Signal Processing 26 (1992) 381 388
Elsevier

Short communication

MA identification using fourth order cumulants

Pierre Comon
Thomson Sintra, Parc Sophia Antipolis, BP 138, F-06561 Valbonne Cedex, France

Received 7 January 1991
Revised 8 August 1991 and 31 October 1991

Abstract. The algorithm proposed aims to identify moving average coefficient matrices of an MA process, not necessarily minimum-phase, driven by an unobserved non-gaussian input. It is assumed that the observation available is of limited duration, and coefficients are estimated from the set of fourth order output cumulants. It is shown that much more equations than unknowns are available, and that robustness for short data records can be obtained by utilizing them all.

Zusammenfassung. Der vorgeschlagene Algorithmus dient dazu, die Koeffizientenmatrizen eines mehrdimensionalen nicht notwendig minimalphasigen MA-Prozesses, der von einem unbeobachteten Nicht-Gaußschen Eingangssignal gespeist wird, zu identifizieren. Es wird angenommen, daß die Beobachtung zeitbegrenzt ist und die Koeffizienten aus einem Satz von Kumulanten vierter Ordnung des Ausgangssignals geschätzt werden. Es wird gezeigt, daß sich mehr Gleichungen als Unbekannte ergeben und daß das Verfahren bei einer geringen Eingangsdatenmenge robust wird, wenn man alle Gleichungen verwendet.

Résumé. L'algorithme proposé a pour but d'identifier les matrices-coefficient d'un processus MA multivariable, pas nécessairement à minimum de phase, et piloté par une entrée non gaussienne non observable. On suppose que l'observation est de durée limitée, et que les coefficients sont estimés à partir d'un ensemble de cumulants d'ordre quatre des sorties. On montre alors qu'il existe plus d'équations que d'inconnues, et que la robustesse de l'identification s'améliore si on les utilise toutes.

Keywords. Blind identification, moving average, non-minimum phase, cumulant, multiple input multiple output, multichannel.

1. Introduction

Consider the p-channel MA model of known order q:

$$y(t) = \sum_{k=0}^{q} B_k w(t-k) + v(t), \tag{1}$$

where coefficients B_k are $p \times p$ matrices, B_0 is known and regular, B_q is non-zero, $w(t)$ is a zero-mean stationary non-gaussian process white up to fourth order with unit covariance matrix, and $v(t)$ is a zero-mean Gaussian noise independent of $w(t)$, possibly colored. It is not assumed that this model is minimum phase. The case where B_0 is unknown regular was first investigated in [3] for instance, and will not be addressed here. The way we are going to deal with MA models is inspired from the works of Giannakis and others in [4, 6], and some extensions and improvements are presented in this paper. For convenience,

we denote vectors with bold lowercases, scalars with plain lowercases, and matrices with plain uppercases. In the rest of the paper, it will be extensively resorted to cumulants and notations borrowed from Kronecker calculus. For this reason it is convenient to start with some definitions and properties.

The Kronecker product of two matrices, A and B, of dimensions $p \times q$ and $r \times s$, respectively, is a $pr \times qs$ block-matrix denoted $A \otimes B$ and defined by the generic $r \times s$ block, $A_{ij}B$. See for instance [1, 5] for more details. Cumulants are known tools in statistics and are usually defined as the coefficients in expansion of the second characteristic function. From a practical point of view, they can be calculated from moments [2]. Let us insist on the fourth order cumulants of random vectors, since this will be our main concern. If $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ and $\boldsymbol{d}$ are four zero-mean random vectors of respective dimension, α, β, γ and δ, then the fourth order cumulant is the $\alpha\beta\gamma\delta \times 1$ vector defined as

