A Closed-Form Identification of Multichannel Moving Average Processes by ESPRIT*

Lang Tong[1] and Rueywen Liu[2]

Abstract. In this paper, a closed-form identification of possibly nonminimum phase multi-channel moving average (MA) processes is derived by exploiting the eigenstructures of the observation cumulant matrices using the ESPRIT algorithm. The proposed approach allows the combination of statistics of different orders for better performance and offers reduced computation complexity when compared with existing iterative approaches. Simulations are also presented to demonstrate the performance of the proposed algorithm.

1. Introduction

Multichannel moving average (MA) modeling is widely used in many scientific and engineering applications. In this paper, we consider the identification and estimation of an Nth-order MA process

$$\mathbf{y}(t) = \sum_{k=0}^{N} \mathbf{H}(k)\mathbf{s}(t-k) + \mathbf{n}(t) . \tag{1}$$

In many literatures, $\mathbf{H}(0)$ is often assumed to be identity. Such an assumption is indeed without loss of generality if $\mathbf{H}(0)$ is known, or if we are only interested in the statistical structure of the observation $\mathbf{y}(t)$, not that of the source $\mathbf{s}(t)$. On the other hand, it is desirable in many practical applications to extract information about $\mathbf{s}(t)$ from the observation process, and $\mathbf{H}(0)$ is often unknown. Among many applications, the widely studied array signal processing formulation assumes a zero-order multichannel MA process [5]. The so-called blind identification requires the identification of $\{\mathbf{H}(k)\}$ *and* $\mathbf{s}(t)$ (see, e.g. [19], [17]). Factor analysis is yet another example [1].

[1] Department of Electrical and Systems Engineering, University of Connecticut, Storrs, Connecticut 06269-3157.
[2] Department of Electrical Engineering, University of Notre Dame, Notre Dame, Indiana 46556.

It is well known that the identification and estimation of a multichannel MA process can only be achieved up to an all-pass factor when the second-order statistics of the observation are used. One of the advantages of parameter estimation using higher-order statistics is that the parameters of a *nonminimum-phase* system can be identified and estimated. Unfortunately, there is a lack of simple estimation algorithms because of the nonlinear nature of the problem. Unlike the single channel case when a *closed-form solution* can be obtained [14], the closed-form solution using cumulants of arbitrary orders, to our best knowledge, has so far not been found. Using the third-order cumulants, Giannakis et al. [3] derived the identification equations, referred to as cumulant-based identification (CBI) equations, for a multichannel MA process and obtained the identifiability results. The CBI equations and identifiability results of using fourth-order and arbitrary-order cumulants can be found in [13], [2], [15], and [6].

Unfortunately, the CBI equations have not been solved satisfactorily. Most existing methods rely on some kind of iterative procedures such as the Newton Raphson type of algorithms, and the convergence properties of these algorithms are unknown and often problematic. In [16], we proposed an eigenstructure-based iterative scheme that *guarantees* to provide a solution of CBI equations in a finite number of iteration steps. By exploiting the eigenstructures of the observation cumulants, we showed that the CBI equations can be solved via singular value decomposition (SVD) of the cumulant matrices. However, as illustrated in the computer simulation [16], a rather large sample size, in the order of 70,000, is necessary to achieve satisfactory accuracy.

In this paper, we present a *closed-form solution* to the parameter identification of multichannel MA processes. By exploiting the eigenstructure in the cumulant matrices, the CBI equations are solved using the ESPRIT algorithm, a technique originally developed for array signal processing [9], [12], [11]. The proposed algorithm enables the combination of higher-order statistics and is less complex than most iterative schemes.

This paper is organized as follows. After a brief introduction of cumulants and basic properties, CBI equations are presented in Section 2. To avoid the computational cost of dealing with high-dimensional matrices, the CBI equations are presented without using Kronecker products. In Section 3, results of identifiability and a closed-form identification are presented. In Section 4, we then outline the proposed identification algorithm, discuss certain implementation issues, and present a computer simulation. We investigate the underlying algebraic problems and give related proofs in the Appendix. The notation here is mostly standard. Symbols for matrices (in capital letters) and vectors are usually in boldface. The notations $(\cdot)^t$ and $(\cdot)^\dagger$ stand for transpose and generalized Moore–Penrose inverse, respectively.

2. Identification equations

Cumulant-based identification (CBI) equations of arbitrary order were obtained

by Swami et al. in [13] using Kronecker products. Although compact and complete, the Kronecker-product form of CBI equations has a rather high dimension and presents a formidable task as far as solving these equations is concerned. In addition, it is difficult to exploit certain symmetrical properties of cumulants. The formulation given here, as a natural extension of the third-order [3] and fourth-order [6] cases, has a much lower dimension. More importantly, it allows us to exploit the symmetrical properties of the cumulants that lead to a closed-form identification.

2.1. Cumulants.

We shall not discuss a cumulant and its properties in detail. A rather complete recent survey can be found in [8], and the readers are also referred to a book by Rosenblatt [10]. Instead, we shall list some basic definitions and properties that are directly related to the later development.

Given n random variables $x_1, \ldots, x_n$, a kth-order cumulant is denoted as

$$\mathrm{cum}(\overbrace{z_1, \ldots, z_1}^{s_1 \text{ times}}, \ldots, \overbrace{z_n, \ldots, z_n}^{s_n \text{ times}}), \tag{2}$$

where $s_1 + \cdots + s_n = k$. A kth-order cumulant can be computed from the corresponding moments and vice versa. For most commonly used third-order and fourth-order cumulants of zero-mean random variables $x_1, \ldots, x_n$, relations between cumulants and moments are given by

$$\mathrm{cum}(x_i, x_j, x_k) = E(x_i x_j x_k), \tag{3}$$

$$\mathrm{cum}(x_i, x_j, x_k, x_l) = E(x_i x_j x_k x_l) - E(x_i x_j) E(x_k x_l) \tag{4}$$

$$- E(x_i x_k) E(x_j x_l) - E(x_i x_l) E(x_j x_k). \tag{5}$$

For a random vector $\mathbf{x} = [x_1, \ldots, x_n]^t$, the $(K+2)$th-order cumulants of $\mathbf{x}$ can be represented by a set of $(n \times n)$ cumulant matrices

