ELSEVIER

Signal Processing 51 (1996) 15–28

SIGNAL PROCESSING

MA blind identification based on order statistics application to binary-driven systems

Jean-Yves Tourneret*, Bernard Lacaze

ENSEEIHT/GAPSE, 2 rue Camichel, 31071 Toulouse cedex, France

Received 27 April 1995; revised 14 September 1995 and 5 February 1996

Abstract

Most commonly used moving average (MA) parameter estimators are approximations of the maximum likelihood estimator (MLE). For large data records, MLE estimates are asymptotically unbiased and minimum variance with zero mean square error. This paper studies a new MA parameter estimator which is based on order statistics. Under certain specific conditions, this estimator has a low probability of error for a finite number of samples. The estimator performance is studied when used to identify minimum and non-minimum phase systems, in the presence of independent additive noise, Gaussian or otherwise.

Zusammenfassung

Die häufigsten angewandten Estimatoren von MA Parametern sind Näherungen zum Estimator von maximaler Wahrscheinlichkeit. Wegen seiner asymptotischen 'unbiased' und minimalen Varianzeigenschaft, der Gebrauch dieses Estimators für Parameterestimation kann für grosse Datenmengen gerechtfertigt werden. Es ist darüberhinaus hinreichend bekannt, dass sein Schätzfehler asymptotisch gegen Null geht. Dieser Artikel befasst sich mit einem neuen Estimator von MA parametern, welcher auf Statistikordnung beruht. Wir erläutern die Umstände unter denen der Fehler dieses Estimators für eine endliche Anzahl von Signalpunkten mit hoher bekannter Wahrscheinlichkeit gegen Null geht. Ausserdem untersuchen wir das Verhalten dieses Estimators bei Gauss'schem und nicht-Gauss'schem Rauschen sowie für die Erkennung von Systemen von nicht-minimaler Phase.

Résumé

La plupart des estimateurs des paramètres MA sont des approximations de l'estimateur du maximum de vraisemblance. L'estimateur du maximum de vraisemblance est asymptotiquement non biaisé et de variance minimale, ce qui permet de justifier son utilisation pour l'estimation de paramètres à partir d'échantillons de grande taille. De plus, il est bien connu que l'estimateur du maximum de vraisemblance possède une erreur asymptotiquement nulle. Le but de cet article est d'étudier un nouvel estimateur MA basé sur la statistique d'ordre. La principale propriété de cet estimateur est qu'il possède, sous certaines conditions qui sont précisées, une erreur nulle pour des nombres de points

* Corresponding author. Tel.: 61 58 83 14; fax: 61 58 82 37; e-mail: tournere@len7.enseeiht.fr.

PII S0165-1684(96)00027-8

finis avec une forte probabilité que nous déterminons. Les performances de cet estimateur, dans le cas de systèmes perturbés par un bruit additif gaussien ou non-gaussien ou dans le cas de systèmes à phase non-minimale, sont ensuite étudiées.

Keywords: Estimation; Parametric modeling; Order statistics; Statistics of extremes

1. Introduction

AutoRegressive Moving Average (ARMA), AutoRegressive (AR) and Moving Average (MA) models have been used successfully in many signal processing applications. These applications include spectral analysis [6], adaptive filtering [3] and pattern recognition [2]. The estimation of the parameters of these models has been studied with increasing interest in recent years. Many methods have been proposed, which use second- or higher-order statistics [6, 8]. These methods account for Gaussian or non-Gaussian signals, the presence of non-linearities and the non-minimum phase property. Two principal methods exist for estimating parameters: (1) the method of moments, which does not always lead to efficient estimators but often is easily implemented and (2) the maximum likelihood method which is often preferred because it is asymptotically unbiased with minimum variance property but may be difficult to compute. Moreover, the Maximum Likelihood Estimator (MLE) has an error, which asymptotically decreases to zero. This paper studies a new MA parameter estimator which is based on order statistics. This estimator, denoted the OS estimator, is restricted to models driven by the special class of white noise input $e(k)$: $e(k)$ is bounded by its maximum $E = \mathrm{Max}\{e(k)\}$ such that $P_E = P[e(k) = E] > 0$. We focus on the special case of binary inputs which leads to the communication channel-equalization problem. The new estimator yields a small probability of error for a finite number of data samples.

This paper is divided into three sections:

(1) MA estimation using order statistics;

(2) OS estimation performance for minimum phase systems with additive noise and comparison with the Durbin estimator;

(3) comparison with the Giannakis–Mendel estimator for non-minimum phase systems.

