Mem. Fac. Sci. Eng. Shimane Univ.
Series A
34, pp. 115-124 (2000)

A Method of Real-world Blind Separation Implemented in Time-frequency Domain

Mitsuru KAWAMOTO[1], Kiyotoshi MATSUOKA[2], Noboru OHNISHI[3]

1. Department of Electronic and Control Systems Engineering, Shimane University
2. Department of Control Engineering, Kyushu Institute of Technology
3. Graduate School of Engineering, Nagoya University

Abstract

In this paper, we propose a new method of real-world blind separation for non-stationary signals (e.g., speech signal, music). The proposed method is implemented in the time-frequency domain and achieves real-world blind separation by minimizing a cost function with the algorithm derived using the natural gradient under a given condition. We carried out computer simulations to demonstrate the validity of our proposed method. Moreover, we apply our method to an experiment which extracts two source signals from their mixtures observed in a normal room with computer and air conditioning noises. The results show that our method can achieve real-world blind separation.

1. Introduction

This paper deals with real-world blind separation which extracts original signals from their mixtures observed by the sensors in a real world. To solve the real-world blind separation, many methods have been proposed until now [e.g., [1]-[8]]. The methods have been implemented in time domain or frequency domain.

In the time domain, the following equation has been used as a transfer function between the sources and sensors.

$$\boldsymbol{x}(t)=\bar{A}(z)\boldsymbol{s}(t) \tag{1}$$

where $\boldsymbol{x}(t)=[x_1(t), \ldots, x_N(t)]^T$ and $\boldsymbol{s}(t)=[s_1(t), \ldots, s_N(t)]^T$. $\bar{A}(z)$ is a matrix which has elements $\bar{a}_{ij}(z)$ $(i, j=1, \ldots, N)$:

$$\bar{a}_{ij}(z)=\sum_{k=-\infty}^{\infty} a_{ij}(k)z^{-k} \quad (i,j=1, \cdots, N) \tag{2}$$

where z^{-k} is a delay operator, i.e., $s_i(t)z^{-k}=s_i(t-k)$. Using a method which can separate $\boldsymbol{s}(t)$ from $\boldsymbol{x}(t)$, real-world blind separation has been solved [e.g., [2]].

On the other hand, in the frequency domain, discrete Fourier transform is applied to $\boldsymbol{x}(t)$:

$$x(\omega, t) = \sum_{t'=t}^{t+M} x(t') \exp(-j\omega t') \tag{3}$$

where $\omega = 2\pi k/M$, $k=0, 1, \ldots, M-1$, and M denotes the window size of the discrete Fourier transform. Using (3), (1) can be written as:

$$x(\omega, t) = A(\omega) s(\omega, t) \tag{4}$$

where $A(\omega)$ can be approximated by applying the discrete Fourier transform to $\bar{A}(z)$, if the window size M is sufficiently long compared to the impulse response of $\bar{A}(z)$, and $s(\omega, t)$ is the discrete Fourier transform of $s(t)$. The separation is implemented at each frequency. Therefore, the treatment of the separation in the frequency domain is easier than that in the time domain. The method of separating $s(\omega, t)$ from $x(\omega, t)$ has been applied to real-world blind separation [e.g., [4]].

In this paper, we deal with real-world blind separation implemented in the frequency (time-frequency) domain. Several methods have been proposed [e.g., [3][4][5]]. An attractive feature of our proposed method, which differs from the previously reported ones, is that non-minimum phase systems can be treated and the signals just before entering the sensors, that is, for example, $\bar{a}_{ij}(z)$ $(i, j=1, \ldots, N)$, can be acquired. The second feature of our method is that it offers good advantage in some applications, for example, speech recognition.

