IFAC Workshop on Adaptation and Learning in Control and
Signal Processing, and IFAC Workshop on Periodic Control
Systems, Yokohama, Japan, August 30 – September 1, 2004

A BLIND SOURCE SEPARATION WITH EXPONENTIALLY WEIGHTED STEPSIZE AND ITS COVERGENCE ANALYSIS IN CONVOLUTIVE MIXTURES WITH REVERBERATIONS

Akihide Horita * **Kenji Nakayama** *
Akihiro Hirano *

* *Dept. of Information and Systems Eng., Faculty of Eng., Kanazawa Univ.*
2-40-20, Kodatsuno, Kanazawa, 920-8667, JAPAN
e-mail:nakayama@t.kanazawa-u.ac.jp

Abstract: A blind source separation (BSS) method with an exponentially weighted (EW) stepsize has been proposed for convolutive mixtures with reverberations. The EW stepsize is also useful for general adaptive filters under the conditions without reverberations. This paper analyzes usefulness of the EW stepsize on the reverberations. In simulations, high-order filters are used in a separation block. Two kinds of conditions for a mixing process, that is with and without reverberations, and two kinds of stepsizes, that is a constant and the EW stepsizes, are taken into account. In the mixing process without reverberations, the EW stepsize can realize fast convergence, however, the final separation results are the same as using the constant stepsize. When reverberations are included, the EW stepsize can provide fast convergence and the good final results. A constant stepsize cannot suppress effects of reverberations. From these results, usefulness of the EW stepsize for reverberations is confirmed.

Keywords: Blind source separation, Convolutive mixture, Reverberation, Stepsize

1. INTRODUCTION

Signal processing including noise cancelation, echo cancelation, equalization of transmission lines, estimation and restoration of signals have been becoming very important technology. In some cases, we do not have enough information about signals and interference. Furthermore, their mixing and transmission processes are not well known in advance. Under these situations, blind source separation (BSS) technology using statistical property of the signal sources have become very important [1]-[7],[13],[14].

Since, in many applications, mixing processes are convolutive mixtures, FIR or IIR filters are required in unmixing processes. Several methods in a time domain and a frequency domain have been proposed. However, when high-order filters are required in the feedbacks, a learning process becomes unstable and separation performance is not enough [8]–[12]. An approach has been proposed taking some practical assumption into account [15]. High-order FIR filters can be used in an unmixing process. Reverberations must be taken into account in real environment, which causes severe condition in BSS. A learning algorithm with an exponentially weighted (EW) stepsize has been proposed [17]. The exponential weighting is automatically adjusted in a learning process. However, the EW stepsize is also useful in general adaptive filters under conditions without reverberations. They should be discriminated under real conditions, in which very high-order filters are required in a separation block.

In this paper, quasi-real conditions are taken into account. 1000 tap FIR filters are used in the separation block. The convolutive mixtures with and without reverberations, and both the EW and constant stepsizes are taken into account.

2. NETWORK STRUCTURE AND EQUATIONS

Figure 1 shows a fully recurrent BSS model proposed by Jutten et all [3]. The mixing stage has convolutive structure. FIR filters are used in feedback circuits of an unmixing block as shown in Fig.2.

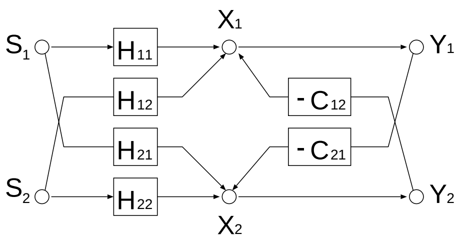

Fig. 1. Block diagram of recurrent BSS.

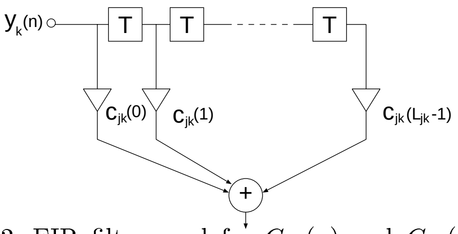

Fig. 2. FIR filter used for $C_{21}(z)$ and $C_{12}(z)$ in feedback.

