2차 Nonstationary 신호 분리: 자연기울기 학습

최희열[0] 최승진
포항공과대학교 컴퓨터공학과
(hychoi[0], seungjin)@postech.ac.kr

Second-order nonstationary source separation: Natural gradient learning

Hee-Youl Choi[0] Seungjin Choi
Dept. Computer Science and Engineering, POSTECH

요 약

Most of source separation methods focus on stationary sources so higher-order statistics is necessary. In this paper we consider a problem of source separation when sources are second-order nonstationary stochastic processes. We employ the natural gradient method and develop learning algorithms for both linear feedback and feedforward neural networks. Thus our algorithms possess equivariant property. Local stability analysis shows that separating solutions are always locally stable stationary points of the proposed algorithms, regardless of probability distributions of sources.

1. Introduction

In the context of source separation, let us assume that the m dimensional vector of measurement signals, $x(t)$, is generated by a linear data model described by

$$x(t) = As(t), \tag{1}$$

where $s(t)$ is the n dimensional vector whose elements are called sources. The matrix $A \in R^{m \times n}$ is called a mixing matrix. The task of source separation is to estimate the mixing matrix A (or its inverse), given only a finite number of measurement signals, $\{x(t)\}$, t = 1, … , N. Source vector $s(t)$ is unknown in advance, but their elements are assumed to be statistically independent.

Most of source separation methods have focused on stationary sources, so higher-order statistics (HOS) is necessary for successful separation, unless sources are temporally correlated. But, many natural signals are *inherently nonstationary stochastic* processes. It was shown in [1] that source separation could be achieved by decorrelation if sources are independent second-order nonstationary stochastic processes.

In this paper we pay our attention to the problem of second-order nonstationary source separation. We develop natural gradient learning algorithms for both linear feedback and feedforward neural networks. Due the natural gradient method, our algorithms converge to separation solutions along the steepest descent direction and possess the equivariant property that was first discovered by Cardoso and Laheld [2]. We also present local stability analysis of our algorithms and show that separating solutions are always locally stable stationary points of our algorithms, regardless of probability distributions of sources.

As in [1], the following assumptions are made throughout this paper:

AS1 The mixing matrix A has full column rank

AS2 Source signals are statistically independent with zero mean.

AS3 $\frac{r_i(t)}{r_j(t)}$ are not constant with time. $r_i(t) = E\{s_i^2(t)\}$

2. Natural Gradient Algorithms

We consider the objective function proposed by Matsuoka *et al.* [1]. Then we employ the natural gradient method which was shown to be efficient for on-line learning [3], [4], [5] and derive on-line learning algorithms for both feedback and feedforward networks. For the sake of simplicity, we only consider the case where there are as many sensors as sources, i.e., $m = n$.

2.1 Objective Function

It was shown in [1] that second-order decorrelation is sufficient for source separation under the assumptions (AS1)-(AS3). The objective function that we consider is given by

$$J(W) = \frac{1}{2}\{\sum_{i=1}^{n} \log E\{y_i^2(t)\} - \log \det(E\{y(t)y^T(t)\})\} \tag{2}$$

where $y(t)$ is the network output vector and det(·) denotes the determinant of a matrix. It takes minima if and only if $E\{y_i(t)y_j(t)\}=0$, for $i,j=1,...,n,(i\neq j)$.

2.2 Feedback Network

In a linear feedback network, the output *y(t)* is described by

$$y(t)=x(t)+Wy(t). \quad (3)$$

We calculate the total differential $dJ(W)$,

$$dJ(W)=J(W+dW)-J(W) \quad (4)$$

due to the change dW

Define a modified differential matrix

$$dV=(I-W)^{-1}dW. \quad (5)$$

Hence, the gradient of the objective function with respect to the modified differential matrix dV is given by

$$dJ(W)/dV=E\{\Lambda^{-1}(t)y(t)y^T(t)\}-I \quad (6)$$

where $\Lambda(t)$ is a diagonal matrix whose *i*th diagonal element is $E\{y_i^2(t)\}$.