To determine $x(\omega, t)$, we use the following equation:

$$x(\omega, t_w) = \sum_{t'=t_w}^{t_w+M} x(t') \exp(-j\omega t') h(t'-t_w) \tag{5}$$

where $h(t)$ is the Hamming window, and $t_w = 0, \Delta T, 2\Delta T, \ldots$ denotes the window position, and ΔT is the shifting interval of the moving window. This transformation is called the windowed-Fourier transform. Then, (4) can be written as

$$x(\omega, t_w) = A(\omega) s(\omega, t_w) \tag{6}$$

The problem is to separate sources $s_i(\omega, t_w)$ from their mixtures $x(\omega, t_w)$ at each frequency. Our method overcomes this by modifying the parameters of a network such that a cost function takes the minimum (zero) at any time. The validity of the proposed method was confirmed by computer simulations and experiments, the results of which show that real-world blind separation can be achieved

2. Separation Process

We assume that source signals are mutually statistically independent and non-stationary signals (e.g., speech signals, music). As for $\bar{A}(z)$ in (1), we assume that $\bar{A}(z)$ does not have poles and zeros on the unit circle $|z| = 1$ and does not depend on the time index t.

2.1 Signal Separation Network

The problem is to extract $s(\omega, t_w)$ from $x(\omega, t_w)$. To this end, we use a network given by the following equation (see Fig. 1).

$$y(\omega, t_w) = B(\omega)^{-1} x(\omega, t_w) \tag{7}$$

where $B(\omega) = [b_{ij}(\omega)]$ is obtained by applying the windowed-Fourier transform to the following equation:

$$\bar{B}(z) = \sum_{k=0}^{M} B(k) z^{-k} \tag{8}$$

Note that the diagonal elements $b_{ii}(\omega)$ $(i=1, \ldots, N)$ of matrix $B(\omega)$ are fixed to 1. $\bar{B}(z)$ does not depend on the time index t.

Substituting (6) into (7), we have

$$\begin{aligned} y(\omega, t_w) &= B(\omega)^{-1} A(\omega) s(\omega, t_w) \\ &= C(\omega) s(\omega, t_w) \end{aligned} \tag{9}$$

where $C(\omega) B(\omega)^{-1} A(\omega)$. For the separated signals, in the blind separation problem, there is an ambiguity of permutation and amplitude. Therefore, if we can find $B_0(\omega)$ such that $C(\omega)$ is equal to $D(\omega)P$, the separation can be achieved. P is a permutation matrix which represents the ambiguity of permutation. $D(\omega)$ is a diagonal matrix which represents the ambiguity of amplitude. In our method, however, since the diagonal elements $b_{ii}(\omega)$ $(i=1, \ldots, N)$ of matrix $B(\omega)$ are fixed to 1, we can determine $D(\omega)$, that is,

$$D(\omega) = \mathrm{diag}[A(\omega) P] \tag{10}$$

where $\mathrm{diag}[X]$ denotes a diagonal matrix which has the diagonal elements of matrix X. Therefore, the desired solution $B_0(\omega)$ is $A(\omega) P^T (\mathrm{diag}[A(\omega) P^T])^{-1}$. Then, the separated signal $y(\omega, t_w)$ becomes

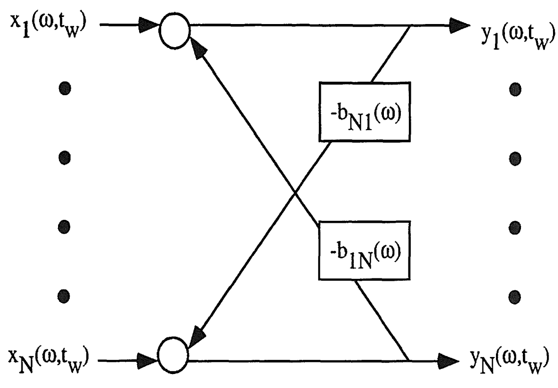

Fig. 1. Signal Separation Network.

$$y(\omega,\ t_w) = \mathrm{diag}[\boldsymbol{A}(\omega)\boldsymbol{P}^T]\boldsymbol{P}\boldsymbol{s}(\omega,\ t_w) \tag{11}$$