2. MA estimation

Let $e(k)$ be the input white noise and $x(k)$ be the output of a qth order MA filter, with parameters $b_0, b_1, \ldots, b_q$. The MA model is given by

$$x(k) = \sum_{i=0}^{q} b_i e(k-i), \quad k \in \mathbb{Z}. \tag{1}$$

A new method is proposed for estimating the parameters $\{b_i\}$ for the special class of white noise input $e(k)$: $e(k)$ is bounded by its maximum $E = \mathrm{Max}\{e(k)\}$ such that

$$P_E = P[e(k) = E| > 0 \quad \forall k \in \mathbb{Z}. \tag{2}$$

The paper focuses on the binary input case, when $e(k) = \pm 1$. However, the approach can be extended to discrete or clipped inputs.

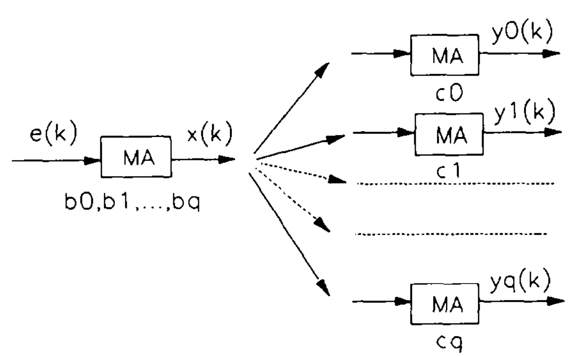

Fig. 1. $q+1$ single-channel model.

2.1. The analysis filter: a convenient tool [10]

A $q+1$ set of first-order MA filters is connected in parallel at the output $x(k)$ as shown in Fig. 1. These filters will be denoted as '*analysis filters*' with parameters $(1, c_j)$, for $j = 0, \ldots, q$. The analysis filter outputs $y_j(k)$ for $j = 0, \ldots, q$ are given by

$$y_j(k) = x(k) + c_j x(k-1)$$
$$= b_0 e(k) + \sum_{i=1}^{q} (b_i + c_j b_{i-1}) e(k-i) + c_j b_q e(k-q-1), \quad j = 0, 1, \ldots, q. \tag{3}$$

The analysis filter output satisfies

$$|y_j(k)| \leqslant |b_0 e(k)| + \sum_{i=1}^{q} |b_i + c_j b_{i-1}| \cdot |e(k-i)| + |c_j b_q| \cdot |e(k-q-1)|. \tag{4}$$

Thus, the maximum of the sequence $\{y_j(k)\}_{k \in \mathbb{Z}}$, denoted by M_j, is

$$M_j = |b_0| + \sum_{i=1}^{q} |b_i + c_j b_{i-1}| + |c_j b_q|. \tag{5}$$

By denoting $S_j(i) = \operatorname{sign}(b_i + c_j b_{i-1})$ for $i = 1, \ldots, q$, $S(0) = \operatorname{sign}(b_0)$ and $S_j(q+1) = \operatorname{sign}(c_j b_q)$, Eq. (5) can be rewritten as

$$M_j = b_0 S(0) + \sum_{i=1}^{q} (b_i + c_j b_{i-1}) S_j(i) + c_j b_q S_j(q+1)$$

or

$$M_j = \sum_{i=0}^{q} b_i [S_j(i) + c_j S_j(i+1)]. \tag{6}$$

The concatenation of (6), for $j = 0, \ldots, q$, yields a linear system of $q+1$ equations with $q+1$ unknowns b_i. This system can be written as

$$\underline{M} = A \underline{b}, \tag{7}$$

with

$$\underline{M} = [M_0, \ldots, M_q]^{\mathrm{T}},$$

$$\underline{b} = [b_0, \ldots, b_q]^{\mathrm{T}},$$

$$A = [a_{ji}] = [S_j(i) + c_j S_j(i+1)].$$

The matrix equation (7) will yield unique MA parameters provided that the $(q+1) \times (q+1)$ matrix A is of full rank $q+1$. The parameters $\{c_j\}$ must be chosen such that the $q+1$ vectors $\underline{L}_j = [S(0) + c_j S_j(1), \ldots, S_j(q) + c_j S_j(q+1)]^{\mathrm{T}}$ are linearly independent (condition A). If condition A is satisfied, (7) provides a new MA parameter estimation method for inputs satisfying the 'bounding' condition in (2). In what follows, a convenient way is presented for choosing filter analysis parameters which satisfy condition A.