$$\begin{aligned}\operatorname{cum}\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}\} = {} & \mathrm{E}\{\boldsymbol{a}\otimes\boldsymbol{b}\otimes\boldsymbol{c}\otimes\boldsymbol{d}\} - \mathrm{E}\{\boldsymbol{a}\otimes\boldsymbol{b}\}\otimes\mathrm{E}\{\boldsymbol{c}\otimes\boldsymbol{d}\} - \mathrm{E}\{\boldsymbol{a}\otimes\mathrm{E}\{\boldsymbol{b}\otimes\boldsymbol{c}\}\otimes\boldsymbol{d}\} \\ & - \mathrm{E}\{\boldsymbol{a}\otimes\mathbf{1}_\beta\otimes\boldsymbol{c}\otimes\mathbf{1}_\delta\} * \mathrm{E}\{\mathbf{1}_\alpha\otimes\boldsymbol{b}\otimes\mathbf{1}_\gamma\otimes\boldsymbol{d}\}, \end{aligned} \tag{2}$$

where $*$ denotes the termwise product (also sometimes called Hadamard product), and $\mathbf{1}_p$ is the p-dimensional vector formed of ones. This expression can be rewritten in a more general (and more complicated) form allowing obvious extension to higher orders from the corresponding expressions in the scalar case:

$$\begin{aligned}\operatorname{cum}\{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}\} = {} & \mathrm{E}\{\mathbf{a}\otimes\boldsymbol{b}\otimes\boldsymbol{c}\otimes\boldsymbol{d}\} - \mathrm{E}\{\boldsymbol{a}\otimes\boldsymbol{b}\otimes\mathbf{1}_\gamma\otimes\mathbf{1}_\delta\} * \mathrm{E}\{\mathbf{1}_\alpha\otimes\mathbf{1}_\beta\otimes\boldsymbol{c}\otimes\boldsymbol{d}\} \\ & - \mathrm{E}\{\boldsymbol{a}\otimes\mathbf{1}_\beta\otimes\mathbf{1}_\gamma\otimes\boldsymbol{d}\} * \mathrm{E}\{\mathbf{1}_\alpha\otimes\boldsymbol{b}\otimes\boldsymbol{c}\otimes\mathbf{1}_\delta\} \\ & - \mathrm{E}\{\boldsymbol{a}\otimes\mathbf{1}_\beta\otimes\boldsymbol{c}\otimes\mathbf{1}_\delta\} * \mathrm{E}\{\mathbf{1}_\alpha\otimes\boldsymbol{b}\otimes\mathbf{1}_\gamma\otimes\boldsymbol{d}\}. \end{aligned} \tag{3}$$

Basic properties of the Kronecker product and related operators are stated in [1, 5]. We shall recall below some of them for the sake of clarity. The operator **vec** is a linear operator mapping any matrix A of size $p \times q$ to a vector of dimension pq by stacking its columns one after the other:

$$\mathbf{vec}(A) = \begin{bmatrix} A_{:1} \\ A_{:2} \\ \vdots \\ A_{:q} \end{bmatrix}. \tag{4}$$

Note that on most computers matrices are actually stored in this format. This definition induces a bijection between the space of $p \times q$ matrices and the space of vectors of size pq. Consequently, it is possible to define the inverse operator as

$$\mathrm{Unvec}_q(\boldsymbol{u}) = U. \tag{5}$$

Since there may be several manners to rearrange a vector in a matrix, the index q indicates the number of columns desired in U, thus avoiding any ambiguity. Now let A, B, C and D be four matrices with compatible dimensions. Then one property of interest is

$$(AB)\otimes(CD) = (A\otimes C)(B\otimes D).$$

By induction, this extends easily to

$$\bigotimes_{i=1}^{r} (B_i w_i) = \left(\bigotimes_{i=1}^{r} B_i\right)\left(\bigotimes_{i=1}^{r} w_i\right). \tag{6}$$

Next, if F, G and X are matrices of size $p \times q$, $r \times s$ and $s \times p$, respectively, the following property holds [1, 4]:

$$[F^{\mathrm{T}} \otimes G]\, \mathbf{vec}\{X\} = \mathbf{vec}\{GXF\}. \tag{7}$$

As a particular case, we have the relation

$$GX = C \quad \Leftrightarrow \quad [I_p \otimes G]\, \mathbf{vec}\{X\} = \mathbf{vec}\{C\} \tag{8}$$

for $F = I_p$, the identity matrix of dimension $p \times p$. If X is an unknown and if G and C are given, this last property allows to solve for it either as a matrix with the left equation or as a vector with the right one, depending on preferences.

