A MODIFICATION OF THE GRADIENT ALGORITHM FOR BLIND SIGNAL SEPARATION

Hyung-Min Park
Brain Science Research Center (BSRC)
and Dept. Electrical Engineering and Computer Science
Korea Advanced Institute of Science and Technology (KAIST)
Daejeon, 305-701, Republic of Korea
hmpark@mail.kaist.ac.kr

Sang-Hoon Oh
Dept. Information Communication Eng.
Mokwon University
Daejeon, 302-729, Republic of Korea

Soo-Young Lee
BSRC and Dept. Biosystems
KAIST
Daejeon, 305-701, Republic of Korea

ABSTRACT

We present a new gradient algorithm to perform blind signal separation (BSS). The algorithm is obtained by taking a trade-off between the ordinary gradient algorithm and the natural gradient algorithm. It provides a better performance than the ordinary gradient algorithm and is free from small-step-size restriction of the natural gradient algorithm. In addition, the algorithm has less computation than the other gradient algorithms. For theoretical support of our algorithm, local stability on desired solutions is proven for a simple network. Simulation results indicate that the algorithm efficiently provides a solution for BSS.

1. INTRODUCTION

Blind signal separation (BSS) is to estimate source signals from their linear mixtures without resorting to any prior information. Although the term 'blind' means that prior information is not available, many BSS algorithms rely on statistical independence of source signals [1, 2]. BSS has received extensive attention in signal and speech processing, machine learning, and neuroscience communities.

Although many researchers have proposed algorithms to perform BSS, most of them are batch-type and not appropriate to on-line separation. In many practical applications, however, all mixing data are not given in advance, and outputs have to be immediately provided for each input sample. Furthermore, most of algorithms have difficulties in separating convolved mixtures or dealing with natural source signals which have correlation among time samples because they have been proposed to separate instantaneous mixtures or whitened signals [3, 4]. Among several approaches to BSS, an ordinary gradient algorithm for entropy maximization is noticeable because of its simple and biologically plausible formulation [5, 6]. However, the parameter space is usually not orthogonal in Riemannian space, and the ordinary gradient does not indicate the steepest ascent direction for a desired solution. As a much more efficient strategy, Amari *et al.* proposed the natural gradient which can consider the relationship between a non-orthogonal space and an orthogonal space [7, 8]. In addition, Cardoso and Laheld independently proposed the same one called 'relative' gradient, and they showed that it has the equivariance property [9].

The ordinary gradient algorithm has slow convergence property and involves matrix inversion which is computationally intensive. Although the natural gradient algorithm is quite efficient and does not involve the matrix inversion, it still has additional computation such as matrix multiplication for instantaneous mixtures or convolution for convolved mixtures. Especially, the natural gradient algorithm has a serious problem to deal 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, one has to approximate the algorithm on the assumption of small filter coefficient changes for some time steps.

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

It forces to use a very small step size especially for a large adaptive filter length and may cause a slow convergence rate.

In order to obtain a better performance than the ordinary gradient algorithm and overcome the disadvantages of the natural gradiet algorithm, we present a new gradient algorithm as a trade-off between the ordinary gradient algorithm and the natural gradient algorithm. The algorithm maintains spatial and temporal independence, and it has less computation than the other gradient algorithms. Simulation results show the efficiency of the algorithm. For theoretical support, local stability on desired solutions of the algorithm is proven for a simple network.

2. BLIND SIGNAL SEPARATION

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

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n), \tag{1}$$

where $\mathbf{A}$ is an unknown full rank mixing matrix. The task is to estimate the source signals by finding an unmixing matrix $\mathbf{W}$, which is a permuted and rescaled version of the inverse of the mixing matrix $\mathbf{A}$. The estimated signals

$$\mathbf{u}(n) = \mathbf{W}\mathbf{x}(n) \tag{2}$$

are the original sources up to permutation and scaling.