The signal sources $s_i(n), i = 1, 2, \cdots, N$ are combined through the unknown convolutive mixture block, which has the impulse response $h_{ji}(m)$, and are sensed at N points, resulting in $x_j(n)$.

$$x_j(n) = \sum_{i=1}^{N} \sum_{m=0}^{M_{ji}-1} h_{ji}(m) s_i(n-m) \quad (1)$$

The output of the unmixing block $y_j(n)$ is given by

$$y_j(n) = x_j(n) - \sum_{\substack{k=1 \\ \neq j}}^{N} \sum_{l=0}^{L_{jk}-1} c_{jk}(l) y_k(n-l) \quad (2)$$

This relation is expressed using vectors and matrices as follows:

$$\boldsymbol{x}(n) = \boldsymbol{H}^T \boldsymbol{s}(n) \quad (3)$$
$$\boldsymbol{y}(n) = \boldsymbol{x}(n) - \boldsymbol{C}^T \tilde{\boldsymbol{y}}(n) \quad (4)$$
$$\boldsymbol{s}(n) = [\boldsymbol{s}_1^T(n), \boldsymbol{s}_2^T(n), \cdots, \boldsymbol{s}_N^T(n)]^T \quad (5)$$
$$\boldsymbol{s}_i(n) = [s_i(n), s_i(n-1), \cdots, s_i(n-M_i+1)]^T \quad (6)$$
$$\boldsymbol{x}(n) = [x_1(n), x_2(n), \cdots, x_N(n)]^T \quad (7)$$
$$\boldsymbol{y}(n) = [y_1(n), y_2(n), \cdots, y_N(n)]^T \quad (8)$$
$$\tilde{\boldsymbol{y}}(n) = [\boldsymbol{y}_1^T(n), \boldsymbol{y}_2^T(n), \cdots, \boldsymbol{y}_N^T(n)]^T \quad (9)$$
$$\boldsymbol{y}_k(n) = [y_k(n), y_k(n-1), \cdots, y_k(n-L_{jk}+1) \quad (10)$$

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{h}_{11} & \boldsymbol{h}_{21} & \ldots & \boldsymbol{h}_{N1} \\ \boldsymbol{h}_{12} & \boldsymbol{h}_{22} & \ldots & \boldsymbol{h}_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{h}_{1N} & \boldsymbol{h}_{2N} & \ldots & \boldsymbol{h}_{NN} \end{bmatrix} \quad (11)$$
$$\boldsymbol{h}_{ji} = [h_{ji}(0), h_{ji}(1), \cdots, h_{ji}(M_{ji}-1)]^T \quad (12)$$

$$\boldsymbol{C} = \begin{bmatrix} 0 & \boldsymbol{c}_{21} & \ldots & \boldsymbol{c}_{N1} \\ \boldsymbol{c}_{12} & 0 & \ldots & \boldsymbol{c}_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{c}_{1N} & \boldsymbol{c}_{2N} & \ldots & 0 \end{bmatrix} \quad (13)$$
$$\boldsymbol{c}_{jk} = [c_{jk}(0), c_{jk}(1), \cdots, c_{jk}(L_{jk}-1)]^T \quad (14)$$
$$(15)$$

Letting $S_i(z)$, $X_j(z)$ and $Y_k(z)$ be z–transform of $s_i(n)$, $x_j(n)$ and $y_k(n)$, respectively, they are related as follows:

$$\boldsymbol{X}(z) = \boldsymbol{H}(z)\boldsymbol{S}(z) \quad (16)$$
$$\boldsymbol{Y}(z) = \boldsymbol{X}(z) - \boldsymbol{C}(z)\boldsymbol{Y}(z) \quad (17)$$
$$\boldsymbol{S}(z) = [S_1(z), S_2(z), \cdots, S_N(z)]^T \quad (18)$$
$$\boldsymbol{X}(z) = [X_1(z), X_2(z), \cdots, X_N(z)]^T \quad (19)$$
$$\boldsymbol{Y}(z) = [Y_1(z), Y_2(z), \cdots, Y_N(z)]^T \quad (20)$$