2. Identification of scalar MA models

The scalar case is really particular insofar as the product is commutative. Moreover, the principles of the identification procedure we propose is much easier to understand in that case. As in [4, 6], the goal is to obviate the minimum phase condition. The observed process satisfies

$$y(t) = \sum_{k=0}^{q} b_k w(t-k) + v(t), \tag{9}$$

where b_0 is a given non-zero scalar. Additionally, we assume that q denotes the order of the model, i.e., that b_q is non-zero. Denote $c_y(i, j, q, 0) = \mathrm{cum}\{y(t+i),\ y(t+j),\ y(t+q),\ y(t)\}$ and $c_w = \mathrm{cum}\{w(t), w(t), w(t), w(t)\}$. Since the process $w(t)$ is white at the fourth order, we have the relation

$$c_y(i, j, q, 0) = b_i b_j b_q b_0 c_w. \tag{10}$$

Refer to [2] for a proof, or to the next section where it is given under more general assumptions. From (10) we deduce the following relations by substitution of c_w [3]:

$$b_i b_j c_y(k, l, q, 0) = b_k b_l c_y(i, j, q, 0) \quad \text{for any } \{i, j, k, l\}.$$

As particular cases, two families of equations follow:

$$b_i c_y(k, j, q, 0) - b_k c_y(i, j, q, 0) = 0 \quad \forall i, j, k, \quad 0 \le i < k \le q, \ 0 \le j \le q, \tag{11}$$

$$b_i^2 c_y(k, k, q, 0) - b_k^2 c_y(i, i, q, 0) = 0 \quad \forall i, k, \quad 0 \le i < k \le q. \tag{12}$$

Equations (11) and (12) provide us with $q(q+1)^2/2$ and $q(q+1)/2$ relations, respectively. Note that (12) can hardly be used because of the sign indetermination. From (11) the vector of coefficients, $\boldsymbol{b} = [b_q, \ldots, b_2, b_1]^{\mathrm{T}}$, may be obtained by solving in the least-squares sense an overdetermined linear system of the form

$$\boldsymbol{M}\boldsymbol{b} = \boldsymbol{d}, \tag{13}$$

where vector $\boldsymbol{d}$ contains only known quantities. In [4, 6], only those q equations where $i = j = 0$ were utilized. We find it very useful to utilize the complete set of equations (11) as show the subsequent simulations. More precisely, simulations suggest to utilize at least the $q(q+1)$ first rows of M (those equations of (11) for which $i = 0$) in order to obtain the best results as demonstrated in Section 4.

As an example, the system (13) obtained for $q=2$ would be of the form

$$\begin{bmatrix} 0 & 0 & 0 & C(0,0) & C(0,1) & C(0,2) & C(1,0) & C(1,1) & C(1,2) \\ C(0,0) & C(0,1) & C(0,2) & 0 & 0 & 0 & -C(2,0) & -C(2,1) & -C(2,2) \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} b_2 \\ b_1 \end{bmatrix}$$

$$= b_0[C(1,0) \quad C(1,1) \quad C(1,2) \quad C(2,0) \quad C(2,1) \quad C(2,2) \quad 0 \quad 0 \quad 0]^{\mathrm{T}}, \tag{14}$$

where for simplicity $C(i,j)$ stands for $c_y(i,j,q,0)$.