The stochastic gradient descent method leads to the updating rule for V that has the form

$$V(t+1)=V(t)+\eta_t\{I-\Lambda^{-1}(t)y(t)y^T(t)\}, \quad (7)$$

where $\eta_t>0$ is the learning rate and $\Lambda(t)$ is a diagonal matrix whose *i*th diagonal element is $\lambda_i(t)$ that can be estimated by

$$\lambda_i(t)=(1-\delta)\lambda_i(t-1)+\delta y_i^2(t) \quad (8)$$

for some small δ (say, $\delta=0.01$).

If follows from the definition (5) that the learning algorithm for W is given by

$$\Delta W(t)=\eta_t(I-W(t))(I-\Lambda^{-1}(t)y(t)y^T(t)). \quad (9)$$

Remarks

1. The algorithm (9) can be viewed as a special form of the robust neural ICA algorithms developed by Cichocki and Unehauen [6].
2. The algorithm leads to a simple form of nonholonomic ICA algorithms proposed by Amari *et al.* [7]

2.3 Feedforward Network

In a linear feedforward network, the output *y(t)* is given by

$$y(t)=Wx(t). \quad (10)$$

Define a modified differential matrix

$$dV=W^{-1}dW. \quad (11)$$

Then, the natural gradient learning algorithm for W has the form

$$\Delta W(t)=\eta_t\Lambda^{-1}(t)(\Lambda(t)-y(t)y^T(t))W(t). \quad (12)$$

Note that two remarks described for the case of linear feedback network also hold in this case.

3. Local Stability Analysis

Stationary points of the algorithm (9) or (12) satisfy

$$E(I-\Lambda^{-1}(t)y(t)y^T(t))=0 \quad (13)$$

which implies that $E\{y_i(t)y_j(t)\}=0$ for $i,j=1,...,n,(i\neq j)$.

In order to show that stationary points of (9) are locally stable, we need to show that the Hessian d^2J is positive. For shorthand notation, we omit the time index t in the following analysis.

Recall that

$$dJ=E\{y^T\Lambda^{-1}dVy\}-tr\{dV\}. \quad (14)$$

Then the Hessian d^2J is

$$d^2J=E\{y^TdV^T\Lambda^{-1}dVy+y^T\Lambda^{-1}dVdVy\}. \quad (15)$$

After some calculation, we have

$$d^2J=\sum_{i,j}[\frac{\lambda_i}{\lambda_j}(dv_{ji})^2+dv_{ij}dv_{ji}]=\sum_{i\neq j}q_{ij}+\sum_i q_{ii} \quad (16)$$

where $q_{ii}=\frac{\lambda_i}{\lambda_i}(dv_{ji})^2+dv_{ij}dv_{ji}$.

One can easily see that the summand in the second term in (16) is always positive. The summand in the first term in (16) can be rewritten as

$$q_{ij}+q_{ji}=\begin{bmatrix}dv_{ij} & dv_{ji}\end{bmatrix}\begin{bmatrix}\frac{\lambda_j}{\lambda_i} & 1\\ 1 & \frac{\lambda_i}{\lambda_j}\end{bmatrix}\begin{bmatrix}dv_{ij}\\ dv_{ji}\end{bmatrix} \quad (17)$$

The equation (17) is always non-negative. Hence d^2J is always positive. The stability of the algorithm (9) does not depend on the probability distributions of sources. Thus our algorithm is always locally stable regardless of the probability distributions of sources.

4. Numerical Examples

We have performed experiments with 3 digitized voice signals, all of which are sampled at 8 kHz. Three mixture signals were generated using the mixing matrix given by

$$A=\begin{bmatrix}0.224 & 0.055 & 0.469\\ 0.162 & 0.505 & 0.476\\ 0.933 & 0.649 & 0.912\end{bmatrix} \quad (18)$$

We evaluate the performance of three algorithms:

Algorithm 1: the Matsuoka's algorithm in [1].

Algorithm 2: the natural gradient algorithm (9) for feedback network.

Algorithm 3: the natural gradient algorithm (12) for feedforward network

The constant learning rate $\eta_t=0.0005$ was used for all three algorithms.

As performance measure, we use the performance index (PI) defined by

$$PI=\sum_{i=1}^{n}\left\{\left(\sum_{k=1}^{n}\frac{|g_{ik}|}{\max_j|g_{ij}|}-1\right)+\left(\sum_{k=1}^{n}\frac{|g_{ki}|}{\max_j|g_{ji}|}-1\right)\right\} \quad (19)$$

where g_{ij} is the *(i,j)*th element of the global system matrix G ($G=(I-W)^{-1}A$ for a recurrent network, $G=WA$ for a feedforward network) and $\max_j g_{ij}$ represents the maximum value among the elements in the *i*th row vector of G, $\max_j g_{ji}$ does the maximum value among the elements in the *i*th column vector of G.

