POLYNOMIAL MATRIX APPROACH TO INDEPENDENT COMPONENT ANALYSIS: (PART I) BASICS

Kenji Sugimoto * **Masuhiro Nitta** *

* *Graduate School of Information Science, Nara Institute of Science and Technology, Nara, 630-0192, JAPAN*

Abstract: This paper proposes a method for blind system identification based on the independence of input signals. Under the assumption that the system is MIMO, square, and represented by a polynomial matrix fraction with constant numerator matrix, the method makes it possible to identify the system without observation of input signals. This rather challenging problem is solved by applying independent component analysis to an augmented state-space representation in order to estimate coefficients of the denominator polynomial matrix and the numerator matrix.

Keywords: blind system identification, vector autoregressive model, independent component analysis, polynomial matrix analysis

1. INTRODUCTION

Independent Component Analysis (ICA) has attracted much attention as a powerful tool for separating signals in terms of statistical independence, and has been applied in various fields (Amari *et al.*, 1996; Bell and Sejnowski, 1995; Cardoso and Souloumiac, 1993; Karhunen *et al.*, 1997; Hyvärinen and Oja, 1997; Yang and Amari, 1997). It is desirable if we can use ICA in system control engineering. In fact, Kano (Kano *et al.*, 2002) has applied ICA to process monitering.

In this paper, together with the companion paper (Sugimoto *et al.*, 2005), we study a polynomial matrix approach to ICA. In Paper (I), we provide a method for identifying a certain class of multi-input multi-output (MIMO) systems with unknown (but mutually independent) input signals. In Paper (II) we then apply the method to a couple of practical issues.

[1] This work is supported in part by Scientific Research Grant-in-Aid 15560379 from the Japan Society for the Promotion of Science.

In ordinary system identification, we use both the input and the output signals. However, in some cases all input signals are not available due to disturbance or input saturation, etc. Even so, it is natural to expect that those input signals arise independently. We achieve this "blind" identification, namely we identify the system without observing input signals.

A similar problem has been studied also in the area of speech processing (Amari *et al.*, 1996; Bell and Sejnowski, 1995; Cardoso and Souloumiac, 1993; Yang and Amari, 1997), in which input signals are mixed through a transfer matrix, and then we retrieve the input signals only from their mixture. This problem is called BSD (Blind Source Deconvolution), while it is called BSS (Blind Source Separation) if the mixing matrix is a constant. The latter is much easier.

In control, we have to treat dynamical systems, and hence have to solve the BSD problem. In (Dapena and Serviere, 2001; Mitianoudis and Davies, 2001), they have proposed to analyze it in the frequency domain thereby reducing the problem to BSS. This is particularly effective in speech processing, since the frequency band is wide enough. In control, however, the band is rather restricted and we may sometimes wish to obtain a model even from transient signals.

In this paper, we assume that the system is described by a polynomial matrix fraction with constant numerator matrix, and reduce the discrete-time convolutive mixture to a static mixture with fixed elements, by using an augmented state-space representation (Kikkawa and Sugimoto, 2002), and then estimate unfixed elements by an ICA method.

In what follows we briefly review what is ICA (§2) and formulate a vector autoregressive model (§3). Then we show our blind identification algorithm (§4) and carry out numerical simulation to show the effectiveness of our method (§5).

2. PRELIMINARIES

Let us first summarize the concept of ICA and an algorithm to solve it for the static mixture case. Consider a linear mixture

$$\eta(t) = As(t) \tag{1}$$

where $s(t) = (s_1(t), \ldots, s_n(t))^T$ is a source signal vector, $\eta(t) = (\eta_1(t), \ldots, \eta_n(t))^T$ is an observed signal vector, and $A = (a_{ij})$ is a mixing matrix with unknown coefficients. Here, T denotes transposition. We assume that A is nonsingular. Our objective is to retrieve the source signal $s(t)$ only from the mixed signal $\eta(t)$ under the assumption that the components of $s(t)$ take values statistically independent.