3. Identification of multichannel MA models

Now we assume that the reader has become familiar with the problem and we shall address a more general case. In this section, we assume that the observed process, $\boldsymbol{y}(t)$, satisfies the model (1) where $\boldsymbol{w}(t)$ and $\boldsymbol{y}(t)$ are unobservable processes satisfying the assumptions below, for some given integer $r>2$:

- $\boldsymbol{w}(t)$ and $\boldsymbol{v}(t)$ are white independent processes at order r, (15)
- $\boldsymbol{w}(t)$ has a finite non-zero cumulant sequence of order r, (16)
- $\boldsymbol{v}(t)$ has a zero cumulant sequence of order r, (17)
- matrix B_0 is known and regular. (18)

In other words, strong whiteness and independence are not required. Denote the cumulant

$$\boldsymbol{c}_y(i_1, i_2, \ldots, i_r) = \mathrm{cum}\{\boldsymbol{y}(t+i_1), \boldsymbol{y}(t+i_2), \ldots, \boldsymbol{y}(t+i_r)\}. \tag{19}$$

Then, from definition (2), property (6), and assumptions (15) and (17):

$$\boldsymbol{c}_y(i_1,i_2,\ldots,i_r) = \sum_{k_1}\sum_{k_2}\cdots\sum_{k_r}\left[\bigotimes_{n=1}^{r} B_{k_n}\right]\boldsymbol{c}_w(i_1-k_1, i_2-k_2, \ldots, i_r-k_r), \tag{20}$$

where the left-hand side is a vector of dimension p^r for every value of the indices. We assume for simplicity that the matrices B_k exist for all k and are null if $k<0$ or if $k>q$. The sums can then be taken over the whole set of integers. Now from assumption (16), the cumulant on the right-hand side is null for all values of the indices that do not satisfy

$$i_1-k_1 = i_2-k_2 = \cdots = i_r-k_r. \tag{21}$$

Consider the particular slice $\{i_{r-1}=q,\ i_r=0\}$. The result above yields then that $k_r=k_{r-1}-q$. But the only possible values of k_r for which B_{k_r} and B_{k_r+q} may be simultaneously non-zero is $k_r=0$. Therefore (21) fixes all the indices:

$$k_1-i_1 = k_2-i_2 = \cdots = k_{r-2}-i_{r-2} = k_{r-1}-q = k_r = 0. \tag{22}$$

We obtain by rewriting (20):

$$\boldsymbol{c}_y(i_1, i_2, \ldots, i_{r-2}, q, 0) = \left[\bigotimes_{n=1}^{r-2} B_{i_n} \otimes B_q \otimes B_0\right]\boldsymbol{c}_w(0, 0, \ldots, 0). \tag{23}$$

This is the generalization of (10) to the multichannel and order r case. Now, we use property (7) with

$$\boldsymbol{c}_w = \mathbf{vec}\{X\}, \qquad F = B_{i_1}^{\mathrm{T}} \quad \text{and} \quad G = \left(\bigotimes_{n=2}^{r-2} B_{i_n} \right) \otimes B_q \otimes B_0,$$

which brings

$$\boldsymbol{c}_y(i_1, i_2, \ldots, i_{r-2}, q, 0) = \mathbf{vec}\{G \, \mathrm{Unvec}_p\{\boldsymbol{c}_w(0, \ldots, 0)\} B_{i_1}^{\mathrm{T}}\}. \tag{24}$$

For the sake of clarity, let us additionally denote

$$\bar{C}_w = \mathrm{Unvec}_p\{\boldsymbol{c}_w(0, \ldots, 0)\}, \tag{25}$$

$$\bar{C}_y(i_1, i_2, \ldots, i_{r-2}, q, 0) = \mathrm{Unvec}_p\{\boldsymbol{c}_y(i_1, i_2, \ldots, i_{r-2}, q, 0)\}. \tag{26}$$

So both sides of (26) are $p^{r-1} \times p$ matrices. Then (24) simplifies by applying the Unvec operator on both sides

$$\bar{C}_y(i_1, i_2, \ldots, i_{r-2}, q, 0) = G \bar{C}_w B_{i_1}^{\mathrm{T}}. \tag{27}$$