From these expressions, a relation between the signal sources and the unmixing outputs becomes

$$\begin{aligned} \boldsymbol{Y}(z) &= (\boldsymbol{I} + \boldsymbol{C}(z))^{-1}\boldsymbol{X}(z) \\ &= (\boldsymbol{I} + \boldsymbol{C}(z))^{-1}\boldsymbol{H}(z)\boldsymbol{S}(z) \end{aligned} \quad (21)$$

In order to evaluate separation performance, the following matrix is defined.

$$\boldsymbol{P}(z) = (\boldsymbol{I} + \boldsymbol{C}(z))^{-1}\boldsymbol{H}(z) \quad (22)$$

If each row and column of $\boldsymbol{P}(z)$ has only a single non-zero element, the signal sources $s_i(n)$ are completely separated at the outputs $y_k(n)$. However, since equalization of $\boldsymbol{H}(z)$ is not guaranteed, the separated signals have the following form.

$$Y_j(z) = P_{ji}(z) S_i(z) \quad (23)$$

3. LEARNING ALGORITHM

The learning algorithm proposed for convolutive BSS is briefly explained here [15]. For simplicity, 2-channel case is taken into account.

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

$$(1)\ \ C_{21}(z)=\frac{H_{21}(z)}{H_{11}(z)} \qquad C_{12}(z)=\frac{H_{12}(z)}{H_{22}(z)} \qquad (24)$$

$$y_1(n)= {}^{T}_{11}\ {}_{1}(n) \qquad y_2(n)= {}^{T}_{22}\ {}_{2}(n) \qquad (25)$$

$$(2)\ \ C_{21}(z)=\frac{H_{22}(z)}{H_{12}(z)} \qquad C_{12}(z)=\frac{H_{11}(z)}{H_{21}(z)} \qquad (26)$$

$$y_1(n)= {}^{T}_{12}\ {}_{2}(n) \qquad y_2(n)= {}^{T}_{21}\ {}_{1}(n) \qquad (27)$$

It is assumed that delay time of $H_{11}(z)$ and $H_{22}(z)$ are shorter than that of $H_{21}(z)$ and $H_{12}(z)$. This means that in Fig.2, the sensor of X_1 is located close to $s_1(n)$, and the sensor of X_2 close to $s_2(n)$. From this assumption, the solutions in the case (1) become causal systems. On the other hand, the solutions in the case (2) are noncausal.

From Eq.(21), the outputs are expressed as

$$\begin{bmatrix} Y_1(z) \\ Y_2(z) \end{bmatrix} = \frac{1}{1-C_{12}(z)C_{21}(z)} \begin{bmatrix} 1 & -C_{12}(z) \\ -C_{21}(z) & 1 \end{bmatrix} \times \begin{bmatrix} H_{11}(z) & H_{12}(z) \\ H_{21}(z) & H_{22}(z) \end{bmatrix} \begin{bmatrix} S_1(z) \\ S_2(z) \end{bmatrix} \qquad (28)$$

$$= \frac{1}{1-C_{12}(z)C_{21}(z)} \times \begin{bmatrix} H_{11}(z)-C_{12}(z)H_{21}(z) & H_{12}(z)-C_{12}(z)H_{22}(z) \\ H_{21}(z)-C_{21}(z)H_{11}(z) & H_{22}(z)-C_{21}(z)H_{12}(z) \end{bmatrix} \times \begin{bmatrix} S_1(z) \\ S_2(z) \end{bmatrix} \qquad (29)$$

Since Eq.(26) cannot be realized using causal circuits, the diagonal elements of Eq.(29) cannot be zero. On the other hand, the non-diagonal elements can be zero. Therefore, a cost function can be defined as follows:

$$J_j(n) = E[q(y_j(n))] \qquad (30)$$

$q()$ is an even function with a single minimum point. By minimizing this cost function, $C_{12}(z)$ and $C_{21}(z)$ can approach to Eq.(24). Instead of $E[q(y_j(n))]$, the instantaneous value $q(y_j(n))$ is used, and the gradient method can be applied.

