Available online at www.sciencedirect.com
ScienceDirect
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

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

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.

In this case, $dJ(\mathbf{W})/dn = 0$ is achieved if and only if $\mathbf{q}_i = 0$ for all i, which is equivalent to the condition that $d\mathbf{W}/dn = 0$.

Equilibrium points of the algorithm can be expressed as

$$E[\Delta\mathbf{W}(n)] \propto \mathbf{I} - E[\varphi(\mathbf{u}(n))\mathbf{u}^{\mathrm{T}}(n)] = \mathbf{0} \tag{3.6}$$

which provide independence for the estimated signals $\{u_i\}$. Note that the algorithm has the same equilibrium points as the ordinary gradient algorithm and the natural gradient algorithm. Moreover, the algorithm does not involve matrix inversion as well as matrix multiplication with $\mathbf{W}$. In particular, it may be useful for hardware implementation owing to its simple form.

From the algorithm, removing the score function $\varphi(\cdot)$ gives a second-order blind decorrelation learning rule [11]. However, it is worth noting that the modified infomax algorithm can obtain independent signals from mixtures using higher-order statistics instead of decorrelated signals. Furthermore, a more general learning rule has been proposed by applying a nonlinear function to $\mathbf{u}^{\mathrm{T}}$ [10]. In this paper, however, we do not use the nonlinear function in order to compare the algorithm with the conventional algorithms directly. Also, the modified infomax algorithm is extended to deal with convolved mixtures and the stability of the equilibrium points of the algorithm is analyzed as shown in the following sections.

4. Extension to convolved mixtures

Now, let us extend these algorithms to BSS of convolved mixtures. If the mixing of source signals involves convolution and time-delays, the observation vector $\mathbf{x}(n)$ can be expressed as

$$\mathbf{x}(n) = \sum_{k=0}^{K-1} \mathbf{A}_k \mathbf{s}(n-k), \tag{4.1}$$

where $\mathbf{A}_k$ denotes a matrix composed of mixing filter coefficients. Let us consider a feedforward network to separate signals from the convolved mixtures as

$$\mathbf{u}(n) = \sum_{k=0}^{K-1} \mathbf{W}_k(n)\mathbf{x}(n-k), \tag{4.2}$$

where adaptive filter matrices $\{\mathbf{W}_k(n), k = 0, 1, \ldots, K-1\}$ make an output vector $\mathbf{u}(n)$ reproduce the source vector $\mathbf{s}(n)$.

Torkkola derived the ordinary gradient algorithm of entropy maximization for convolved mixtures as [29]

$$\Delta\mathbf{W}_k(n) \propto (\mathbf{W}_0^{\mathrm{T}}(n))^{-1}\delta_k - \varphi(\mathbf{u}(n))\mathbf{x}^{\mathrm{T}}(n-k). \tag{4.3}$$

On the other hand, the natural gradient algorithm is given by [3,8]

$$\Delta\mathbf{W}_k(n) \propto \mathbf{W}_k(n) - \varphi(\mathbf{u}(n))\mathbf{r}_k^{\mathrm{T}}(n), \tag{4.4}$$

where $\mathbf{r}_k(n) = \sum_{l=0}^{K-1} \mathbf{W}_l^{\mathrm{T}}(n)\mathbf{u}(n-k+l)$. Unfortunately, the natural gradient algorithm shows that the update of $\mathbf{W}_k(n)$ depends on future outputs $\mathbf{u}(n-k+l)$, $k-l<0$, through $\mathbf{r}_k(n)$. In addition, it involves very intensive computation to compute all $\mathbf{r}_k(n)$, $k = 0, \ldots, K-1$, at each time step. Practically, the algorithm is approximated by introducing a $K-1$ sample delay to remove the non-causal terms and by reusing past results assuming that $\mathbf{W}_k(n) \approx \mathbf{W}_k(n-1) \approx \cdots \approx \mathbf{W}_k(n-2K+2)$ and $\mathbf{r}_k(n) \approx \mathbf{r}_0(n-k)$. With this approximation, the algorithm becomes

$$\Delta\mathbf{W}_k(n) \propto \mathbf{W}_k(n) - \varphi(\mathbf{u}(n-K+1))\mathbf{r}^{\mathrm{T}}(n-k), \tag{4.5}$$

where $\mathbf{r}(n) = \sum_{l=0}^{K-1} \mathbf{W}_{K-1-l}^{\mathrm{T}}(n)\mathbf{u}(n-l)$. With the assumption $\mathbf{W}_k(n) \approx \cdots \approx \mathbf{W}_k(n-2K+2)$, the algorithm has to use a very small step size, especially for a large number of adaptive filter coefficients, in order to converge on a proper solution. Therefore, this may result in performance degradation by the approximation error and slow convergence by the small-step-size restriction.

To deal with convolved mixtures, extending the modified infomax algorithm in Eq. (3.1) gives

$$\Delta\mathbf{W}_k(n) \propto \mathbf{I}\delta_k - \varphi(\mathbf{u}(n))\mathbf{u}^{\mathrm{T}}(n-k). \tag{4.6}$$

Since equilibrium points of the algorithm satisfy

$$E[\Delta\mathbf{W}_k(n)] \propto \mathbf{I}\delta_k - E[\varphi(\mathbf{u}(n))\mathbf{u}^{\mathrm{T}}(n-k)] = \mathbf{0}, \tag{4.7}$$

the algorithm has the same equilibrium points as the ordinary gradient algorithm and the natural gradient algorithm. Also, spatial and temporal independence for the estimated signals is achieved at these points. In addition, contrary to the natural gradient algorithm, it is not necessary to use a very small step size or to compute the additional convolution $\mathbf{r}(n)$. In the natural gradient algorithm, M^2K multiplications are approximately required to compute $\mathbf{r}(n)$ per time instant for an $(M \times M)$-dimensional matrix $\mathbf{W}_k(n)$ [3]. Furthermore, it does not involve matrix inversion such as the zero-delay weight update of the ordinary gradient algorithm.

The above algorithms have indeterminacy of the estimated signals up to permutation and arbitrary filtering. Entropy maximization attempts to make the outputs temporally whitened, which may degrade outputs in many applications such as BSS of natural signals. Whitening the estimated outputs can be avoided by forcing direct filters, $W_{ii}(z)$, to scaling factors [29]. Here, $W_{ii}(z)$ are the filters composed of diagonal elements of adaptive filter matrices $\mathbf{W}_k(n)$.

5. Stability analysis

To check the stability of the equilibrium points, we rewrite Eq. (4.6) with the step size μ as

$$\mathbf{W}_k(n+1) = \mathbf{W}_k(n) + \mu\{\mathbf{I}\delta_k - \varphi(\mathbf{u}(n))\mathbf{u}^{\mathrm{T}}(n-k)\}. \tag{5.1}$$

Without loss of generality, we can assume that $\mathbf{u}^{\mathbf{o}}(n) = \mathbf{s}(n)$ because of scale indeterminacy. Here, $\mathbf{u}^{\mathbf{o}}(n)$ denotes $\mathbf{u}(n)$ in a separating equilibrium point, in which $E[\varphi(\mathbf{s}(n))\mathbf{s}^{\mathrm{T}}(n-k)] = \mathbf{I}\delta_k$. If the unmixing matrix is perturbed from the equilibrium point, $\mathbf{W}_k(n) = \mathbf{W}_k^{\mathbf{o}} + \widetilde{\mathbf{W}}_k(n)$ and $\mathbf{u}(n) = \mathbf{s}(n) + \widetilde{\mathbf{u}}(n)$, where $\widetilde{\mathbf{W}}_k(n)$ is the error matrix. In addition,

$\varphi(\mathbf{u}(n))$ can be approximated to the first order as

$$\varphi(\mathbf{u}(n)) \approx \varphi(\mathbf{s}(n)) + \mathbf{D}_{\varphi'}(n)\tilde{\mathbf{u}}(n), \tag{5.2}$$

where the score function $\varphi(\mathbf{s}(n))$ can be represented as $\left[-\frac{\partial p_1(s_1(n))/\partial s_1(n)}{p_1(s_1(n))}, \ldots, -\frac{\partial p_M(s_M(n))/\partial s_M(n)}{p_M(s_M(n))}\right]^{\mathrm{T}}$, and $\mathbf{D}_{\varphi'}(n)$ is a diagonal matrix composed of the derivatives of $\varphi_i(s_i(n))$.