$$\mathbf{C_x}(\alpha) = \begin{pmatrix} \mathrm{cum}(x_1, x_1, x_{\alpha_1}, \ldots, x_{\alpha_K}) & \cdots & \mathrm{cum}(x_1, x_n, x_{\alpha_1}, \ldots, x_{\alpha_K}) \\ \vdots & \vdots & \vdots \\ \mathrm{cum}(x_n, x_1, x_{\alpha_1}, \ldots, x_{\alpha_K}) & \cdots & \mathrm{cum}(x_n, x_n, x_{\alpha_1}, \ldots, x_{\alpha_K}) \end{pmatrix}, \tag{6}$$

where $\alpha = [\alpha_1, \ldots, \alpha_K]^t$ is a vector of integer indices ranging from 1 to K. For a zero-mean n-dimensional random vector $\mathbf{x}$, a third-order cumulant $\mathbf{C_x}(\alpha)$ is related to the third-order moments by

$$\mathbf{C_x}(\alpha) = [\mathrm{cum}(x_i, x_j, x_{\alpha_1})], \tag{7}$$

$$= E(\mathbf{x}\mathbf{x}^t x_{\alpha_1}), \quad \alpha_1 = 1, 2, \ldots, n. \tag{8}$$

Similarly, the fourth-order cumulants of $\mathbf{x}$ are represented by a set of n^2 cumulant matrices and they are related to the moments by

$$\mathbf{C_x}(\alpha) = [\mathrm{cum}(x_i, x_j, x_{\alpha_1}, x_{\alpha_2})], \tag{9}$$

$$=E\left(\mathbf{x}\mathbf{x}^{t}x_{\alpha_{1}}x_{\alpha_{2}}\right)-E\left(\mathbf{x}\mathbf{x}^{t}\right)E\left(x_{\alpha_{1}}x_{\alpha_{2}}\right) \tag{10}$$

$$-E\left(\mathbf{x}x_{\alpha_{1}}\right)E\left(\mathbf{x}^{t}x_{\alpha_{2}}\right)-E\left(\mathbf{x}x_{\alpha_{2}}\right)E\left(\mathbf{x}^{t}x_{\alpha_{1}}\right),\quad \alpha_{1},\alpha_{2}=1,2,\ldots,n. \tag{11}$$

We list here several properties of cumulants that are important in the later development. Their proofs are mostly immediate from the definitions of cumulants and can be found in [10].

Properties. *Given a real random vector* $\mathbf{x}=[x_{1},\ldots,x_{n}]^{t}$,

1. $\mathrm{cum}(a_{1}x_{1},\ldots,a_{n}x_{n})=a_{1}\cdots a_{n}\,\mathrm{cum}(x_{1},\ldots,x_{n})$, *for any real* $a_{1},\ldots,a_{n}$;
2. *If* $\mathbf{x}$ *is Gaussian, then for any* α, $\mathbf{C}_{\mathbf{x}}(\alpha)=\mathbf{0}$;
3. *If a random vector* $\mathbf{y}=[y_{1},\ldots,y_{n}]^{t}$ *is independent of* $\mathbf{x}$, *then*

$$\mathrm{cum}(y_{1}+x_{1},\ldots,y_{n}+x_{n})=\mathrm{cum}(y_{1},\ldots,y_{n})+\mathrm{cum}(x_{1},\ldots,x_{n}). \tag{12}$$

2.2. Assumptions.

We impose the following assumptions throughout this paper.

Assumptions. *For the multichannel MA process given by* (1),

- **A1.** $\{s(t)\}$ *is an i.i.d., non-Gaussian vector sequence with independent components and nonvanishing kth-order cumulants for some* $k>2$.

- **A2.** *The noise process* $\mathbf{n}(t)$, *independent of* $\mathbf{s}(t)$, *is Gaussian.*

- **A3.** *Both* $\mathbf{H}(0)$ *and* $\mathbf{H}(N)$ *are of full column rank.*

The assumption of $s(t)$ having nonvanishing kth-order cumulants can be relaxed to the assumption that each component of $\mathbf{s}(t)$ has a nonvanishing higher-order cumulant.

2.3. Cumulant-based identification equations.

Under the above assumptions and cumulant properties, one obtains a relation among a $(K+2)$th-order output cumulant, the MA coefficients, and the sources cumulants

$$\mathrm{cum}(y_{p}(t),y_{q}(t+\tau_{0}),y_{\alpha_{1}}(t+\tau_{1}),\ldots,y_{\alpha_{K}}(t+\tau_{K}))$$

$$=\sum_{k=0}^{N}\sum_{i=1}^{m}c_{s_{i}}^{(K+2)}h_{pi}(k)h_{qi}(\tau_{0}+k)\prod_{j=1}^{K}h_{a_{j}i}(\tau_{j}+k), \tag{13}$$

where $y_{i}(t)$ is the ith component of the vector $\mathbf{y}(t)$, $h_{ij}(k)$ is the (i,j)th entry of $\mathbf{H}(k)$, and

$$c_{s_{i}}^{(K+2)}=\mathrm{cum}(\overbrace{s_{i},\ldots,s_{i}}^{(K+2)\text{ times}}). \tag{14}$$

To abbreviate the notation, denote $\tau = [\tau_1, \ldots, \tau_K]^t$, and $\alpha = [\alpha_1, \ldots, \alpha_K]^t$. For any fixed α, τ_0, and τ, define an $(n \times n)$ cumulant matrix $\mathbf{C_y}(\tau_0; \tau, \alpha)$ whose (p, q)th entry is $\text{cum}(y_p(t), y_q(t + \tau_0), y_{\alpha_1}(t + \tau_1), \ldots, y_{\alpha_K}(t + \tau_K))$, i.e.,

$$\mathbf{C_y}(\tau_0; \tau, \alpha) = [\text{cum}(y_p(t), y_q(t + \tau_0), y_{\alpha_1}(t + \tau_1), \ldots, y_{\alpha_K}(t + \tau_K))] . \quad (15)$$

We then obtain the matrix form of the equations (13).

$$\mathbf{C_y}(\tau_0; \tau, \alpha) = \begin{cases} \displaystyle\sum_{k=0}^{N} \mathbf{H}(k)\mathbf{D}_k(\tau, \alpha)\mathbf{C_s}^{(K+2)}\mathbf{H}^t(\tau_0 + k) ; & \text{if } -N \leq \tau_0, \\ & \ldots, \tau_K \leq N \\ \mathbf{0} & \text{otherwise,} \end{cases} \quad (16)$$

where

$$\mathbf{D}_k(\tau, \alpha) = \text{diag}\left(\prod_{j=1}^{K} h_{\alpha_j 1}(\tau_j + k), \ldots, \prod_{j=1}^{K} h_{\alpha_j m}(\tau_j + k)\right) \quad (17)$$

$$\mathbf{C_s}^{(K+2)} = \text{diag}\left(c_{s_1}^{(K+2)}, \ldots, c_{s_m}^{(K+2)}\right) . \quad (18)$$

Evaluating $\mathbf{C_y}(\tau_0; \tau, \alpha)$ at $\tau = [N, \ldots, N]^t \overset{\Delta}{=} \tau(N)$, we obtain the so-called cumulant-based identification (CBI) equations.