Here, let us consider the separated signals. The signals just before entering the microphones can be defined as

$$\bar{a}_{1p_1}(z)s_{p_1}(t),\ \bar{a}_{2p_2}(z)s_{p_2}(t),\ \cdots,\ \bar{a}_{Np_N}(z)s_{p_N}(t) \tag{12}$$

where $\{p_1, \ldots, p_N\}$ is an arbitrary permutation of $\{1, \ldots, N\}$. When the signals of (11) are applied to the inverse Fourier transform, the signals are as follows.

$$y(t) = \mathrm{diag}[\bar{A}(z)\boldsymbol{P}^T]\boldsymbol{P}\boldsymbol{s}(t) \tag{13}$$

Therefore, one can see that if the matrix $\boldsymbol{B}_0(\omega)$ can be determined, the signals of (12) can be acquired.

Our aim now is to develop a method which can acquire the matrix $\boldsymbol{B}_0(\omega)$.

2.2 Separation Method

In order to find the matrix $\boldsymbol{B}_0(\omega)$, we use the following function:

$$Q(t_w, \boldsymbol{B}(\omega)) = \frac{1}{2}\left\{\left(\sum_{i=1}^{N} \log \mathrm{E}\,[y_i(\omega, t_w)y_i(\omega, t_w)^*]\right) - \log\det E[\boldsymbol{y}(\omega, t_w)\boldsymbol{y}(\omega, t_w)^H]\right\} \tag{14}$$

where $y_i(\omega, t_w)^*$ is the conjugate value of $y_i(\omega, t_w)$, and the superscript H denotes the Hermitian operator, and $\det X$ denotes the determinant of the matrix X. The same function as that in (14) has been used by the authors in the case of the time domain [2]. We applied the function used in the time domain to the frequency domain. Wu, et al. [4] have used the same cost function as in (14). They used the summation of the cost function with respect to time t_w. However, we advocate that real-world blind separation can be achieved by using only the function in (14).

We showed that real-world blind separation in the time domain could be achieved by minimizing the same function as in (14) [2]. We apply the same method for the frequency domain. Namely, $\boldsymbol{B}_0(\omega)$ is found by minimizing the function $Q(t_w, \boldsymbol{B}(\omega))$. With respect to the minimization method, we use the natural gradient under the constraint where the diagonal elements $b_{ii}(\omega)$ of $\boldsymbol{B}(\omega)$ are not modified, that is, $b_{ii}(\omega)$ $(i=1, \ldots, N)$ are fixed to 1. The following equation is the minimization algorithm of $\boldsymbol{B}(\omega)$ derived from the natural gradient under that constraint.

$$\begin{aligned}\Delta\boldsymbol{B}(\omega) = \alpha\Bigg[&\frac{\partial Q(t_w, \boldsymbol{V}(\omega))}{\partial \boldsymbol{V}(\omega)}\boldsymbol{V}(\omega)^H\boldsymbol{B}(\omega)\\ &-\left\{\mathrm{diag}\left[\frac{\partial Q(t_w, \boldsymbol{V}(\omega))}{\partial \boldsymbol{V}(\omega)}\boldsymbol{V}(\omega)^H\boldsymbol{B}(\omega)\right]\right\}(\mathrm{diag}\,[\boldsymbol{B}(\omega)^H\boldsymbol{B}(\omega)])^{-1}\boldsymbol{B}(\omega)^H\boldsymbol{B}(\omega)\Bigg]\end{aligned} \tag{15}$$

where α is a small positive constant, and $\boldsymbol{V}(\omega)$ denotes $\boldsymbol{B}(\omega)^{-1}$. (15) is used to modify only the non-diagonal elements of $\boldsymbol{B}(\omega)$. In order to derive (15), we used the following relation.

$$\frac{\partial Q(t_w, \boldsymbol{B}(\omega))}{\partial \boldsymbol{B}(\omega)} = -\frac{\partial Q(t_w, \boldsymbol{V}(\omega))}{\partial \boldsymbol{V}(\omega)} \boldsymbol{V}(\omega)^H \boldsymbol{V}(\omega) \tag{16}$$