Therefore, the error matrix $\widetilde{\mathbf{W}}_k(n)$ can be approximated as

$$\begin{aligned}\widetilde{\mathbf{W}}_k(n+1) \approx \widetilde{\mathbf{W}}_k(n) + \mu\{&\mathbf{I}\delta_k - \varphi(\mathbf{s}(n))\mathbf{s}^{\mathrm{T}}(n-k) \\ &- \mathbf{D}_{\varphi'}(n)\tilde{\mathbf{u}}(n)\mathbf{s}^{\mathrm{T}}(n-k) \\ &- \varphi(\mathbf{s}(n))\tilde{\mathbf{u}}^{\mathrm{T}}(n-k)\}.\end{aligned} \tag{5.3}$$

Since $\tilde{\mathbf{u}}(n) = \sum_{k=0}^{K-1} \widetilde{\mathbf{W}}_k(n) \sum_{l=0}^{K-1} \mathbf{A}_l \mathbf{s}(n-k-l)$ and $\widetilde{\mathbf{W}}_k(n) = \mathbf{0}, \mathbf{A}_k = \mathbf{0}, \forall k < 0$, the expectation of the zero-delay error matrix is

$$\begin{aligned}E[\widetilde{\mathbf{W}}_0(n+1)] \approx E[\widetilde{\mathbf{W}}_0(n)] &- \mu\{E[\mathbf{D}_{\varphi'}(n)\widetilde{\mathbf{W}}_0(n)\mathbf{A}_0\mathbf{s}(n)\mathbf{s}^{\mathrm{T}}(n)] \\ &+ E[\varphi(\mathbf{s}(n))\mathbf{s}^{\mathrm{T}}(n)\mathbf{A}_0^{\mathrm{T}}\widetilde{\mathbf{W}}_0^{\mathrm{T}}(n)]\},\end{aligned} \tag{5.4}$$

and the expectation of a delay error matrix is

$$\begin{aligned}&E[\widetilde{\mathbf{W}}_k(n+1)] \\ &\approx E[\widetilde{\mathbf{W}}_k(n)] - \mu E\left[\mathbf{D}_{\varphi'}(n)\sum_{l=0}^{k}\widetilde{\mathbf{W}}_{k-l}(n)\mathbf{A}_l\mathbf{s}(n-k)\right. \\ &\left.\times\mathbf{s}^{\mathrm{T}}(n-k)\right] \quad \forall k \neq 0.\end{aligned} \tag{5.5}$$

In Eqs. (5.4) and (5.5), we assume that the source signals $\{s_i, i = 1, 2, \ldots, M\}$ are i.i.d., which gives

$$E[\mathbf{s}(n)\mathbf{s}(n-k)] = \mathbf{0} \quad \forall k \neq 0, \tag{5.6}$$

$$E[\varphi(\mathbf{s}(n))\mathbf{s}(n-k)] = \mathbf{0} \quad \forall k \neq 0. \tag{5.7}$$

Assuming that only a diagonal element of $\mathbf{W}_0(n)$, $w_{ii0}(n)$, is perturbed and $E[\widetilde{w}_{ii0}(n)]$ is uncorrelated with $\mathbf{s}(n)$ through the independence assumption [12], Eq. (5.4) can be written as

$$\begin{aligned}E[\widetilde{w}_{ii0}(n+1)] \approx E[\widetilde{w}_{ii0}(n)] &- \mu\{E[\varphi_i'(s_i(n))a_{ii0}s_i^2(n)] \\ &+ E[\varphi_i(s_i(n))s_i(n)a_{ii0}]\}E[\widetilde{w}_{ii0}(n)].\end{aligned} \tag{5.8}$$

Here, a_{ii0} denotes the ith diagonal element of the mixing matrix $\mathbf{A}_0$. In case that only a k-delay diagonal element, $w_{iik}(n)$, is perturbed, Eq. (5.5) is

$$\begin{aligned}E[\widetilde{w}_{iik}(n+1)] \approx E[\widetilde{w}_{iik}(n)] &- \mu E[\varphi_i'(s_i(n))a_{ii0}s_i^2(n-k)] \\ &\times E[\widetilde{w}_{iik}(n)] \quad \forall k \neq 0.\end{aligned} \tag{5.9}$$

In order to converge the diagonal element $\widetilde{w}_{ii0}(n)$ and $\widetilde{w}_{iik}(n)$ to 0 with an appropriate step size μ, $E[\varphi_i'(s_i(n))a_{ii0}s_i^2(n)] + E[\varphi_i(s_i(n))s_i(n)a_{ii0}]$ and $E[\varphi_i'(s_i(n))a_{ii0}s_i^2(n-k)]$ should be positive, respectively. Note that it can be achieved by $a_{ii0} > 0$ on the assumption that the score function $\varphi_i(s_i)$ is odd and increases monotonically.

Repeating the derivation for off-diagonal elements, perturbation of a zero-delay element $w_{ij0}(n)$ and a k-delay element $w_{ijk}(n)$ can be described as

$$\begin{aligned}E[\widetilde{w}_{ij0}(n+1)] \approx E[\widetilde{w}_{ij0}(n)] &- \mu E[\varphi_i'(s_i(n))a_{jj0}s_j^2(n)] \\ &\times E[\widetilde{w}_{ij0}(n)] \quad \forall i \neq j,\end{aligned} \tag{5.10}$$

$$\begin{aligned}E[\widetilde{w}_{ijk}(n+1)] \approx E[\widetilde{w}_{ijk}(n)] &- \mu E[\varphi_i'(s_i(n))a_{jj0}s_j^2(n-k)] \\ &\times E[\widetilde{w}_{ijk}(n)] \quad \forall i \neq j,\end{aligned} \tag{5.11}$$

respectively. In this case, the perturbation of the off-diagonal elements can be removed by $a_{jj0} > 0$.

Therefore, the common constraint is that zero-delay diagonal elements of the mixing filter coefficients should be positive. However, the mixing filters are convolved with the source signal $\mathbf{s}(n)$ to provide the observations $\mathbf{x}(n)$. By changing the signs of the source signals if needed, we can always generate the same observations with a mixing system that consists of positive zero-delay diagonal elements because of scale indeterminacy. Therefore, this constraint does not limit usefulness of the proposed algorithm in real-world applications.

Local stability for instantaneous mixtures can be analyzed in the same way. For two instantaneous mixtures mixed from two Laplace-distributed source signals, we can prove local stability in a more straightforward way, as shown in the Appendix.

