LETTER

# Robust Blind Equalization Algorithms Based on the Constrained Maximization of a Fourth-Order Cumulant Function

**Kiyotaka KOHNO**[†a)], ***Student Member***, **Mitsuru KAWAMOTO**[††*b)], ***Member***, **Asoke K. NANDI**[†††c)], ***Nonmember***, ***and*** **Yujiro INOUYE**[†**d)], ***Member***

**SUMMARY** The present letter deals with the blind equalization problem of a single-input single-output infinite impulse response (SISO-IIR) channel with additive Gaussian noise. To solve the problem, we propose a new criterion for maximizing constrainedly a fourth-order cumulant. The algorithms derived from the criterion have such a novel property that even if Gaussian noise is added to the output of the channel, an effective zero-forcing (ZF) equalizer can be obtained with as little influence of Gaussian noise as possible. To show the validity of the proposed criterion, some simulation results are presented.

***key words:*** *blind equalization, Gaussian noise, single-input single-output infinite impulse response (SISO-IIR) systems, fourth-order cumulant functions, eigenvector algorithms*

## 1. Introduction

This letter deals with a blind equalization problem in which a single-input single-output infinite impulse response (SISO-IIR) channel with additive Gaussian noise is considered. To solve the blind equalization problem, we use an idea of the constrained maximization of a fourth-order cumulant function (e.g., [2], [4]).

Fourth-order (or higher-order) cumulant based algorithms are very popular and attractive for solving the blind equalization problem. Among these algorithms, however, there exist some algorithms which employ second-order cumulants (e.g., [1]–[4]). Such types of algorithms have a major drawback that they are very sensitive to Gaussian noise (this will be shown in Sect. 4).

In this letter, we propose a new criterion for overcoming the drawback. The algorithms derived from the criterion consist of only the fourth-order cumulants of the output of the channel; hence they become robust to Gaussian noise. This is a novel key point in this research. Computer simulations are presented to demonstrate the validity of the criterion.

Manuscript received September 26, 2005.
Manuscript revised December 5, 2005.
Final manuscript received January 18, 2006.

[†]The authors are with the Department of Electronic and Control Systems Engineering, Shimane University, Matsue-shi, 690-8504 Japan.
[††]The author is with the National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba-shi, Japan.
[†††]The author is with the Department of Electrical Engineering and Electronics, The University of Liverpool, UK.
[*]The author is also with Bio-Mimetic Control Research Center, RIKEN, Nagoya-shi, 463-0003 Japan.
[**]The author is also with the Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Wako-shi, 351-0198 Japan.
a) E-mail: kohno@yonago-k.ac.jp
b) E-mail: m.kawamoto@aist.go.jp
c) E-mail: aknandi@liverpool.ac.uk
d) E-mail: inouye@riko.shimane-u.ac.jp
DOI: 10.1093/ietfec/e89–a.5.1495

## 2. Problem Formulation and Assumptions

We consider a single-input single-output (SISO) channel system with an additive noise as described by

$$y(t) = \sum_{k=-\infty}^{\infty} h^{(k)} s(t-k) + n(t), \tag{1}$$

where $\{s(t)\}$ is an unobserved input sequence generated from a discrete-time stationary random process, $\{h^{(k)}\}$ is the impulse response of an unknown time-invariant system defined by $H(z) = \sum_{k=-\infty}^{\infty} h^{(k)} z^k$, $y(t)$ denotes the output of the channel system, and $n(t)$ denotes a Gaussian noise. Figure 1 illustrates a diagram of the basic problem.

In the blind equalization problem we want to find the equalizer $W(z)$ such that $G(z) := W(z)H(z)$ becomes

$$\hat{G}(z) = \hat{W}(z)H(z) = cz^{k_1}, \tag{2}$$

even if the Gaussian noise $n(t)$ is included in the output $y(t)$, where $c$ in $cz^{k_1}$ is a nonzero value allowing for a scale change, and the superscript "$k_1$" of $z^{k_1}$ denotes an integer corresponding for a constant delay. The notation "$z$" is used instead of the commonly used $z^{-1}$ in the $z$-transform throughout this letter.

To design such an equalizer $\hat{W}(z)$ in (2), we make the following assumptions on the system, the input signal, and the equalizer.