We can prove that the relation in (16) is satisfied. (see Appendix A)

Calculating the right-hand side of (15), we obtain

$$\begin{aligned}\Delta \boldsymbol{B}(\omega) = \alpha[\{(\mathrm{diag}\,[E[\boldsymbol{y}(\omega, t_w)\boldsymbol{y}(\omega, t_w)^H]])^{-1}E[\boldsymbol{y}(\omega, t_w)\boldsymbol{y}(\omega, t_w)^H]^{-1}\}\boldsymbol{B}(\omega) \\ -\mathrm{diag}\,[\{(\mathrm{diag}\,[E[\boldsymbol{y}(\omega, t_w)\boldsymbol{y}(\omega, t_w)^H]])^{-1}E[\boldsymbol{y}(\omega, t_w)\boldsymbol{y}(\omega, t_w)^H]^{-1}\}\boldsymbol{B}(\omega)] \\ \times (\mathrm{diag}\,[\boldsymbol{B}(\omega)^H\boldsymbol{B}(\omega)])^{-1}\boldsymbol{B}(\omega)^H\boldsymbol{B}(\omega)]\end{aligned} \tag{17}$$

(17) is the algorithm proposed by us to update $\boldsymbol{B}(\omega)$ ($\omega = 2\pi k/M$, $k=0, 1, \ldots M-1$).

3. Experimental Results

We carried out some computer simulations and experiments to confirm the validity of our method. This section describes one simulation result and two experimental results.

Example 1 (Computer simulation) In this case, the source signals were three; two male voices, and music (the sound of a drum). The channel matrix (z) was given as

$$\bar{A}(z) = \begin{bmatrix} z^{-1}+0.4z^{-2} & 0.4+0.4z^{-1} & 0.25z^{-2} \\ 0.4+0.3z^{-1} & 1+0.4z^{-1}+z^{-2} & 0.4+0.4z^{-1} \\ 0.25z^{-3} & 0.5z^{-2} & z^{-1}+0.5z^{-2} \end{bmatrix}$$

The poles of $\bar{A}(z)^{-1}$ are 5.19, -0.67, $-0.32-0.52\mathrm{I}$, $-0.32+0.52\mathrm{I}$, $0.055-0.24\mathrm{I}$, $0.055+0.24\mathrm{I}$. Therefore, $\bar{A}(z)^{-1}$ has one pole outside the unit circle $|z|=1$, from which we can deduce that $\bar{A}(z)^{-1}$ is a non-minimum phase system. The value for parameter M in (8) was set at 31. The parameter for the learning algorithm was chosen as $\alpha = 0.001$ (see (17)). The initial values of $b_{ij}(\omega)$ ($i, j=1, 2, 3$; $i \neq j$; $\omega = 2\pi k/M$, $k=0, 1, \ldots, M-1$) were set to 0. The estimation of $E[y(\omega, t_w)y(\omega, t_w)^H]$ was replaced by its instantaneous value, that is, $y(\omega, t_w)y(\omega, t_w)^H$. To estimate $\mathrm{diag}[E[y(\omega, t_w)y(\omega, t_w)^H]]$, we use the following moving average:

$$\phi_i(\omega, t) = \beta\phi_i(\omega, t-1) + (1-\beta)y_i(\omega, t_w)y_i(\omega, t_w)^* \quad (0<\beta<1) \tag{18}$$

The initial values of $\phi_i(\omega, t)$ ($i=1, 2, 3$; $\omega = 2\pi k/M$, $k=0, 1, \ldots, M-1$) and β were set to 1 and 0.9, respectively.

The separation is implemented by using the following process.

- Calculation of $\boldsymbol{x}(t)$ using (1)
- Calculation of $\boldsymbol{x}(\omega, t_w)$ using (5)
- Calculation of $y(\omega, t_w)$ using (7)
- Modification of $\boldsymbol{B}(\omega)$ using (17) and (18)
- Repeat first four steps until learning convergence.

