A blind source separation cascading separation and linearization for low-order nonlinear mixtures

A BLIND SOURCE SEPARATION CASCADING SEPARATION AND LINEARIZATION FOR LOW-ORDER NONLINEAR MIXTURES

Takayuki Nishiwaki *Kenji Nakayama* *Akihiro Hirano*

Dept. of Information Systems Eng., Faculty of Eng., Kanazawa Univ.
2-40-20, Kodatsuno, Kanazawa, 920-8667, Japan
nakayama@t.kanazawa-u.ac.jp

ABSTRACT

A network structure and its learning algorithm have been proposed for blind source separation applied to nonlinear mixtures. Nonlinearity is expressed by low-order polynomials, which are acceptable in many practical applications. A separation block and a linearization block are cascaded. In the separation block, the cross terms are suppressed, and the signal sources are separated in each group, which include its high-order components. The high-order components are further suppressed through the linearization block. A learning algorithm minimizing the mutual information is applied to the separation block. A new learning algorithm is proposed for the linearization block. Simulation results, using 2-channel speech signals, instantaneous mixtures, and 2nd-order post nonlinear functions, show good separation performance.

1. INTRODUCTION

In practical applications of a blind signal source separation (BSS), processes of generating, mixing and sensing signals include nonlinearity, caused by loud speakers, microphones, amplifiers and so on. Statistical independency is not enough to separate the signal sources, some additional prior knowledge are required. Furthermore, since a unique solution is not guaranteed, some regularization techniques are required [6]. Post-nonlinear (PNL) mixtures, in which the signal sources are first linearly mixed, and they are transferred through nonlinear functions. For the PNL mixtures, a mirror structure BSS, in which a nonlinear process and a linear unmixing process are cascaded in this order, has been mainly used [7]. Nonlinear distortion is suppressed in the first stage assuming some prior conditions. Spline nonlinear functions or spline neural networks have been applied to the linearization process [3], [4]. Furthermore, a maximum likelihood estimator has been applied [5]. Also, neural networks have been applied [8].

In this paper, the nonlinearity is limited to low-order polynomial expressions, which are acceptable in many practical applications. A BSS model is proposed, in which a separation block and a linearization block are cascaded in this order. In the first block, signal groups, including high-order components, are separated. In the second block, the high-order components are suppressed. Simulation using 2-channel speech signals and 2nd-order nonlinearity will be shown to confirm usefulness of the proposed method.

2. NOLINEAR MIXTURES

In this paper, the nonlinearity is expressed by polynomials. Thus, the observed signals include the high-order terms of the signal sources and the cross terms among the different signal sources. Letting s_i be the signal sources, and nonlinearity be a 2nd-order function, the observed signal $x_j(n)$ is expressed as

$$x_k = \sum_{i=1}^{n} a_{ki}s_i + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{k,ij}s_is_j \qquad (1)$$

Thus, x_j contains the original s_i, the high-order terms s_i^2, and the cross terms $s_is_j, i \neq j$. Nonlinearity is not limited to post-nonlinearrity, rather it can be included in the processes of generating, transmitting, and sensing signals. Order of nonlinearity is limited to 2nd or 3rd-order. However, in many practical applications, linear processing is a main part, and nonlinearity is parasitic phenomena, which can be approximated by low-order nonlinear functions.

If the signal sources are statistically independent, then $a_is_i + b_is_i^2$ and $a_js_j + b_js_j^2, i \neq j$ are also statistically independent, and can be separated by minimizing the mutual information [1]. The cross term $s_is_j, i \neq j$ has some correlation with both s_i and s_j, then it can be suppressed through the above learning process.

3. SEPARATION BLOCK

3.1. Network Structure

A proposed cascade form BSS is shown in Fig.1. The post-nonlinear (PNL) mixture model is used here [3], [4], [5].

However, the proposed approach is not limited to the PNL mixtures. First, the signal sources s_i are mixed through linear combination resulting in u_j. After that, they are transmitted through nonlinear functions F_k resulting in x_k.

