PROCEEDINGS OF ISAP2000, FUKUOKA, JAPAN

CHARACTERISTICS OF BLIND DECONVOLUTION WITH NATURAL GRADIENT METHOD

Kaoru GOTOH and Masashi HAYAKAWA
The University of Electro-Communications
1-5-1 Chofugaoka Chofu Tokyo, 182-8585, Japan
E-mail gotoh@whistler.ee.uec.ac.jp

1. Introduction

Blind separation is a very important problem of considerable interest in a wide range of applications as an inverse problem including seismology, wireless communication, pattern recognition, non-invasive medical diagnosis. These problems consist of reconstruction of a set of statistically independent source signals from a vector of received signals. Recently many heuristic neural algorithms for blind signal processing have been reported in different fields. Most of these works are devoted to instantaneous mixture of stochastic independent signals. Only few works have been treated the more realistic conditions where received signals were convolved or mixed with some delay. In fact, in real-life applications wide-band signals are filtered and delayed before reaching sensors. The present paper uses a MA(moving average) convolution model as propagation system and minimizes the cost function using natural gradient. The natural gradient search method which is proposed by Amari has emerged as a particularly useful technique, when the parameter space is not a uniform Euclidean but has some structure. We show the deconvolution result of the computer simulation and compare it with the convergence property of the natural gradient and normal gradient method.

2. Blind deconvolution problem

In this section, we study the measurement model and specify our purpose. Let us assume $\boldsymbol{x}(k) = [x_1(k) \cdots x_m(k)]^T$ be a vector of m source signals whose components are zero-mean independent, i.i.d.(independent and identically distributed) signal. The observed signal vector $\boldsymbol{y}(k) = [y_1(k) \cdots y_n(k)]^T$ at the time k is described by the model.

$$\boldsymbol{y}(k) = \sum_{p=0}^{\infty} \mathbf{H}_p \boldsymbol{x}(k-p), \tag{1}$$

where $\mathbf{H}_p$ which indicates the characteristic of the propagation channel is $m \times n$ full-rank matrix. In order to reconstruct the source signal, we use the feed-forward model.

$$\boldsymbol{s}(k) = \sum_{p=0}^{\infty} \boldsymbol{W}_p(k) \boldsymbol{y}(k-p). \tag{2}$$

In an operator form, the input and output of the equalized system are given by the following equations.

$$\boldsymbol{y}(k) = \mathbf{H}(z)[\boldsymbol{x}(k)], \tag{3}$$

$$\boldsymbol{s}(k) = \boldsymbol{W}(z,k)[\boldsymbol{y}(k)] = \boldsymbol{C}(z,k)[\boldsymbol{s}(k)], \tag{4}$$

where

$$\mathbf{H}(z) = \sum_{p=0}^{\infty} \mathbf{H}_p z^{-p}, \quad \boldsymbol{W}(z,k) = \sum_{p=0}^{\infty} \boldsymbol{W}_p(k) z^{-p}. \tag{5}$$

$$\boldsymbol{C}(z,k) = \boldsymbol{W}(z,k)\mathbf{H}(z) \tag{6}$$

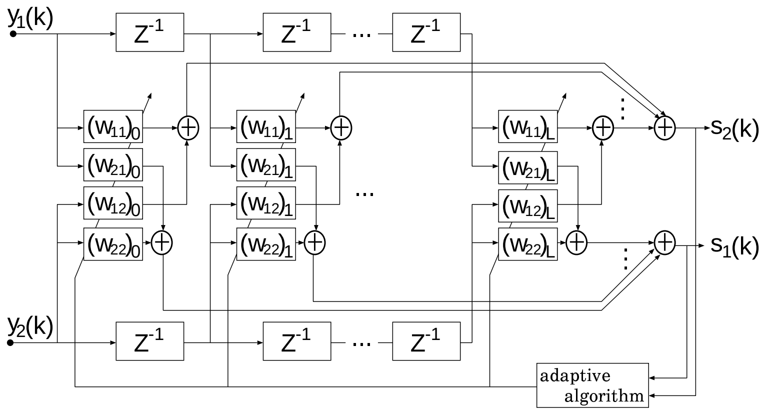

Figure 1: Blind deconvolution feed-forward model

The goal of Blind deconvolution problem is to obtain the rescaled and time-delayed estimates of the source signals $\boldsymbol{s}(k) = [s_1(k) \cdots s_m(k)]$ without any information about the sources and propagation coefficient such that

$$\lim_{k \to \infty} \boldsymbol{C}(z,k) = \mathbf{PD}(z), \tag{7}$$

by adjusting the estimator $\boldsymbol{W}(z,k)$. Because the problem is ill-posed, we cannot estimate $\mathbf{H}^{-1}$ directly and the estimates involve the following two ambiguities; $\mathbf{P} \in \mathbf{R}^{n \times n}$, a permutation matrix and $\mathbf{D}(z) \in \mathbf{R}^{n \times n}$, a diagonal matrix with (i,i)th entries $d_i z^{(-\Delta_i)}$ (c_i is a scaling factor and Δ_i is an integer delay value).

