

$$\frac{\mathrm{d}\mathbf{W}}{\mathrm{d}n} = \eta[\mathbf{I} - \varphi(\mathbf{u})\mathbf{u}^{\mathrm{T}}] = -\eta\frac{\partial J}{\partial\mathbf{W}}\mathbf{W}^{\mathrm{T}}, \tag{3.4}$$

where η is positive, and Eq. (3.3) is

$$\begin{aligned}\frac{\mathrm{d}J(\mathbf{W})}{\mathrm{d}n} &= -\eta\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{\partial J}{\partial w_{ij}}\sum_{k=1}^{M}\frac{\partial J}{\partial w_{ik}}w_{jk} \\ &= -\eta\sum_{i=1}^{M}\mathbf{q}_i^{\mathrm{T}}\mathbf{W}\mathbf{q}_i,\end{aligned} \tag{3.5}$$

where $\mathbf{q}_i$ denotes the ith column vector of $\partial J/\partial\mathbf{W}$. Therefore, when $\mathbf{W}$ is positive definite, $\mathrm{d}J(\mathbf{W})/\mathrm{d}n$ is not positive, which leads that the cost function $J(\mathbf{W})$ is a Lyapunov function of the modified infomax algorithm.