A LEARNING ALGORITHM WITH DISTORTION FREE CONSTRAINT AND COMPARATIVE STUDY FOR FEEDFORWARD AND FEEDBACK BSS

Akihide HORITA, Kenji NAKAYAMA, Akihiro HIRANO, and Yasuhiro DEJIMA

Graduate School of Natural Science and Technology, Kanazawa Univ.
Kakuma-machi, Kanazawa, 920-1192, JAPAN
phone: + (81) (0)76-234-4896, fax: + (81) (0)76-234-1900,
email: horita@leo.ec.t.kanazawa-u.ac.jp, nakayama@t.kanazawa-u.ac.jp

ABSTRACT
Source separation and signal distortion are theoretically analyzed for the FF-BSS systems implemented in both the time and frequency domains and the FB-BSS system. The FF-BSS systems have some degree of freedom, and some source signal distortion. The FB-BSS has a unique solution for complete separation and signal distortion free. Next, the condition for complete separation and signal distortion free is derived for the FF-BSS systems. This condition is applied to the learning algorithms. Computer simulations by using speech signals are carried out for the conventional methods and the new learning algorithms employing the proposed distortion free constraint. The proposed method can drastically suppress signal distortion, while maintaining high separation performance. The FB-BSS system also demonstrates good performances. The FF-BSS systems and the FB-BSS system are compared based on the transmission time difference in the mixing process. Location of the signal sources and the sensors are rather limited in the FF-BSS system.

1. INTRODUCTION
Signal processing, including noise cancellation, echo cancellation, equalization of transmission lines, estimation and restoration of signals have become a very important research area. In some cases, we do not have enough information about signals and their interference. Furthermore, their mixing and transmission processes are not well known in advance. In these kind of situations, blind source separation (BSS) technology using statistical properties of signal sources have become very important [1], [2].

Since, in many applications mixing processes are convolutive mixtures, several methods in the time domain and the frequency domain have been proposed. Two kinds of proposed network structures are feedforward (FF) and feedback (FB) structures. Separation performance is highly dependent on the signal sources and the transfer functions in the mixture [4], [7], [9], [10].

BSS learning algorithms make the output signals to be statistically independent. This direction cannot always guarantee distortion free separation. A method in which the distance between the observed signals and the separated signals is added to the cost function has been proposed[5]. However, since the observations include many kinds of signal sources, it is difficult to suppress signal distortion. Furthermore, even though signal distortion in BSS systems is an important problem, it has not been addressed well up to now. Therefore, we have discussed an evaluation measure of signal distortion. Conditions for source separation and signal distortion free have been derived. Based on these conditions, convergence properties have been analyzed[11].

In this paper, a new learning algorithm with distortion free constraint for the FF-BSS systems is proposed. The proposed method is compared to the conventional FF-BSS and FB-BSS systems through computer simulations. Although the FB-BSS system has good performance in both source separation and distortion free, it has some limitation depending on transmission time difference. This limitation is also analyzed.

2. FF-BSS SYSTEMS FOR CONVOLUTIVE MIXTURE

2.1 Network Structure and Equations
For simplicity, 2 signal sources and 2 sensors are used. A block diagram is shown in Fig.1. The observations and the output signals are given by:

\[ x_j(n) = \sum_{i=1}^{2} \sum_{t=0}^{K_j-1} h_{ji}(t) s_i(n-t), j = 1, 2 \] (1)

\[ y_k(n) = \sum_{j=1}^{2} \sum_{t=0}^{K_k-1} w_{kj}(t) x_j(n-t), k = 1, 2 \] (2)

2.2 Learning Algorithm in Time Domain
The learning algorithm is derived following a natural gradient algorithm using a mutual information as a cost function [3].

\[ w_{kj}(n+1, l) = w_{kj}(n, l) + \eta \left[ \varphi(y_k(n)) y_k(n-l+q) w_{kj}(n, q) \right] \] (3)

\[ \varphi(y_k(n)) = \frac{1 - e^{-y_k(n)}}{1 + e^{-y_k(n)}} \] (4)

The learning rate is represented by \( \eta \).

2.3 Learning Algorithm in Frequency Domain
Filter coefficients in the separation block are trained according [3], [6], [8],

\[ W(r+1, m) = W(r, m) + \eta \left( \Phi(Y(r, m)) Y(r, m) \right) W(r, m) \] (5)

\[ \Phi(Y(r, m)) = \frac{1}{1 + e^{-Y^2(r, m)}} + \frac{j}{1 + e^{-Y^2(r, m)}} \] (6)
make the output signals to be statistically independent. Estimation of the mixing process is not taken into account. Especially, in convolutive mixtures, the output signals are not guaranteed to approach to the sources. Therefore, the signal sources observed at the sensors are taken into account as a criterion for the signal distortion [2].