In order to make it possible to exploit this result, we consider the particular slice $i_1 = 0$, use assumption (18) and get

$$\bar{C}_y(0, i_2, \ldots, i_{r-2}, q, 0) B_0^{-\mathrm{T}} = G \bar{C}_w,$$

which yields finally by substituting $G\bar{C}_w$ back in (27):

$$\bar{C}_y(i_1, i_2, \ldots, i_{r-2}, q, 0) = \bar{C}_y(0, i_2, \ldots, i_{r-2}, q, 0) B_0^{-\mathrm{T}} B_{i_1}^{\mathrm{T}}. \tag{28}$$

This relation holds for any indices satisfying

$$1 \leqslant i_1 \leqslant q \quad \text{and} \quad 0 \leqslant i_s \leqslant q, \quad \text{with } 1 < s < r-1.$$

For a fixed value i of i_1, (28) may be seen as a set of $(q+1)^{r-3}$ linear systems, each of size $p^{r-1} \times p$. Therefore, matrices B_i can be obtained by solving in the least-squares (LS) sense the following overdetermined system:

$$M B_i^{\mathrm{T}} = D_i, \quad 1 \leqslant i \leqslant q,$$

where M is a $(q+1)^{r-3} p^{r-1} \times p$ matrix. Since the same matrix M is involved for each B_i, it may be cheaper regarding computational complexity to compute the pseudo-inverse of M once for all, M^-, and to compute the products $M^- D_i$ afterwards.

Now, possible improvements can be obtained by using total least-squares (TLS) techniques instead of LS. With this goal a last operation is necessary, namely to arrange this set of $(q+1)^{r-3}$ linear systems in a single large linear system that will be soluble by standard computer algorithms. For this purpose, we resort to property (8) and get the linear system

$$[I_p \otimes (\bar{C}_y(0, j_1, \ldots, j_{r-3}, q, 0) B_0^{-\mathrm{T}})] \, \mathbf{vec}\{B_i^{\mathrm{T}}\} = \mathbf{c}_y(i, j_1, \ldots, j_{r-3}, q, 0). \tag{29}$$

Each unknown, $\boldsymbol{b}_i = \mathbf{vec}\{B_i^{\mathrm{T}}\}$, is a vector of dimension p^2, and the left-hand side matrix is of size $p^r \times p^2$. In other words, each unknown $\boldsymbol{b}_i$ satisfies an overdetermined linear system of equations of the form

$$N \boldsymbol{b}_i = \boldsymbol{d}_i, \quad 1 \leqslant i \leqslant q,$$

where N is of dimension $(q+1)^{r-3} p^r \times p^2$. The TLS solution will minimize $\|\mathrm{N}\boldsymbol{b}_i - \boldsymbol{d}_i\|$ subject to a constraint like $\|\boldsymbol{b}_i\|^2 + \|\boldsymbol{d}_i\|^2 = 1$ in the L^2-norm.

In general there is no advantage to go to large orders, except when assumptions (15) to (18) are not satisfied for lower orders. Taking $r=3$ may produce an ill-conditioned system if $w(t)$ is almost symmetrically distributed. The next value is $r=4$ and is very satisfactory in most situations. In that case (28) becomes

$$\bar{C}_y(0,j,q,0)B_0^{-\mathrm{T}}B_i^{\mathrm{T}}=\bar{C}_y(i,j,q,0), \quad 1\leqslant i\leqslant q, \ 0\leqslant j\leqslant q. \tag{30}$$

Thus relation (28), which is eventually the key relation of the paper, provides when $r=4q+1$ overdetermined matrix-equations for each B_i. They can be noticed to be much fewer than the $q(q+1)^2/2$ equations available in the scalar case. In fact, even if (23) is the multichannel analog of (10), it is not possible to build a multichannel analog of (11), due to non-commutativity of the Kronecker and matrix products.

4. Simulation results

In this section, we compare three different methods obtained by solving linear systems of equations more and more overdetermined:

Scalar MA models

Method 1. Use of the equations in (11) for which $i=j=0$ and $1\leqslant k\leqslant q$. This is the system of Giannakis et al. [4], using q equations (square system of linear equations).