**A1)** The unknown system $H(z)$ is a stable, possibly nonminimum phase, linear time-invariant filter whose inverse (which may be noncausal and stable) $H(z)^{-1}$ exists.

**A2)** The input sequence $\{s(t)\}$ is a zero-mean, non-Gaussian random process. Moreover, the process $\{s(t)\}$ is an i.i.d. process with nonzero variance $\sigma_s^2 \neq 0$ and nonzero fourth-order

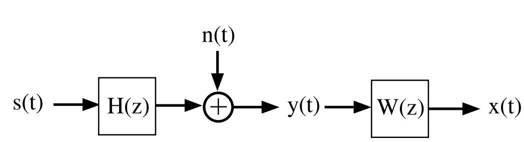


**Fig. 1** The composite system of an unknown channel and a filter.

cumulant, $\gamma_s$ defined as

$$\gamma_s = \text{cum}\{s(t), s(t), s(t), s(t)\} \neq 0. \tag{3}$$

**A3)** The equalizer $W(z) = \sum_{k=L_1}^{L_2} w^{(k)} z^k$ is an FIR system of sufficient length $L = L_2 - L_1 + 1$ so that the truncation effect can be ignored.

It is assumed for the sake of simplicity in this letter that all the signals and all the systems are real-valued.

## 3. Fourth-Order Cumulant Based Algorithms (FOCAs)

### 3.1 Conventional FOCAs

Almost all the conventional FOCAs may be used together with second-order cumulants. For example, Jelonnek et al. [2] have proposed an eigenvector algorithm by solving the following problem:

$$\text{Maximize} \quad |\gamma_{xr}| \qquad \text{subject to} \quad \sigma_x^2 = \sigma_s^2. \tag{4}$$

where $\sigma_x^2 := \text{cum}\{x(t), x(t)\}$, which is the second-order cumulant of $x(t)$, and $\gamma_{xr} = \text{cum}\{x(t), x(t), r(t), r(t)\}$, which is the fourth-order cross-cumulant of $x(t)$ and $r(t)$. Here, the signal $r(t)$ is a *reference signal* given by the calculation of $F(z)y(t)$, where $F(z)$ is a filter denoted by $F(z) = \sum_k f^{(k)} z^k$ (see Fig. 2). Differently from the fourth-order auto-cumulant of $x(t)$ as in [4], that is, $\gamma_x = \text{cum}\{x(t), x(t), x(t), x(t)\}$, $\gamma_{xr}$ becomes a quadratic function with respect to $w^{(k)}$. Hence they showed that since the variance $\sigma_x^2$ and the fourth-order cumulant $\gamma_{xr}$ can be respectively uniquely expressed in terms of the equalizer coefficient vector $\boldsymbol{w}$ as $\sigma_x^2 = \boldsymbol{w}^T \boldsymbol{R} \boldsymbol{w}$ and $\gamma_{xr} = \boldsymbol{w}^T \boldsymbol{C}_{yr} \boldsymbol{w}$, then by the Lagrangean method, the constrained maximization in (4) leads to a closed-form expression as the following generalized eigenvector problem:

$$\boldsymbol{C}_{yr} \boldsymbol{w} = \hat{\lambda} \boldsymbol{R} \boldsymbol{w}, \tag{5}$$

where $\boldsymbol{w}$ is the $L$-column equalizer-coefficient vector, that is, $\boldsymbol{w} = [w^{(L_1)}, w^{(L_1+1)}, \cdots, w^{(L_2)}]^T$, the matrix $\boldsymbol{C}_{yr}$ is an $L \times L$ matrix whose $(r_1, r_2)$th element is $\text{cum}\{y(t-r_1), y(t-r_2), r(t), r(t)\}$, and $\boldsymbol{R} = \text{E}[\boldsymbol{y}(t)\boldsymbol{y}^T(t)]$ is the covariance matrix of $L$-column vector $\boldsymbol{y}(t)$ with elements $y(t-l)$, $l = L_1, (L_1+1), \cdots, L_2$.