$$\hat{J}_j(n) = q(y_j(n)) \qquad (31)$$

The gradient of $\hat{J}_j(n)$ becomes

$$\frac{\partial \hat{J}_j(n)}{\partial c_{jk}(l)} = \frac{\partial q(y_j(n))}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial c_{jk}(l)} = \dot{q}(y_j(n))y_k(n-l) \qquad (32)$$

$$y_j(n) = x_j(n) - \sum_{l=0}^{L_{jk}-1} c_{jk}(l)y_k(n-l) \qquad (33)$$

$\dot{q}()$ is a partial derivative, which is an odd function. If $k = 1$, then $j = 2$, and vice versa. Therefore, the update equation of $c_{jk}(l)$ is given by

$$c_{jk}(n+1,l) = c_{jk}(n,l) + \Delta c_{jk}(n,l) \qquad (34)$$

$$\Delta c_{jk}(n,l) = \mu \dot{q}(y_j(n))y_k(n-l) \qquad (35)$$

The probability density function (pdf) of the signal sources are assumed to be even functions. Furthermore, the signal sources are statistically independent to each other. Then, they satisfy

$$E[f(s_1(n))g(s_2(n))] = E[f(s_1(n))]E[g(s_2(n))] \qquad (36)$$

$$= 0 \qquad (37)$$

$$f(), g()\ :\ \text{odd functions}$$

If a very small stepsize μ is used in Eq.(35), the correction term can be regarded as $E[\dot{q}(y_j(n))y_k(n-l)]$. Since, $\dot{q}(y_j(n))$ and $y_k(n-l)$ are also odd functions, then Eq.(37) can be held. This means that as the correction terms are reduced, $y_1(n)$ and $y_2(n)$ can approach to $\boldsymbol{h}_{11}^T\boldsymbol{s}_1(n)$ and $\boldsymbol{h}_{22}^T\boldsymbol{s}_2(n)$, respectively, .

4. A LEARNING ALGORITHM FOR CONVOLUTIVE BSS WITH REVERBERATIONS

4.1 Convergence Analysis

When reverberations occur, the assumption on the transmission delay in the mixing process cannot be held. A model including reverberations is shown in Fig.3. $H'_{11}(z)$ and $H'_{22}(z)$ express transfer functions caused by reverberation, which has a long transmission delay. $H'_{12}(z)$ and $H'_{21}(z)$ are not shown here for simplicity. By using the learning algorithm described in the previous section, the following two terms can be reduced at X_1.

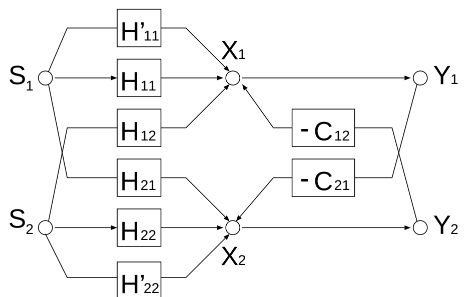

Fig. 3. Convolutive BSS model with reverberations $H'_{11}(z)$ and $H'_{22}(z)$.

$$H'_{11}(z)S_1(z) - C_{12}(z)H_{21}(z)S_1(z) \to 0 \quad (38)$$

$$C_{12}(z) \to \frac{H'_{11}(z)}{H_{21}(z)} \quad (39)$$

$$H_{12}(z)S_2(z) - C_{12}(z)H_{22}(z)S_2(z) \to 0 \quad (40)$$

$$C_{12}(z) \to \frac{H_{12}(z)}{H_{22}(z)} \quad (41)$$

$H'_{11}(z)/H_{21}(z)$ can be also causal. In other words, not only the $S_2(z)$ component but also the $S_1(z)$ component can be cancelled by the signal through the path of $-C_{12}(z)$. However, the optimum forms of $C_{12}(z)$ for canceling $S_2(z)$ and $S_1(z)$ are different.