Definition 1. The random vector $\xi = (\xi_1, \ldots, \xi_n)^T$ is said to be *independent* if its joint probability density function (p. d. f) is equal to the product of marginal p. d. f. of all entries $\xi_i, i = 1, \ldots, n$. Namely,

$$p_\xi(\xi) = \prod_{i=1}^{n} p_{\xi_i}(\xi_i), \tag{2}$$

where p_{ξ_i} denotes the p. d. f. of ξ_i. □

If we find $W = A^{-1}$, then the source signal can be retrieved as

$$\hat{s}(t; W) = W\eta(t). \tag{3}$$

In reality, however, A is unknown, so that we have to search W such that $\hat{s}(t)$ is independent. [2]

[2] The independence holds even if the components of $s(t)$ are in an incorrect order or have incorrect magnitude. Namely, we can only obtain $W = \Gamma P A^{-1}$ up to some diagonal matrix Γ and permutation matrix P.

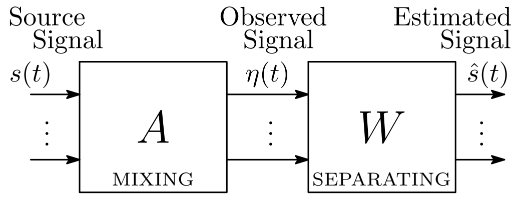

Fig. 1. Block diagram of standard ICA.

This is achieved by evaluating the independence. Yang et al. (Yang and Amari, 1997) have proposed to use Kullback-Leibler Divergence (KLD)

$$I(\hat{s}; W) = \int p_{\hat{s}}(\hat{s}; W) \log \frac{p_{\hat{s}}(\hat{s}; W)}{\prod_{i=1}^{n} p_{\hat{s}_i}(\hat{s}_i; W)} \mathrm{d}\hat{s}_i. \tag{4}$$

It is clear that $I \geq 0$ and the equality holds iff $\hat{s}$ is independent. Hence we need only to update W so that $I(\hat{s}; W)$ decreases.

By means of the entropy defined below, we have

$$I(\hat{s}; W) = \sum_{i=1}^{n} H(\hat{s}_i; W) - H(\hat{s}; W) \geq 0. \tag{5}$$

Definition 2. *(Differential) entropy* $H(x)$ is defined by

$$H(x) = -\int p(x) \log p(x) \mathrm{d}x =: \mathrm{E}_p(-\log p(x)). \tag{6}$$

□

In view of $\hat{s} = W\eta$, the equation (5) becomes

$$I(\hat{s}; W) = \sum_{i=1}^{n} H(\hat{s}_i; W) - H(\eta) - \log|W|, \tag{7}$$

Let us compute partial derivatives of the terms of (7) with respect to W. First we have

$$\frac{\partial}{\partial W} H(\hat{s}_k; W) = \mathrm{E}_{p_\eta}\left(\phi(\hat{s}_k) e_k \eta^T\right), \tag{8}$$

where $e_k = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ is the natural basis whose k-th component is unity, and $\phi(\hat{s}_k)$ is defined by

$$\phi(\hat{s}_k) = -\frac{\mathrm{d} \log p(\hat{s}_k)}{\mathrm{d}\hat{s}_k}, \tag{9}$$

Hence in the first term of (7), we have

$$\frac{\partial}{\partial W} \sum_{k=1}^{n} H(\hat{s}_k; W) = \mathrm{E}_{p_\eta}\left(\phi(\hat{s}) \eta^T\right), \tag{10}$$

where we interpret $\phi(\hat{s})$ as

$$\phi(\hat{s}) = (\phi(\hat{s}_1), \phi(\hat{s}_2), \ldots, \phi(\hat{s}_n))^T. \tag{11}$$

In the third term of (7), we have

$$\frac{\partial}{\partial W} \log|W| = (W^T)^{-1} =: W^{-T}. \tag{12}$$

Thus the gradient becomes

$$
\begin{aligned}
\frac{\partial I(\hat{s};W)}{\partial W} &= \mathrm{E}_{p_\eta}\left(\phi(\hat{s})\eta^T\right) - W^{-T} \\
&= \left(\mathrm{E}_{p_{\hat{s}}}\left(\phi(\hat{s})\hat{s}^T\right) - I\right) W^{-T}. \quad (13)
\end{aligned}
$$

Since $\phi(\hat{s})$ is unknown, we usually substitute it for $\phi(x) = \tanh x$ or x^3. Amari et al.(Amari *et al.*, 1996) has proposed the following update law using the natural gradient.

$$
\begin{aligned}
\Delta W &= -\frac{\partial I(\hat{s};W)}{\partial W} W^T W \\
&= \left(I - \mathrm{E}_{p_{\hat{s}}}\left(\phi(\hat{s})\hat{s}^T\right)\right) W. \quad (14)
\end{aligned}
$$