It follows from (5) that the vector $\boldsymbol{w}$ obtained by choosing the eigenvector of $\boldsymbol{R}^{-1}\boldsymbol{C}_{yr}$ associated with the maximum (absolute) magnitude eigenvalue $|\hat{\lambda}_m|$ is given as a solution for solving the blind equalization problem except for a pathological case (see [2] for the details of the pathological case), which is referred to as an *eigenvector algorithm*

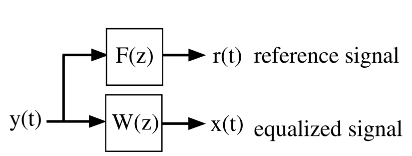


**Fig. 2** The equalized signal and the reference signal.

(EVA) [2]. Note that this situation can be attained in the case that there is no additive noise to the output $y(t)$. In the case that there is Gaussian noise in the output $y(t)$, however, since $\boldsymbol{R}$ is the covariance matrix of $\boldsymbol{y}(t)$, the solution obtained from the eigenvector of $\boldsymbol{R}^{-1}\boldsymbol{C}_{yr}$ is affected by Gaussian noise. That is, we conclude that any algorithm derived by solving the problem (4) has such a property, because the matrix $\boldsymbol{R}$ is included into the derived algorithms.

### 3.2 The Proposed FOCA

In this subsection, we shall show a constraint, subject to which an algorithm derived from the maximization of fourth-order cumulants (e.g., $\gamma_{xr}$ and $\gamma_x$) becomes robust against Gaussian noise. To this end, we use the following value, instead of $\sigma_x^2$ in (4):

$$\gamma_{xy} := \sum_{l=-\infty}^{\infty} \text{cum}\{x(t), x(t), y(t-l), y(t-l)\}. \tag{6}$$

From the assumption A2), the relation $\gamma_{xy} = \alpha\gamma_s \sum_k g^{(k)2}$ holds, where $\alpha = \sum_l h^{(l)2} > 0$. If there is no noise in the output $y(t)$, $\sigma_x^2$ can be expressed as $\sigma_x^2 = \sigma_s^2 \sum_k g^{(k)2}$. Hence one can see that if we put $\gamma_{xy} = \alpha\gamma_s$ as a constraint for the maximization of fourth-order cumulants, the constraint has the same role as the one in (4).

Here, let us show an algorithm derived by solving the following problem:

$$\text{Maximize} \quad |\gamma_{xr}| \qquad \text{subject to} \quad \gamma_{xy} = \alpha\gamma_s. \tag{7}$$

We note that we may choose an appropriate positive value for $\alpha$ in (7) if its true value is not available, because this choice affects only the length of the equalizer coefficient vector $\boldsymbol{w}$. By the same way as in [2], it can be proved that the problem (7) attains to the solution (2) except for a pathological case. The cumulant $\gamma_{xy}$ can be uniquely expressed in terms of the equalizer coefficient vector $\boldsymbol{w}$ as $\gamma_{xy} = \boldsymbol{w}^T \boldsymbol{Q} \boldsymbol{w}$, where $\boldsymbol{Q} = \sum_l \boldsymbol{C}_{y,l}^{(4)}$ with $\boldsymbol{C}_{y,l}^{(4)}$ being an $L \times L$ matrix whose $(r_1, r_2)$th element is $\text{cum}\{y(t-r_1), y(t-r_2), y(t-l), y(t-l)\}$. Therefore, by the Lagrangean method, the problem (7) leads to the following generalized eigenvector problem:

$$\boldsymbol{C}_{yr} \boldsymbol{w} = \lambda \boldsymbol{Q} \boldsymbol{w}. \tag{8}$$

Then, it can be seen from (5) that the eigenvector of $\boldsymbol{Q}^{-1}\boldsymbol{C}_{yr}$ associated with the maximum magnitude eigenvalue $|\lambda_m|$ is given as a solution for solving the blind equalization problem except for a pathological case. Note that there exists the inverse of the matrix $\boldsymbol{Q}$ (see [5]). The solution obtained by (8) is less sensitive to Gaussian noise, because $\boldsymbol{Q}^{-1}\boldsymbol{C}_{yr}$ consists of only fourth-order cumulants. This is a novel key point of our proposed FOCA. The algorithm (8) is referred to as a *robust eigenvector algorithm* (REVA).

**Remark 1:** In order to calculate the matrix $\boldsymbol{Q}$, we should carry out the summation over the infinite interval $(-\infty, \infty)$ in theory. However, it is impossible; hence in practice we should approximate a finite interval to the infinite one. In Sect. 4, we will choose interval $[L_1, L_2]$ for this approximation.