Bell and Sejnowski proposed to learn the unmixing matrix $\mathbf{W}$ 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 [5]. Ordinary gradient for maximizing the entropy leads the following learning rule:

$$\Delta\mathbf{W} \propto [\mathbf{W}^T]^{-1} - \varphi(\mathbf{u})\mathbf{x}^T, \qquad \varphi(\mathbf{u}) = \left[-\frac{\frac{\partial p_1(u_1)}{\partial u_1}}{p_1(u_1)}, \cdots, -\frac{\frac{\partial p_M(u_M)}{\partial u_M}}{p_M(u_M)}\right]^T \tag{3}$$

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

A much more efficient way to learn the unmixing matrix is to follow the natural gradient [7, 9]. In this case, the natural gradient rescales the ordinary gradient by post-multiplying it with $\mathbf{W}^T\mathbf{W}$ giving

$$\Delta\mathbf{W} \propto [\mathbf{I} - \varphi(\mathbf{u})\mathbf{u}^T]\mathbf{W}. \tag{4}$$

It is known that the natural gradient provides the most efficient direction to update the unmixing matrix. Because the natural gradient algorithm does not involve computationally intensive matrix inversion, it provides a smaller computational load than the ordinary gradient algorithm.

3. A NEW GRADIENT ALGORITHM

We propose to use a 'modified' gradient algorithm,

$$\Delta\mathbf{W} \propto \mathbf{I} - \varphi(\mathbf{u})\mathbf{u}^T, \tag{5}$$

which takes a trade-off between the ordinary gradient algorithm and the natural gradient algorithm. It corresponds to the ordinary gradient algorithm multipled by $\mathbf{W}^T$. Note that the algorithm maintains independence among the estimated signals $\{u_i\}$ in eq. (5). Moreover, the algorithm does not involve matrix inversion as well as matrix multiplication with $\mathbf{W}$. Especially, it may be useful for hardware implementation because of its simple form.

From the algorithm, removing the score function $\varphi(\cdot)$ gives a second-order blind decorrelation learning rule [10]. However, it is worthy of note that the modified gradient algorithm can obtain independent signals from mixtures using higher-order statistics instead of decorrlated signals.

There are 16 equilibrium points of the modified gradient algorithm for 2 estimated independent signals and 2 observations assuming that the pdfs of the estimated signals are Laplace functions. Because the estimated signals are Laplace-distributed, $\mathrm{sgn}(\cdot)$ is used as the score function $\varphi(\cdot)$. The equilibrium points, at which $E[\Delta\mathbf{W}] = \mathbf{0}$, are

$$t_{11} = \pm l_1,\ t_{12} = 0,\ t_{21} = 0,\ t_{22} = \pm l_2, \tag{6}$$

$$t_{11} = 0,\ t_{12} = \pm l_2,\ t_{21} = \pm l_1,\ t_{22} = 0, \tag{7}$$

$$t_{11} = c_{11}\frac{2}{3}l_1,\ t_{12} = c_{12}\frac{2}{3}l_2,\ t_{21} = c_{21}\frac{2}{3}l_1,\ t_{22} = c_{22}\frac{2}{3}l_2, \qquad c_{ij} = \pm 1,\ \prod_{i=1}^{2}\prod_{j=1}^{2} c_{ij} = -1, \tag{8}$$

where t_{ij} are elements of the overall matrix $\mathbf{T} = \mathbf{W}\mathbf{A}$, and the source pdfs are $p_1(s_1) = \frac{l_1}{2}e^{-l_1|s_1|}$ and $p_2(s_2) = \frac{l_2}{2}e^{-l_2|s_2|}$. Eq. (6) and (7) are separating states whereas eq. (8) is not.