3. MODEL DESCRIPTION AND AUGMENTED STATE-SPACE REPRESENTATION

Consider the discrete-time system

$$y(t) = G(z)u(t), \quad (15)$$

where

$$u(t) = (u_1(t), \ldots, u_m(t))^T \in \mathbf{R}^m, \quad (16)$$
$$y(t) = (y_1(t), \ldots, y_m(t))^T \in \mathbf{R}^m, \quad (17)$$

are respectively input and output signal vectors with $m > 1$, and $G(z)$ is assumed to be a stable and minimum phase transfer matrix of full normal rank. Here t means discrete-time and z is the unit time shift operator meaning $zu(t) = u(t+1)$. We assume that $G(z)$ and $u(t)$ are both unknown, but that $u(t)$ is independent in each time.

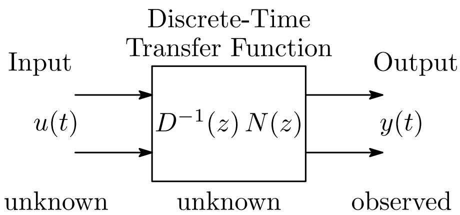

Fig. 2. Dynamical system to be identified.

This is a blind identification problem, where we estimate the parameter of $G(z)$ and the signal $u(t)$ only from $y(t)$ (Fig. 2). At first sight, it appears to be solvable by means of existing ICA methods. However, the algorithm in §2 is valid only when $G(z)$ is a constant as in (1). To overcome this difficulty, we confine ourselves to the case where the system to be identified is represented by a left coprime factorization

$$G(z) = D^{-1}(z)N(z), \quad (18)$$

$$D(z) = \begin{pmatrix} d_1(z) \\ \vdots \\ d_m(z) \end{pmatrix}, \quad (19)$$

$$d_i(z) = e_i z^{\mu_i} + d_i^1 z^{\mu_i - 1} + \cdots + d_i^{\mu_i},$$

$$N(z) = \begin{pmatrix} z^{\mu_1} & & 0 \\ & \ddots & \\ 0 & & z^{\mu_m} \end{pmatrix} N_0, \quad (20)$$

where the row degrees $\mu_1, \cdots, \mu_m$ of polynomial matrices are assumed to be known, d_i^k's are arbitrary row vectors of dimension m and $\{e_1, \cdots, e_m\}$ is the natural basis. Note that $D(z)$ is in the so-called row-reduced form with $[D]_r = I$.

Note that $N_0 = (n_{ij}^0)$ is nonsingular by assumption. We further assume that

$$n_{ii}^0 = 1 \quad \text{and} \quad |n_{ij}^0| < 1 \quad (i \neq j), \quad (21)$$

in order to avoid indefiniteness in magnitude when applying ICA. Now we have

$$D(z)y(t) = N(z)u(t), \quad (22)$$

which is equivalent to

$$
\begin{aligned}
y_i(t) = &-d_i^1 y(t-1) - \cdots - d_i^{\mu_i} y(t-\mu_i) \\
&+ e_i N_0 u(t), \quad (23)
\end{aligned}
$$

for $i = 1, \cdots, m$. We further reduce it to the following "augmented state-space representation".

$$
\underbrace{\begin{pmatrix} y(t-\mu) \\ \vdots \\ y(t-1) \\ y(t) \end{pmatrix}}_{\eta(t)} = \underbrace{\begin{pmatrix} I & O \\ -\tilde{D} & N_0 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} y(t-\mu) \\ \vdots \\ y(t-1) \\ u(t) \end{pmatrix}}_{s(t)}, \quad (24)
$$

$$\tilde{D} = \begin{pmatrix} O & & d_1^{\mu_1} & \cdots & d_1^1 \\ \vdots & & & & \vdots \\ O & d_m^{\mu_m} & \cdots & & d_m^1 \end{pmatrix}, \quad (25)$$

with O denoting zero matrices or zero vectors of appropriate dimensions, I denoting the identify matrix, and $\mu := \max \mu_i$. Note that $\tilde{D}$ is an $m \times m\mu$ matrix.

4. IDENTIFICATION ALGORITHM

In later development we confine ourselves to the case of uniform row degrees; i.e., $\mu_1 = \cdots = \mu_m$ for simplicity.