The signal distortion is evaluated by several measures, based on the signals, the transfer functions and their amplitude responses. The transfer functions from the j-th source to the k-th output Akj(\(e^{j\omega}\)) is compared to that from the i-th source to the j-th sensor \(H_j(e^{j\omega})\). The output signals \(A_{ki}(e^{j\omega})S_i(e^{j\omega})\) is compared to the observed signal \(H_j(e^{j\omega})S_i(e^{j\omega})\). Four kinds of measures are shown below.

\[
\sigma_{dwa} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H_j(e^{j\omega})S_i(e^{j\omega}) - A_{ki}(e^{j\omega})S_i(e^{j\omega})|^2 d\omega
\]  
\[
\sigma_{dwb} = \frac{1}{2\pi} \int_{-\pi}^{\pi} (|H_j(e^{j\omega})|^2 - |A_{ki}(e^{j\omega})|^2) d\omega
\]  
\[
\sigma_1 = \frac{1}{\pi} \int_{-\pi}^{\pi} |H_j(e^{j\omega})|^2 d\omega
\]  
\[
SD_{1x} = 10\log_{10} \frac{\sigma_{dwa}}{\sigma_1}, x = a, b
\]  
\[
\sigma_{dxx} = \frac{1}{2\pi} \int_{-\pi}^{\pi} |H_j(e^{j\omega})|^2 d\omega
\]  
\[
SD_{2x} = 10\log_{10} \frac{\sigma_{dxx}}{\sigma_2}, x = a, b
\]

Since FF-BSS systems cannot control the output signal level, the output signal level might differ from the criteria. In order to neglect this scaling effect in the calculation of \(SD_{1x}\) and \(SD_{2x}\), average power of \(H_j(z)S_i(z)\), \(A_{ki}(z)S_i(z)\), \(H_j(z)\), and \(A_{ki}(z)\) are normalized.

5. SOURCE SEPARATION AND SIGNAL DISTORTION IN FF-BSS SYSTEMS

5.1 Source Separation and Signal Distortion

For simplicity, an FF-BSS system with 2-sources and 2-sensors, shown in Fig.1, is used. Furthermore, \(S_i(z)\) is assumed to be separated at the output \(Y_i(z)\). This does not lose generality. Taking the signal distortion criterion into account, the condition for distortion-free source separation can be expressed as follows:

\[
W_{11}(z)H_{11}(z) + W_{12}(z)H_{21}(z) = H_{11}(z)
\]  
\[
W_{11}(z)H_{12}(z) + W_{12}(z)H_{22}(z) = 0
\]  
\[
W_{21}(z)H_{11}(z) + W_{22}(z)H_{21}(z) = 0
\]  
\[
W_{21}(z)H_{12}(z) + W_{22}(z)H_{22}(z) = H_{22}(z)
\]

The above equations imply two conditions. First, complete source separation, that is the non-diagonal elements are all zero, as shown in Eqs.(20) and (21). Secondly, signal distortion free, that is the diagonal elements are \(H_i(z)\) as shown in Eqs.(19) and (22).

The conventional learning algorithm given by Eqs.(3)-(7) satisfies only Eqs.(20) and (21). Equations (19) and (22) are not satisfied. Therefore, signal distortion free is not guaranteed in general.
5.2 Signal Distortion in FF-BSS Systems Trained in Frequency Domain

In the frequency domain, there is a weighting effect. From Eqs. (5), (7), it follows that the correction of the weights are proportional to the product of the outputs $Y_i(r,m)Y_j(r,m)$. In the early stage of the learning process, the output $Y_i(r,m)$ includes both $S_i(r,m)$ and $S_j(r,m)$. Therefore, the correction of the weights are proportional to $S_i(r,m) + S_j(r,m)$. On the other hand, as the learning makes progress, the output $Y_i(r,m)$ includes mainly $S_i(r,m)$. Therefore, the correction of the weights are proportional to $S_i(r,m) 	imes S_j(r,m)$. If the signal sources are all speech, their spectra are similar to each other. In this case, the spectra of $S_i(r,m) + S_j(r,m)$ and $S_i(r,m) 	imes S_j(r,m)$ are similar to those of the signal sources. If the signal sources locate in different frequency bands, e.g. in music, their spectra are not similar. In this case, $S_i(r,m) + S_j(r,m)$ and $S_i(r,m) 	imes S_j(r,m)$ have many peaks. Therefore, the learning process amplifies some part of the spectrum of the signal source, which is not dominant. It can be expected that weighting will cause signal distortion, and makes source separation more difficult.