Let us examine local stability of these equilibrium points. It is known that an equilbrium point is locally stable if the eigenvalues of $\mathbf{J}$ have negative real parts [11, 12], where

$$\mathbf{J} = \begin{bmatrix} \frac{\partial}{\partial w_{11}} E[1-\varphi(u_1)u_1] & \frac{\partial}{\partial w_{12}} E[1-\varphi(u_1)u_1] & \frac{\partial}{\partial w_{21}} E[1-\varphi(u_1)u_1] & \frac{\partial}{\partial w_{22}} E[1-\varphi(u_1)u_1] \\ \frac{\partial}{\partial w_{11}} E[-\varphi(u_1)u_2] & \frac{\partial}{\partial w_{12}} E[-\varphi(u_1)u_2] & \frac{\partial}{\partial w_{21}} E[-\varphi(u_1)u_2] & \frac{\partial}{\partial w_{22}} E[-\varphi(u_1)u_2] \\ \frac{\partial}{\partial w_{11}} E[-\varphi(u_2)u_1] & \frac{\partial}{\partial w_{12}} E[-\varphi(u_2)u_1] & \frac{\partial}{\partial w_{21}} E[-\varphi(u_2)u_1] & \frac{\partial}{\partial w_{22}} E[-\varphi(u_2)u_1] \\ \frac{\partial}{\partial w_{11}} E[1-\varphi(u_2)u_2] & \frac{\partial}{\partial w_{12}} E[1-\varphi(u_2)u_2] & \frac{\partial}{\partial w_{21}} E[1-\varphi(u_2)u_2] & \frac{\partial}{\partial w_{22}} E[1-\varphi(u_2)u_2] \end{bmatrix}. \quad (9)$$

Here, w_{ij} are elements of the mixing matrix $\mathbf{W}$. Determinant of $\lambda\mathbf{I} - \mathbf{J}$ is

$$\begin{aligned} &\det(\lambda\mathbf{I} - \mathbf{J}) \\ &= \lambda^4 + \lambda^3\{E[k_1u_2x_2] + E[k_2u_1x_1]\} \\ &+ \lambda^2\{E[k_1u_2x_2]E[k_2u_1x_1]\} \\ &+ \lambda\{E[k_1u_2x_2]E[k_2u_1x_1]E[m_1x_1] \\ &\quad + E[k_1u_2x_2]E[k_2u_1x_1]E[m_2x_2] \\ &\quad - E[k_1u_2x_2]E[k_2u_1x_2]E[m_2x_1] \\ &\quad - E[k_1u_2x_1]E[k_2u_1x_1]E[m_1x_2]\} \\ &+ E[k_1u_2x_2]E[k_2u_1x_1]E[m_1x_1]E[m_2x_2] \\ &- E[k_1u_2x_2]E[k_2u_1x_2]E[m_1x_1]E[m_2x_1] \\ &- E[k_1u_2x_1]E[k_2u_1x_1]E[m_1x_2]E[m_2x_2] \\ &+ E[k_1u_2x_1]E[k_2u_1x_2]E[m_1x_2]E[m_2x_1]. \quad (10) \end{aligned}$$

In this equation, $k_i = \lim_{q\to\infty} q(1-\tanh^2(qu_i))$ and $m_i = \lim_{q\to\infty}\tanh(qu_i)$ since $\mathrm{sgn}(u)$ can be replaced by $\lim_{q\to\infty}\tanh(qu),\ q>0$. Without finding roots of $\det(\lambda\mathbf{I}-\mathbf{J}) = 0$, we can check with Routh-Hurwitz criterion whether the eigenvalues of $\mathbf{J}$ have negative real parts [13].