To obtain the separated signals, we must calculate $\bar{\boldsymbol{B}}(z)^{-1}$. We obtain $\bar{\boldsymbol{B}}(z)^{-1}$ by calculating the inverse Fourier transform of $\boldsymbol{B}(\omega)^{-1}$ ($\omega = 2\pi k/M$, $k = 0, 1, \ldots, M-1$). Since $\bar{\boldsymbol{A}}(z)^{-1}$ is a non-minimum phase system, we need to rotate the leading weights of the filters of $\bar{\boldsymbol{B}}(z)^{-1}$ to the middle of the filters. We implemented this calculation using MATLAB (command name, Shift()).

Figure 2 shows the elements $\bar{c}_{ij}(i, j=1, 2, 3)$ of the matrix $\bar{\boldsymbol{C}}(z) = \bar{\boldsymbol{B}}_0(z)^{-1}\bar{\boldsymbol{A}}(z)$, where $\bar{\boldsymbol{B}}_0(z)^{-1}$ is the inverse Fourier transform of $\boldsymbol{B}(\omega)^{-1}$ ($\omega = 2\pi k/M$, $k=0, 1, \ldots, M-1$). The matrix $\bar{\boldsymbol{C}}(z)$ is a transfer function between the sources and output signals. The elements $\bar{c}_{ij}(z)$ are denoted as

$$\bar{c}_{ij}(z) = \sum_{k=0}^{34} c_{ij}(k) z^{-k} \quad (i, j = 1, 2, 3).$$

In Fig. 2, the horizontal axis is the number k ($k=0, \ldots, 34$). It can be seen that the non-diagonal elements are nearly equal to zero and the diagonal ones, $\bar{c}_{11}(z)$, $\bar{c}_{22}(z)$, and $\bar{c}_{33}(z)$ are equal to $\bar{a}_{11}(z)$, $\bar{a}_{22}(z)$, and $\bar{a}_{33}(z)$, respectively. From this result, one can see that our method can be used successfully to achieve the separation and to acquire the signals just before entering the microphones.

Example 2 (Experimental result I)

In example 2, 3, we deal with an experiment which extracts two source signals from their mixtures observed in a normal room with air conditioning and computer noises. The source signals $s_1(t)$ and $s_2(t)$ were male voices. And they were input at the same time to two speaker

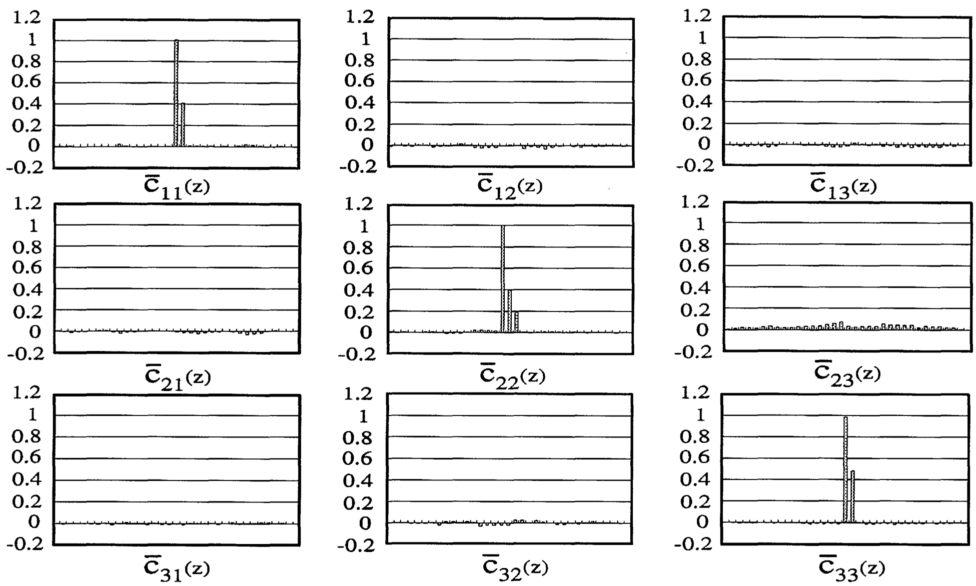

Fig. 2. The elements $\bar{c}_{ij}(z)$ of the matrix $\bar{\boldsymbol{C}}(z)$.