Method 2. Use of the equations in (11) for which $i=0$. This system is formed of the $q(q+1)$ first equations in (11).

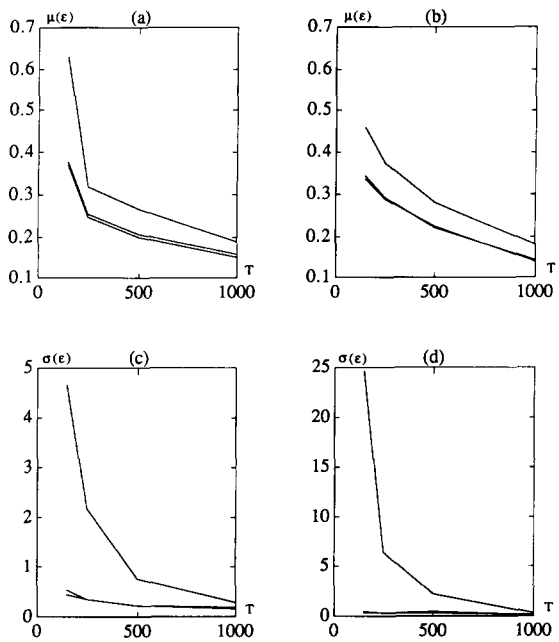

Fig. 1. Identification of a scalar MA model of order 2: statisics of the error $\|\boldsymbol{b}-\hat{\boldsymbol{b}}\|$ are plotted as functions of the data length, T. The three curves correspond to the use of 3, 6 and 9 equations, respectively. (a) Median of the error without gaussian noise and (b) in presence of gaussian noise, (c) standard deviation of the error without gaussian noise and (d) in presence of gaussian noise. It is clear that the more equations we use, the better results, especially for short data lengths. Performances are however about the same for 6 and 9 equations.

Method 3. Use of all the $q(q+1)^2/2$ equations defined by (11).

In these simulations, we generate a scalar MA process of order $q=2$ with coefficients $\boldsymbol{b}=[b_2, b_1, b_0]=[0.74, -0.86, 1.0]$; the driving process $w(t)$ was a zero-mean uniform white noise of unit variance. Denote $\hat{\boldsymbol{b}}$ the estimated $(q+1)$-vector of coefficients. The performance is measured with the help of the error $\varepsilon=\|\boldsymbol{b}-\hat{\boldsymbol{b}}\|$, which was computed for various data lengths, T, and various independent realizations $\{\boldsymbol{w}^{[n]}(t) \mid 1\leqslant t\leqslant T\}$, $1\leqslant n\leqslant 100$. The median, $\mu(\varepsilon)$, and the standard deviation

$$\sigma(\varepsilon)=\left(\frac{1}{100}\sum_{n=1}^{100}\{[\varepsilon^{[n]}-\mu(\varepsilon)]^2\}\right)^{1/2} \tag{31}$$

of the error ε were computed over the 100 independent realizations $\varepsilon^{[n]}$. Figure 1 shows the results obtained in absence of gaussian noise (Figs. 1(a) and 1(c)), and in the presence of a gaussian noise $v(t)$ of variance $1/100$ (Figs. 1(b) and 1(d)). Values have been computed for data lengths 150, 250, 500 and 1000.

For short data lengths, Estimators 2 and 3 perform really better than Estimator 1, whereas for longer data records, all methods tend to be equivalent.

The use of median instead of arithmetic mean has shown to be necessary, especially for short data lengths, because of the large variability of the results obtained. When gaussian noise was present however, arithmetic mean and median give roughly the same result; in other words, the gaussian noise has a 'stabilizing' effect.