6. Experimental results

6.1. Simulations on instantaneous mixtures

To compare the modified infomax algorithm with others, two Laplace-distributed source signals have been generated and mixed with randomly generated 2×2 mixing matrices. Here, the 2000 mixing matrices are sorted according to the condition number which indicates closeness to a singular matrix by the ratio of the largest singular value to the smallest, and the results are shown in Fig. 1(a). Each source signal consisted of 160,000 i.i.d. samples. Since the sources were Laplace-distributed signals, sgn(·) was used as the score function $\varphi(\cdot)$. As a performance measure of the algorithms, we used the performance index PI, which was defined by [2]

$$PI = \sum_{i=1}^{M}\left(\sum_{j=1}^{M}\frac{|t_{ij}|}{\max_k|t_{ik}|} - 1\right) + \sum_{j=1}^{M}\left(\sum_{i=1}^{M}\frac{|t_{ij}|}{\max_k|t_{kj}|} - 1\right), \tag{6.1}$$

where t_{ij} is the (i,j)th element of the overall matrix $\mathbf{T} = \mathbf{WA}$. Figs. 1(b)–(e) show the performance indices of several algorithms. To compare the performance under various mixing conditions for a fixed number of data samples, we display the performance indices after one sweep training of the unmixing matrix. Successful separation rates with the criterion "$PI < 0.1$" are also displayed in

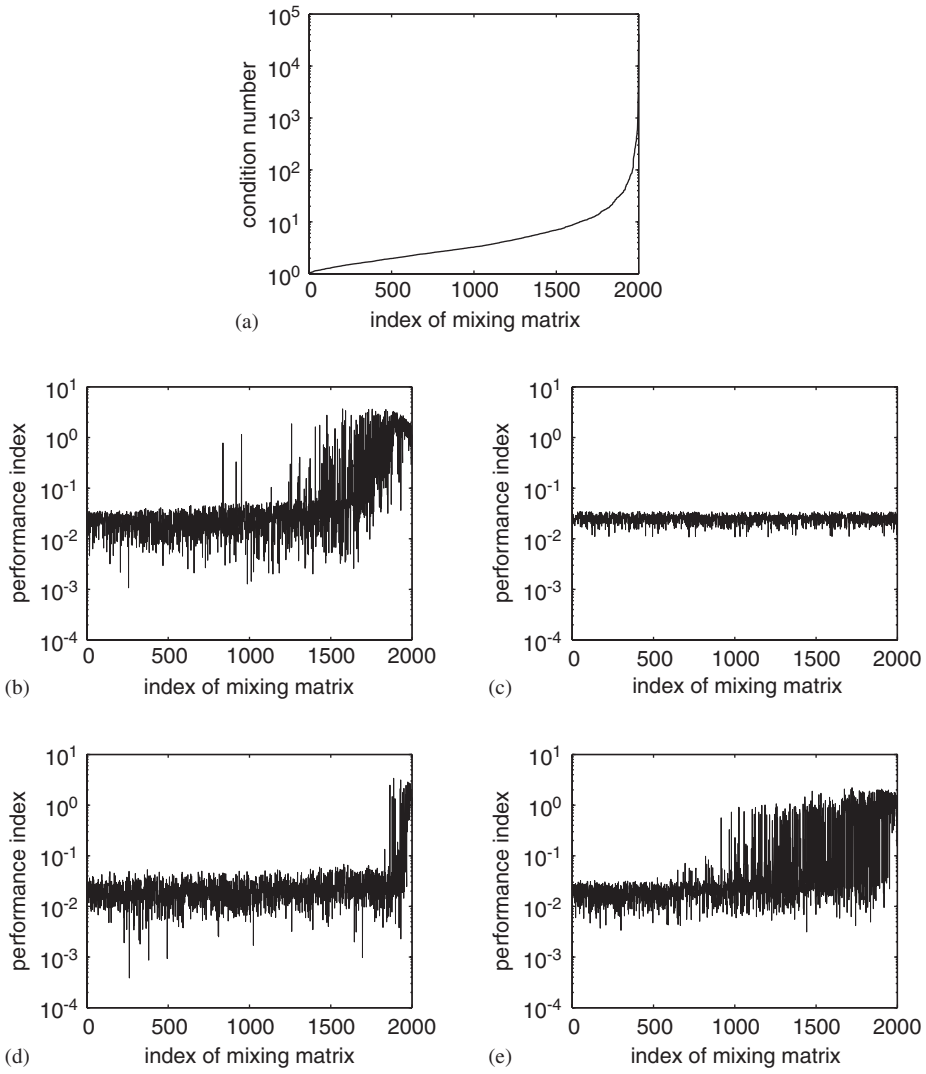


Fig. 1. Performance indices of the BSS algorithms for instantaneous mixtures from two Laplace-distributed source signals: (a) codition number of mixing matrix, (b) ordinary gradient algorithm, (c) natural gradient algorithm, (d) modified infomax algorithm, (e) $\Delta w_{ij} \propto -\mathrm{sgn}(u_i)u_j, i \neq j$.

Table 1
Successful separation rates for instantaneous mixtures (criterion: performance index <0.1)

Ordinary gradient	Natural gradient	Modified infomax	$\Delta w_{ij} \propto -\mathrm{sgn}(u_i)u_j, i \neq j$
81.9%	100.0%	96.5%	80.5%

Table 1. We have chosen the identity matrix for the initial unmixing matrix.

As shown in Fig. 1(c), the natural gradient algorithm provided satisfactory performance for all simulated mixing matrices. This supports the equivariance property of the natural gradient [7]. Comparing Figs. 1(b) and (d), we find that the modified infomax algorithm failed to separate signals within one sweep of training for some mixing matrices with very large condition numbers, but it could separate a much larger number of mixing signals than the ordinary gradient algorithm. The poor performance mostly arose from very ill-conditioned mixing matrices. Fig. 1(e) shows the performance indices for the algorithm proposed by Ling et al. [22]. The algorithm is similar to the Hérault–Jutten algorithm [17] and corresponds to the modified infomax algorithm with fixed diagonal elements of the unmixing matrix $\mathbf{W}$. Although Ling et al. did not use $\mathrm{sgn}(\cdot)$ as the score function, we used it for its efficiency. However, the algorithm failed to separate signals much more frequently than the modified infomax algorithm. Note that the modified infomax algorithm provides a simpler formulation than the other algorithms and successfully separated signals within 160,000 samples except for some severely ill-conditioned mixing matrices.

With the modified infomax algorithm, Fig. 2 displays the performance indices and the derivatives of the cost function computed by Eq. (3.5) with $\eta = 1$ for two mixing matrices. The condition numbers of the mixing matrices were 4.796 and 19.437. For the mixing 1 which gave the relatively small condition number, the modified infomax