Theorem 1 (Cumulant-Based Identification Equation). *For any integer $K > 0$ and an integer index vector α, the $(K + 2)$th-order output cumulant matrix $\mathbf{C_y}(k; \tau(N), \alpha)$ of an m-input n-output, Nth-order MA process is related to the process coefficients $\{\mathbf{H}(k)\}$ and input cumulants $\{c_{s_k}^{(K+2)}\}$ by the following CBI equation*

$$\mathbf{C_y}(k; \tau(N), \alpha) = \mathbf{H}(0)\mathbf{D}_0(N), \alpha)\mathbf{C_s}^{(K+2)}\mathbf{H}^t(k) . \quad (19)$$

It turns out that (19) contains all the information necessary for the identification of $\{\mathbf{H}(k)\}$, a fact to be shown in Section 3. The key results rest on the recognition of the special structures of (19). Before we proceed to discuss the issue of identifiability, a closer look at (19) for special cases is in order.

Third-order CBI equations. When third-order cumulants are used; i.e., $K = 1$, $\tau(N) = N$, and for any α, $1 \leq \alpha \leq n$, (19) becomes

$$\mathbf{C_y}(k; \tau(N), \alpha) = \mathbf{H}(0) \text{ diag}\left(h_{\alpha 1}(N)c_{s_1}^{(3)}, \ldots, h_{\alpha m}(N)c_{s_m}^{(3)}\right) \mathbf{H}^t(k) , \quad (20)$$

an equation presented in [3].

Fourth-order CBI equations. For the fourth-order case; i.e., $K = 2$, $\tau(N) = [N, N]^t$, and any $\alpha = [\alpha_i, \alpha_j]^t$, $1 \leq \alpha_i, \alpha_j \leq n$,

$$\mathbf{C_y}(k; \tau(N), \alpha) = \mathbf{H}(0) \text{ diag}\left(h_{\alpha_i 1}(N)h_{\alpha_j 1}(N)c_{s_1}^{(4)},\right.$$
$$\left.\ldots, h_{\alpha_i m}(N)h_{\alpha_j m}(N)c_{s_m}^{(4)}\right) \mathbf{H}^t(k) . \quad (21)$$

The above equation was derived in [6].

3. A closed-form solution of CBI equations

In this section, we concentrate on solving the CBI equations given in (19). We first address the issue of uniqueness. A closed-form solution is then presented.

3.1. Identifiability.

It is not difficult to see that, under the given assumptions, the parameter matrices $\{\mathbf{H}(k)\}$ can not be determined uniquely from observation alone. For example, the MA process generated from $\{\mathbf{H}(k)\}$ driven by $s(t)$ is the same as the process generated from $\{\mathbf{H}(k)\mathbf{P}\}$ driven by $\mathbf{P}^t s(t)$ for any permutation matrix $\mathbf{P}$. Similarly, the MA process generated from $\{\mathbf{H}(k)\}$ driven by $\mathbf{s}(t)$ is the same as the process generated from $\{\mathbf{H}(k)\mathbf{D}\}$ driven by $\mathbf{D}^{-1}\mathbf{s}(t)$ for any nonsingular diagonal matrix $\mathbf{D}$. The main result of this section is to show that, except for the above two types of ambiguity, the parameter matrices $\{\mathbf{H}(k)\}$ are uniquely determined by the CBI equations.

The following lemma plays a key role in establishing the identifiability and deriving the closed-form solution.

Lemma 1. *For almost any real sequence* $\{\delta_\alpha\}$,

$$\text{rank}\left(\sum_\alpha \delta_\alpha \mathbf{D}_0(\tau(N), \alpha)\right) = m . \tag{22}$$

Proof. The key of the proof is to relate $\mathbf{D}_0(\tau(N), \alpha)$ to the column vectors of the Kronecker product $\mathbf{H}^{[k]}(N)$, where

$$\mathbf{H}^{[K]}(N) = \overbrace{\mathbf{H}(N) \otimes \mathbf{H}(N) \cdots \otimes \mathbf{H}(N)}^{K \text{ times}} . \tag{23}$$

From (17), we have

$$\mathbf{D}_0(\tau(N), \alpha) = \text{diag}\left(\prod_{j=1}^K h_{\alpha_j 1}(N), \ldots, \prod_{j=1}^K h_{\alpha_j m}(N)\right) . \tag{24}$$

The diagonal entries of matrix $\mathbf{D}_0(\tau(N), \alpha)$ are in fact the (element-wise) products of row vectors of $\mathbf{H}(N)$. Given an index vector $\alpha = [\alpha_1, \ldots, \alpha_K]$, $1 \le \alpha_i \le n$, assign

$$\mathbf{v}_k \left[\prod_{j=1}^K h_{\alpha_j 1}(N), \ldots, \prod_{j=1}^K h_{\alpha_j m}(N)\right]^t , \tag{25}$$

where

$$k = \alpha_1 + (\alpha_2 - 1)n + \cdots + (\alpha_K - 1)n^{K-1} . \tag{26}$$

It can then be seen that the rows of matrix

$$V = [v_1, \ldots, v_{n^K}] \tag{27}$$

are made of columns of $H^{[K]}(N)$. Because $H(N)$ has a full column rank, $H^{[K]}(N)$ has a full column rank, and V has a full row rank. Consequently, for almost any weighting vector δ, $V\delta$ is a vector with nonzero components. When the weighting vector δ is applied to the diagonal matrix $D_0(\tau(N), \alpha)$, $\sum_\alpha \delta_\alpha D_0(\tau(N), \alpha)$ is nonsingular.   □