In the same manner, at X_2,

$$H'_{22}(z)S_2(z) - C_{21}(z)H_{12}(z)S_2(z) \to 0 \quad (42)$$

$$C_{21}(z) \to \frac{H'_{22}(z)}{H_{12}(z)} \quad (43)$$

$$H_{21}(z)S_1(z) - C_{21}(z)H_{11}(z)S_1(z) \to 0 \quad (44)$$

$$C_{21}(z) \to \frac{H_{21}(z)}{H_{11}(z)} \quad (45)$$

Eqs.(41) and (45) are the ideal solutions. However, $C_{12}(z)$ and $C_{21}(z)$ cannot approach to these solutions due to the reverberations given by Eqs.(39) and (43).

4.2 A Learning Algorithm with Exponential Scaling

Reverberations have a long delay, and from Eqs.(39) and (43), effects of reverberations appear at the latter part of the impulse responses. For this reason, the correction in the latter part is suppressed. This can be done by controlling the stepsize μ exponentially along a delay line in the FIR filters. The update equation is modified as follows:

$$c_{jk}(n+1,l) = c_{jk}(n,l) + \mu(l)f(y_j(n))g(y_k(n-l)) \quad (46)$$

$$\mu(l) = \mu_0 r^l, \qquad 0 < r < 1 \quad (47)$$

$\mu(l)$ should be proportional to the ideal solution. However, it is not known beforehand. Therefore, the exponential scaling is proposed here. μ_0 is the initial stepsize and r^l is an exponential part.

5. SIMULATION

5.1 Simulation Conditions

Two channel blind source separation was simulated. The stepsize is optimized for each combination of the input signals and with or without reverberations. They include (a-1) white noise and no reverberations, (a-2) white noise and reverberations, (b-1) speech signal and no reverberations, (b-2) speech signal and reverberations. A constant stepsize and the exponential stepsize are set to (a-1) $\mu = 0.0001$, $\mu_0 = 0.00078$, $r_{12} = 0.9933$, $r_{21} = 0.9918$, (a-2) $\mu = 0.0003$, $\mu_0 = 0.00078$, $r_{12} = 0.9883$, $r_{21} = 0.9822$, (b-1) $\mu = 0.00005$, $\mu_0 = 0.00015$, $r_{12} = 0.9933$, $r_{21} = 0.9918$, (b-2) $\mu = 0.0002$, $\mu_0 = 0.00015$, $r_{12} = 0.9883$, $r_{21} = 0.9822$.

The separation performance is evaluated by the following Noise Reduction Ratio (NRR), defined by using $\boldsymbol{H}(z)$ and $\boldsymbol{P}(z)$ in Eqs.(16) and (22)

$$\sigma_{xs}^2 = \sum_{i=1}^{2} \frac{1}{2\pi} \int_{-\pi}^{\pi} |H_{ii}(e^{j\omega T})|^2 d\omega T \quad (48)$$

$$\sigma_{xc}^2 = \sum_{j \neq i} \frac{1}{2\pi} \int_{-\pi}^{\pi} |H_{ji}(e^{j\omega T})|^2 d\omega T \quad (49)$$

$$SNR_x = 10 \log \frac{\sigma_s^2}{\sigma_c^2} \quad [\mathrm{dB}] \quad (50)$$

$$\sigma_{ys}^2 = \sum_{i=1}^{2} \frac{1}{2\pi} \int_{-\pi}^{\pi} |P_{ii}(e^{j\omega T})|^2 d\omega T \quad (51)$$

$$\sigma_{yc}^2 = \sum_{j \neq i} \frac{1}{2\pi} \int_{-\pi}^{\pi} |P_{ji}(e^{j\omega T})|^2 d\omega T \quad (52)$$

$$SNR_y = 10 \log \frac{\sigma_s^2}{\sigma_c^2} \quad [\mathrm{dB}] \quad (53)$$

$$NRR = SNR_y - SNR_x \quad [\mathrm{dB}] \quad (54)$$

σ_{xs}^2 and σ_{ys}^2 express power of the selected signals and σ_{xc}^2 and σ_{ys}^2 are that of the cross components.