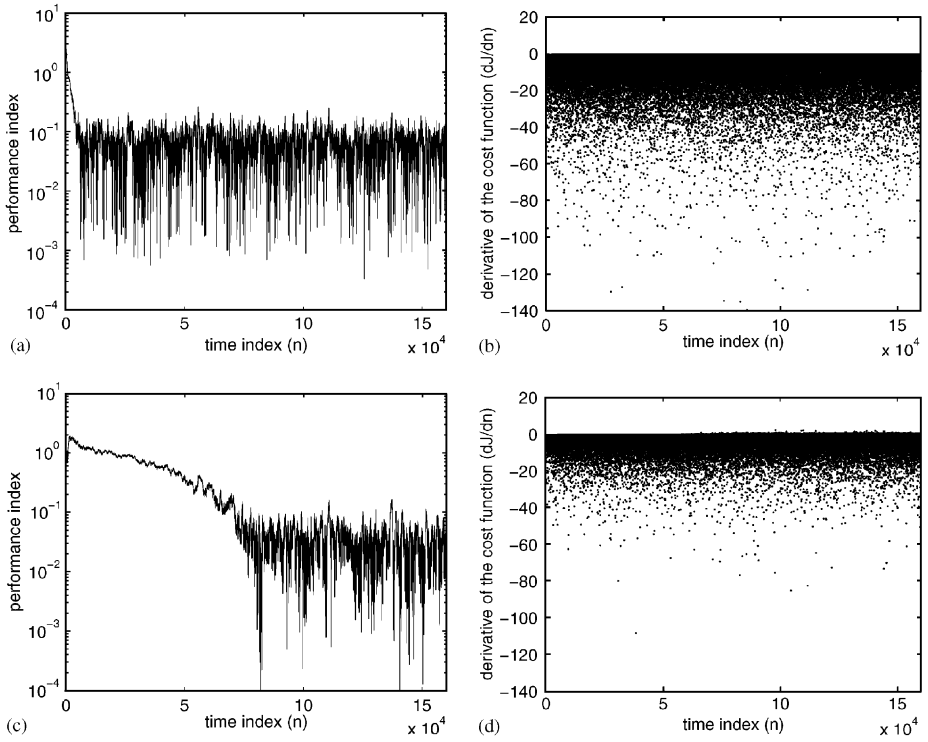


Fig. 2. Performance indices and derivatives of the cost function for two mixing matrices: (a) performance index for mixing 1, (b) derivative of the cost function for mixing 1, (c) performance index for mixing 2, (d) derivative of the cost function for mixing 2.

algorithm converged after a short time period, and the derivative of the cost function did not have a positive value during the learning period. However, for the mixing 2, the algorithm did not provide positive derivative values during some time period, but it did at some time indices after that period. Although the modified infomax algorithm did not guarantee positive definiteness of the unmixing matrix, the negative derivative was kept for the mixing 1. This means that the algorithm found out a solution of a positive definite unmixing matrix from initialization with the identity matrix with resort to the scaling and permutation indeterminacy. Moreover, note that the algorithm successfully obtained a solution for the mixing 2 even though some derivative values were positive. Recall that positive definiteness of the unmixing matrix is a rigorous condition for the convergence, but failing to meet the condition does not necessarily mean that the algorithm cannot find out a solution.

Fig. 3 shows the first time indices when the positive derivative of the cost function occurred for the 2000 mixing matrices. The probability that the positive derivative occurred during the learning period is roughly proportional to the condition number of the mixing matrix. A mixing matrix with a high condition number is relatively close to a singular matrix, so the corresponding unmixing matrix is usually far from the identity matrix for initialization. Therefore, the unmixing matrix cannot keep positive definiteness more probably for an ill-conditioned mixing matrix. The rate not to have the positive derivative during the learning period for the 2000 mixing matrices was 91.8%

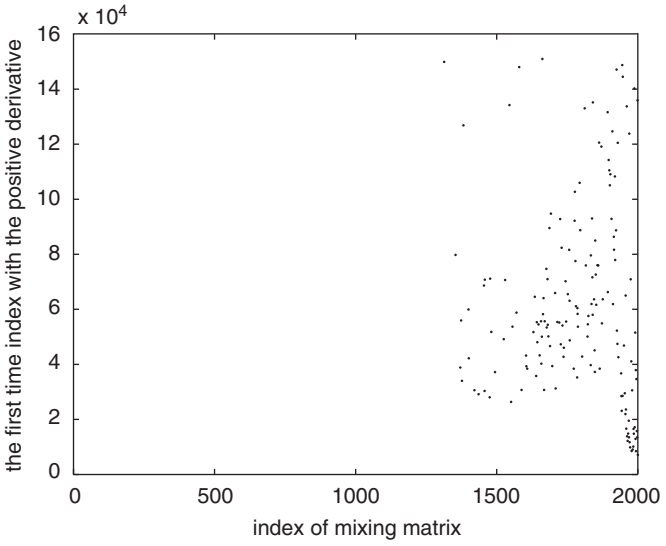


Fig. 3. The first time indices with the positive derivative of the cost function.

which was a great portion in view of the fact that the algorithm did not ensure positive definiteness of the unmixing matrix. Furthermore, recalling that the separation rate was 96.5%, the modified infomax algorithm successfully found out a solution for a large part of the cases where the unmixing matrices failed to keep positive definiteness.

In order to analyze the effect of the score function assumption on the performance of the modified infomax algorithm, the performance indices are displayed in Fig. 4

when tanh(·) was used as the score function for the same source signals and mixing matrix as in Fig. 2(a). In addition, the stability of the equilibrium points needs a positive slope of the score function as shown in the previous section, but $\mathrm{sgn}(\cdot) = \lim_{q\to\infty} \tanh(q\cdot)$ does not fulfill this condition. Since tanh(·) is monotonically

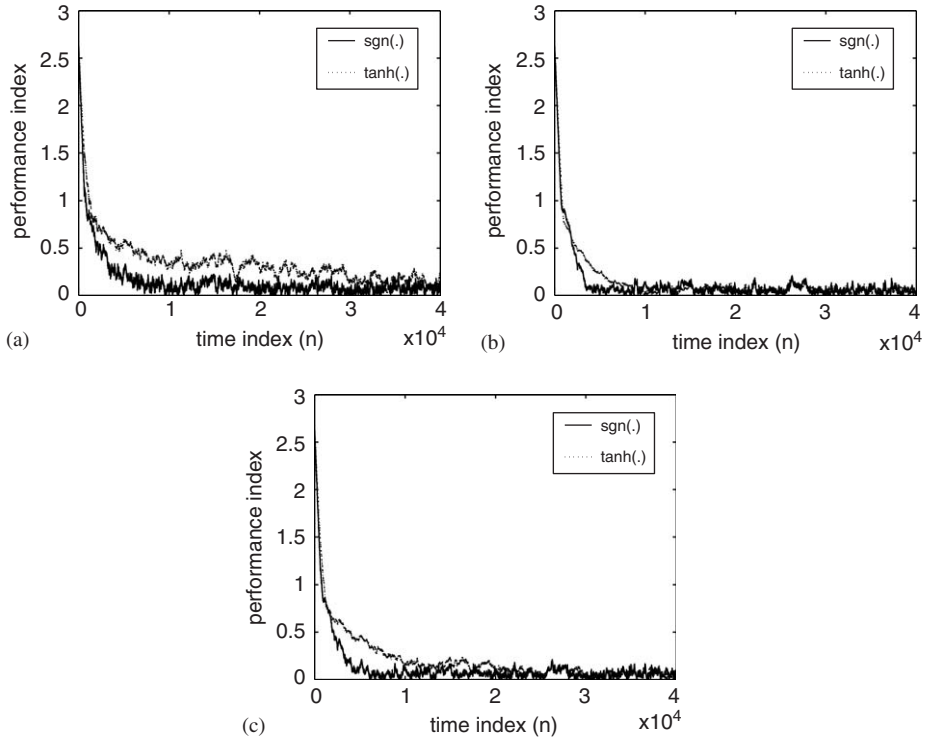


Fig. 4. Performance indices of the BSS algorithms for the score function tanh(·) and two Laplace-distributed source signals: (a) ordinary gradient algorithm, (b) natural gradient algorithm, (c) modified infomax algorithm.