Let $\{\delta_\alpha\}$ be such a sequence satisfying (22), and define

$$K_{\delta,k} = \sum_\alpha \delta_\alpha C_y(k; \tau(N), \alpha) , \tag{28}$$

$$D_\delta = \sum_\alpha \delta_\alpha D_0(\tau(N), \alpha) . \tag{29}$$

From (19), we have

$$K_{\delta,k} = H(0) D_\delta C_s^{(K+2)} H^t(k) . \tag{30}$$

From our Lemma along with the assumption that both $H(0)$ and $H(N)$ are of full column rank, it follows that, in parallel to the derivation in [3] for the third-order case,

$$H(k) = L_k H(0) , \tag{31}$$

where

$$L_k = \begin{cases} I , & k = 0 ; \\ K_{\delta,k}^t (K_{\delta,0} K_{\delta,0}^t)^\dagger K_{\delta,0} , & k = 1, \ldots, N . \end{cases} \tag{32}$$

where $(\cdot)^\dagger$ stands for the Moore–Penrose generalized inverse.

It is now clear from (31) that the central issue is solving $H(0)$. The uniqueness of solving $\{H(k)\}$ hinges on the uniqueness of solving $H(0)$. The following theorem shows that $H(0)$ is uniquely specified by the CBI equations (19) up to a *generalized permutation matrix*.[3] Because of (31), all parameter matrices are uniquely specified by the CBI equations up to the *same* generalized permutation matrix.

Theorem 2 (Identifiability). *Under the given assumptions, if* $(\{H(k)\}, C_s^{(K+2)})$ *satisfies the CBI equations, then* $(\{\tilde{H}(k)\}, C_{\tilde{s}}^{(K+2)})$ *also satisfies the CBI equations if and only if there exists a generalized permutation matrix* P *such that*

$$\tilde{H}(k) = H(k) P , \quad k = 0, 1, \ldots, N . \tag{33}$$
$$C_{\tilde{s}}^{(K+2)} = P^{-1} C_s^{(K+2)} P^{-t} . \tag{34}$$

---

[3]   A generalized permutation matrix is defined as a product of a permutation matrix and a nonsingular diagonal matrix.

Proof. With our Lemma, Theorem 2 is a direct consequence of Theorem 4 in the Appendix. □

The important implication of Theorem 2 is that the coefficients of MA processes are specified uniquely by the output cumulants up to a *constant* generalized permutation matrix, a significant reduction of ambiguity in comparing to the identification using second-order output statistics where the identification can only be achieved up to an all-pass system. The ambiguity represented by the generalized permutation matrix can be further reduced with a prior knowledge of the sources such as the variance of the inputs.

3.2. Closed-form identification.

With the identifiability result, we are only interested in finding one solution of the CBI equations because all the solutions are related by generalized permutation matrices. The main result here is a closed-form solution of the CBI equations. Such a solution is obtained by exploiting the eigenstructure of the CBI equations and the application of the ESPRIT techniques originally developed for sensor array processing [9].

Theorem 3 (Closed-Form Identification). *Given the $(k + 2)$th-order output cumulants $C_y(0; \tau(N), \alpha)$ of a multichannel moving average process satisfying our Assumptions, form almost any weighting coefficients $\{\delta_\alpha, \gamma_\alpha, \phi_\alpha\}$ there exists a generalized permutation matrix P such that*

$$H(k) = L_k U_s E^{-t} P , \qquad (35)$$

where U_s is obtained from the Schur decomposition of $K_{\delta,0}$:

$$K_{\delta,0} = \sum_\alpha \delta_\alpha C_y(0; \tau(N), \alpha) , \qquad (36)$$

$$= [U_s U_n] \, \text{diag}(\Delta, 0) [U_s U_n]^t . \qquad (37)$$

Matrix E is the generalized eigenmatrix [4] obtained from the generalized eigendecomposition of $\{U_s^t K_{\gamma,0} U_s, U_s^t K_{\phi,0} U_s\}$, where

$$K_{\gamma,0} = \sum_\alpha \gamma_\alpha C_y(0; \tau(N), \alpha) , \qquad (38)$$

$$K_{\phi,0} = \sum_\alpha \phi_\alpha (C_y(0; \tau(N), \alpha) , \qquad (39)$$

$$L_k = \begin{cases} I , & k = 0 ; \\ K_{\delta,k}^t (K_{\delta,0} K_{\delta,0}^t)^\dagger K_{\delta,0} , & k = 1, \ldots, N . \end{cases} \qquad (40)$$

Proof. With our Lemma, the proof is immediate after the proof of Theorem 4 given in the Appendix. □

Under Theorem 2, $\{\mathbf{L}_k \mathbf{U}_s \mathbf{E}^{-t}\}$ is a solution of the CBI equations, and it differs from the "true" parameter coefficients by, at most, a generalized permutation matrix. Note also that, except for the generalized permutation matrix $\mathbf{P}$, all quantities on the right side of equation (35) are related to the observation process $\mathbf{y}(t)$, and hence can be estimated from the data.

As a special case when $m = n$, i.e., the number of sources is equal to the number of received signals, Theorem 3 takes a simpler form.

Corollary 1 (Closed-Form Identification — A Special Case). *Given the* $(k+2)$*th-order output cumulants* $\mathbf{C}_\mathbf{y}(0, \tau(N), \alpha)$ *of a multichannel moving average process satisfying our Assumptions and almost any* $\{\delta_\alpha, \gamma_\alpha, \phi_\alpha\}$, *there exists a generalized permutation matrix* $\mathbf{P}$ *such that*

$$\mathbf{H}(k) = \mathbf{K}_{\delta,k}^t \mathbf{K}_{\delta,0}^{-t} \mathbf{E}^{-t} \mathbf{P} , \tag{41}$$

where matrix $\mathbf{E}$ *is the generalized eigenmatrix obtained from* $\{\mathbf{K}_{\gamma,0}, \mathbf{K}_{\phi,0}\}$, *where*

$$\mathbf{K}_{\gamma,0} = \sum_\alpha \gamma_\alpha \mathbf{C}_\mathbf{y}(0; \tau(N), \alpha) , \tag{42}$$

$$\mathbf{K}_{\phi,0} = \sum_\alpha \phi_\alpha \mathbf{C}_\mathbf{y}(0; \tau(N), \alpha) , \tag{43}$$

$$\mathbf{K}_{\delta,0} = \sum_\alpha \delta_\alpha \mathbf{C}_\mathbf{y}(0; \tau(N), \alpha) . \tag{44}$$

4. An identification algorithm and a computer simulation example

Theorem 3 clearly suggests a computational algorithm for the identification of multichannel MA processes. Here we present the details of the algorithm and discuss certain implementation issues followed by results of a Monte Carlo simulation.