5.3 Learning Algorithm in Time Domain Suppressing Signal Distortion by Using Observation as Criteria

A learning algorithm for reducing signal distortion has been proposed [5]. The cost function includes the distance between the observed signals and the output signals. This means that the output signals are forced to approach to the observed signals. The update equation is given by

$$w(n+1,l) = w(n,l) + \alpha \sum_{m=0}^{K} \left[ |I(n-m) - \langle F(y(n)) \rangle y^T(n-l+m) | - \beta |y(n) - \Sigma(n)| y^T(n-l+m) \right] w(n,m)$$

$$\varphi(y(n)) = \frac{1-e^{-|y(n)|}}{1+e^{-|y(n)|}}$$

In this method, the output signals $Y_i(z) = A_i(z)S_i(z) + A_j(z)S_j(z)$ tend to approach to the observed signals $X_i(z) = H_{i1}(z)S_i(z) + H_{i2}(z)S_j(z)$. When $S_i(z)$ and $S_j(z)$ are statistically independent, $A_i(z)$ and $A_j(z)$ are able to approach to $H_{i1}(z)$ and $H_{i2}(z)$, respectively. The former guarantees distortion free, however, the latter avoids source separation.

6. SOURCE SEPARATION AND SIGNAL DISTORTION IN FF-BSS SYSTEMS

There are two cases, in which possible solutions for perfect separation exist, as shown below:

$$C_{21}(z) = \frac{H_{21}(z)}{H_{11}(z)} C_{12}(z) = \frac{H_{12}(z)}{H_{22}(z)}$$

$$C_{21}(z) = \frac{H_{21}(z)}{H_{11}(z)} C_{12}(z) = \frac{H_{12}(z)}{H_{22}(z)}$$

It is assumed that the delay times of $H_{ij}(z)$ are shorter than those of $H_{ij}(z)$. This means that in Fig. 2, the sensor of $X_i$ is located close to $S_i$. From this assumption, the solutions in case (1) become causal systems. On the other hand, the solutions in case (2) are noncausal. When $C_{ij}(z)$ satisfy the separation conditions Eqs. (25), the output signals are given by

$$Y_i(z) = H_{11}(z)S_i(z)$$

$$Y_j(z) = H_{22}(z)S_j(z)$$

They are exactly the same as the criteria of the signal distortion discussed in Sec. 4. Therefore, the FF-BSS systems have a unique solution, which satisfies both source separation as well as the signal distortion free simultaneously. Thus, in the FF-BSS systems, if complete signal separation is achieved, signal distortion free is also automatically satisfied.

7. DISTORTION FREE CONDITION AND ITS APPLICATION TO LEARNING ALGORITHM FOR FF-BSS SYSTEMS

7.1 FF-BSS Systems Trained in Frequency Domain

From the relations of Eqs. (20) and (21), $H_{ij}(z)$ are expressed.

$$H_{12}(z) = -\frac{W_{12}(z)}{W_{11}(z)} H_{22}(z)$$

$$H_{21}(z) = -\frac{W_{21}(z)}{W_{22}(z)} H_{11}(z)$$

By substituting the above equations into Eqs. (19) and (22), $H_{ij}(z)$ can be removed, and the following equations consisting only of $W_{kj}(z)$ can be obtained.

$$W_{11}(z)W_{22}(z) - W_{12}(z)W_{21}(z) = W_{22}(z)$$

$$W_{11}(z)W_{22}(z) - W_{12}(z)W_{21}(z) = W_{11}(z)$$

From these equations, $W_{11}(z) = W_{22}(z)$ is derived. Therefore, the above equations result in

$$W_{22}(z) = \frac{1}{2} \sqrt{1 + 4W_{12}(z)W_{21}(z)}$$

This 2nd-order equation expresses the condition for both complete source separation and signal distortion free. This equation is solved for $W_{11}(z)$ and $W_{22}(z)$ as follows:

$$W_{22}(z) = \frac{1}{2} \sqrt{1 + 4W_{12}(z)W_{21}(z)}$$

This constraint can be included in the learning processes for the FF-BSS systems in the frequency domain. Eqs. (5) and (7). Furthermore, $W_{ij}(z)$, which are determined from Eq. (34) are taken into account to some extent, as shown in the following.