### 3.3 How to Process the REVA

In this subsection, we shall show how the REVA is carried out and the reference signal is selected. In this letter, the vector $\boldsymbol{w}$ is iteratively modified by the REVA. Hence the eigenvector $\boldsymbol{w}$ obtained by the REVA is newly rewritten by $\boldsymbol{w}(t_i)$. Then (8) can be expressed as

$$\boldsymbol{C}_{yr}(t)\boldsymbol{w}(t_i) = \lambda \boldsymbol{Q}(t)\boldsymbol{w}(t_i), \tag{9}$$

where $t_i$ denotes an iteration number. The reference signal $r(t)$ used to calculate the matrix $\boldsymbol{C}_{yr}(t)$ is given by $r(t) = \boldsymbol{w}^T(t_i - 1)\boldsymbol{y}(t)$. Namely, when the eigenvector $\boldsymbol{w}(t_i)$ is calculated by (9), the coefficients $f^{(k)}$ of the filter $F(z)$ (see Fig. 2) are the elements of the previous eigenvector $\boldsymbol{w}(t_i - 1)$. The matrices $\boldsymbol{C}_{yr}(t)$ and $\boldsymbol{Q}(t)$ are estimated by using the moving averages shown in Sect. 4, before calculating the eigenvector $\boldsymbol{w}(t_i)$ by (9). Consequently, the process of the REVA is summarized as follows:

Set initial values: $\boldsymbol{w}(0), \boldsymbol{C}_{yr}(0), \boldsymbol{Q}(0)$
  for $t_i = 1 : t_{i_{all}}$ (10)
    for $t = t_d(t_i - 1) + 1 : t_d t_i$ (11)
      $r(t) = \boldsymbol{w}^T(t_i - 1)\boldsymbol{y}(t)$
      Calculate $\boldsymbol{C}_{yr}(t)$ and $\boldsymbol{Q}(t)$ by (12) and (15)
    end
    Calculate the eigenvector $\boldsymbol{w}(t_i)$ associated with
    $|\lambda_m|$ from (9)
  end

Here $t_{i_{all}}$ denotes the total number of iterations and $t_d$ denotes the number of data samples for estimating the matrices $\boldsymbol{C}_{yr}(t)$ and $\boldsymbol{Q}(t)$.

## 4. Simulation Results

In this section, the validity of the proposed criterion of maximizing constrainedly the fourth-order cumulant will be demonstrated by using the proposed REVA. The unknown system $H(z)$ was set to be a filter of length 7 with the impulse response (0.4, 1, −0.7, 0.6, 0.3, −0.4, 0.1) (as in [5]). We used an equalizer of length $L = 16$ (where $L_1 = 0$ and $L_2 = 15$) which was initialized to $\boldsymbol{w}(0) = [0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]^T$. The initial values of the matrices $\boldsymbol{C}_{yr}(t)$ and $\boldsymbol{Q}(t)$ were set to be zero matrix. Their moving averages were defined as follows:

$$\boldsymbol{C}_{yr}(t) := \beta_1 \boldsymbol{C}_{yr}(t-1) \\ +(1-\beta_1)\{r(t)^2\boldsymbol{V}(t) - 2\tilde{\boldsymbol{v}}(t)\tilde{\boldsymbol{v}}^T(t) - r(t)^2\tilde{\boldsymbol{V}}(t)\}, \tag{12}$$

where $\tilde{\boldsymbol{v}}(t)$ is a column vector, which is a moving average of $\boldsymbol{v}(t) = r(t)\boldsymbol{y}(t)$ calculated by

$$\tilde{\boldsymbol{v}}(t) = \beta_2\tilde{\boldsymbol{v}}(t-1) + (1-\beta_2)r(t)\boldsymbol{y}(t), \tag{13}$$

and $\tilde{\boldsymbol{V}}(t)$ is a moving average of $\boldsymbol{V}(t) := \boldsymbol{y}(t)\boldsymbol{y}^T(t)$, which is