4.1. Summary of a closed-form identification (CFI) algorithm and discussions.

CFI algorithm.

1. Choose the order of the cumulants to be used in the identification. For most practical applications, fourth-order $(K = 2)$ cumulants are often sufficient.

2. Choose a set of K-dimensional index vector α's and estimate cumulant matrices $\hat{\mathbf{C}}_\mathbf{y}(0; \tau(N), \alpha)$ by, for example, estimating the higher-order moments. When the fourth-order cumulants are used, for any $\alpha = [i, j], 1 \le i, j \le n$,

$$\hat{\mathbf{C}}_\mathbf{y}(0; \tau(N), \alpha) = \frac{1}{T} \sum_{t=1}^{t=T} \left\{ \mathbf{y}(t)\mathbf{y}^t(t)y_i(t-N)y_j(t-N) \right\}$$

$$- \mathbf{y} \frac{1}{T} \sum_{t=1}^{t=T} \left\{ \mathbf{y}(t) \mathbf{y}^t(t) \right\} \frac{1}{T} \sum_{t=1}^{t=T} \left\{ y_i(t-N) y_j(t-N) \right\}$$

$$- \frac{1}{T} \sum_{t=1}^{t=T} \left\{ \mathbf{y}(t) y_i(t-N) \right\} \frac{1}{T} \sum_{t=1}^{t=T} \left\{ \mathbf{y}^t(t) y_j(t-N) \right\}$$

$$- \frac{1}{T} \sum_{t=1}^{t=T} \left\{ \mathbf{y}(t) y_j(t-N) \right\} \frac{1}{T} \sum_{t=1}^{t=T} \left\{ \mathbf{y}^t(t) y_i(t-N) \right\} \quad (45)$$

3. Choose a weighting vector δ and compute

$$\hat{\mathbf{K}}_{\delta,0} = \sum_\alpha \delta_\alpha \hat{\mathbf{C}}_\mathbf{y}(0; \tau(N), \alpha) \quad (46)$$

and its Schur decomposition

$$\hat{\mathbf{K}}_{\delta,0} = [\hat{\mathbf{U}}_s \hat{\mathbf{U}}_n] \, \text{diag}(\hat{\Delta}_1, \hat{\Delta}_2) [\hat{\mathbf{U}}_s \hat{\mathbf{U}}_n]^t , \quad (47)$$

where $\hat{\mathbf{U}}_s$ consists of the singular vectors associated with the m largest singular values. The criterion of choosing the weighting vector δ is to make $\hat{\mathbf{K}}_{\delta,0}$ well conditioned in the sense that the ratio between the largest singular value and the mth singular value is as small as possible.

4. Choose weighting vectors γ and ϕ and compute

$$\hat{\mathbf{K}}_{\gamma,0} = \sum_\alpha \gamma_\alpha \hat{\mathbf{C}}_\mathbf{y}(0; \tau(N), \alpha) \quad (48)$$

$$\hat{\mathbf{K}}_{\phi,0} = \sum_\alpha \phi_\alpha \hat{\mathbf{C}}_\mathbf{y}(0; \tau(N), \alpha) . \quad (49)$$

5. Find the generalized eigenmatrix $\hat{\mathbf{E}}$ from the generalized eigen-decomposition of $\left\{ \hat{\mathbf{U}}_s^t \hat{\mathbf{K}}_{\gamma,0} \hat{\mathbf{U}}_s, \ \hat{\mathbf{U}}_s^t \hat{\mathbf{K}}_{\phi,0} \hat{\mathbf{U}}_s \right\}$. The criteria of choosing γ and ϕ is that the generalized eigenvalues are well separated.

6. The identification of $\{\mathbf{H}(k)\}$ is given by

$$\hat{\mathbf{H}}(0) = \hat{\mathbf{U}}_s \hat{\mathbf{E}}^{-t} \mathbf{P} , \quad (50)$$

$$\hat{\mathbf{H}}(k) = \hat{\mathbf{K}}_{\delta,k}^t (\hat{\mathbf{K}}_{\delta,0} \hat{\mathbf{K}}_{\delta,0}^t)^\dagger \hat{\mathbf{K}}_{\delta,0} \hat{\mathbf{H}}(0) , \quad k = 1, \ldots, N . \quad (51)$$

Discussions.

• The algorithm can be simplified if all the parameter matrices are square as suggested in our Corollary. The singular value decomposition of $\hat{\mathbf{K}}_{\delta,0}$ is no longer necessary. The parameter matrices $\{\mathbf{H}(K)\}$ are estimated according to (41).

• The selection of $\{\delta_\alpha, \gamma_\alpha, \phi_\alpha\}$ is almost arbitrary theoretically. One can choose the weighting vectors to emphasize or de-emphasize the estimates of the cumulants. The selection, however, affects the performance of the estimation.

In choosing $\{\gamma_\alpha, \phi_\alpha\}$, the rule of thumb is, for numerical reasons, to make the generalized eigenvalues of $\left\{\hat{\mathbf{U}}_s^t \hat{\mathbf{K}}_{\gamma,0} \hat{\mathbf{U}}_s, \hat{\mathbf{U}}_s^t \hat{\mathbf{K}}_{\phi,0} \hat{\mathbf{U}}_s\right\}$ have the greatest separation (in order to obtain accurate estimates of the eigenmatrix). Note that these criteria can be verified once $\{\delta_\alpha, \gamma_\alpha, \phi_\alpha\}$ is chosen.

- When m is unknown; i.e., the number of sources is unknown, by evaluating the singular values of $\mathbf{K}_{\delta,0}$, one can in fact estimate m by observing the most significant singular values. See, for example, [7] and [20].