$$W_{ij}(r+1,m) = (1-a)W(r+1,m)$$

$$+ \frac{1}{2} \sqrt{1 + 4W_{12}(z)W_{21}(z)}$$

7.2 Learning Algorithm with Constraint in Time Domain

In this section, a new learning algorithm for FF-BSS systems, trained in the time domain, is proposed. The constraint given by Eq. (34) is taken into account in the learning process. Equation (34) is rewritten as follows:

$$W_{ij}(z) - 1 = 4W_{12}(z)W_{21}(z)$$

This constraint is used in the learning process as follows: Given $W_{12}(z)$ and $W_{21}(z)$, $W_{ij}(z)$ are obtained so as to approximate the relation of Eq. (36).

The condition for the distortion free source separation holds complete separation and signal distortion free. However, the learning of the separation block starts from an initial guess. Therefore, in the early stage of the learning process, the signal sources are not well separated. Taking this situation into account, the constraint of Eq. (36) is gradually
imposed as the learning process progresses. The following learning algorithm is proposed.

\[ w_{kj}(n+1, l) = w_{kj}(n) + \eta \{ w_{kj}(n) - \sum_{o=0}^{\kappa_k} - \sum_{p=1}^{L} \phi(y_k(n)) y_p(n - o + p) w_{kp}(n, o) \} \]

\[ w_{jj}(N + 1, l) = (1 - \alpha) w_{jj}(n + 1, l) + \alpha \hat{w}_{jj}(n + 1) \]

\[ \hat{w}_{jj}(n + 1) \]

is determined so as to approximate the relation of Eq.(36). \( \alpha \) is set to a small positive number.

8. SIMULATION AND DISCUSSION

8.1 Learning Methods and Their Abbreviation

In this paper, many kinds of learning methods will be compared. They are summarized in Table 1.

8.2 Simulation Conditions

Two sources and two sensors are used. The transfer function of the cross paths are related to the direct paths as \( H_{ij}(z) = 0.9z^{-1} H_{ij}(z) \). Speeches are used as sources. The FFT size is 256 points in the frequency domain training. FIR filters with 256 taps are used in the FF-BSS system, trained in the time domain and in the FB-BSS system. The initial guess of the separation block are \( W_{11}(z) = W_{22}(z) = 1 \) and \( W_{ij}(z) = 0, k \neq j \), in the FF-BSS system, and \( C_{1}(z) = C_{21}(z) = 0 \) in the FB-BSS system.

Source separation is evaluated by the following two signal-to-interference ratios \( SIR_1 \) and \( SIR_2 \). Here, \( S_1(z) \) and \( S_2(z) \) are assumed to be separated in \( Y_1(z) \) and \( Y_2(z) \), respectively. However, it does not lose generality:

\[ \sigma_{s1} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( |A_{11}(e^{j\omega})| S_1(e^{j\omega})^2 \right) d\omega \]

\[ \sigma_{s2} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( |A_{12}(e^{j\omega})| S_2(e^{j\omega})^2 \right) d\omega \]

\[ SIR_1 = \text{log}_{10} \frac{\sigma_{s1}}{\sigma_{s1}} \]

\[ SIR_2 = \text{log}_{10} \frac{\sigma_{s2}}{\sigma_{s2}} \]

8.3 BSS Performance for Speech Signals

Evaluation measures are summarized in Table 2. We set \( i = j = k \) in Eqs.(11) to Eq.(18) with respect to signal distortion evaluations, because \( S_1(z) \) and \( S_2(z) \) are assumed to be separated in \( Y_1(z) \) and \( Y_2(z) \), respectively.

For FF-BSS time(MDP) has the worst performance in signal distortion \( SD_{le} \). FF-BSS time(DF) can improve signal distortion. However, as discussed in Sec.5.3, due to the residual cross terms, the signal to interference ratio \( SIR_i \) is not good. FF-BSS freq(1) and also \( SIR_i \). Compared to FF-BSS time, \( SIR_i \) is slightly reduced. However, FF-BSS time gains \( SIR_i \) by signal distortion. This means the frequency component in mainly the high frequency band is amplified, and signal distorted, at the same time, the signal power is increased resulting in relatively high \( SIR_i \).