Impulse responses of the convolutive mixture with reverberations are shown in Fig.4. Furthermore, those of the case without reverberations are shown in Fig.5. In this simulation, almost the same length FIR filters are used in the separation block in both cases with and without reverberations. For this purpose, the impulse responses in two cases are different.

The ideal filter coefficients are shown in Figs.6 and 7. Thus, 1000 length FIR filters are used in the separation block.

5.2 Separation Performance

NRR with several conditions are shown in Fig.8. The signals sources are two different white noises. In the first case without reverberations, even though the EW stepsize can realize fast convergence, the final NRR obtained by both the EW and constant stepsize are the same. Therefore, in this case, usefulness of the EW stepsize is to make convergence fast.

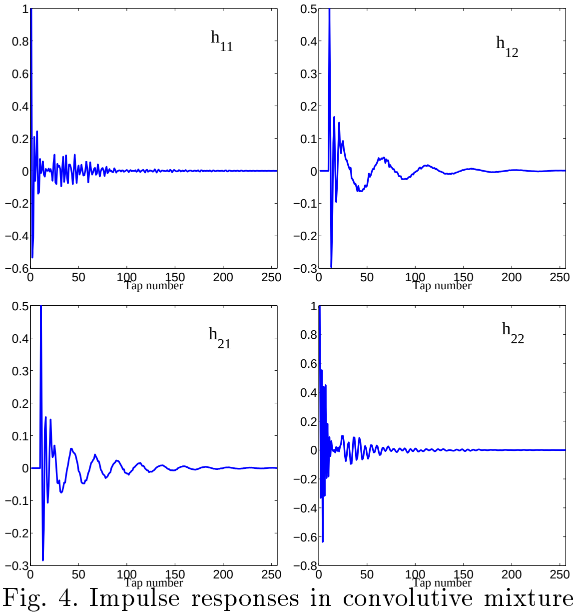

Fig. 4. Impulse responses in convolutive mixture with reverberations.

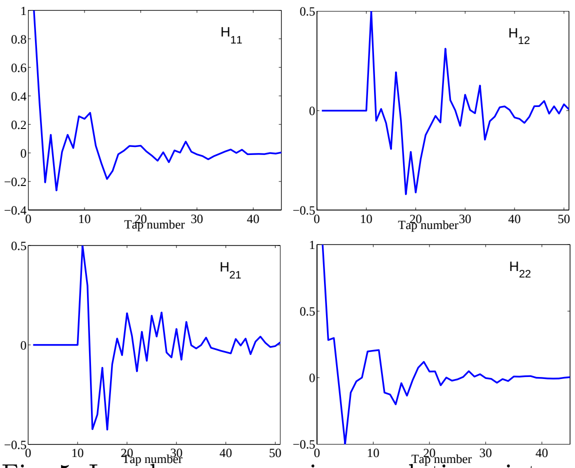

Fig. 5. Impulse responses in convolutive mixture without reverberations.

In the second case with reverberations, the EW stepsize can realize fast convergence like in the first case. However, the learning curve with the constant stepsize shows slow convergence and the lower NRR compared with the EW stepsize. This means, degradation of separation performance can be suppressed by using the EW stepsize.

Figure 9 shows NRR for speech signals sources. The almost same trend are shown.