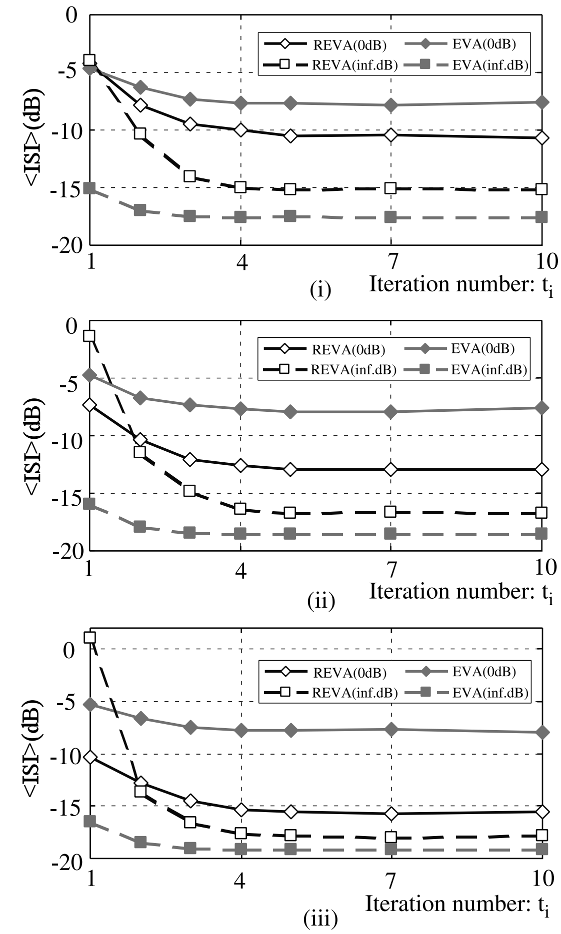


**Fig. 3** The performances of the REVA and the EVA with varying number of iterations, in the three cases of (i) 5000 samples, (ii) 10000 samples, (iii) 30000 samples, in each of which SNR = 0 dB and ∞ dB.

calculated by

$$\tilde{\boldsymbol{V}}(t) = \beta_2\tilde{\boldsymbol{V}}(t-1) + (1-\beta_2)\boldsymbol{V}(t), \tag{14}$$

in which the initial values of $\tilde{\boldsymbol{V}}(t)$ and $\tilde{\boldsymbol{v}}(t)$ were set to be zero matrix and vector, respectively. [See Appendix A for a brief derivation of (12).]

$$\boldsymbol{Q}(t) := \beta_1\boldsymbol{Q}(t-1) + (1-\beta_1)\boldsymbol{U}(t)\boldsymbol{V}(t), \tag{15}$$

where $\boldsymbol{U}(t)$ is given by

$$\boldsymbol{U}(t) := \boldsymbol{V}(t) - 2\tilde{\boldsymbol{V}}(t) - \mathrm{tr}\{\tilde{\boldsymbol{V}}(t)\}\boldsymbol{I}, \tag{16}$$

where tr$\{X\}$ denotes the trace of the matrix $X$, $\boldsymbol{I}$ denotes the identity matrix. [See Appendix B for a brief derivation of (15).] Note that in the matrix $\boldsymbol{Q}(t)$ calculated by (15), the interval for $l'$s was set to be [0,15]. The input $s(t)$ of the system $H(z)$ was binary signal which takes one of two values, −1 and 1 with equal probability 1/2. Then the value of $\gamma_s$ is −2. The Gaussian noise $n(t)$ with its variance $\sigma_n^2$ was included in the output $y(t)$ at various SNR levels. The SNR is, for convenience, defined as SNR := 10 $\log_{10}(\sigma_s^2/\sigma_n^2)$, where $\sigma_s^2$ is the variance of $s(t)$ and is equal to 1. As a measure of

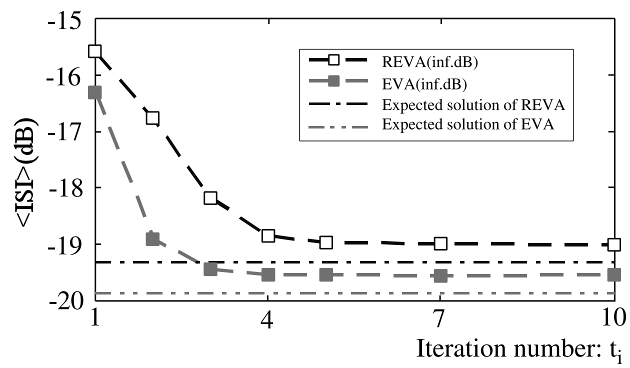


**Fig. 4** The average performances of the REVA and the EVA with varying number of iterations, in the cases of SNR = $\infty$ dB and $t_d$=10,000 samples, where the simulation conditions, except for $\boldsymbol{Q}$, were the same as the ones in the examples shown in Fig. 3, and the ISIs of the expected solutions of the REVA and the EVA.