Regarding the frequency domain implementation, FF-BSS freq(2) is not good in signal distortion due to the identity matrix in Eq.(5), which causes whitening as discussed in Sec.2.3. On the other hand, FF-BSS freq(2) can improve signal distortion due to the weighting effects on the filter coefficient update as discussed in Sec.5.2. However, by using the proposed distortion free constraint, FF-BSS freq(1+DF) and FF-BSS freq(2+DF) can improve more from each counterpart. \( SIR_i \) of FF-BSS freq(2) and FF-BSS freq(2+DF) are almost the same. However, \( SIR_i \) with the lower \( SD_{le} \) has the higher accuracy.

On the contrary, FB-BSS demonstrates good performances in \( SIR_i \) and \( SD_{le} \) due to a unique solution for complete separation and signal distortion free, as discussed in Sec.6. However, its performances are slightly lower than those of FF-BSS freq(2+DF).

The residual cross terms are more investigated here. As discussed in Sec.5.3, in the learning process of FF-BSS time(MDP), \( A_{jj}(z) \) and \( A_{ij}(z) \) tend to approach to \( H_{ij}(z) \) and \( H_{ji}(z) \), respectively. The former is evaluated in \( SD_{le} \) and the latter is evaluated in \( SIR_i \). Here, difference between \( H_{ij}(z) \) and \( A_{ij}(z) \) is evaluated in detail. Table 3 shows the numerical data, where \( SD_{le} \) are calculated from Eqs.(11) through (18), setting the indices to be \( i \neq j \). Small values mean they are similar to each other. Since FF-BSS time(MDF) has the smallest values, especially in \( SD_{le} \), that is in the frequency band, where the signal components are mainly located, then the theoretical discussion in Sec.5.3 can be supported.

8.4 Comparison between FF-BSS and FB-BSS

From the above comparison in \( SIR_i \) and \( SD_{le} \), the FB-BSS system can always provide good performance compared to
Table 3: Difference between $H_{ij}(z)$ and $A_{ij}(z)$ for eight kinds of BSS systems for speech signals.

<table>
<thead>
<tr>
<th>Methods</th>
<th>$SD_{1w}$</th>
<th>$SD_{1b}$</th>
<th>$SD_{2w}$</th>
<th>$SD_{2b}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF-BSS time</td>
<td>2.44</td>
<td>-0.25</td>
<td>2.03</td>
<td>-0.63</td>
</tr>
<tr>
<td>FF-BSS time (DF)</td>
<td>-1.14</td>
<td>-5.00</td>
<td>5.83</td>
<td>-8.12</td>
</tr>
<tr>
<td>FF-BSS time (MDP)</td>
<td>-0.36</td>
<td>-10.3</td>
<td>8.53</td>
<td>-9.39</td>
</tr>
<tr>
<td>FF-BSS freq (1)</td>
<td>3.89</td>
<td>-4.87</td>
<td>-0.42</td>
<td>-0.55</td>
</tr>
<tr>
<td>FF-BSS freq. (1+DF)</td>
<td>-0.21</td>
<td>6.51</td>
<td>-4.08</td>
<td>-5.11</td>
</tr>
<tr>
<td>FF-BSS freq (2)</td>
<td>3.89</td>
<td>-5.57</td>
<td>-0.31</td>
<td>-1.99</td>
</tr>
<tr>
<td>FF-BSS freq. (2+DF)</td>
<td>3.99</td>
<td>-5.39</td>
<td>-0.03</td>
<td>-0.76</td>
</tr>
<tr>
<td>FB-BSS</td>
<td>0.38</td>
<td>-2.70</td>
<td>-0.59</td>
<td>-4.67</td>
</tr>
</tbody>
</table>

Figure 3: Effects of transmission time difference on $SIR_1$ in FF-BSS time (DF) and FB-BSS.

In this paper, source separation and signal distortion have been theoretically analyzed for the FF-BSS systems implemented in both the time and frequency domains and the FB-BSS system. The FF-BSS systems have some degree of freedom, and cause some signal distortion. The FB-BSS has a unique solution for complete separation and signal distortion free. Next, the condition for complete separation and signal distortion free has been derived for the FF-BSS systems. This condition has been applied to the learning algorithms. Computer simulations by using speech signals have been carried out for the conventional methods and the new learning algorithms employing the proposed distortion free constraint. The proposed method can drastically suppress signal distortion, while maintaining high separation performance. The FF-BSS system also has demonstrated good performances. The FF-BSS systems and the FB-BSS system have been compared based on the transmission time difference in the mixing process. As a result, location of the signal sources and the sensors are rather limited in the FB-BSS system.

10. CONCLUSIONS

REFERENCES