3. Learning algorithm of the inverse model

In order to deconvolute these signals the inverse network must adapt without a supervisor which is a rather difficult type of learning. ICA(independent component analysis) of a random vector consists of searching for a linear transformation that minimizes the statistical dependence between its component. The dependency of estimated signals is evaluated by the KL(Kullback-Leibler) distance.

$$KL(p(\boldsymbol{s}); q(\boldsymbol{s})) = \int p(\boldsymbol{s}) \log \frac{p(\boldsymbol{s})}{q(\boldsymbol{s})} d\boldsymbol{s}, \tag{8}$$

where $p(\boldsymbol{s})$ is joint probability density function of estimates and $q(\boldsymbol{s}) = \prod_i (s_i)$ is marginal probability density function. It can be shown that KL information is the expectation of the following cost function.

$$l(\boldsymbol{s}, \boldsymbol{W}) = -\frac{1}{2} \log[\det(\boldsymbol{W} W^T)] - \sum_{i=1}^{m} \log q_i(s_i). \tag{9}$$

We now determine an algorithm for minimizing $l(\boldsymbol{s}, \boldsymbol{W})$ with respect to estimator $\boldsymbol{W}(k)$ by the use of temporal iterations. The stochastic gradient method is a popular learning method in the general optimization framework,

$$\frac{d\boldsymbol{W}_p(k)}{dt} = -\mu(t) \frac{\partial l(\boldsymbol{s}|\boldsymbol{W})}{\partial \boldsymbol{W}_p}, \tag{10}$$

where $\mu(t) > 0$ is the learning rate. It is well known that the stochastic gradient method for parameterized system sometimes suffers from a slow convergence. Additionally the parameter space of estimator $\boldsymbol{W}(k)$ is not Euclidean, but has Riemannian structure in many cases.

Therefore the ordinary gradient does not give the steepest direction of the parameter space. We propose to use the natural gradient approach introduced first by Amari for minimizing the cost function such that

$$\frac{d\boldsymbol{W}_p(k)}{dt} = -\mu(t)\frac{\partial l(\boldsymbol{s}|\boldsymbol{W})}{\partial \boldsymbol{W}_p}\boldsymbol{B}_p \tag{11}$$

where $\boldsymbol{B}_k = \boldsymbol{W}_k^T\boldsymbol{W}_k$ is nonsingular for each t. We then get the following learning formula.

$$\frac{d\boldsymbol{W}_p}{dt} = \mu(t)[\boldsymbol{I}\delta_p - \boldsymbol{f}[\boldsymbol{s}(t)]\boldsymbol{s}_p^T(t)]\boldsymbol{W}_p(t). \tag{12}$$

The optimal choice of $\boldsymbol{f}(\boldsymbol{s}) = [f_1(s_1)\cdots f_n(s_n)]^T$ depends on source signal.

$$f_i(s_i) = -\frac{d\log(q_i(s_i))}{ds_i} = -\frac{q_i'}{q_i}. \tag{13}$$

For sub-Gaussian source, typical choice is $f_i(s_i) = \alpha_i s_i + s_i|s_i|^2$ and we can choose for super-Gaussian source $f_i(s_i) = \alpha_i s_i + \tanh(\beta_i s_i)$, $\alpha \geq 0$ and $\beta \geq 2$.

4. Simulations

We shall show some computer simulation results to check the validity of the algorithm. It can be shown that the deconvolution results using the learning rule which is described by eq.(12) and the effectiveness of natural gradient method.

Example 1. Deconvolution result using by feed-forward model

Let us consider that three source signals were mixed using a uniformly random full rank matrix $\mathbf{H}_p$.

$$x_m(k) = \sum_{i=1}^{100} a_{mi}\sin(2\pi f_i k + \theta_{mi}) \qquad (m = 1, 2, 3). \tag{14}$$

$$\boldsymbol{y}(t) = \mathbf{H}_0\boldsymbol{x}(k) + \mathbf{H}_5\boldsymbol{x}(k-5) + \mathbf{H}_{10}\boldsymbol{x}(k-10) \tag{15}$$

We consider that the frequency band is 0.001-0.018Hz. The spectral density of x_1,x_2 and x_3 are; uniform distribution for x_1, Gaussian distribution($\mu = 0.01, \sigma = 0.003$) for x_2, and Gaussian distribution($\mu = 0.01, \sigma = 0.0003$) for x_3 , respectively. The initial phases θ_{ni} are selected to be uniformly random in the range of $-\frac{\pi}{4} \leq \theta_{mi} \leq \frac{\pi}{4}$ (m=1), $0 \leq \theta_{mi} \leq 2\pi$ (m=2,3). Fig.2 shows the simulation results with $f(s) = \tanh(s)$, $\mu = c$ for $k \geq 2000$, $\mu = c\exp(-0.005t)$ for $k > 2000$, $c = 0.001$. Although the received vector includes some time delay, we can obtain the good reconstruction signals in a few iterations.