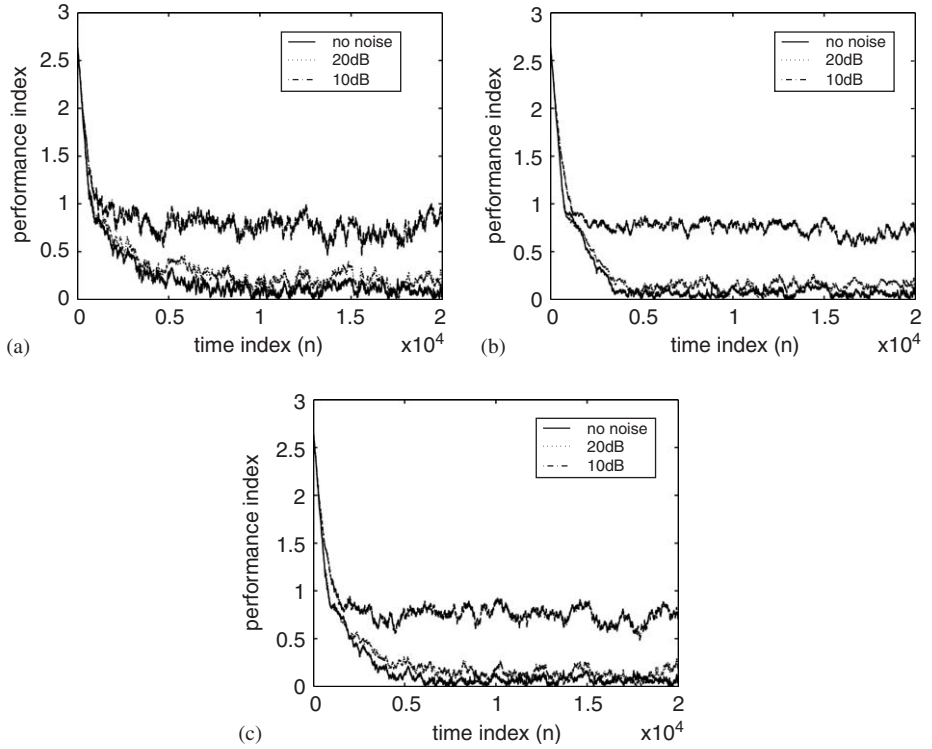


Fig. 5. Performance indices of the BSS algorithms for mixtures corrupted by white Gaussian noise: (a) ordinary gradient algorithm, (b) natural gradient algorithm, (c) modified infomax algorithm.

increasing and differentiable everywhere, it also can be used for estimating this influence. From the standpoint of comparing convergence curves, any significant difference among the algorithms could not be found, although the error on the score function reduced convergence speeds. Also, the successful separation rate within 160,000 samples for the 2000 mixing matrices was 91.9%. Considering slow convergence caused by the error on the score function, it can be reasoned that this convergence property was not seriously affected by mixing condition as well. Additionally, sgn(·) as the score function provided faster convergence than tanh(·) even though the former did not meet the condition that the score function should have a positive slope. This result might support that this condition is not indispensable.

When the mixtures were corrupted by white Gaussian noise, performance indices are also shown in Fig. 5. For the same signal-to-noise ratio, all the algorithms gave the almost same performance indices, and the noise effect for the modified infomax algorithm was not notably different from that for the other algorithms.

As a final experiment for instantaneous mixtures, three independent source signals generated from different distributions have been considered. The distributions were uniform, Laplace, and the function obtained by summing two Gaussian distributions. To deal with the variously distributed signals, the score function was obtained in the same way as in the extended infomax algorithm [21]. Because the number of parameters was increased and the extended algorithm required additional parameter estimation to obtain appropriate score functions, the adaptation has been repeated 10 times for 160,000 i.i.d. samples. Fig. 6 shows the performance indices for 100 randomly generated 3×3 mixing matrices. This result demonstrated that the modified infomax algorithm could successfully separate various kinds of sources within 10 sweeps except for some ill-conditioned mixing cases.

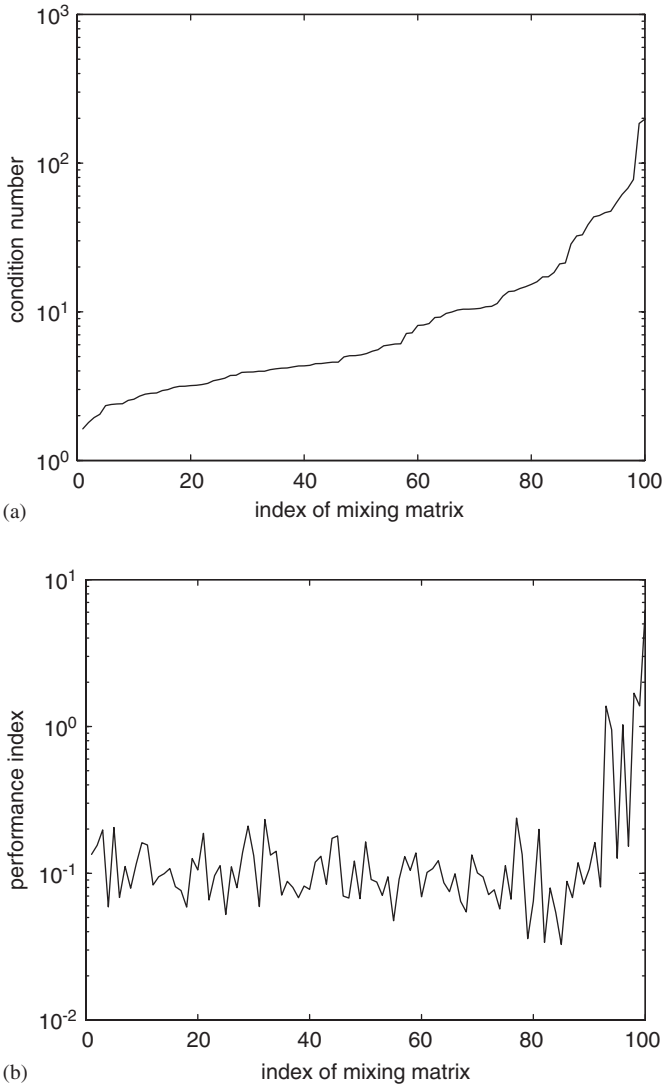


Fig. 6. Performance indices for instantaneous mixtures from three differently distributed source signals: (a) condition number of mixing matrix, (b) performance indices.

6.2. Simulations on convolved mixtures

To perform experiments for convolved mixtures, we have mixed real-recorded speech signals. Each speech signal had 10 second length at 16 kHz sampling rate. It is known that speech signal approximately follows Laplacian distribution. Therefore, sgn(·) was used as the score function $\varphi(\cdot)$. Experimental results were compared in terms of interference reduction ratio (IRR), which was defined as the difference between the signal-to-interference ratios (SIRs) of the final and the initial unmixing systems. The SIR is a ratio of the signal power to the interference power at the outputs given by

$$\mathrm{SIR}(\mathrm{d}B) = \frac{1}{2} \cdot \left| 10 \log \left(\frac{\langle (u_{1,s_1}(n))^2 \rangle}{\langle (u_{1,s_2}(n))^2 \rangle} \cdot \frac{\langle (u_{2,s_2}(n))^2 \rangle}{\langle (u_{2,s_1}(n))^2 \rangle} \right) \right| \qquad (6.2)$$

for 2×2 mixing/unmixing system [27]. In Eq. (6.2), $u_{j,s_i}(n)$ denotes the jth output of the cascaded mixing/unmixing system when only $s_i(n)$ is active.

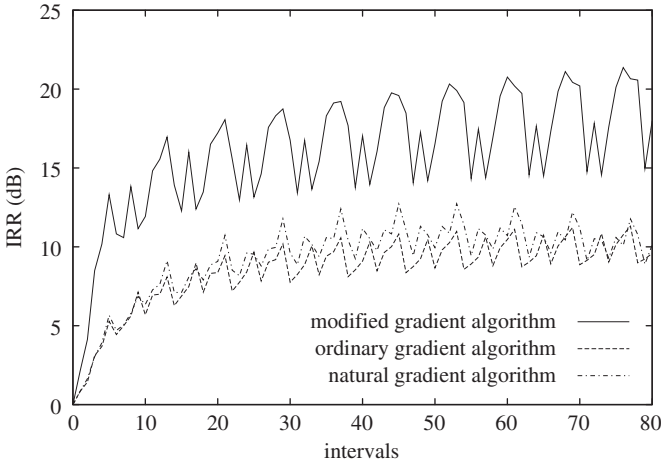


Fig. 7. IRRs of the BSS algorithms for mixtures generated from a simple mixing environment.