2.2. Choice of filter analysis parameters

First, some rough MA parameter estimates are obtained using a conventional method such as Durbin [6]. These estimates will be denoted b_i^* for $i = 0, \ldots, q$. Let $r_i^* = -b_i^*/b_{i-1}^*$ for $i = 1, \ldots, q$ and $r_{q+1}^* = 0$.

Secondly, the r_i^* are arranged in increasing order. This gives the r_i^* *order-statistic vector* $[s_1^*, \ldots, s_{q+1}^*]$ such that

$$s_1^* \leqslant s_2^* \leqslant \cdots \leqslant s_{q+1}^*.$$

The parameters c_j, for $j = 0, \ldots, q$, are then chosen in the following way:

$$c_0 < s_1^* < c_1 < s_2^* < \cdots < s_q^* < c_q < s_{q+1}^*. \tag{8}$$

Note that (8) does not take into account cases where some s_i^* are equal. This particular case does not occur in most applications, because of numerical effects for instance.

Under some specific conditions, the next part shows that this choice of analysis filter parameters, leads to $q+1$ linearly independent vectors $\underline{L}_0, \ldots, \underline{L}_q$. Note that $\underline{L}_j$ can be written as

$$\underline{L}_j = \underline{U}_j + c_j \underline{V}_j,$$

with $\underline{U}_j = [S(0), S_j(1), \ldots, S_j(q)]^{\mathrm{T}}$ and $\underline{V}_j = [S_j(1), \ldots, S_j(q+1)]^{\mathrm{T}}$. For $j = 0, \ldots, q$, $\underline{U}_j$ and $\underline{V}_j$ are vectors whose components are ± 1 (note that both vectors $\underline{U}_j$ and $\underline{V}_j$ depend on c_j values).

Assume that the estimates of $S_j(i)$, obtained from parameters b_i^*, are equal to $S_j(i)$ (this is generally the case because $\operatorname{sign}(b_i^* + c_j b_{i-1}^*) = \operatorname{sign}(b_i + c_j b_{i-1})$ as discussed in the next section). Appendix

A then shows that analysis filter parameters, satisfying (8), lead to $q+1$ linearly independent vectors $\underline{V}_0, \dots, \underline{V}_q$. Hence

$$\mathrm{Det}(\underline{V}_0, \dots, \underline{V}_q) \neq 0. \tag{9}$$

The determinant of the $q+1$ vectors $\underline{L}_0, \dots, \underline{L}_q$ can be considered as a multivariate polynomial $P = P(c_0, \dots, c_q)$ with $q+1$ unknowns $c_0, \dots, c_q$ such that

$$\begin{aligned} P(c_0 \dots, c_q) &= \mathrm{Det}(\underline{L}_0, \dots, \underline{L}_q) \\ &= c_0 c_1 \cdots c_q \mathrm{Det}(\underline{V}_0, \dots, \underline{V}_q) + \cdots . \end{aligned} \tag{10}$$

Inequality (9) implies that this multivariate polynomial differs from the zero polynomial (the coefficient of $c_0 \cdots c_q$ differs from zero). The following property can then be obtained:

If the filter analysis parameters differ from the zeros of the multivariate polynomial P, the vectors $\underline{L}_0, \dots, \underline{L}_q$ are linearly independent.

2.3. Practical implementation

For a practical implementation, the largest values M_j are estimated using K analysis filter output samples $y_j(0), y_j(1), \dots, y_j(K-1)$ and denoted $\hat{M}_j$. The different $S_j(i)$ sign estimates are obtained from the rough MA parameter estimates. The analysis filter parameters are chosen as

$$c_i = \frac{s_i^* + s_{i+1}^*}{2}, \quad i = 1, \dots, q, \qquad c_0 < s_1^*.$$