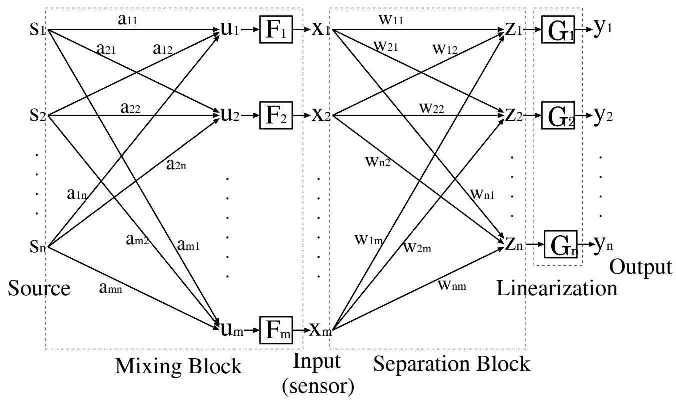

Fig. 1. Network structure of proposed cascade form BSS.

3.2. Number of Sensors

In this model, in order to cancel the cross terms and to separate the signal source groups, the number of the sensors is increased. One example is shown below. Two signal sources are received by four sensors.

$$x_1 = b_{11}s_1 + b_{12}s_2 + b_{13}s_1^2 + b_{14}s_1s_2 + b_{15}s_2^2 \quad (2)$$
$$x_2 = b_{21}s_1 + b_{22}s_2 + b_{23}s_1^2 + b_{24}s_1s_2 + b_{25}s_2^2 \quad (3)$$
$$x_3 = b_{31}s_1 + b_{32}s_2 + b_{33}s_1^2 + b_{34}s_1s_2 + b_{35}s_2^2 \quad (4)$$
$$x_4 = b_{41}s_1 + b_{42}s_2 + b_{43}s_1^2 + b_{44}s_1s_2 + b_{45}s_2^2 \quad (5)$$

x_k are treated as a constant. From these linear equations, the cross term s_is_j can be cancelled, resuting in

$$x'_1 = c_{11}s_1 + c_{12}s_2 + c_{13}s_1^2 + c_{15}s_2^2 \quad (6)$$
$$x'_2 = c_{21}s_1 + c_{22}s_2 + c_{23}s_1^2 + c_{25}s_2^2 \quad (7)$$
$$x'_3 = c_{31}s_1 + c_{32}s_2 + c_{33}s_1^2 + c_{35}s_2^2 \quad (8)$$

Furthermore, s_2^2 is cancelled as,

$$x''_1 = d_{11}s_1 + d_{12}s_2 + d_{13}s_1^2 \quad (9)$$
$$x''_2 = d_{21}s_1 + d_{22}s_2 + d_{23}s_1^2 \quad (10)$$

Furthermore, s_2 can be cancelled, and the s_1 group, which includes s_1 and s_1^2, is separated. At the same time, s_1^2 is cancelled as,

$$x'''_1 = e_{11}s_1 + e_{12}s_2 + e_{15}s_2^2 \quad (11)$$
$$x'''_2 = e_{21}s_1 + e_{22}s_2 + e_{25}s_2^2 \quad (12)$$

Furthermore, s_1 can be cancelled, and the s_2 group, which includes s_2 and s_2^2, is separated.

These processes are equivalent to multiplying a vector $[x_1, x_2, x_3, x_4]^T$ by a 2×4 linear matrix $\boldsymbol{W} = \{w_{lk}\}$.

3.3. Learning Algorithm

In this block, the signal sources are separated based on their statistical independency. Therefore, the conventional learning algorithm, that is likelihood estimation minimizing the mutual information can be applied [1].

$$\boldsymbol{W}(n+1) = \boldsymbol{W}(n) + \eta[\boldsymbol{\Lambda}(t) - \varphi(\boldsymbol{z}(n))\boldsymbol{z}^T(n)]\boldsymbol{W}(n) \quad (13)$$

η is a learning rate, $\boldsymbol{\Lambda}(t)$ is a diagonal matrix, and $\varphi()$ is a nonlinear function [2].