6. CONCLUSIONS

Usefulness of the exponentially weighted (EW) stepsize for the convolutive mixture with reverberations has been comparatively investigate. In the first case without reverberations, the EW stepsize can realize fast convergence, the learning curve of the constant stepsize can reach that of the EW stepsize. In the case with reverberations, the EW stepsize can provide fast convergence and a good separation performance. On the contrary,

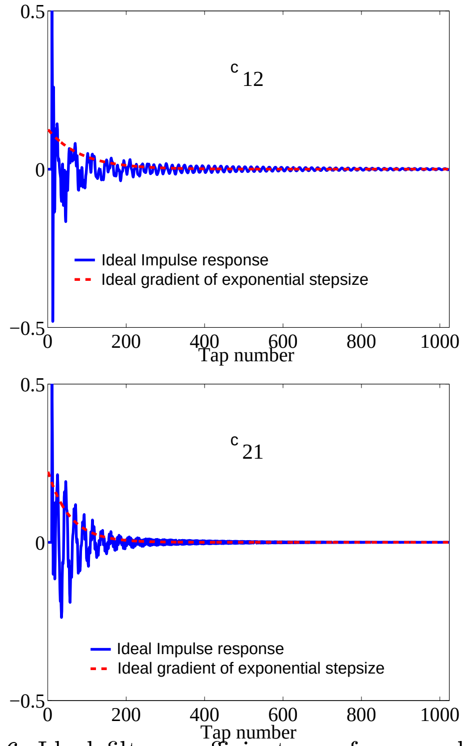

Fig. 6. Ideal filter coefficients c_{ij} for convolutive mixture with reverberations.

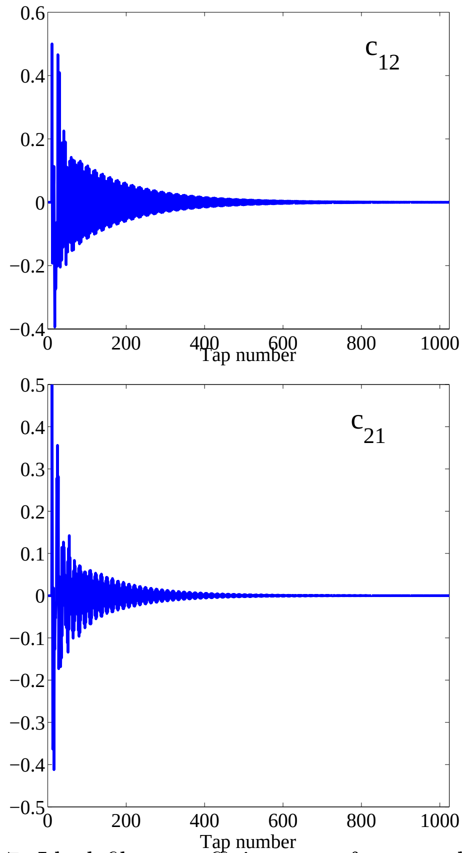

Fig. 7. Ideal filter coefficients c_{ij} for convolutive mixture without reverberations.

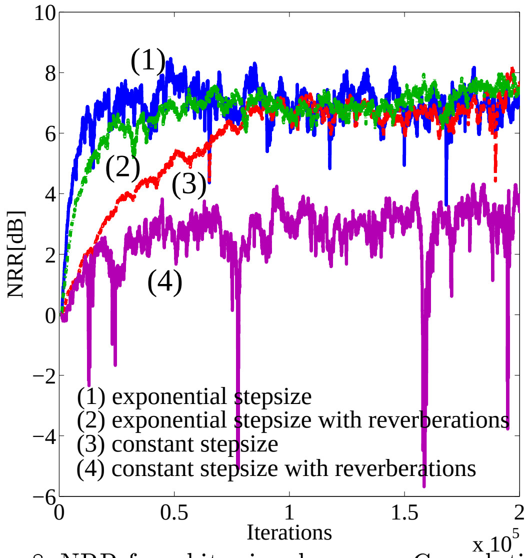

Fig. 8. NRR for white signal sources. Convolutive mixtures with and without reverberations, and constant and EW stepsizes are used.