The matrix system (7) then leads to

$$A\underline{\hat{b}} = \underline{\hat{M}}, \tag{11}$$

with $A = [a_{ji}] = [S_j(i) + c_j S_j(i+1)]$, $\underline{\hat{b}} = [\hat{b}_0, \dots, \hat{b}_q]^{\mathrm{T}}$ and $\underline{\hat{M}} = [\hat{M}_0, \dots, \hat{M}_q]^{\mathrm{T}}$. The solution of (11) yields the MA parameter estimates $\hat{b}_i$. Note, especially, a fundamental property of the proposed estimator: *When the $q+1$ absolute maxima M_j are reached (i.e. when $\hat{M}_j = M_j$ for $j = 0, \dots, q$) and the $S_j(i)$ estimates are correct, the MA parameter estimations are equal to $b_0, b_1, \dots, b_q$ and have a zero error.* The assumption that the signs $S_j(i)$ are correct may seem surprising. An estimate of the corresponding probability can be obtained as follows. Let us assume that the Durbin estimates b_i^* are Gaussian with mean b_i. The Gaussian assumption for Durbin estimations seems to be reasonable for large data records. It cannot be easily justified theoretically. The asymptotic normality results for the maximum likelihood estimator cannot be used when the range of the likelihood function depends on the true parameters to be estimated [6,7]. However, normality tests (such as the Kolmogorov–Smirnov test) have been applied to real data and have shown that this assumption is realistic.

In the case of Gaussian Durbin estimates, the variables $b_i^* + c_j b_{i-1}^*$ are Gaussian with mean $b_i + c_j b_{i-1}$ and variance σ_{ij}^2. When $b_i + c_j b_{i-1} > 0$, the probability that the $S_j(i)$ estimate is correct is the probability that $b_i^* + c_j b_{i-1}^* > 0$, which is given by

$$P[b_i^* + c_j b_{i-1}^* > 0] = \frac{1}{\sqrt{2\pi}} \int_{\frac{-(b_i + c_j b_{i-1})}{\sigma_{ij}}}^{+\infty} \mathrm{e}^{-u^2/2}\, \mathrm{d}u.$$

As soon as $(b_i + c_j b_{i=1})/\sigma_{ij} > 4$ or 5, $P[b_i^* + c_j b_{i-1}^* > 0] = 1 - \varepsilon$, with $\varepsilon < 5e-5$. The sign $S_j(i)$ is then correct with the probability $1-\varepsilon$.

The aim of the next section is to show that, for binary inputs $e(K)$, the equalities $\hat{M}_j = M_j$ for $j = 0, \dots, q$ are reached with a high known probability depending on the number of samples K and the MA model order q.

2.4. Probability of error

For binary inputs such that $P[e(k) = 1] = P_E > 0$, the property $\hat{M}_j = M_j$ is verified if there exists an integer $k \in \{0, \dots, K-1\}$ such that

$$\begin{aligned} & b_0 e(k) + \sum_{i=1}^{q} (b_i + c_j b_{i-1}) e(k-i) \\ & + c_j b_q e(k-q-1) \\ & = |b_0| + \sum_{i=1}^{q} |b_i + c_j b_{i-1}| + |c_j b_q|. \end{aligned} \tag{12a}$$

Alternatively,

$$[e(k), e(k-1), \dots, e(k-q), e(k-q-1)] = [S(0), S_j(1), \dots, S_j(q), S_j(q+1)]. \tag{12b}$$

Note that the number of parameters $S_j(i)$ is $q+2$. $\hat{M}_j = M_j$ if there exists an integer $k \in \{0, \dots, K-1\}$ such that the vector $[e(k), e(k-1), \dots, e(k-q), e(k-q-1)]$ is equal to a binary vector. Define P_k^{q+2} as the probability of having at least one integer $k \in \{0, \dots, K-1\}$ such that (12b) is satisfied. P_k^{q+2} can be recursively determined, for a qth MA filter and K samples [10]. For instance, when the maximum is obtained for $e(k) = e(k-1) = \cdots = e(k-q-1)$ (which is the worst case i.e. leads to the lowest probability), P_k^{q+2} is given by

$$P_K^{q+2} = (P_E)^{q+2} + (P_E)^{q+3}(K-q-2), \quad q+2 \leqslant K < 2q+5, \tag{13}$$

$$P_K^{q+2} = (P_E)^{q+2} + (P_E)^{q+3}\left[K-q-2-\left(\sum_{i=q+2}^{K-q-3} P_i^{q+2}\right)\right], \quad K \geqslant 2q+5. \tag{14}$$

Consider the special case of a binary input with equally likely values ± 1. Then

$$P_E = P[e(k) = +1] = P[e(k) = -1] = \tfrac{1}{2}.$$

For this particular case, P_K^{q+2} is shown in Fig. 2 as a function of q and K. For instance, for a $q=7$ order MA model and for $K = 10\,000$, $P_{10\,000}^{9} \approx 0.99$. When (12a) and (12b) are satisfied for every $j = 0, \dots, q$, MA parameter estimates have a zero error. The accuracy is then limited only by the computer's last significant digit.