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

$$\begin{aligned} &a_{11}a_{22} \neq 0,\ \frac{a_{12}a_{21}}{a_{11}a_{22}} < 1,\ w_{11} > 0,\ w_{22} > 0 \\ &\quad \text{for the points} \\ &\qquad t_{11} = \pm l_1,\ t_{12} = 0,\ t_{21} = 0,\ t_{22} = \pm l_2, \quad (11) \\ &a_{12}a_{21} \neq 0,\ \frac{a_{11}a_{22}}{a_{12}a_{21}} < 1,\ w_{11} > 0,\ w_{22} > 0 \\ &\quad \text{for the points} \\ &\qquad t_{11} = 0,\ t_{12} = \pm l_2,\ t_{21} = \pm l_1,\ t_{22} = 0, \quad (12) \end{aligned}$$

where a_{ij} are elements of the mixing matrix $\mathbf{A}$. Note that any mixing matrix belongs to one of the conditions in eq. (11) and (12). In addition, one can always find an unmixing matrix in which $w_{11} > 0$ and $w_{22} > 0$ because of scale indeterminacy. Therefore, the modified gradient algorithm can obtain locally stable equilibrium points which correspond to separating states. In the same way, one can obtain conditions for local stability of the equilibrium points which are not separating states. The conditions show that the points are not locally stable.

4. EXTENSION TO CONVOLVED MIXTURES

One can extend the modified gradient algorithm to BSS of convolved mixtures. If the mixing of source signals involves convolution and time-delays,

$$\mathbf{x}(n) = \sum_{k=0}^{K-1} \mathbf{A}_k(n)\mathbf{s}(n-k), \quad (13)$$

where $\mathbf{x}(n)$ and $\mathbf{s}(n)$ are an observation vector and an unknown source vector, respectively. $\mathbf{A}_k(n)$ denotes a matrix composed of mixing filter coefficients. Let us consider a feedforward network to separate signals from convolved mixtures as

$$\mathbf{u}(n) = \sum_{k=0}^{K-1} \mathbf{W}_k(n)\mathbf{x}(n-k), \quad (14)$$

where adaptive filter matrices $\mathbf{W}_k(n)$ supposedly make an output vector $\mathbf{u}(n)$ reproduce the source vector $\mathbf{s}(n)$.

Torrkola derived the ordinary gradient algorithm of entropy maximization for convolved mixtures as [6]

$$\Delta\mathbf{W}_k(n) \propto (\mathbf{W}_0^T(n))^{-1}\delta_k - \varphi(\mathbf{u}(n))\mathbf{x}^T(n-k). \quad (15)$$

On the other hand, the natural gradient algorithm, which is the steepest ascent direction in Riemannian space, is given by [8]

$$\Delta\mathbf{W}_k(n) \propto \mathbf{W}_k(n) - \varphi(\mathbf{u}(n))\mathbf{r}_k^T(n), \quad (16)$$

where $\mathbf{r}_k(n) = \sum_{l=0}^{K-1}\mathbf{W}_l^T(n)\mathbf{u}(n-k+l)$. However, it involves non-causal terms and very intensive computation for computing all $\mathbf{r}_k(n)$ at each time step. Practically, the algorithm is modified by introducing a $K-1$ sample delay to remove the non-causal terms and reusing past results such that $\mathbf{r}_k(n) \approx \mathbf{r}_0(n-k)$, assuming that $\mathbf{W}_k(n) \approx \mathbf{W}_k(n-1) \approx \cdots \approx \mathbf{W}_k(n-2K+2)$. With these changes, the algorithm is

$$\Delta\mathbf{W}_k(n) \propto \mathbf{W}_k(n) - \varphi(\mathbf{u}(n-K+1))\mathbf{r}^T(n-k), \quad (17)$$

where $\mathbf{r}(n) = \sum_{l=0}^{K-1}\mathbf{W}_{K-1-l}^T(n)\mathbf{u}(n-l)$. With the assumption of $\mathbf{W}_k(n) \approx \cdots \approx \mathbf{W}_k(n-2K+2)$, the algorithm has to use a very small step size especially for a large number of adaptive filter taps in order to

converge on a proper solution stably. Therefore, it may cause a slow convergence speed.