We first rewrite (24) as

$$\begin{pmatrix} \tilde{y}(t) \\ y(t) \end{pmatrix} = A \begin{pmatrix} \tilde{y}(t) \\ u(t) \end{pmatrix}, \; A := \begin{pmatrix} I & O \\ -\tilde{D} & N_0 \end{pmatrix}, \quad (26)$$

where we put

$$\tilde{y}(t) = \begin{pmatrix} y(t-\mu) \\ y(t-\mu+1) \\ \vdots \\ y(t-1) \end{pmatrix}. \quad (27)$$

Note that we cannot apply the method in §2 to $\eta(t)$ in (24), since in this case we obtain incorrect $\hat{s}(t;W) = W\eta(t)$ due to the fixed-element structure in the mixing matrix. To remedy this, we first consider the first term of (5). $H(\hat{s}_i;W)$ should not be affected by W for $i = 1,\ldots,\mu m$

Thus, from (8) we put

$$\phi(\hat{s}_i) = 0, \quad (i = 1,\ldots,\mu m) \quad (28)$$

for the known signals.

In the augmented state-space model, it is enough to take

$$\phi_\Pi(\hat{s}) = \Pi\phi(\hat{s}), \quad (29)$$

where

$$\Pi = \mathrm{block\,diag}(O,\ldots,O,I). \quad (30)$$

Now we proceed to $H(\hat{s};W)$ in the second term of (5). The change of $\hat{s}$ by W is

$$\hat{s} + \mathrm{d}\hat{s} = (I + \mathrm{d}WW^{-1})\hat{s},$$

hence in ordinary ICA context the change of $H(\hat{s};W)$ is

$$\begin{aligned} \mathrm{d}H &= \mathrm{E}(-\log p(\hat{s}+\mathrm{d}\hat{s})) - \mathrm{E}(-\log p(\hat{s})) \\ &= \mathrm{E}\big(-\log p(\hat{s})/|I+\mathrm{d}WW^{-1}|\big) - \mathrm{E}(-\log p(\hat{s})) \\ &= \log|I+\mathrm{d}WW^{-1}| \simeq \mathrm{tr}\,(\mathrm{d}WW^{-1}). \end{aligned} \quad (31)$$

On the other hand, if we express

$$\mathrm{d}W = \begin{pmatrix} \mathrm{d}W_{11} & \mathrm{d}W_{12} \\ \mathrm{d}W_{21} & \mathrm{d}W_{22} \end{pmatrix}, \quad (32)$$

in the augmented state-space model, the seperation matrix should be

$$W_\Pi = A^{-1} = \begin{pmatrix} I & O \\ N_0^{-1}\tilde{D} & N_0^{-1} \end{pmatrix}, \quad (33)$$

so that we have to consider

$$\mathrm{d}W_\Pi = \begin{pmatrix} O & O \\ \mathrm{d}W_{21} & \mathrm{d}W_{22} \end{pmatrix} = \Pi\mathrm{d}W. \quad (34)$$

Therefore we have

$$\mathrm{d}H_\Pi = \mathrm{tr}\,(\mathrm{d}W_\Pi W^{-1}) = \mathrm{tr}\,(\mathrm{d}WW^{-1}\Pi), \quad (35)$$

and hence the gradient is given by

$$\frac{\partial}{\partial W}H_\Pi = (W^{-1}\Pi)^T = \Pi W^{-T}. \quad (36)$$

We thus obtain the natural gradient

$$\begin{aligned} \Delta W_\Pi &= -\left(\mathrm{E}_{p_{\hat{s}}}\big(\phi_\Pi(\hat{s})\eta^T\big) - \Pi W^{-T}\right)W^TW \\ &= \Pi\left(I - \mathrm{E}_{p_{\hat{s}}}\big(\phi(\hat{s})\hat{s}^T\big)\right)W, \end{aligned} \quad (37)$$

correspondingly to (24), and the gradient is given by $\Delta W_\Pi/\|\Delta W_\Pi\|$. In actural implementation, we simply adopt the sample mean value instead of $\mathrm{E}_{p_{\hat{s}}}(\cdot)$ in order to avoid large estimation cost.

Proposed Algorithm

(1) Obtain observation signal $y(t)$ as discrete-time sequence data from $t = 0$ to $M-1$ and give the degree μ of the autoregressive mode.

(2) For each time t, obtain the output signal vector $\eta(t)$ in the augmented state-space model (24).