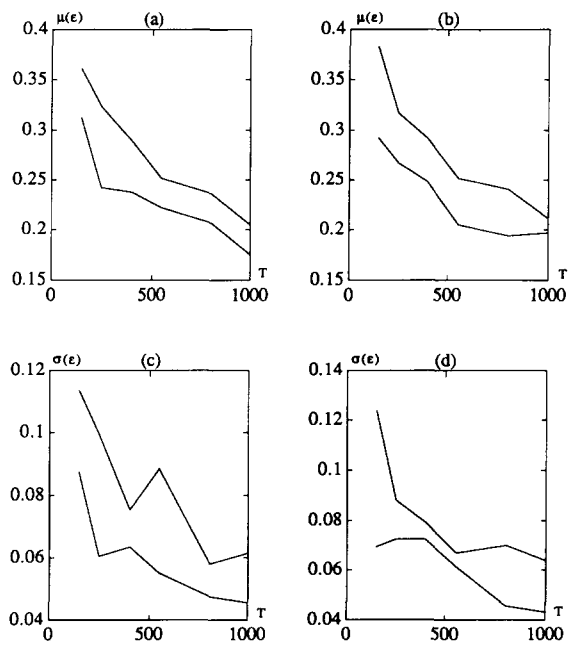

Fig. 2. Identification of an MA model of order 2 and dimension 2: statistics of the error (32) are plotted as functions of the data length. The two curves correspond to the use of 8 and 24 equations, respectively. (a) Median of the error without gaussian noise and (b) in presence of gaussian noise, (c) standard deviation of the error without gaussian noise and (d) in presence of gaussian noise.

Multichannel MA models

Method 1. Use of the equations in (30) for which $j=0$. This system has p^2q equations, and as many unknowns.

Method 2. Use of all equations in (30). This system is formed of $p^2q(q+1)$ equations, and contains p^2q unknowns.

In these simulations, we generate a scalar MA process of order $q=2$ and of dimension $p=2$. The coefficients were $B_0=I$, $B_1=-\text{Diag}[0.86, 0.77]$, $B_2=\text{Diag}[0.74, 0.59]$. The driving input $\boldsymbol{w}(t)$ was composed of two independent components, each zero-mean, unit-variance, and uniformly distributed. For each realization and each data length the error was defined as

$$\varepsilon=\left(\frac{\sum_{k=1}^{q}\|\hat{B}_k-B_k\|^2}{\sum_{k=1}^{q}\|B_k\|^2}\right)^{1/2}. \tag{32}$$

The interest in normalizing the error is to make it less sensitive to the dimension p (comparisons will be possible for different values of p). The results presented in Fig. 2 consist of the median and the standard deviation of ε, estimated over 100 independent trials as in the scalar case. Here ε was computed for data length $T\in\{150, 250, 400, 550, 800, 1000\}$.

As in the scalar case, Figs. 2(a) and 2(b) are plots of the median with and without gaussian noise, and Figs. 2(c) and 2(d) are plots of the corresponding standard deviations. The gaussian noise $\boldsymbol{v}(t)$ had a covariance $1/100\,I$.

Acknowledgment

The author thanks J.F. Podevin for his help in running the computer simulations in the multichannel case. The comments of one referee have been especially useful as an aid to improve the clarity of the paper.

References

[1] J.W. Brewer, "Kronecker products and matrix calculus in system theory", *IEEE Trans. Circuits and Systems*, Vol. 25, No. 9, September 1978, pp. 772–781.

[2] D.R. Brillinger, *Time Series Data Analysis and Theory*, Holden-Day, San Fancisco, CA, 1981.

[3] P. Comon, "Separation of sources using high-order cumulants", *SPIE Conf. Advanced Algorithms and Architecture*, Vol. Real-Time Sig. Proc. XII, San Diego, 8–10 August 1989, pp. 170–181.

[4] G. Giannakis, Y. Inouye and J. Mendel, "Cumulant-based identification of multichannel moving-average models", *IEEE Trans. Automat. Control*, Vol. 34, No. 7, July 1989, pp. 783–787.

[5] C.R. Rao, *Linear Statistical Inference and Its Applications*, Wiley, New York, 1973.

[6] A. Swami, G.B. Giannakis and J.M. Mendel, "A unified approach to modeling multichannel ARMA processes", *IEEE Internat. Conf. Acoust. Speech Signal Process.*, Glasgow, 1989, pp. 2182–2185.