4. LINEARIZATION BLOCK

4.1. Linearization Based on Solving Equations

At the outputs of the separation block, it is assumed that the signal sources are completely separated as follows:

$$z_1 = f_{11}s_1 + f_{12}s_1^2 \quad (14)$$
$$z_2 = f_{21}s_2 + f_{22}s_2^2 \quad (15)$$

Since z_1 and z_2 include only s_1 and s_2, respectively, they can be linearized through the following nonlinear functions.

$$y_1 = G_1(z_1) = \frac{-f_{11} \pm \sqrt{f_{11}^2 + 4f_{12}z_1}}{2f_{12}} \quad (16)$$
$$y_2 = G_2(z_2) = \frac{-f_{21} \pm \sqrt{f_{21}^2 + 4f_{22}z_2}}{2f_{22}} \quad (17)$$

Finaly, the separated and linearized signal sources are obtained.

$$y_1 = g_1s_1 \quad (18)$$
$$y_2 = g_2s_2 \quad (19)$$

4.2. Learning Algorithm

Transformations in the linearization block are given by Eqs.(16) and (17). However, in real applications, the coefficients f_{ij} are not known. So, they should be adjusted through an iterative method. Equations (16) and (17) can be expressed by using two parameters as follows:

$$y_i(n) = -\frac{\alpha_i}{2} \pm \sqrt{\frac{\alpha_i^2}{4} + \frac{z_i(n)}{\beta_i}} \quad (20)$$
$$\alpha_i = \frac{f_{i1}}{b_{i2}}, \qquad \beta_i = \frac{1}{f_{i2}} \quad (21)$$

α_i and β_i are adjusted through an iterative method.

Error Function:
In this paper, 2nd-order nonlinearity is assumed. Thus, after the linear source separation, the outputs include 1st-order and 2nd-order terms of the signal sources. Furthermore, if

we take speech and music signals into account, their average is almost zero. Therefore, the output average can be used as a cost function.

$$E_i(n) = \frac{1}{M}\sum_{l=0}^{M-1} y_i(n-l) \tag{22}$$

The gradient descent algorithm is used for adjusting the parameters.

$$\alpha_i(n) = \alpha_i(n-1) - \eta\frac{\partial E_i(n)}{\partial \alpha_i(n)} \tag{23}$$

$$\beta_i(n) = \beta_i(n-1) - \eta\frac{\partial E_i(n)}{\partial \beta_i(n)} \tag{24}$$

$$\begin{aligned}\frac{\partial E_i(n)}{\partial \alpha_i(n)} &= \frac{1}{M}\sum_{l=0}^{M-1}\frac{\partial y_i(n-l)}{\partial \alpha_i(n)} \\ &= \frac{1}{M}\sum_{l=0}^{M-1}(-\frac{1}{2} \pm \frac{\alpha_i(n)}{4}(\frac{\alpha_i^2(n)}{4} \\ &\quad +\frac{1}{\beta_i(n)}z(n-l))^{-\frac{1}{2}})\end{aligned} \tag{25}$$

$$\begin{aligned}\frac{\partial E_i(n)}{\partial \beta_i(n)} &= \frac{1}{M}\sum_{i=0}^{M-1}\frac{\partial y_i(n-l)}{\partial \beta_i(n)} \\ &= \frac{1}{M}\sum_{l=0}^{M-1}(\mp\frac{z(n-l)}{2\beta_i^2}(\frac{\alpha_i(n)^2}{4} \\ &\quad +\frac{1}{\beta_i(n)}z(n-l))^{-\frac{1}{2}})\end{aligned} \tag{26}$$

Porality Control:
In the above update equations, there is a freedom of polarity. It should be judged which polarity should be used. For this purpose, the following conditions are introduced. These conditions do not lose generality in real applications.

1. A linear component is greater than a nonlinear component.
2. The signal source level is limited. Say, for instance $|s_i(n)| < 1$.