We first show results for a simple mixing environment, which is [29]

$$A_{11}(z) = 1 - 0.4z^{-25} + 0.2z^{-45},$$
$$A_{12}(z) = 0.4z^{-20} - 0.2z^{-28} + 0.1z^{-36},$$
$$A_{21}(z) = 0.5z^{-10} + 0.3z^{-22} + 0.1z^{-34},$$
$$A_{22}(z) = 1 - 0.3z^{-20} + 0.2z^{-38}. \quad (6.3)$$

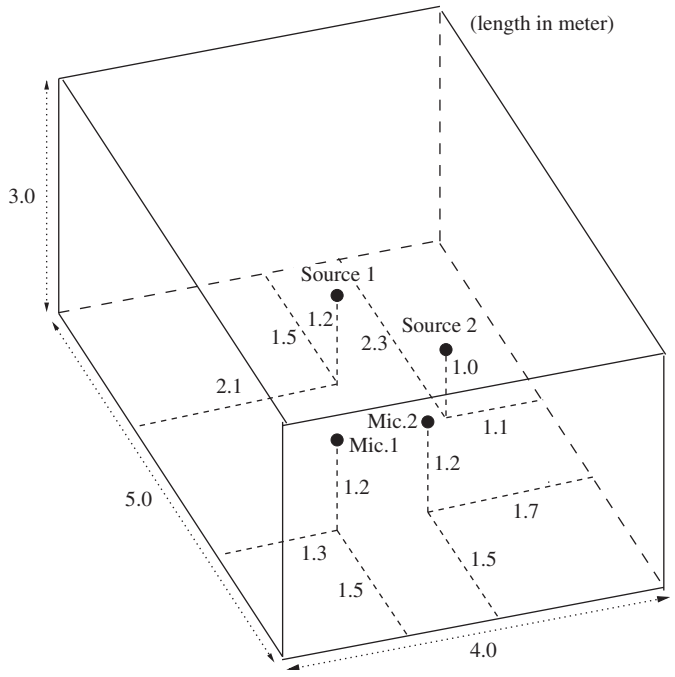


Fig. 8. Virtual room to simulate impulse responses from two speaker points to two microphone points.

As an unmixing system, we used a 2×2 filter system whose initialization was an identity, and the length of each filter was set to 100 taps for the mixing environment.

Fig. 7 displays the IRRs of the three algorithms for the mixing environment. Each signal was divided into eight intervals, each of which consisted of 20,000 samples. We repeated the adaptation with the same mixtures 10 times. Therefore, there were 80 intervals on the horizontal axis of the figure. The speech signal had many silent sections, and the SIR could not be improved in these sections. Therefore, the intervals had different IRRs depending on whether the intervals included the silent sections. The modified infomax algorithm showed better IRRs than the other algorithms. The natural gradient algorithm did not provide good performance because it had to use a very small step size in order to converge on a desired solution stably and it accumulated errors of the adaptive unmixing filter coefficients to $\mathbf{r}(n)$. In accordance with the case of instantaneous mixtures, the ordinary gradient algorithm displayed inferior performance relative to the modified infomax algorithm.

To obtain mixtures from a more complex environment, we constructed 10 different 2×2 mixing systems with the commercial software 'Room Impulse Response v2.5' [14], which calculates impulse responses using a time-domain image expansion method in a rectangular enclosure. We used a virtual room with the dimensions 4 m × 5 m × 3 m, and the reverberation time was 160 ms. Fig. 8 shows one of the mixing systems to simulate the impulse responses from two speaker points to two microphone points, and Fig. 9 shows the resulting impulse responses. To avoid excessive computation, each unmixing filter consisted of 1024 taps. In this experiment, we repeated the adaptation with the

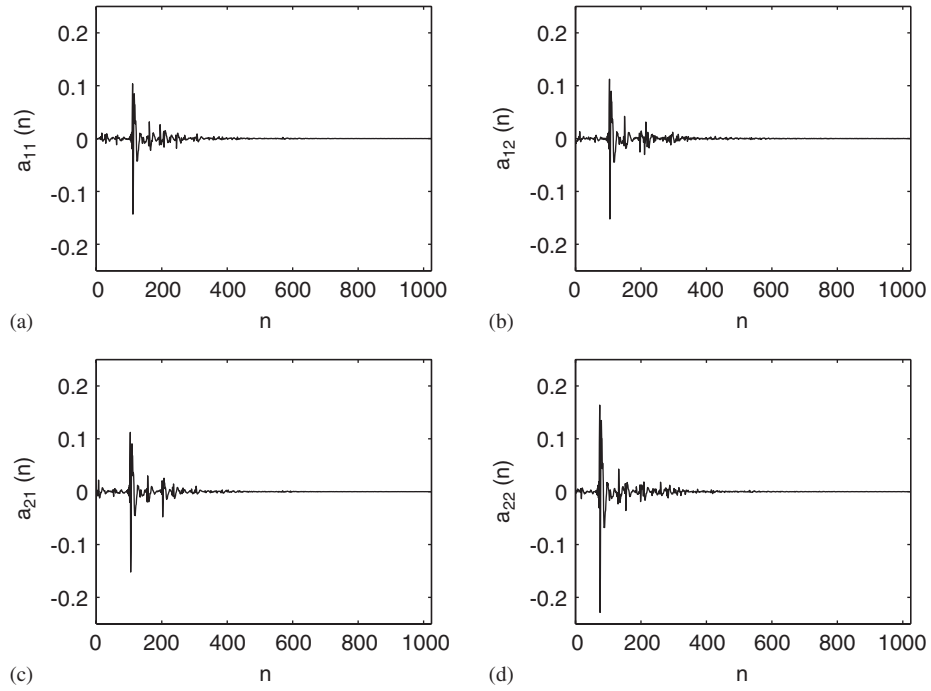
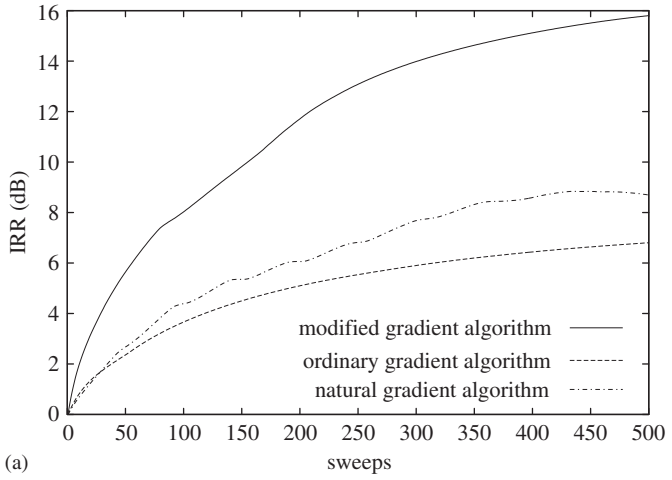


Fig. 9. Simulated room impulse responses for the virtual room in Fig. 8: (a) a_{11}, (b) a_{12}, (c) a_{21}, (d) a_{22}.