**Table 1** The parameters $\beta_1$ and $\beta_2$.

| | $t_d$=5,000 | $t_d$=10,000 | $t_d$=30,000 |
|---|---|---|---|
| $\beta_1$ | 0.9998 | 0.9999 | 0.99997 |
| $\beta_2$ | 0.998 | 0.999 | 0.9995 |

performance, we used the intersymbol interference (ISI) defined in [5], and hence a large negative value of ISI (shown in Fig. 3 and Fig. 4) indicates that the corresponding vector $\boldsymbol{g}$ is very close to the proximity to the desired solution.

For comparison, the EVA in (5) was used, in which (14) was used as an estimation of $\boldsymbol{R}$. Figure 3 shows the average performances of the proposed REVA and the EVA, in the case where the SNR levels were taken to be 0 dB ($\sigma_n^2$=1) and $\infty$ dB ($\sigma_n^2$=0), the total number of iterations $t_{i_{all}}$ (see (10)) was set to be 10, and there were three kinds of $t_d$ (see (11)), $t_d$ = 5,000 samples (Fig. 3(i)), 10,000 samples (Fig. 3(ii)), and 30,000 samples (Fig. 3(iii)). The parameters $\beta_1$ and $\beta_2$ for each case were summarized in Table 1. The vertical and horizontal axes in Fig. 3 represent the average ISI denoted by <ISI> and the number of iterations, respectively, where each average was calculated by the results of 100 Monte Carlo trials.

From Fig. 3, it can be seen that for each $t_d$, in the case of SNR = 0 dB, as the number of data samples increases, the REVA provides better performance, whereas the performance of the EVA hardly changes. In the case of SNR = $\infty$ dB, however, it can be seen from Fig. 3 that the performances of the REVA become worse than those of the EVA. One of the reasons to such a problem may be that it is difficult to estimate the fourth-order cumulants of $\boldsymbol{Q}$ for the REVA compared with the second-order cumulants of $\boldsymbol{R}$ for the EVA. However, even if $\boldsymbol{Q}(t)$ in (9) has its theoretical value, the performance of the REVA becomes worse than that of the EVA (see Fig. 4). Indeed, the difference between the ISIs of the expected solutions of the EVA and the REVA, using the theoretical values of $\boldsymbol{C}_{yx}$, $\boldsymbol{R}$, and $\boldsymbol{Q}$ was about 0.56 dB (see Fig. 4).

From these results, we conclude that the algorithms derived from the criterion of maximizing fourth-order cumulants using $\gamma_{xy}$ can work well for the blind equalization in the presence of additive Gaussian noise. However, there exists such a case that for the blind equalization in the case where the power of additive Gaussian noise is small, it is better to use the conventional algorithms using the constraint $\sigma_x^2 = \sigma_s^2$.

## 5. Conclusion

In order to solve the blind equalization problem, we have proposed a new criterion of maximizing constrainedly a fourth-order cumulant function. Also, as one example of the algorithms derived from the criterion, the *robust eigenvector algorithm* (REVA) has been introduced. The REVA is robust against Gaussian noise, which means that the REVA can be used to obtain the solution (2), even if Gaussian noise is added to the output of the system (see Fig. 1). This is a novel property of any algorithm derived from the criterion of maximizing fourth-order cumulants using $\gamma_{xr}$, not possessed by the conventional algorithms using the constraint $\sigma_x^2 = \sigma_s^2$ in (4). The computer simulations have shown that the novel property of the proposed criterion works effectively for the blind equalization in the presence of Gaussian noise.

## Acknowledgment

The authors would like to thank the reviewer for his useful comments.