Under these conditions, in the linear separation output,

$$z_i(n) = f_{i1}s_i(n) + f_{i2}s_i^2 \tag{27}$$

the following inequality is always held.

$$|f_{i1}s_i(n)| > |f_{i2}s_i^2(n)| \tag{28}$$

This means the porality of $z_i(n)$ is equal to that of $f_{i1}s_i(n)$. So, except for the polarity of f_{i1}, that of the output $y_i(n)$ can be controlled so as to be the same as that of $z_i(n)$. The polarity of f_{i1} does not affect separation performance. Because in blind source separation, constant scaling inherently remains.

5. SIMULATIONS AND DISCUSSIONS

5.1. Simulation Conditions

Two signal sources and four observations are used. The signal sources are male speech signals. The mixing matrix is

$$A = \begin{bmatrix} 1 & -2 \\ -3 & 2 \\ 2 & -1 \\ 1 & 2 \end{bmatrix}$$

The learning rate is $\eta = 0.001$. The nonlinear functions in the mixing block are

$$\begin{aligned} F_1(u_1) &= u_1 + 0.4u_1^2 \\ F_2(u_2) &= u_2 + 0.2u_2^2 \\ F_3(u_3) &= u_3 - 0.6u_3^2 \\ F_4(u_4) &= u_4 + 0.3u_4^2 \end{aligned}$$

5.2. Separation Block

Separation and linearization performances are evaluated based on the following SNR. Assuming s_i is dominant in z_j, and letting σ_{s1}^2 and σ_{n1}^2 be the power of s_i and s_i^2 in z_j, and the power of z_j except for s_i and s_i^2, furthermore, σ_{s2}^2 and σ_{n2}^2 be the power of s_i in z_j, and the power of z_j except for s_i, SNR is defined by

$$SNR_i = 10\log_{10}\frac{\sigma_{si}^2}{\sigma_{ni}^2} \tag{29}$$

SNR_1 includes the high-order components, which are not suppressed in the separation block. The learning curve is shown in Fig.2. The vertical axis indicates SNR_1 in dB, the horizontal axis is iteration number. In this figure, the solid line (2) indicates the learning curve, and the dashed line (1) shows SNR_1 obtained by using the trained w_{ij} for reference. From these curves, the training converges at 17,000 iterations.

5.3. Linearization Block

Error valuation
In this process, separation performance is evaluated by SNR_2 defined by Eq.(29). However, the same formula cannot be used. The s_i component and the other components are discriminated as follows:

$z_i(n)$ is linearized through

$$y_i(n) = -\frac{\alpha_i}{2} + \sqrt{\frac{\alpha_i^2}{4} + \frac{z_i(n)}{\beta_i}} \tag{30}$$

Let

$$\sqrt{\frac{\alpha_i^2}{4} + \frac{z_i(n)}{\beta_i}} = \sqrt{a_i s_i^2(n) + b_i(n)s_i(n) + c_i(n)} \tag{31}$$

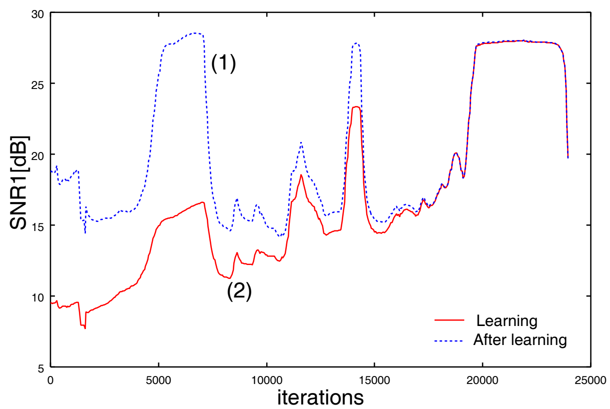

Fig. 2. Learning curve for separation block.

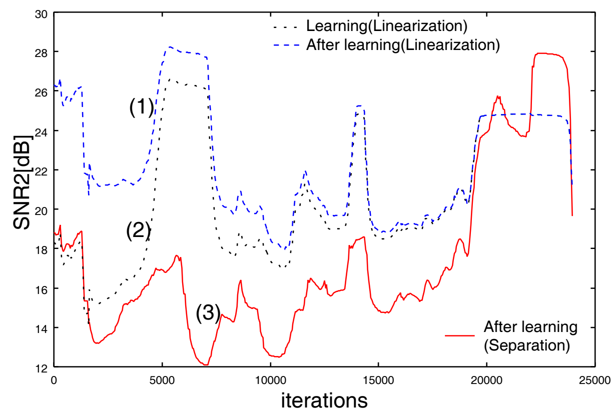

Fig. 3. Learning curve for linearization block.