(3) Construct the matrix

$$Y = (\eta(0),\ldots,\eta(M-1)) \in \mathbf{R}^{(\mu+1)n\times M}$$

and normalize it so that the time mean value of each row equals 0, and put it as Y again.

(4) Give the identity I as an initial value of separation matrix W. This is in accordance with the fact that the mixing matrix A in (26) has fixed block elements I.

(5) Renew the separation matrix W as:

$$\begin{aligned} \hat{S} &= WY, \\ \Delta W &= \Pi\left(I - \phi(\hat{S})\hat{S}^T/M\right)W, \\ W &\leftarrow W + \alpha\Delta W/\|\Delta W\|. \end{aligned}$$

Here α is a learning factor (step size parameter) which is positive and decreases with repetition.

(6) The learning process on W converges if

$$\left\|\Pi\left(I - \phi(\hat{S})\hat{S}^T/M\right)\right\| \to 0.$$

Otherwise go to Step (5).

(7) Retrieve the mixing matrix by $A = W^{-1}$.

(8) Normalize the column vectors and the amplitude so that mixing matrix A is equal to that of (24).

(9) From A, obtain the corresponding $\{D_i\}$ and N_0.

The normalization in Step (8) means that for $A = (a_{ij})$, we put $a_{ik} = a_{ik}/a_{kk}$ for the k-th column and $\hat{S}_k = a_{kk}\hat{S}_k$ for the source signal. Since the diagonal element in A is restricted to 1, we can retrieve the magnitude of the source signal by this normalization.

Concerning $\phi(x)$, it is said in literature that $\tanh x$ and x^3 are respectively effective if the input signal is super-Gaussian and sub-Gaussian. Since we do not know their real distribution, we need to switch those functions accordingly to the sign of 4-th order cumulant estimation.

Discussion on Convergence

Let us derive a condition under which the above algorithm gives the true separation matrix (33) as a local minimum. In view of (37), it is enough to see when $\Delta := \mathit{\Pi}\left(I - \mathrm{E}_{p_{\hat{s}}}\left(\phi(\hat{s})\hat{s}^T\right)\right)$ is a zero matrix.

Denoting $x_k := x(t-k)$ for simplicity, the estimated signal is written as $\hat{s}_0 = (y_\mu^T, \ldots, y_1^T, u_0^T)^T$. Then we have

$$\Delta = \begin{bmatrix} O & & & O \\ & \ddots & & \\ & & O & O \\ \mathrm{E}\left(\phi(u_0)y_\mu^T\right) & \ldots & \mathrm{E}\left(\phi(u_0)y_1^T\right) & I - \mathrm{E}\left(\phi(u_0)u_0^T\right) \end{bmatrix}. \tag{38}$$

Hence $\Delta = O$ iff $\mathrm{E}\left(\phi(u_0)y_k^T\right) = O \quad (k = 1, \ldots, \mu)$ and $I - \mathrm{E}\left(\phi(u_0)u_0^T\right) = O$.

From the former equality, u_0 must be a white noise in order that we have $\mathrm{E}\left(\phi(u_0)y_k^T\right) = O$ for $k \neq 0$.

Concerning the latter equality, since u_0 is independent, we see that the each component u_0^i, u_0^j $(i \neq j)$ are uncorrelated and zero mean, and hence we have

$$\mathrm{E}\left(\phi(u_0^i)u_0^j\right) = \mathrm{E}\left(\phi(u_0^i)\right)\mathrm{E}\left(u_0^j\right) = 0. \tag{39}$$

On the other hand, for $i = j$ we have

$$\mathrm{E}\left(\phi(u_0^i)u_0^i\right) = -\int_{-\infty}^{\infty} \frac{\mathrm{d}p(u_0^i)}{\mathrm{d}u_0^i} u_0^i \mathrm{d}u_0^i = 1, \tag{40}$$

and hence $I - \mathrm{E}\left(\phi(u_0)u_0^T\right) = O$.

Therefore, in order that (33) is a local minimum, the input signal $u(t)$ must be a white signal.

5. NUMERICAL SIMULATION

In what follows we assume the sampling frequency 100 Hz and we represent estimated values up to the second order below the decimal point.