### References

[1] F. Herrmann and A.K. Nandi, "Reduced computation blind super-exponential equaliser," Electron. Lett., vol.34, no.23, pp.2208–2209, Nov. 1998.
[2] B. Jelonnek and K.D. Kammeyer, "A closed-form solution to blind equalization," Signal Process., vol.36, no.3, pp.251–259, 1994.
[3] A.K. Nandi, Blind Estimation Using Higher-Order Statistics, Kluwer Academic Publishers, Dordrecht, 1999.
[4] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase system (channels)," IEEE Trans. Inf. Theory, vol.36, no.2, pp.312–321, 1990.
[5] O. Shalvi and E. Weinstein, "Super-exponential methods for blind deconvolution," IEEE Trans. Inf. Theory, vol.39, no.2, pp.504–519, 1993.

## Appendix A: The Derivation of (12)

The matrix $\boldsymbol{C}_{yr}$ in the case of $2 \times 2$ is

$$\boldsymbol{C}_{yr} = \begin{bmatrix} \mathrm{cum}\{y_1, y_1, r, r\} & \mathrm{cum}\{y_1, y_2, r, r\} \\ \mathrm{cum}\{y_2, y_1, r, r\} & \mathrm{cum}\{y_2, y_2, r, r\} \end{bmatrix}, \qquad \text{(A· 1)}$$

where $y_1 = y(t)$, $y_2 = y(t-1)$, and $r = r(t)$. Then (A· 1) can be expressed as

$$\boldsymbol{C}_{yr} = \begin{bmatrix} \mathrm{E}[y_1^2 r^2] & \mathrm{E}[y_1 y_2 r^2] \\ \mathrm{E}[y_2 y_1 r^2] & \mathrm{E}[y_2^2 r^2] \end{bmatrix} - 2\begin{bmatrix} \mathrm{E}[y_1 r]\mathrm{E}[y_1 r] & \mathrm{E}[y_1 r]\mathrm{E}[y_2 r] \\ \mathrm{E}[y_2 r]\mathrm{E}[y_1 r] & \mathrm{E}[y_2 r]\mathrm{E}[y_2 r] \end{bmatrix}$$

$$- \begin{bmatrix} \mathrm{E}[y_1^2]\mathrm{E}[r^2] & \mathrm{E}[y_1 y_2]\mathrm{E}[r^2] \\ \mathrm{E}[y_2 y_1]\mathrm{E}[r^2] & \mathrm{E}[y_2^2]\mathrm{E}[r^2] \end{bmatrix}$$

$$= \mathrm{E}[r^2 \boldsymbol{y}\boldsymbol{y}^T] - 2\mathrm{E}[\boldsymbol{y}r]\mathrm{E}[\boldsymbol{y}^T r] - \mathrm{E}[r^2]\mathrm{E}[\boldsymbol{y}\boldsymbol{y}^T]. \quad \text{(A· 2)}$$

Therefore, from (A· 2), we can estimate $\boldsymbol{C}_{yr}$ by using (12).

## Appendix B: The Derivation of (15)

To simplify the notation of the matrix $\boldsymbol{Q}$, we consider a $2 \times 2$ matrix case, that is, $\boldsymbol{y} := [y(t), y(t-1)]^T$ and the summation variable $l$ is set to be $l = 0$ and $l = 1$. Then, the matrix $\boldsymbol{Q}$ becomes

$$\boldsymbol{Q} = \sum_{i=1}^{2} \begin{bmatrix} \mathrm{cum}\{y_1, y_1, y_i, y_i\} & \mathrm{cum}\{y_1, y_2, y_i, y_i\} \\ \mathrm{cum}\{y_2, y_1, y_i, y_i\} & \mathrm{cum}\{y_2, y_2, y_i, y_i\} \end{bmatrix} \quad \text{(A· 3)}$$

where $y_1 = y(t)$ and $y_2 = y(t-1)$. Moreover, $\boldsymbol{Q}$ can be expressed as