Furthermore,

$$\sqrt{a_i s_i^2(n) + b_i s_i(n) + c_i(n)} = d_i s_i(n) + e_i(n) \quad (32)$$

$$a_i s_i^2(n) + b_i s_i(n) + c_i(n)$$
$$= d_i^2 s_i^2(n) + 2 d_i s_i(n) e_i(n) + e_i(n)^2 \quad (33)$$

Comparing the coefficients, the following relations are obtained.

$$d_i^2 = a_i \quad (34)$$
$$2 d_i e_i(n) = b_i(n) \quad (35)$$
$$e_i^2(n) = c_i(n) \quad (36)$$

a_i and $c_i(n)$ are calculated using α_i, β_i and $z_i(n)$ at each iteraton. SNR_2 is calculated by

$$SNR_2 = 10 \log \frac{p(n)}{q(n)} \quad (37)$$

$$p(n) = \frac{1}{M} \sum_{i=0}^{M-1} (y_i(n) + \frac{\alpha_i}{2} - e_i(n))^2 \quad (38)$$

$$q(n) = \frac{1}{M} \sum_{i=0}^{M-1} (-\frac{\alpha_i}{2} + e_i(n))^2 \quad (39)$$

The learning curve of SNR_2, defined by Eq.(37), is shown in Fig.3 with a dotted line (2). In this figure, SNR_2 after the separation, defined by Eq.(29), and after the linearinzation, defined by Eq.(37), are shown with a solid line (3) and a dashed line (1), respectively. By linearinzation, SNR_2 can be improved by 6 dB. Approximately, $SNR_2 = 20$ dB is achieved, which is good signal source separation performance.

6. CONCLUSIONS

In this paper, a blind source separation method has been proposed for instantaneous nonlinear mixtures. It consists of the separation block and the linearization block in a cascade form. Both blocks are separately trained. The conventional learning algorithm and the new learning algorithm have been proposed for both blocks, respectively. Simulation, using two speech signals and 2nd-order nonlinearity, shows usefulness of the proposed method.

7. REFERENCES

[1] S.Amari, T.Chen and A.Cichocki, "Stability analysis of learning algorithms for blind source separation", Neural Networks, vol.10, no.8, pp.1345-1351, 1997.

[2] K.Nakayama, A.Hirano and T.Sakai, "An adaptive nonlinear function controlled by estimated output PDF for blind source separation", Proc. ICA'2003, Nara, pp.427-432, April 2003.

[3] M.Solazzi, F.Piazza and A.Uncini, "Nonlinear blind source separation by spline neural networks", IEEE Proc. ICASSP'2001, Salt Lake City, MULT-P3.4, May 2001.

[4] F.Milani, M.Solazzi and A.Uncini, "Blind source separation of convolutive nonlinear mixtures by flexible spline nonlinear functions", IEEE Proc. ICASSP'2002, Orlando, pp.1641-1644, May 2002.

[5] A.Koutras,"Blind separation of non-linear convoluved speech mixtures", IEEE Proc. ICASSP'2002, Orlando, pp.913-916, May 2002.

[6] C.Jutten and J.Karhunen, "Advance in nonlinear blind source separation", ICA03, Nara, pp.245-256, April 2003.

[7] A.Ziehe, M.Kawanabe, S.Harmeling and K.R.Muller, "Blind separation of post-nonlinear mixtures using Gaussianizing transformations and temporal decorrelation", Proc. ICA'03, Nara, pp.269-274, April 2003.

[8] R.M.Clemente, S.H.Mellado, J.I.Acha, F.Rojas and C.G.Puntone., "MLP-based source separation for MLP-like nonlinear mixtures", Proc. ICA'03, Nara, pp.155-160, April 2003.