To separate convolved mixtures, extending the modified gradient algorithm in eq. (5) gives

$$\Delta \mathbf{W}_k(n) \propto \mathbf{I}\delta_k - \varphi(\mathbf{u}(n))\mathbf{u}^T(n-k). \qquad (18)$$

The algorithm maintains spatial and temporal independence, and it does not involve matrix inversion. In addition, it does not have to use a very small step size and compute additional convolution such as computation of $\mathbf{r}(n)$ in the natural gradient algorithm.

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

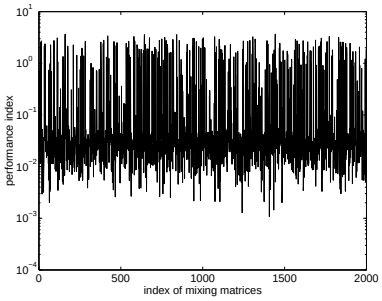
5. EXPERIMENTAL RESULTS

5.1. Simulations on instantaneous mixtures

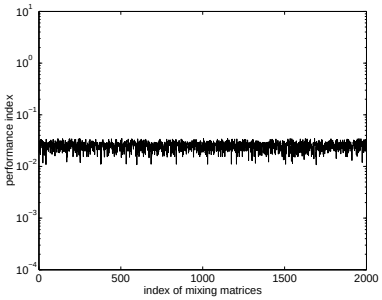
The modified gradient algorithm was compared with others for instantaneous mixtures. To measure performances of the algorithms, we have used the performance index which gives a lower value for a better performance [7]. Fig. 1 shows the performance indices for 2000 randomly generated 2×2 mixing matrices with Laplace-distributed source signals, each of which consists of 160000 i.i.d. samples. We have chosen identity matrix for the initial unmixing matrix.

The natural gradient algorithm provided satisfactory performances for all simulated mixing matrices. This supports the equivariance property of the natural gradient [9]. The modified gradient algorithm failed to separate signals for some mixing matrices, but it could separate much more signals than the ordinary gradient algorithm. The poor performances mostly came from ill-conditioned mixing matrices. Note that the modified gradient algorithm successfully separates signals except some ill-conditioned mixing matrices and provides a simpler formulation than the other gradient algorithms.

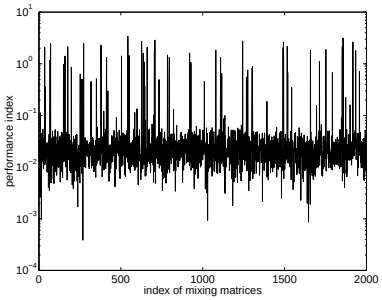
Fig. 1(d) shows the performance indices for the algorithm proposed by Ling *et al.* [14]. The algorithm is similar to the Jutten-Hérault algorithm [15] and corresponds to the modified gradient algorithm with fixed diagonal elements of the unmixing matrix $\mathbf{W}$. Although Ling *et al.* had not used $\mathrm{sgn}(\cdot)$ as the score function, we used it for its efficiency. However, the algorithm failed to separate signals much more frequently than the modified gradient algorithm.



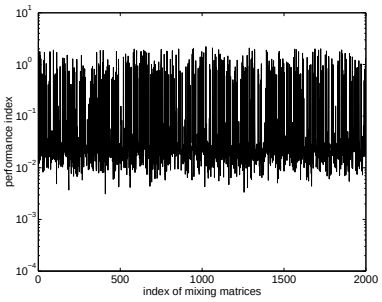
(a) Ordinary gradient algorithm



(b) Natural gradient algorithm



(c) Modified gradient algorithm



(d) $\Delta w_{ij} \propto -\mathrm{sgn}(u_i)u_j,\ i \neq j$

Figure 1: The performance indices of the algorithms for instantaneous mixtures

5.2. Simulations on convolved mixtures

To perform experiments for convolved mixtures, real-recorded speech signals were mixed with two different mixing systems. One of them was [1]

$$\begin{aligned} A_{11}(z) &= 0.9 + 0.5z^{-1} + 0.3z^{-2}, \\ A_{12}(z) &= -0.7z^{-5} - 0.3z^{-6} - 0.2z^{-7}, \\ A_{21}(z) &= 0.5z^{-5} + 0.3z^{-6} + 0.2z^{-7}, \\ A_{22}(z) &= 0.8 - 0.1z^{-1}. \end{aligned} \quad (19)$$