- The computation complexity involved here is substantially reduced as compared with the method presented in [16].

- The algorithm can be extended to incorporate statistics of different orders. Note that (19) holds for all $K \geq 1$, and more importantly, they all have the same structures as those equations in (54) of the Appendix. Therefore, by taking linear combinations of $\mathbf{K}_y(0; \tau(N), \alpha)$'s, with respect to α's of different orders, one can easily obtain a closed-form identification involving statistics of different orders. Such a solution may provide better estimates of the process coefficients. However, determining how to choose the weighting parameters to obtain better estimates seems to be an open problem.

4.2. A simulation example.

We consider a second-order, 3-input, 7-output MA process

$$
\mathbf{H}(0) = \begin{pmatrix} 0.0012 & 0.3701 & 0.4330 \\ 0.2330 & 0.4115 & 0.5879 \\ 0.0917 & 0.4494 & 0.1438 \\ 0.5530 & 0.3354 & 0.3375 \\ 0.4192 & 0.4203 & 0.4272 \\ 0.5998 & 0.0925 & 0.2304 \\ 0.3098 & 0.4379 & 0.3109 \end{pmatrix},
$$

$$
\mathbf{H}(1) = \begin{pmatrix} 0.2689 & 0.2201 & 0.5784 \\ 0.6033 & 0.2622 & 0.4719 \\ 0.0662 & 0.0707 & 0.2523 \\ 0.1770 & 0.4841 & 0.1815 \\ 0.2926 & 0.5403 & 0.0606 \\ 0.2981 & 0.5428 & 0.3646 \\ 0.5945 & 0.2383 & 0.4561 \end{pmatrix}, \tag{52}
$$

$$
\mathbf{H}(2) = \begin{pmatrix} 0.3434 & 0.1765 & 0.1799 \\ 0.4331 & 0.1870 & 0.3149 \\ 0.1960 & 0.3417 & 0.4988 \\ 0.3467 & 0.3470 & 0.0669 \\ 0.2110 & 0.6137 & 0.5516 \\ 0.6075 & 0.5334 & 0.5362 \\ 0.3498 & 0.1885 & 0.1532 \end{pmatrix}. \tag{53}
$$

The source signals were one-sided exponentially distributed. In comparison with the finite-step global convergence (FSGC) algorithm [16], only the third-order cumulants were used. The signal-to-noise ratio (SNR) was set at 20 dB. Monte Carlo simulations of 100 runs were conducted with 5000 samples used to estimate the third-order cumulants. Sample means and root mean square errors of the two estimators are shown in Tables 1 through 6. In the simulation, uniform weighting factors were used. The closed-form identification was clearly superior than the iterative method in this simulation, both in computation complexity and in estimation accuracy.

Table 1. The comparison of estimation means of $\hat{\mathbf{H}}(0)$. 100 Monte Carlo runs at SNR $= 20$ dB. 5000 samples were used.

Estimation Mean of $\hat{\mathbf{H}}(0)$ by FSGC [16]			Estimation Mean of $\hat{\mathbf{H}}(0)$ by CFI		
-0.0029	0.3299	0.4219	0.0016	0.3675	0.4343
0.2310	0.3753	0.5855	0.2339	0.4046	0.5845
0.0847	0.4496	0.0940	0.0872	0.4607	0.1486
0.5498	0.3141	0.3258	0.5551	0.3236	0.3363
0.4209	0.4155	0.3971	0.4159	0.4236	0.4190
0.5925	0.0768	0.2263	0.5946	0.0762	0.2265
0.3023	0.4152	0.2886	0.3084	0.4296	0.3193

Table 2. The comparison of estimation means of $\hat{\mathbf{H}}(1)$. 100 Monte Carlo runs at SNR $= 20$ dB. 5000 samples were used.

Estimation Mean of $\hat{\mathbf{H}}(1)$ by FSGC [16]			Estimation Mean of $\hat{\mathbf{H}}(1)$ by CFI		
0.2715	0.2070	0.5609	0.2701	0.2135	0.5659
0.5976	0.2429	0.4535	0.5995	0.2475	0.4661
0.0653	0.0490	0.2446	0.0651	0.0606	0.2512
0.1760	0.4546	0.1520	0.1768	0.4892	0.1926
0.2793	0.4880	0.0234	0.2872	0.5330	0.0805
0.2939	0.4883	0.3324	0.2966	0.5361	0.3743
0.5935	0.2247	0.4420	0.5924	0.2276	0.4466

Table 3. The comparison of estimation means of $\hat{\mathbf{H}}(2)$. 100 Monte Carlo runs at SNR $= 20$ dB. 5000 samples were used.

Estimation Mean of $\hat{\mathbf{H}}(2)$ by FSGC [16]			Estimation Mean of $\hat{\mathbf{H}}(2)$ by CFI		
0.3395	0.1650	0.1635	0.3402	0.1665	0.1799
0.4323	0.1682	0.3069	0.4297	0.1787	0.3118
0.1956	0.3066	0.4933	0.1987	0.3374	0.4922
0.3344	0.3456	0.0356	0.3409	0.3439	0.0751
0.2030	0.5783	0.5229	0.2082	0.6173	0.5515
0.6056	0.5092	0.5200	0.6099	0.5281	0.5308
0.3469	0.1819	0.1346	0.3466	0.1816	0.1555

Table 4. The comparison of root mean square errors of $\hat{H}(0)$. 100 Monte Carlo runs at SNR = 20 dB. 5000 samples were used.

RMSE of $\hat{H}(0)$ by FSGC [16]			RMSE $\hat{H}(0)$ by CFI		
0.0741	0.1300	0.0607	0.0499	0.0364	0.0190
0.0628	0.1544	0.0545	0.0329	0.0429	0.0265
0.0411	0.1271	0.2033	0.0370	0.0600	0.0317
0.0384	0.0964	0.0539	0.0308	0.0410	0.0236
0.0425	0.1059	0.0857	0.0297	0.0291	0.0217
0.0382	0.1272	0.0792	0.0270	0.0719	0.0488
0.0435	0.0957	0.1077	0.0249	0.0257	0.0221

Table 5. The comparison of root mean square errors of $\hat{H}(1)$. 100 Monte Carlo runs at SNR = 20 dB. 5000 samples were used.