Consider the case

$$D(z) = z^2 I + D_1 z + D_2,$$
$$D_1 = \begin{pmatrix} 0.0 & 1.9 \\ -0.6 & -0.2 \end{pmatrix}, D_2 = \begin{pmatrix} -1.4 & -0.6 \\ -0.3 & -0.1 \end{pmatrix},$$
$$N_0 = \begin{pmatrix} 1.0 & -0.7 \\ 0.4 & 1.0 \end{pmatrix}.$$

As an independent but unknown input, we have applied a random noise and M-sequence shown in Fig. 3 whose data consist of 256 points from time 0 s to 2.55 s. As a result, the system response shown in Fig. 4 has been observed.

By taking $\phi(x) = x^3$ and as learning factor α an exponential decreasing function having values from 10^{-1} to 10^{-3}, we have obtained:

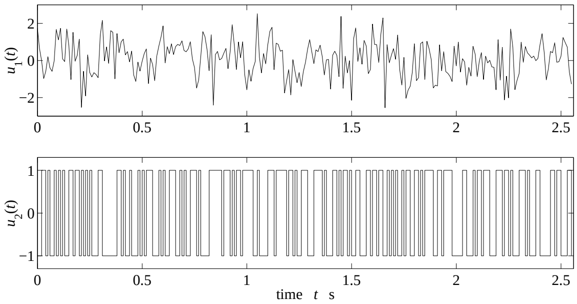

Fig. 3. Input signal.

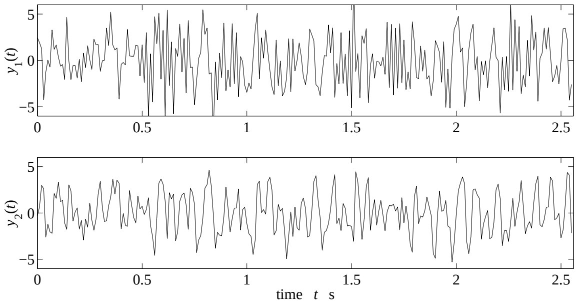

Fig. 4. Observed signal.

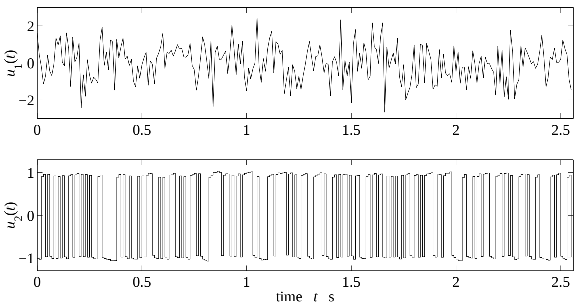

Fig. 5. Retrieved input signal.

$$A = \begin{pmatrix} 1.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 \\ 1.44 & 0.60 & 0.02 & -1.88 & 1.00 & -0.78 \\ 0.34 & 0.12 & 0.62 & 0.18 & 0.41 & 1.00 \end{pmatrix}.$$

We have thus obtained the following estimation.

$$D_1 = \begin{pmatrix} -0.02 & 1.88 \\ -0.62 & -0.18 \end{pmatrix}, \; D_2 = \begin{pmatrix} -1.44 & -0.60 \\ -0.34 & -0.12 \end{pmatrix},$$
$$N_0 = \begin{pmatrix} 1.00 & -0.78 \\ 0.41 & 1.00 \end{pmatrix}.$$

Together with parameter estimation we can retrieve the input signal. Fig. 5 indicates that the input and retrieved signals are in good coincidence.

Fig. 6 shows the Bode diagram of each entry of the obtained transfer matrix. The true values are in solid line and the estimated values are in dotted line, but they are almost identical.