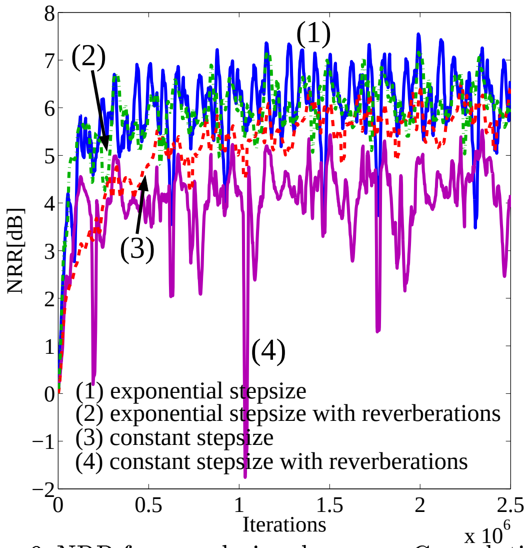

Fig. 9. NRR for speech signal sources. Convolutive mixtures with and without reverberations, and constant and EW stepsizes are used.

the constant stepsize cannot suppress effects of reverberations on the signal separation, and its learning curve shows slow convergence and lower separation performance.

REFERENCES

[1] C.Jutten, J.Herault and A.Guerin, "IIN.C.A: An independent components analyzer based on an adaptive neuromimetic network", in: J.Degmongeot, T.Herve, V.Raille and C.Roche, eds., Artificial Intelligence and Cognitive Science, Manhester Univ. Press, Manchester, 1988.

[2] J.F.Cardoso, "Eigen structure of the 4th order cumulant tensor with application to the blind source separation problem", ICASSP Proc. pp. 2655-1658.

[3] C.Jutten and Jeanny Herault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture", Signal Processing, 24, pp.1-10, 1991.

[4] P.Comon, C.Jutten and J.Herault, "Blind separation of sources, Part II: Problems statement", Signal Processing, 24, pp.11-20, 1991.

[5] 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.

[6] K.Nakayama, A.Hirano and M.Nitta, "A constraint learning algorithm for blind source separation", Proc. IJCNN'2000, pp.24-27, July, 2000.

[7] K,Nakayama, A.Hirano and T.Sakai, "A pair-channel learning algorithm with constraints for multi-channel blind separation", Proc. IJCNN'01, July 2001.

[8] H.L.Nguyen Thi and C.Jutten, "Blind source separation for convolutive mixtures", Signal Processing, vol.45, no.2, pp.209–229, March 1995.

[9] C.Simon, G.d' Urso, C.Vignat, Ph.Loubaton and C.Jutten, "On the convolutive mixture source separation by the decorrelation approach", Proc. ICASSP'98, pp.IV-2109–2112, May 1998.

[10] S.Cruces and L.Castedo, "A Gauss-Newton methods for blind source separation of convolutive mixtures", ICASSP'98, pp.IV2093–2096, May 1998.

[11] S.Araki, S.Makino, T.Nishikawa and H.Saruwatari, "Fundamental limitation of frequency domain blind source separation for convolutive mixture of speech", Proc. ICASSP'01, MULT-P2.3, May 2001.

[12] I.Kopriva, Z.Devcic and H.Szu, "An adaptive short-time frequency domain algorithm for blind separation of nonstationary convolved mixtures", Proc. IJCNN'01, pp.424–429, July 2001.

[13] H.Mathis and S.C.Douglas, "On optimal and universal nonlinearities for blind signal separation", Proc. ICASSP'01, MULT-P3.3, May 2001.

[14] K.Nakayama, A.Hirano and T.Sakai, "An adaptive nonlinear function controlled by kurtosis for blind source separation", Proc. IJCNN'2002, pp.1234-1239, May 2002.

[15] K.Nakayama, A.Hirano and A.Horita, "A learning algorithm for convolutive blind source separation with transmission delay constraint", Proc. IJCNN'2002, pp.1287-1292, May 2002.

[16] S.Makino, Y.Kaneda, N.Koizumi, "Exponentially weighted stepsize NLMS adaptive filter based on the statistics of a room impulse response", IEEE Trans. Speech and Audio Processing, vol.1, no.1, pp.101-108, Jan. 1993.

[17] K.Nakayama, A.Hirano and A.Horita, "A learning algorithm with adaptive exponential stepsize for blind source separation of convolutive mixtures with reverberations", IEEE&INNS Proc. IJCNN'03, Portland, Oregon, July 2003.