RMSE of $\hat{H}(1)$ by FSGC [16]			RMSE $\hat{H}(1)$ by CFI		
0.0571	0.2022	0.0851	0.0249	0.0818	0.0546
0.0313	0.1753	0.0505	0.0285	0.0767	0.0294
0.0558	0.1174	0.0485	0.0395	0.0592	0.0322
0.0473	0.0944	0.1449	0.0445	0.0379	0.0412
0.0635	0.1747	0.2063	0.0493	0.0614	0.0872
0.0525	0.1327	0.1314	0.0484	0.0316	0.0426
0.0341	0.1654	0.0433	0.0311	0.0722	0.0342

Table 6. The comparison of root mean square errors of $\hat{H}(2)$. 100 Monte Carlo runs at SNR = 20 dB. 5000 samples were used.

RMSE of $\hat{H}(2)$ bye FSGC [16]			RMSE $\hat{H}(2)$ by CFI		
0.0358	0.1057	0.1074	0.0280	0.0500	0.0367
0.0341	0.1578	0.0979	0.0302	0.0656	0.0354
0.0628	0.1565	0.0817	0.0267	0.0390	0.0510
0.0365	0.1286	0.1585	0.0233	0.0478	0.0539
0.0886	0.1035	0.1022	0.0513	0.0424	0.0301
0.0547	0.0991	0.0657	0.0452	0.0292	0.0347
0.0333	0.1010	0.0874	0.0303	0.0483	0.0477

5. Conclusion

A closed-form identification of multichannel MA processes has been presented by exploiting eigenstructures of the output cumulant matrices. An estimation algorithm is then derived from the closed-form identification. The method is computationally simpler and offers better performance than previously proposed schemes. It also allows one to use cumulants of different orders and weight the estimates of cumulants to achieve better performance.

There are several open issues that need to be addressed in future work. The assumption that both $\mathbf{H}(0)$ and $\mathbf{H}(N)$ have full column ranks may be too restrictive for some applications. Also there is a need to select the weighting factors in a systematic way.

Appendix. On solving a class of nonlinear matrix equations

As shown in Section 2, the CBI equations (19) take a special form. The underlying algebraic problem of multichannel MA process identification can be reduced to solving a set of special nonlinear matrix equations having the following special algebraic structure

$$\mathbf{H}\mathbf{D}_i\mathbf{H}^t = \mathbf{K}_i \quad i = 1, 2, \ldots, M , \tag{54}$$

where $\mathbf{H}$ is an $(n \times m)$-matrix of full column rank, and $\mathbf{D}_i$ is an $(m \times m)$ *diagonal* matrix not necessarily of full rank.

We ask the following questions:

- *Given $\mathbf{K}_i$, to what extent is $\mathbf{H}$ determined?*
- *Given $\mathbf{K}_i$, how does one solve for $\mathbf{H}$?*

These questions can be answered by the following theorem (see also [18]).

Theorem 4 (Closed-Form Solution). *Suppose that $\{\mathbf{H}, \mathbf{D}_i\}$ satisfies equations (54). Let matrix $\mathbf{D}$ be constructed from $\mathbf{D}_k = \mathrm{diag}(d_{1k}, \ldots, d_{mk})$ by*

$$\mathbf{D} = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1M} \\ d_{21} & d_{22} & \cdots & d_{2M} \\ \vdots & \vdots & \vdots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mM} \end{pmatrix} . \tag{55}$$

If $\mathrm{rank}(\mathbf{D}) = m$, then

1. If $\mathbf{H}$ is square, i.e., $m = n$, then for almost any $\{\gamma_i, \phi_i\}$ there exists a generalized permutation matrix $\mathbf{P}$ such that

$$\mathbf{H} = \mathbf{E}^{-t}\mathbf{P} , \tag{56}$$

where matrix $\mathbf{E}$ is a generalized eigenmatrix obtained from $\{\mathbf{K}_\gamma, \mathbf{K}_\phi\}$, where

$$\mathbf{K}_\gamma = \sum_i \gamma_k \mathbf{K}_i , \tag{57}$$

$$\mathbf{K}_\phi = \sum_i \phi_i \mathbf{K}_i . \tag{58}$$

2. If $n > m$, then for almost any $\{\delta_i, \gamma_i, \phi_i\}$ there exists a generalized permutation matrix $\mathbf{P}$ such that

$$\mathbf{H} = \mathbf{U}_s\mathbf{E}^{-t}\mathbf{P} , \tag{59}$$

where $\mathbf{U}_s$ is obtained from the Schur decomposition of $\mathbf{K}_\delta$

$$\mathbf{K}_\delta = \sum_i \delta_i \mathbf{K}_i , \tag{60}$$

$$= [\mathbf{U}_s \mathbf{U}_n] \operatorname{diag}(\Delta, \mathbf{0}) [\mathbf{U}_s \mathbf{U}_n]^t ; \tag{61}$$

matrix $\mathbf{E}$ is the generalized eigenmatrix obtained from the generalized eigendecomposition of $\{\mathbf{U}_s^t \mathbf{K}_\gamma \mathbf{U}_s, \mathbf{U}_s^t \mathbf{K}_\phi \mathbf{U}_s\}$ ($\mathbf{K}_\gamma$ and $\mathbf{K}_\phi$ are defined in equations (57) and (58)).

Proof. The proof of this theorem follows a technique used in deriving the ESPRIT algorithm [9].

1. Because matrix $\mathbf{D}$ is of full rank, then for almost any weighting vectors $\{\gamma_i, \phi_i\}$, $(\mathbf{K}_\gamma, \mathbf{K}_\phi)$ has distinct generalized eigenvalues, where $\mathbf{K}_\gamma$ and $\mathbf{K}_\phi$ are defined by (57) and (58). Suppose $\{\lambda_k\}$ and $\{\mathbf{e}_k\}$ are the generalized eigenvalues and eigenvectors, respectively; i.e.,

$$\mathbf{K}_\gamma \mathbf{e}_k = \lambda_k \mathbf{K}_\phi \mathbf{e}_k . \tag{62}$$

We then have

$$\mathbf{H}(\mathbf{D}_\gamma - \lambda_k \mathbf{D}_\phi) \mathbf{H}^t \mathbf{e}_k = \mathbf{0} . \tag{63}$$

Because $(\mathbf{D}_\gamma - \lambda_k \mathbf{D}_\phi)$ is a diagonal matrix with rank $m-1$ and matrix $\mathbf{H}$ is of full column rank, vector $\mathbf{e}_k$ is orthogonal to all but one column vector of $\mathbf{H}^t$. Let

$$\mathbf{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_m] . \tag{64}$$

Then, because of the orthogonality of the eigenvectors with respect to the column vectors of $\mathbf{H}$, there is a generalized permutation matrix $\mathbf{P}$ such that

$$\mathbf{H}^t \mathbf{E} = \mathbf{P}^t . \tag{65}$$

Hence,

$$\mathbf{H} = \mathbf{E}^{-t} \mathbf{P} , \tag{66}$$

and a solution of equation (54) is obtained.