$$\boldsymbol{Q} = \sum_{i=1}^{2} \left\{ \begin{bmatrix} \mathrm{E}[y_1^2 y_i^2] & \mathrm{E}[y_1 y_2 y_i^2] \\ \mathrm{E}[y_2 y_1 y_i^2] & \mathrm{E}[y_2^2 y_i^2] \end{bmatrix} \right.$$
$$- 2 \begin{bmatrix} \mathrm{E}[y_1 y_i]\mathrm{E}[y_1 y_i] & \mathrm{E}[y_1 y_i]\mathrm{E}[y_2 y_i] \\ \mathrm{E}[y_2 y_i]\mathrm{E}[y_1 y_i] & \mathrm{E}[y_2 y_i]\mathrm{E}[y_2 y_i] \end{bmatrix}$$
$$\left. - \begin{bmatrix} \mathrm{E}[y_1^2]\mathrm{E}[y_i^2] & \mathrm{E}[y_1 y_2]\mathrm{E}[y_i^2] \\ \mathrm{E}[y_2 y_1]\mathrm{E}[y_i^2] & \mathrm{E}[y_2^2]\mathrm{E}[y_i^2] \end{bmatrix} \right\}$$

$$= \begin{bmatrix} \mathrm{E}[y_1^4] + \mathrm{E}[y_1^2 y_2^2] & \mathrm{E}[y_1^3 y_2] + \mathrm{E}[y_1 y_2^3] \\ \mathrm{E}[y_1^3 y_2] + \mathrm{E}[y_1 y_2^3] & \mathrm{E}[y_1^2 y_2^2] + \mathrm{E}[y_2^4] \end{bmatrix}$$
$$- 2 \begin{bmatrix} \mathrm{E}[y_1^2]\mathrm{E}[y_1^2] + \mathrm{E}[y_1 y_2]\mathrm{E}[y_1 y_2] & \mathrm{E}[y_1^2]\mathrm{E}[y_1 y_2] + \mathrm{E}[y_1 y_2]\mathrm{E}[y_2^2] \\ \mathrm{E}[y_1 y_2]\mathrm{E}[y_1^2] + \mathrm{E}[y_2^2]\mathrm{E}[y_1 y_2] & \mathrm{E}[y_1 y_2]\mathrm{E}[y_1 y_2] + \mathrm{E}[y_2^2]\mathrm{E}[y_2^2] \end{bmatrix}$$
$$- \begin{bmatrix} \mathrm{E}[y_1^2](\mathrm{E}[y_1^2] + \mathrm{E}[y_2^2]) & \mathrm{E}[y_1 y_2](\mathrm{E}[y_1^2] + \mathrm{E}[y_2^2]) \\ \mathrm{E}[y_2 y_1](\mathrm{E}[y_1^2] + \mathrm{E}[y_2^2]) & \mathrm{E}[y_2^2](\mathrm{E}[y_1^2] + \mathrm{E}[y_2^2]) \end{bmatrix} \quad \text{(A· 4)}$$

$$= \mathrm{E}\left[ \begin{bmatrix} y_1^2 & y_1 y_2 \\ y_2 y_1 & y_2^2 \end{bmatrix} \begin{bmatrix} y_1^2 & y_1 y_2 \\ y_2 y_1 & y_2^2 \end{bmatrix} \right]$$
$$- 2 \begin{bmatrix} \mathrm{E}[y_1^2] & \mathrm{E}[y_1 y_2] \\ \mathrm{E}[y_2 y_1] & \mathrm{E}[y_2^2] \end{bmatrix} \begin{bmatrix} \mathrm{E}[y_1^2] & \mathrm{E}[y_1 y_2] \\ \mathrm{E}[y_2 y_1] & \mathrm{E}[y_2^2] \end{bmatrix}$$
$$- (\mathrm{E}[y_1^2] + \mathrm{E}[y_2^2]) \begin{bmatrix} \mathrm{E}[y_1^2] & \mathrm{E}[y_1 y_2] \\ \mathrm{E}[y_2 y_1] & \mathrm{E}[y_2^2] \end{bmatrix}$$
$$= \mathrm{E}[\boldsymbol{y}\boldsymbol{y}^T \boldsymbol{y}\boldsymbol{y}^T] - 2\mathrm{E}[\boldsymbol{y}\boldsymbol{y}^T]\mathrm{E}[\boldsymbol{y}\boldsymbol{y}^T] - tr\{\mathrm{E}[\boldsymbol{y}\boldsymbol{y}^T]\}\mathrm{E}[\boldsymbol{y}\boldsymbol{y}^T]. \quad \text{(A· 5)}$$

Therefore, from (A· 5), we can estimate $\boldsymbol{Q}$ by using (15).