devices. The observed signals $x_1(t)$ and $x_2(t)$ were detected by two directional microphones (Sony ECM-670). Two speakers and two microphones were placed like Figure 3. Parameter M of (8) was set to 511. The parameters of the learning algorithm were chosen as $\alpha=0.01$ (see (17)) and $\beta=0.9$ (see (18)). The initial values of $b_{ij}(\omega)$ ($\omega=0$, $(1/M)2\pi, \ldots$, $((M-1)/M)2\pi$; $i, j=1, 2$; $i\neq j$) and $\phi_i(\omega, k)$ were set to 0 and 1, respectively.

Fig. 4 shows the plots of $s_i(t)$, $x_i(t)$, and $y_i(t)$ ($i=1, 2$). It can be seen that the output signals $y_1(t)$ and $y_2(t)$ are close to the original speech signals $s_1(t)$ and $s_2(t)$, respectively. Therefore, one can see that our method could separate the source signals from their mixtures observed in a normal room.

Example 3 (Experimental result II)

In this example, source signals $s_1(t)$ and $s_2(t)$ were music and a male voice, respectively. The configuration of two speakers and two microphones is the same as the case of example 2. We used the same parameters (M, α, β) and the same initial values of $b_{12}(\omega)$, $b_{21}(\omega)$, and $\phi_i(\omega, k)$ as in example 2.

Fig. 5 shows the plots of $s_i(t)$, $x_i(t)$, and $y_i(t)$ ($i=1, 2$). It can be seen that the output signals $y_1(t)$ and $y_2(t)$ are close to the original signals $s_1(t)$ and $s_2(t)$, respectively.

4. Conclusions

We have presented a new method of real-world blind separation which has been derived from the natural gradient method under a constraint. An attractive feature of our proposed method is that non-minimum phase systems can be treated and the signals just before entering the sensors can be acquired.

We have shown the result of a computer simulation. In the example, we have dealt with the problem which separates three source signals from three observed signals, and we have demon-

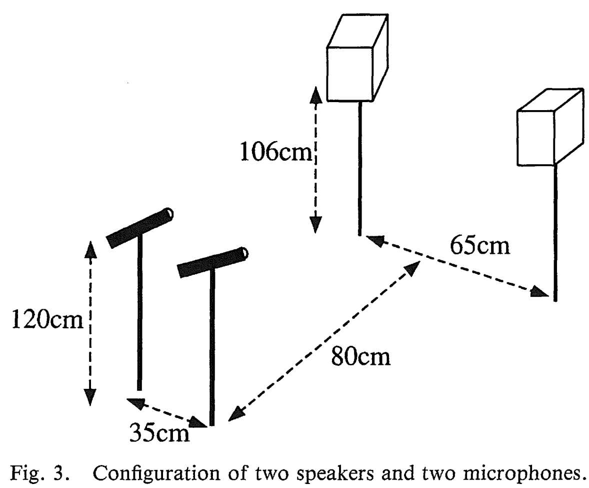

Fig. 3. Configuration of two speakers and two microphones.