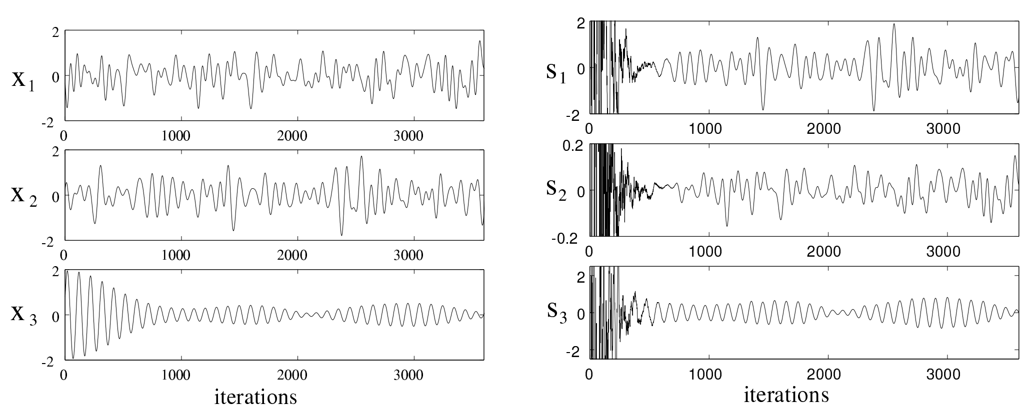

Figure 2: Deconvolution results of the feed-forward inverse model

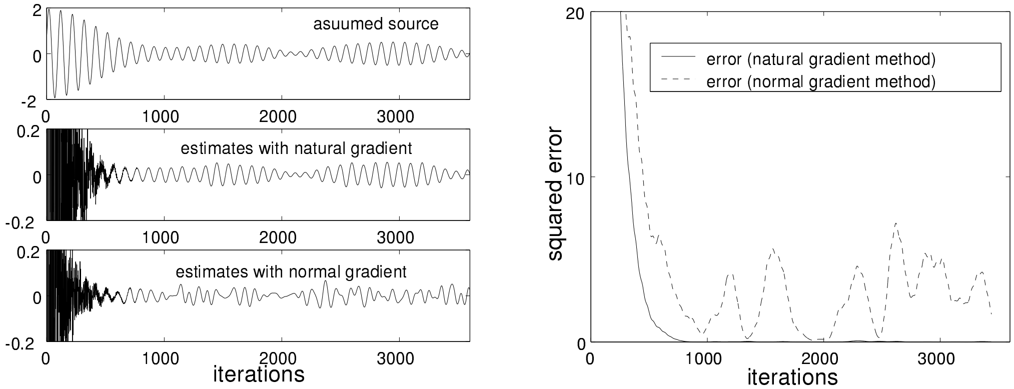

Figure 3: A comparison of reconstruction results by the natural gradient method and normal gradient method

Example 2. Effect of natural gradient method

Now, we make a comparison between with convergence property of normal gradient method and natural gradient method. We will show the two reconstruction results using natural gradient method and gradient method. The received signals are mixed with no delay for simplicity, and the remixed signals are shown in Fig.3. Even if we use the same simulation parameters, the result with natural gradient method is better than that with gradient one. The right panel of Fig.3 illustrates the squared errors of normalized estimates.

5. Conclusion and future work

It is shown that the blind deconvolution algorithm with natural gradient method, gives us some effective results. We have confirmed the validity of natural gradient method and this method can be applied as well to the actually observed data. Although we can obtain good results, there are many problems to be studied. For example, (1) in the case of nonstationary environment, (2)the stability and convergence analysis of learning algorithm, (3) how to evaluate the error, being related with (2).

References

[1] S.I.Amari and A.Cichochi, Adaptive blind signal processing - neural network approaches, Proceeding of the IEEE, vol.86,No.10,October,1998

[2] A.Cichochi, S.I.Amari and J.Cao, Neural network models for blind separation of time delayed and convolved signals, IEICE Trans. Fundamentals, vol.E80-A,No.9,September,1997

[3] P.Comon, Independent component analysis, A new concept?, Signal Processing vol.36,1994

[4] A.Bell and T.J.Sejnowski, An information-maximization approach to blind separation and blind deconvolution,Neural Computation,vol.7,1995

[5] M.Girolami, An alternative perspective on adaptive independent component analysis algorithms, Neural Computation,vol.10,1998

[6] S.I.Amari, Natural gradient works efficiently in learning, Neural Computation,vol.10,1998

[7] S.I.Amari, Differential geometry of a parametric family of invertible linear systems-Riemannian metric,dual affine connections,and divergence, Mathematical System Theory,vol.20,1987

[8] L.Q.Zhang, A.Cichochi and S.I.Amari, Natural gradient Algorithm for blind separation of overdetermined mixture with additive noise,IEEE Signal processing Letters,vol.6,No.11,1999