The other was [6]

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

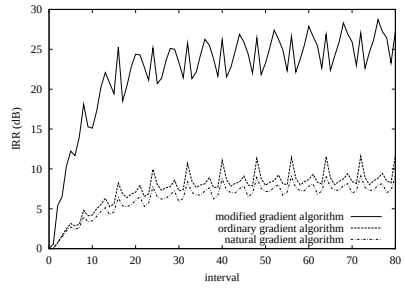
To deal with the mixing systems, 40 and 100 taps were used for unmixing of the first and the second systems, respectively. In addition, to avoid whitening of the estimated signals, we have forced direct filters of the unmixing systems, $W_{ii}(z)$, to scaling factors. Each speech signal had 10 second length at 16kHz sampling rate. It is known that speech signal approximately follows Laplacian distribution. Therefore, $\mathrm{sgn}(\cdot)$ was used as the score function $\varphi(\cdot)$. Experimental results were compared in terms of interference reduction ratio (IRR), which is defined as difference between signal-to-interference ratios (SIRs) of the final and the initial unmixing systems. The SIR is a ratio between the signal source power over the interference source power at the outputs,

$$\mathrm{SIR} = \frac{1}{2} \cdot 10 \log \left| \frac{< (u_{1,s_1})^2 >}{< (u_{1,s_2})^2 >} \cdot \frac{< (u_{2,s_2})^2 >}{< (u_{2,s_1})^2 >} \right| \quad (21)$$

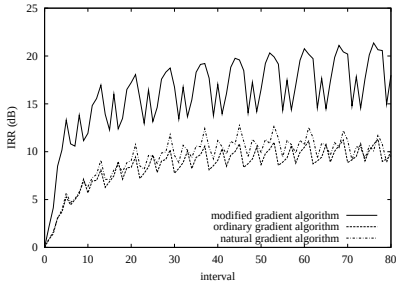
for 2×2 mixing/unmixing system [16]. In eq. (21), u_{j,s_i} denotes the jth output of the cascaded mixing/unmixing system when only s_i is active. As the initial unmixing system, we have chosen identity system without delay. All experiments were conducted with several step sizes, and the best performance is shown.

Fig. 2 displays the IRRs of the three gradient algorithms for the two different mixing systems. Each signal was divided into 8 intervals, each of which had 20000 samples. We repeated adaptation with the input mixed signals for 10 times. Therefore, there were 80 intervals on the horizental axis of each figure.

For the two experiments, the modified gradient algorithm showed better IRRs than the other gradient algorithms. The natural gradient algorithm did not provide good performances because it had to use a very small step size in order to converge on a desired solution stably and accumulated errors of the adaptive



(a) Mixing system in [1]



(b) Mixing system in [6]

Figure 2: The IRRs of the algorithms for convolved mixtures

unmixing filter coefficients to $\mathbf{r}(n)$. As the case of instantaneous mixtures, the ordinary gradient algorithm gave inferior performances to the modified gradient algorithm. It is worthy of note that the modified gradient algorithm has less computation and better performances than the other gradient algorithms.

6. CONCLUSION

In this paper, we proposed a modified gradient algorithm to perform BSS. By taking a trade-off between the ordinary gradient algorithm and the natural gradient algorithm, the algorithm provided a better performance than the ordinary gradient algorithm and overcame the disadvantages of the natural gradient algorithm such as restriction to a small step size. Furthermore, the algorithm gave a simpler formulation than the other gradient algorithms. We also proved that the algorithm was locally stable for a 2×2 network to separate instantaneous mixtures.