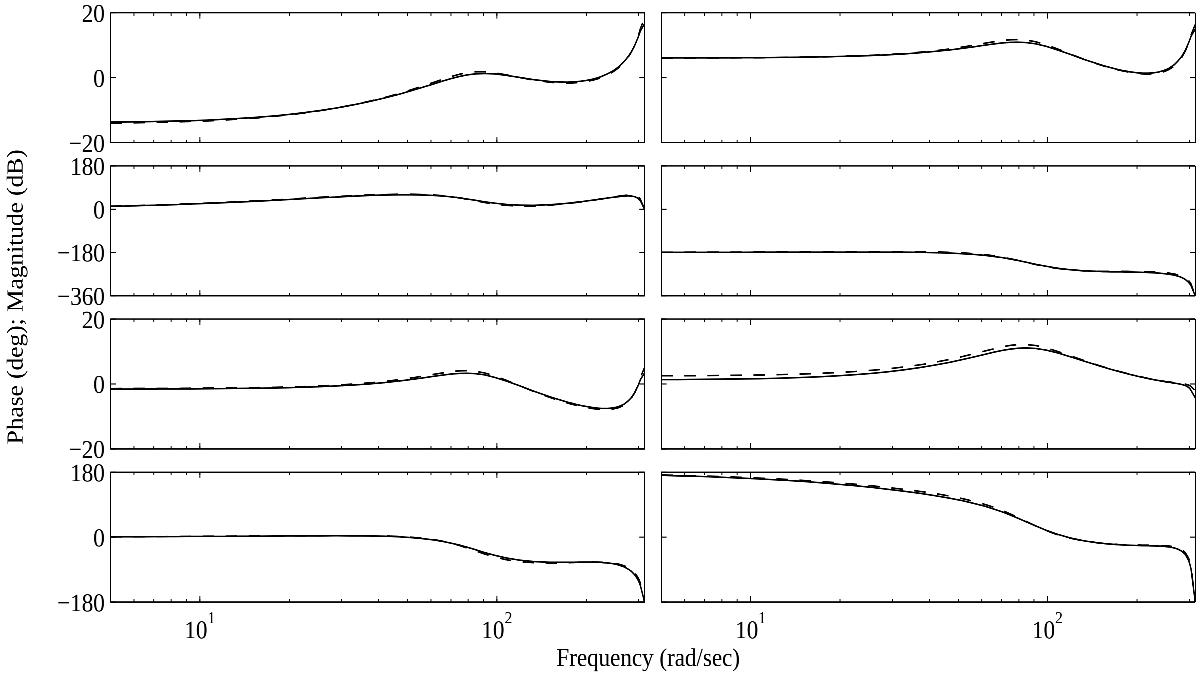

Fig. 6. Bode diagram (gain and phase) of element transfer functions (each entry of 2×2 matrix) .

6. CONCLUSION

To conclude, a few remarks are given below.

We have assumed the structure (20) in the numerator polynomial matrix. In other words, we have treated a kind of vector AR (Autoregressive) model. This is because resorting to the augmented state-space representation is equivalent to using a Moving Average (MA) model as a demixing system, but we omit a detail here. Relaxing this assumption is by no means trivial, but the authors are currently tackling this problem.

Blind identification is also possible by conventional methods such as the LS (Least Squares) method. However, we can not obtain N_0 by this method. Also, numerical simulation indicates that the proposed method is superior in accuracy to the LS method.

REFERENCES

Amari, S., A. Cichocki and H. Yang (1996). *A New Learning Algorithm for Blind Signal Separation.* MIT Press. Cambridge.

Bell, A. and T. Sejnowski (1995). An information maximisation approach to blind separation and blind deconvolution. *Neural Computation* **7**(6), 1129–1159.

Cardoso, J. and A. Souloumiac (1993). Blind beamforming for non gaussian signals. *IEE Proceedings F* **140**(6), 362–370.

Dapena, A. and C. Serviere (2001). A simplified frequency-domain approach for blind separation of convolutive mixtures. *3rd International Conference on Independent Component Analysis and Blind Source Separation (ICA2001).*

Hyvärinen, A. and E. Oja (1997). A fast fixed-point algorithm for independent component analysis. *Neural Computation* **9**(7), 1483–1492.

Kano, M., S. Tanaka, H.Maruta, S. Hasebe, I. Hashimoto and H. Ohno (2002). Statistical process monitoring with external analysis and independent component analysis. *SICE* **38**(11), 958–965.

Karhunen, J., E. Oja, L. Wang, R. Vigario and J. Joutsensalo (1997). A class of neural networks for independent component analysis. *IEEE Trans. on Neural Networks* **8**(3), 486–503.

Kikkawa, Y. and K. Sugimoto (2002). Independent component analysis for signals mixed by a class of dynamical systems. *ICASE/SICE Joint Workshop.*

Mitianoudis, N. and M. Davies (2001). New fixed-point solutions for convolved mixtures. *3rd International Conference on Independent Component Analysis and Blind Source Separation (ICA2001).*

Sugimoto, K., A. Suzuki, M. Nitta and N. Adachi (2005). Polynomial matrix approach to independent component analysis: (part II) application. *16th IFAC World Congress.*

Yang, H. and S. Amari (1997). Adaptive online learning algorithms for blind separation: Maximum entropy and minimum mutual information. *Neural Computation* **9**(7), 1457–1482.