2. When $\mathbf{H}$ is not square; i.e., $n > m$, it is no longer true that equation (63) implies that $\mathbf{e}_k$ is orthogonal to all but one column vector of $\mathbf{H}$. However, this problem can be easily circumvented by a simple transformation. Again, for almost any $\{d_i\}$, $\operatorname{rank}(\mathbf{K}_\delta) = m$, where

$$\mathbf{K}_\delta = \sum_i \delta_i \mathbf{K}_i . \tag{67}$$

Let $\mathbf{K}_\delta$ have a singular value decomposition of the following form,

$$\mathbf{K}_\delta = [\mathbf{U}_s \mathbf{U}_n] \operatorname{diag}(\Delta, \mathbf{0}) [\mathbf{U}_s \mathbf{U}_n]^t . \tag{68}$$

It is evident that

$$\mathbf{H} = \mathbf{U}_s \mathbf{T} , \tag{69}$$

where $\mathbf{T}$ is a nonsingular (square) matrix. If a linear transform is applied to $\mathbf{K}_\gamma$ and $\mathbf{K}_\phi$ by

$$\tilde{\mathbf{K}}_\gamma = \mathbf{U}_s^t \mathbf{K}_\gamma \mathbf{U}_s , \tag{70}$$

$$= \mathbf{T} \mathbf{D}_\gamma \mathbf{T}^t , \tag{71}$$

$$\tilde{\mathbf{K}}_\phi = \mathbf{U}_s^t \mathbf{K}_\phi \mathbf{U}_s , \tag{72}$$

$$= \mathbf{T} \mathbf{D}_\phi \mathbf{T}^t , \tag{73}$$

we can apply the previous theorem to the above equations, and obtain

$$\mathbf{T} = \tilde{\mathbf{E}}^{-t} \tilde{\mathbf{P}} . \tag{74}$$

Consequently,

$$\mathbf{H} = \mathbf{U}_s \tilde{\mathbf{E}}^{-t} \tilde{\mathbf{P}} . \tag{75}$$

$\square$

Acknowledgments

References

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley & Sons, New York, 1984.

[2] G. Giannakis, On the identifiability of non-Gaussian ARMA models using cumulants, *IEEE Trans. Automat. Contr.*, 35(1):18–26, January 1990.

[3] G. Giannakis, Y. Inouye, and J. Mendel, Cumulant based identification of multichannel moving-average models, *IEEE Trans. Automat. Contr.*, 34(7):783–787, July 1989.

[4] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, MD, 1990.

[5] S. Haykin, *Array Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1984.

[6] Y. Inouye, L. Tong, and R. Liu, Identification and estimation of multichannel non-Gaussian moving-average processes using fourth-order cumulants, in *Proc. 28th Allerton Conference on Communication, Control and Computing*, October 1990.

[7] K. Konstantinides and K. Yao, Statistical analysis of effective singular values in matrix rank determination, *IEEE Trans. ASSP*, ASSP-36(5):757–763, May 1988.

[8] J. M. Mendel, Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications, *Proc. IEEE*, 79(3):278–305, March 1991.

[9] A. Paulraj, R. Roy, and T. Kailath, Estimation of signal parameter via rotational invariance techniques — ESPRIT, in *Proc. 19th Asilomar Conf. on Circuits, Systems and Comp.*, Asilomar, CA, November 1985.

[10] M. Rosenblatt, *Stationary Sequences and Random Fields*, Birkhäuser, Boston, MA, 1985.

[11] R. Roy and T. Kailath, ESPRIT — Estimation of Signal parameters via rotational invariance techniques, *IEEE Trans. ASSP*, 32(7):984–995, July 1989.

[12] R. Roy, A. Paulraj, and T. Kailath, ESPRIT — A subspace rotation approach to estimation of parameters of cissoids in noise, *IEEE Trans. Acoust. Speech, Signal Processing*, ASSP-34(10):1340–1342, October 1986.

[13] A. Swami, G. B. Giannakis, and J. Mendel, A unified approach to modeling multichannel ARMA processes, in *Proc. ICASSP89 Conf.*, vol. 4, 2182–2185, Glasgow, Scotland, May 1989.

[14] A. Swami and J. M. Mendel, Closed-form recursive estimation of MA coefficients using autocorrelation, *IEEE Trans. Acoust. Speech, Signal Processing*, 36(11):1794–1795, November 1989.

[15] A. Swami and J. M. Mendel, Time and lag recursive computation of cumulants from a state-space model, *IEEE Trans. Automat. Contr.*, 35(1):4–18, January 1990.

[16] L. Tong, Y. Inouye, and R. Liu, A finite-step global convergence algorithm for the parameter estimation of multichannel MA processes, *IEEE Trans. Signal Processing*, 40(10): 2547–2558, October 1992.

[17] L. Tong, Y. Inouye, and R. Liu, Waveform preserving blind estimation of multiple independent sources, *IEEE Trans. Signal Processing*, 41(7): 2461–2470, July 1993.

[18] L. Tong and R. Liu, On solving simultaneous identification equations, Technical Report 90-06-02, University of Notre Dame, Notre Dame, IN, June 1991.

[19] L. Tong, V. Soon, Y. F. Huang, and R. Liu, Indeterminacy and identifiability of blind identification, *IEEE Trans. Circuits and Systems*, 38(5):499–509, May 1991.

[20] M. Wax and T. Kailath, Detection of signals by information theorem criteria, *IEEE Trans. ASSP*, ASSP-33(2):387–392, April 1985.