7. REFERENCES

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

[2] J.-F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, 1998.

[3] J.-F. Cardoso, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, no. 1, pp. 157–192, 1999.

[4] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[5] A. J. Bell and T. J. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[6] K. Torkkola, "Blind separation of convolved sources based on information maximization," in *IEEE Int. Workshop on Neural Networks for Signal Processing*, 1996.

[7] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind source separation," in *Advances in Neural Information Processing Systems 8*. 1996, pp. 757–763, MIT Press.

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

[9] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, no. 12, pp. 3017–3030, 1996.

[10] S. C. Douglas and A. Cichocki, "Neural networks for blind decorrelation of signals," *IEEE Trans. Signal Processing*, vol. 45, no. 11, pp. 2829–2842, 1997.

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

[12] E. Sorouchyari, "Blind separation of sources, part III: stability analysis," *Signal Processing*, vol. 24, pp. 21–29, 1991.

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

[14] X.-T. Ling, Y.-F. Huang, and R.-W. Liu, "A neural network for blind signal separation," in *IEEE Int. Symp. Circuit and Systems*, 1994, pp. 69–72.

[15] C. Jutten and J. Hérault, "Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1–10, 1991.

[16] D. Schobben, K. Torrkola, and P. Smaragdis, "Evaluation of blind signal separation methods," in *Int. Conf. ICA and BSS*, 1999, pp. 261–266.

A. APPENDIX

From a standpoint, the modified gradient algorithm can be derived from a risk function using Kullback-Leibler divergence. In this case, the loss function $L(\mathbf{W})$ is given by

$$L(\mathbf{W}) = -\log|\det(\mathbf{W})| - \sum_{i=1}^{M} \log\, p_i(u_i). \qquad (A.1)$$

An infinitesimal increment of the loss function for an increment $d\mathbf{W}$ is

$$dL(\mathbf{W}) = L(\mathbf{W}+d\mathbf{W}) - L(\mathbf{W}) = -tr(d\mathbf{V}) + \varphi^T(\mathbf{u})d\mathbf{V}\mathbf{u}. \qquad (A.2)$$

where $d\mathbf{V} = d\mathbf{W}\mathbf{W}^{-1}$ is called as the modified differential matrix composed of transfer matrix $\mathbf{W}$ between input vector $\mathbf{x}$ and output vector $\mathbf{u}$. Minimizing the loss function for the modified differential matrix gives the modified gradient algorithm

$$\Delta\mathbf{W} \propto -\frac{dL(\mathbf{W})}{d\mathbf{V}} = \mathbf{I} - \varphi(\mathbf{u})\mathbf{u}^T. \qquad (A.3)$$

Extending the derivation to convolved mixtures, one obtains the modified gradient algorithm for convolved mixtures. A feedforward network is used for separating signals as

$$\mathbf{u}(n) = \sum_{k=0}^{K-1} \mathbf{W}_k(n)\mathbf{x}(n-k) = \mathbf{W}(z,n)\mathbf{x}(n), \qquad (A.4)$$

where $\mathbf{W}(z,n)$ is a polynomial matrix in the time-shift operator z^{-1}. Here, the loss function becomes [8]

$$L(\mathbf{W}(z,n)) = -\log|\det(\mathbf{W}_0(n))| - \sum_{i=1}^{M} \log\, p_i(u_i(n)). \qquad (A.5)$$

As the case of instantaneous mixtures, the modified gradient algorithm is obtained by minimizing the loss function for the modified differential matrix which is $d\mathbf{V}(z,n) = d\mathbf{W}(z,n)\mathbf{W}^{-1}(z,n)$ as follows:

$$\Delta\mathbf{W}_k(n) \propto -\frac{dL(\mathbf{W}(z,n))}{d\mathbf{V}_k(n)} = \mathbf{I}\delta_k - \varphi(\mathbf{u}(n))\mathbf{u}^T(n-k). \qquad (A.6)$$