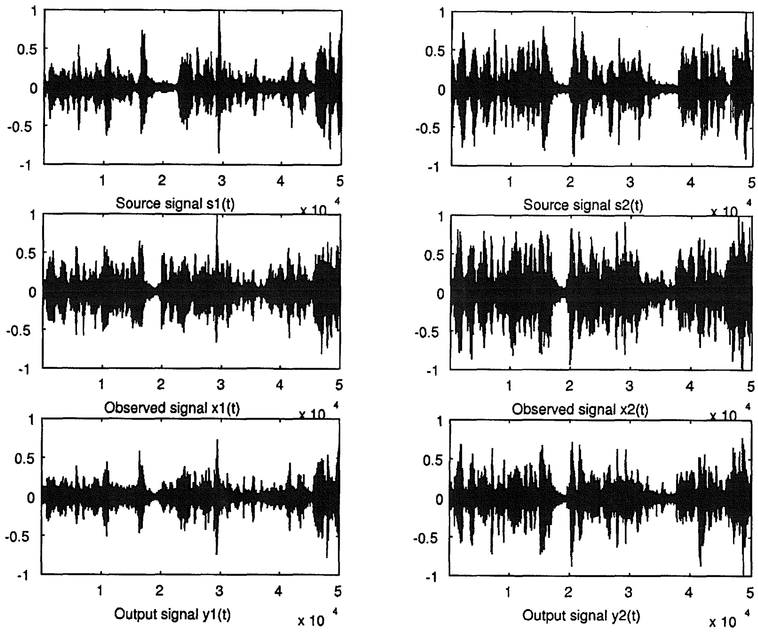

Fig. 4. The plots of $s_i(t)$, $x_i(t)$, $y_i(t)$ $(i=1, 2)$.

strated that our method makes it possible to acquire the signals just before entering the microphones. Moreover, we have conducted some experiments on separating two source signals from their mixtures observed in a normal room, and they have shown that the proposed method can be used successfully to achieve real-world blind separation.

When real-world blind separation is implemented in the time-frequency domain, one must solve the problem of ambiguity of permutation $\boldsymbol{P}$. We have not obtained the solution to this problem. In our computer simulation, however, our algorithm was not influenced by the ambiguity of permutation $\boldsymbol{P}$. In a future work, we will investigate the problem of ambiguity of permutation $\boldsymbol{P}$.

Acknowledgments

This work was supported by a Grant-in Aid for the Scientific Research by the Ministry of Education, Science and Culture of Japan, No. 11750406.

Appendix A: Proof of (16)

The following equations are always valid.

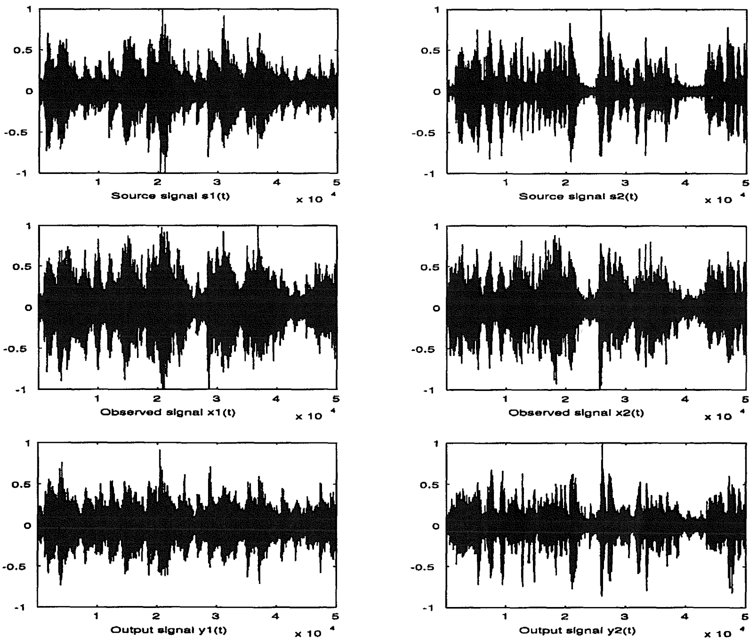

Fig. 5. The plots of $s_i(t)$, $x_i(t)$, $y_i(t)$ $(i=1, 2)$.