In addition to the performance measure (19), we also calculated the Signal to Interference Ratio Improvement (SIRI) defined by

$$SIRI_i = \frac{E\{(x_i - s_i)^2\}}{E\{(y_i - s_i)^2\}} \qquad (20)$$

Numerical experimental results are shown in [Fig 1] and are summarized in [Table 1]. Poor performance of Algorithm 1 might result from the simplification approximation made in [1] which is not reasonable for the case of $n > 2$.
Moreover, our algorithms possess the equivariant property, thus they give satisfactory results even for the case of ill-conditioned mixing. More details can be found in [8].

Type of Algorithm	SIRI
Matsuoka [1]	$SIRI_1$= 16.0 dB
	$SIRI_2$= 14.8 dB
	$SIRI_3$= 36.5 dB
Feedback (9)	$SIRI_1$= 68.1 dB
	$SIRI_2$= 61.5 dB
	$SIRI_3$= 65.1 dB
Feedforward (12)	$SIRI_1$= 68.0 dB
	$SIRI_2$= 61.5 dB
	$SIRI_3$= 65.2 dB

[Table 1]. Performance Comparison in Terms of SIRI

5. Conclusions

Two natural gradient learning algorithms which perform second-order nonstationary source separation. We also presented local stability analysis of the algorithms and showed that separating solutions are always locally stable stationary points of the proposed algorithms, regardless of probability distributions of sources. Numerical experimental results confirmed the high performance of the algorithms.

6. Acknowledgments

This work was supported by Korea Ministry of Science and Technology under Brain Science and Engineering Research Program and by Korea Ministry of Information and Communication under Advanced backbone IT technology development project and by Ministry of Education of Korea for its financial support toward the Electrical and Computer Engineering Division at POSTECH through its BK21 program.

7. Reference

[1] K. Matsuoka, M. Ohya, and M. Kawamoto, " A neural net for blind separation of nonstationary signals," *Neural Networks*, vol. 8, no. 3, pp. 411-419, 1995.

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

[3] S. Amari, " Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, no. 2, pp.251-276, 1998.

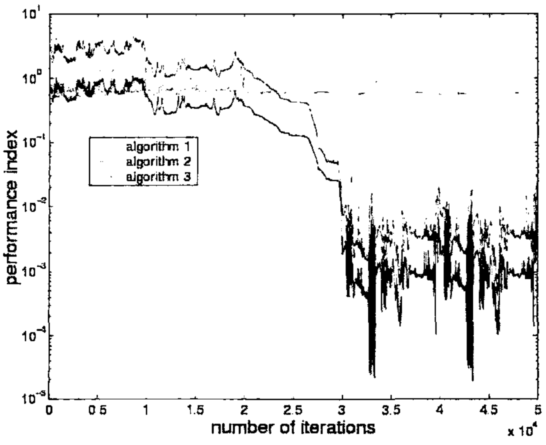
[4] S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, " Multi channel blind deconvolution and equalization using the natural gradient," in *Proc. SPAWC*, (Paris, France), pp. 101-104, 1997

[5] S. Choi, S. Amari, and A. Cichocki, " Natural gradient learning for spatio-temporal decorrelation: Recurrent network," *IEICE Trans. Fundamentals*, vol. E83-A, pp. 2715-2722, Dec. 2000.

[6] A. Cichocki and R. Unbehauen, " Robust neural networks with on-line learning for blind identification and blind separation of sources," *IEEE Trans. Circuits and Systems - I*: Fundamental Theory and Applications, vol. 43, pp. 894-906, 1996.

[7] S. Amari, T. P. Chen, and A. Cichocki, " Nonholonomic orthogonal learning algorithms for blind source separation," *Neural Computation*, vol. 12, no. 6, pp. 1463-1484, 2000.

[8] S. Choi, A. Cichocki, and S. Amari, " Equivariant nonstationary source separation," *Neural Networks*, 2002, to appear.



[Fig 1]. The evolution of performance index is shown