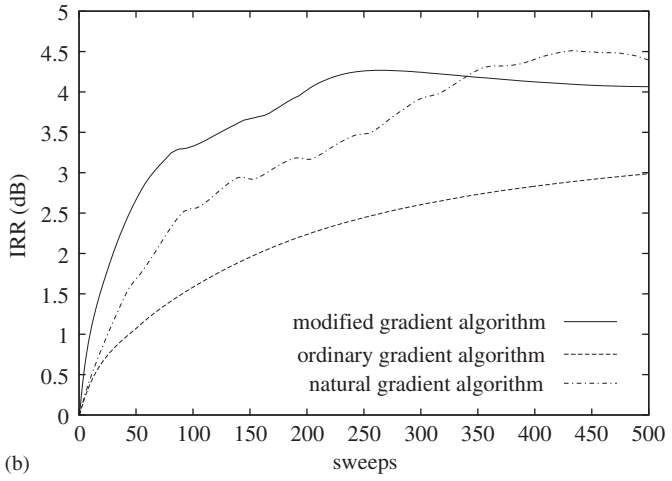


Fig. 10. IRRs of the BSS algorithms for mixtures generated from simulated room impulse responses.

mixed signals 500 times to train the unmixing system, because the mixing system was much more complex than the previous mixing environment. Other settings were the same as those of the previous experiments. Fig. 10 shows the mean values of the IRRs for the 10 mixing systems. We also display the standard deviation values of the IRRs to show the sensitivity for specific mixing systems. The three algorithms did not show particularly outstanding values in terms of the standard deviation. Because the mixing system was much more complex, the overall convergence was slower than in the previous experiment. In particular, the natural gradient algorithm showed quite slower convergence because of the harder small-step-size restriction with larger adaptive filter lengths. Above all, it should be noted that the modified infomax algorithm showed superior IRRs with less computational complexity relative to the other algorithms.

7. Conclusion

In this paper, a modified infomax algorithm to perform BSS was presented. The algorithm is a compromise between the ordinary gradient algorithm and the natural gradient algorithm. We showed the proof of local stability of the desired equilibrium points for BSS by using monotonically increasing and odd score functions. The algorithm gave a simpler formulation than the other algorithms, and it overcame the disadvantages of the natural gradient algorithm such as the approximation error and the restriction to a small step size for convolved mixtures. Simulations on convolved mixtures showed that the modified infomax algorithm provided better performance than the other algorithms. For instantaneous mixtures, furthermore, our algorithm attained good performance during learning period except for some very ill-conditioned mixings.

Acknowledgments

This work was supported by the Brain Neuroinformatics Research Program sponsored by the Korean Ministry of Science and Technology.

Appendix A. Proof of local stability for two instantaneous mixtures from two Laplace-distributed source signals

For BSS of instantaneous mixtures, equilibrium points of the modified infomax algorithm satisfy

$$E[\Delta \mathbf{W}(n)] \propto \mathbf{I} - E[\varphi(\mathbf{u}(n))\mathbf{u}^{\mathrm{T}}(n)] = \mathbf{0}. \tag{A.1}$$

Assuming Laplace-distributed source signals, there are 16 equilibrium points for two estimated independent signals and two observations. In this case, the score function $\varphi(\cdot)$ is given by $\mathrm{sgn}(\cdot)$. Omitting the time index n, the equilibrium points are

$$t_{11} = \pm l_1, \quad t_{12} = 0, \quad t_{21} = 0, \quad t_{22} = \pm l_2, \tag{A.2}$$

$$t_{11} = 0, \quad t_{12} = \pm l_2, \quad t_{21} = \pm l_1, \quad t_{22} = 0, \tag{A.3}$$

$$\begin{aligned} &t_{11} = c_{11}\tfrac{2}{3}l_1, \quad t_{12} = c_{12}\tfrac{2}{3}l_2, \\ &t_{21} = c_{21}\tfrac{2}{3}l_1, \quad t_{22} = c_{22}\tfrac{2}{3}l_2, \\ &c_{ij} = \pm 1, \ \prod_{i=1}^{2}\prod_{j=1}^{2} c_{ij} = -1, \end{aligned} \tag{A.4}$$

where t_{ij} is the (i,j)th element of the overall matrix $\mathbf{T} = \mathbf{WA}$, and the source pdfs are $p_1(s_1) = (l_1/2)\mathrm{e}^{-l_1|s_1|}$ and $p_2(s_2) = (l_2/2)\mathrm{e}^{-l_2|s_2|}$. Eqs. (A.2) and (A.3) are proper separating states whereas Eq. (A.4) is not.

Let us examine local stability on these equilibrium points. It is known that an equilibrium point is locally stable if the eigenvalues of $\mathbf{J}$ have negative real parts

[23,24,28], where

$$\mathbf{J}=\begin{bmatrix} \frac{\partial}{\partial w_{11}}E[1-\varphi(u_1)u_1] & \frac{\partial}{\partial w_{12}}E[1-\varphi(u_1)u_1] & \cdots & \frac{\partial}{\partial w_{22}}E[1-\varphi(u_1)u_1] \\ \frac{\partial}{\partial w_{11}}E[-\varphi(u_1)u_2] & \frac{\partial}{\partial w_{12}}E[-\varphi(u_1)u_2] & \cdots & \frac{\partial}{\partial w_{22}}E[-\varphi(u_1)u_2] \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial w_{11}}E[1-\varphi(u_2)u_2] & \frac{\partial}{\partial w_{12}}E[1-\varphi(u_2)u_2] & \cdots & \frac{\partial}{\partial w_{22}}E[1-\varphi(u_2)u_2] \end{bmatrix}. \quad \text{(A.5)}$$

Here, w_{ij} is the (i,j)th element of the mixing matrix $\mathbf{W}$. Without finding the roots of $\det(\lambda\mathbf{I}-\mathbf{J})=0$, we can check with the Routh–Hurwitz criterion whether the eigenvalues of $\mathbf{J}$ have negative real parts [18,26].

Conditions for local stability of the separating equilibrium points are as follows:

$$a_{11}a_{22}\neq 0,\quad \frac{a_{12}a_{21}}{a_{11}a_{22}}<1,\quad w_{11}>0,\quad w_{22}>0$$
$$\text{for the points } t_{11}=\pm l_1,\ t_{12}=0,$$
$$t_{21}=0,\ t_{22}=\pm l_2, \quad \text{(A.6)}$$

$$a_{12}a_{21}\neq 0,\quad \frac{a_{11}a_{22}}{a_{12}a_{21}}<1,\quad w_{11}>0,\quad w_{22}>0$$
$$\text{for the points } t_{11}=0,\ t_{12}=\pm l_2,$$
$$t_{21}=\pm l_1,\ t_{22}=0, \quad \text{(A.7)}$$

where a_{ij} is the (i,j)th element of the mixing matrix $\mathbf{A}$. Note that any mixing matrix belongs to one of the conditions in Eqs. (A.6) and (A.7). In addition, one can always find an unmixing matrix such that $w_{11}>0$ and $w_{22}>0$ because of scale indeterminacy. Therefore, the modified infomax algorithm can obtain locally stable equilibrium points which correspond to separating states. In the same way, it can be proved that the non-separating equilibrium points are locally unstable.

References

[1] S. Amari, Natural gradient works efficiently in learning, Neural Comput. 10 (2) (1998) 251–276.

[2] S. Amari, A. Cichocki, H.H. Yang, A new learning algorithm for blind source separation, in: Advances in Neural Information Processing Systems, vol. 8, MIT Press, Cambridge, MA, 1996, pp. 757–763.

[3] S. Amari, S.C. Douglas, A. Cichocki, H.H. Yang, Multichannel blind deconvolution and equalization using the natural gradient, in: IEEE International Workshop on Wireless Communication, 1997, pp. 101–104.

[4] A.J. Bell, T.J. Sejnowski, An information-maximisation approach to blind separation and blind deconvolution, Neural Comput. 7 (1995) 1129–1159.