$$\mathrm{Tr}\left\{\frac{\partial Q(t_w, \boldsymbol{B}(\omega))}{\partial \boldsymbol{B}(\omega)} d\boldsymbol{B}(\omega)^H\right\} = \mathrm{Tr}\left\{\frac{\partial Q(t_w, \boldsymbol{V}(\omega))}{\partial \boldsymbol{V}(\omega)} d\boldsymbol{V}(\omega)^H\right\}, \tag{A1}$$

$$d\boldsymbol{V}(\omega) = -\boldsymbol{B}(\omega)^{-1} d\boldsymbol{B}(\omega)\boldsymbol{B}(\omega)^{-1} = -\boldsymbol{V}(\omega)\, d\boldsymbol{B}(\omega)\, \boldsymbol{V}(\omega) \tag{A2}$$

where Tr $\{X\}$ denotes the trace of matrix X. Substituting (A2) into (A1), we have

$$\begin{aligned}\mathrm{Tr}\left\{\frac{\partial Q(t_w, \boldsymbol{B}(\omega))}{\partial \boldsymbol{B}(\omega)} d\boldsymbol{B}(\omega)^H\right\} &= -\mathrm{Tr}\left\{\frac{\partial Q(t_w, \boldsymbol{V}(\omega))}{\partial \boldsymbol{V}(\omega)} \boldsymbol{V}(\omega)^H d\boldsymbol{B}(\omega)^H \boldsymbol{V}(\omega)^H\right\}, \\ &\quad -\mathrm{Tr}\left\{\frac{\partial Q(t_w, \boldsymbol{V}(\omega))}{\partial \boldsymbol{V}(\omega)} \boldsymbol{V}(\omega)^H \boldsymbol{V}(\omega)^H d\boldsymbol{B}(\omega)^H\right\},\end{aligned} \tag{A3}$$

From (A3), we obtain the following relationship.

$$\frac{\partial Q(t_w, \boldsymbol{B}(\omega))}{\partial \boldsymbol{B}(\omega)} = -\frac{\partial Q(t_w, \boldsymbol{V}(\omega))}{\partial \boldsymbol{V}(\omega)} \boldsymbol{V}(\omega)^H \boldsymbol{V}(\omega)^H$$

(Q.E.D.)

References

[1] A. Cichoki, S. Amari, and J. Cao, "Blind separation of delayed and convolutive signals with self-adaptive learning rate," Proc. Of 1996 International Symposium on Nonlinear Theory and its Appli-

cations, Kouchi, pp. 229–232, 1996.

[2] M. Kawamoto, A. K. Barros, A. Mansour, K. Matsuoka, and N. Ohnishi, "Real world blind separation of convolved non-stationary signals," Proc. Of First International Workshop on Independent Component Analysis and Signal Separation, France, pp. 347–352, January, 1999.

[3] S. Ikeda and N. Murata, "A method of ICA in time-frequency domain," Proc. Of First International Workshop on Independent Component Analysis and Signal Separation, France, pp. 365–370, January, 1999.

[4] H. C. Wu and J. C. Principe, "Simultaneous diagonalization in the frequency domain (SDIF) for source separation," Proc. Of First International Workshop on Independent Component Analysis and Signal Separation, France, pp. 245–250, January 1999.

[5] C. Mejuto and J. C. Principe, "A second-order method for blind source separation of convolutive mixtures," Proc. Of First International Workshop on Independent Component Analysis and Signal Separation, France, pp. 395–400, January 1999.

[6] J. Xavier, V. Barroso, and J. M. F. Moura, "Closed Form Blind Identification of MIMO Channels," Proc. ICASSP98, Vol. 6, pp. 3165–3168, Seattle WA, 1998.

[7] K. L. Diamantaras and A. Petropulu, "Blind Two-Input-Two-Output FIR Channel Estimation and Source Separation," Proc. IEEE Information Theory Workshop on Detection, Estimation, Classification and Imaging (DECI–99), Santa Fe, NM (USA), February 24–26, 1999.

[8] L. Parra, C. Spense, and B. de Vries, "Convolutive blind source separation based on multiple decorrelation," Proc. 8th IEEE Workshop Neural Networks for Signal Processing, Cambridge, UK, Sept. 1998.