[5] J.-F. Cardoso, Blind signal separation: statistical principles, Proc. IEEE 86 (10) (1998) 2009–2025.

[6] J.-F. Cardoso, High-order contrasts for independent component analysis, Neural Comput. 11 (1) (1999) 157–192.

[7] J.-F. Cardoso, B. Laheld, Equivariant adaptive source separation, IEEE Trans. Signal Process. 44 (12) (1996) 3017–3030.

[8] S. Choi, A. Cichocki, An unsupervised hybrid network for blind separation of independent non-gaussian source signals in multipath environment, J. Commun. Networks 1 (1) (1999) 19–25.

[9] S. Choi, A. Cichocki, L. Zhang, S. Amari, Approximate maximum likelihood source separation using the natural gradient, IEICE Trans. Fundam. E86-A (1) (2003) 206–214.

[10] A. Cichocki, W. Kasprzak, S. Amari, Multi-layer neural networks with a local adaptive learning rule for blind separation of source signals, in: International Symposium on Nonlinear Theory and its Applications, 1995, pp. 61–65.

[11] S.C. Douglas, A. Cichocki, Neural networks for blind decorrelation of signals, IEEE Trans. Signal Process. 45 (11) (1997) 2829–2842.

[12] S. Haykin, Adaptive Filter Theory, Prentice-Hall, New Jersey, NJ, 1996.

[13] S. Haykin (Ed.), Unsupervised Adaptive Filtering, Wiley, New York, NY, 2000.

[14] Homepage of Room Impulse Response v2.5: ⟨http://www.dspalgorithms.com/room/room25.html⟩.

[15] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, IEEE Trans. Neural Networks 10 (3) (1999) 626–634.

[16] A. Hyvärinen, E. Oja, A fast fixed-point algorithm for independent component analysis, Neural Comput. 9 (7) (1997) 1483–1492.

[17] C. Jutten, J. Hérault, Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture, Signal Process. 24 (1991) 1–10.

[18] B.C. Kuo, Automatic Control Systems, Prentice-Hall, Englewood Cliffs, NJ, 1991.

[19] T.-W. Lee, Independent Component Analysis, Kluwer Academic Publishers, Boston, MA, 1998.

[20] T.-W. Lee, M. Girolami, A.J. Bell, T.J. Sejnowski, A unifying information-theoretic framework for independent component analysis, Comput. Math. Appl. 31 (11) (2000) 1–21.

[21] T.-W. Lee, M. Girolami, T.J. Sejnowski, Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources, Neural Comput. 11 (2) (1999) 417–441.

[22] X.-T. Ling, Y.-F. Huang, R.-W. Liu, A neural network for blind signal separation, in: IEEE International Symposium on Circuit and Systems, 1994, pp. 69–72.

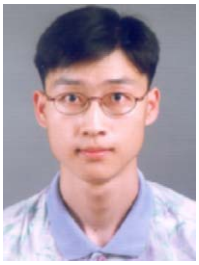
[23] O. Macchi, E. Moreau, Self-adaptive source separation, part I: convergence analysis of a direct linear network controlled by the Hérault–Jutten algorithm, IEEE Trans. Signal Process. 45 (4) (1997) 918–926.

[24] N. Minorsky, Nonlinear Oscillations, R.E. Krieger Pub. Co., Malabar, 1983.

[25] L. Molgedey, G. Schuster, Separation of a mixture of independent signals using time delayed correlations, Phys. Rev. Lett. 72 (23) (1994) 3634–3637.

[26] N.S. Nise, Control Systems Engineering, The Benjamin/Cummings Pub. Co., Redwood City, CA, 1992.

[27] D. Schobben, K. Torkkola, P. Smaragdis, Evaluation of blind signal separation methods, in: International Conference on Independent Component Analysis and Blind Signal Separation, 1999, pp. 261–266.
[28] E. Sorouchyari, Blind separation of sources, part III: stability analysis, Signal Process. 24 (1991) 21–29.
[29] K. Torkkola, Blind separation of convolved sources based on information maximization, in: IEEE International Workshop on Neural Networks for Signal Processing, 1996.
[30] K. Torkkola, Blind separation for audio signals—are we there yet?, in: International Conference on Independent Component Analysis and Blind Signal Separation, 1999, pp. 239–243.



Hyung-Min Park received the B.S., M.S., and Ph.D. degrees in Electrical Engineering and Computer Science from Korea Advanced Institute of Science and Technology, Daejeon, Korea, in 1997, 1999, and 2003, respectively. From 2003 to early 2005, he was a Post-Doc. at the Department of Biosystems, Korea Advanced Institute of Science and Technology. Currently, he is working in the Language Technologies Institute, Carnegie Mellon University, as a visiting researcher. His current research interests include the theory and applications of independent component analysis, blind source separation, adaptive noise canceling, and binaural or multi-microphone processing for noise-robust speech recognition.

Sang-Hoon Oh received his B.S. and M.S. degrees in Electronics Engineering from Pusan National University in 1986 and 1988, respectively. He received his Ph.D. degree in Electrical Engineering from Korea Advanced Institute of Science and Technology in 1999. From 1988 to 1989, he worked for the LG semiconductor, Ltd., Korea. From 1990 to 1998, he was a senior researcher in Electronics and Telecommunications Research Institute (ETRI), Korea. From 1999 to 2000, he was with Brain Science Research Center, KAIST. In 2000, he was with Brain Science Institute, RIKEN in Japan. In 2001, he was an R&D manager of Extell Technology Corporation, Korea. Since 2002, he has been with the Department of Information Communication Engineering, Mokwon University, Daejon, Korea. His research interests are supervised/unsupervised learning for intelligent information processing, speech processing, and pattern recognition.

Soo-Young Lee received B.S., M.S., and Ph.D. degrees from Seoul National University in 1975, Korea Advanced Institute of Science in 1977, and Polytechnic Institute of New York in 1984, respectively. From 1977 to 1980 he worked for the Taihan Engineering Co., Seoul, Korea. From 1982 to 1985 he also worked for General Physics Corporation at Columbia, MD, USA. In early 1986, he joined the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, as an Assistant Professor and now is a Full Professor at the Department of BioSystems and also department of Electrical Engineering and Computer Science. In 1997, he established Brain Science Research Center, which is the main research organization for the Korean Brain Neuroinformatics Research Program. The research program is one of the Korean Brain Research Promotion Initiatives sponsored by Korean Ministry of Science and Technology from 1998 to 2008, and currently about 35 Ph.D. researchers have joined the research program from many Korean universities. He is a Past-President of Asia-Pacific Neural Network Assembly, and has contributed to International Conference on Neural Information Processing as Conference Chair (2000), Conference Vice Co-Chair (2003), and Program Co-Chair (1994, 2002). Dr. Lee is the Editor-in-Chief of the newly established online/offline journal with double-blind review process, Neural Information processing—Letters and Reviews, and is on Editorial Board for two international journals, i.e., Neural Processing Letters and Neurocomputing. He received Leadership Award and Presidential Award from International Neural Network Society in 1994 and 2001, respectively, and APPNA Service Award in 2004. His research interests have resided in artificial brain, the human-like intelligent systems based on biological information processing mechanism in our brain. He has worked on the auditory models from the cochlea to the auditory cortex for noisy speech processing, information-theoretic binaural processing models for sound localization and speech enhancement, the unsupervised pro-active developmental models of human knowledge with multi-modal man–machine interactions, and the top–down selective attention models for superimposed pattern recognitions. His research scope covers the mathematical models, neuromorphic chips, and real-world applications. Also, he had recently extended his research into brain–computer interfaces with simultaneous fMRI and EEG measurements.