The determination of the OS estimator bias and variance is a difficult task because the variables $y_j(k)$, for $j = 0, \dots, q+1$, are not independent. However, the following results can be obtained. When the $S_j(i)$ are known, the matrix A of (11) is deterministic and

$$E[\underline{\hat{b}}] - \underline{b} = A^{-1}(E[\underline{\hat{M}}] - \underline{M}), \tag{15a}$$

$$E[(\hat{b}-b)(\hat{b}-b)^{\mathrm{T}}] = A^{-1} C_{\hat{M}} (A^{-1})^{\mathrm{T}}, \tag{15b}$$

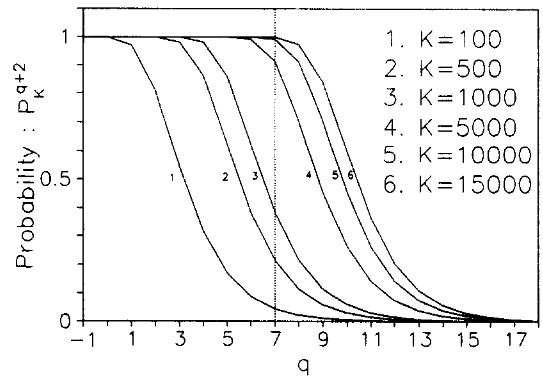

Fig. 2. Probability P_K^{q+2} as a function of q and of the number of samples K.

$C_{\hat{M}}$ being the $\hat{M}$ covariance matrix and $E[\cdot]$ denoting the mathematical expectation. In the case of a binary input, the analysis filter output $y_j(k)$ is a discrete variable. $\hat{M}_j = \mathrm{Max}_k\, y_j(k)$ is then a discrete variable with different values denoted by m_j^i, with $m_j^j = M_j$. The bias of $\hat{M}_j$ then satisfies

$$\begin{aligned}|E[\hat{M}_i - M_i]| &\leqslant \sum_{j\neq i} |m_j^i - M_i| P[\hat{M}_i = m_j^i] \\ &\leqslant \underset{j\neq i}{\mathrm{Max}}\, |m_j^i - M_i| \sum_{j \neq i} P[\hat{M}_i = m_j^i] \\ &\leqslant \underset{j\neq i}{\mathrm{Max}}\, |m_j^i - M_i|(1 - P_K^{q+2}).\end{aligned}$$

For any q, $\lim_{K\to+\infty} P_K^{q+2} = 1$. Thus, *the estimators* $\underline{\hat{M}}$ and $\underline{\hat{b}}$ *are asymptotically unbiased.*

The $\hat{M}_i$ mean square error (MSE) satisfies the following equation:

$$\begin{aligned}E[(\hat{M}_i - M_i)^2] &= \sum_{j\neq i} (m_j^i - M_i)^2 P[\hat{M}_i = m_j^i] \\ &\leqslant \underset{j\neq i}{\mathrm{Max}}\, (m_j^i - M_i)^2 \sum_{j\neq i} P[\hat{M}_i = m_j^i] \\ &\leqslant \underset{j\neq i}{\mathrm{Max}}\, (m_j^i - M_i)^2 (1 - P_K^{q+2}).\end{aligned}$$

Hence

$$\lim_{K\to+\infty} E[(\hat{M}_i - M_i)^2] = 0.$$

The OS estimator is consistent.

Table 1
Comparison between OS and Durbin estimator bias for different numbers of samples

K	$\hat{b}_1$ bias (Durbin)	$\hat{b}_1$ bias (OS)	$\hat{b}_2$ bias (Durbin)	$\hat{b}_2$ bias (OS)
50	1.16E − 1	2.72E − 1	5.56E − 2	9.94E − 2
100	6.40E − 2	5.47E − 2	2.75E − 2	2.24E − 2
150	2.97E − 2	1.51E − 2	1.59E − 2	4.51E − 3
200	2.89E − 2	6.74E − 4	1.38E − 2	1.98E − 3
250	9.50E − 3	1.10E − 14	3.64E − 3	1.90E − 15
500	9.37E − 3	1.11E − 14	3.67E − 3	1.89E − 15

Table 2
Comparison between OS and Durbin estimator variance for different numbers of samples

K	$\hat{b}_1$ variance (Durbin)	$\hat{b}_1$ variance (OS)	$\hat{b}_2$ variance (Durbin)	$\hat{b}_2$ variance (OS)
50	1.40E − 1	4.50	5.55E − 2	4.50
100	6.30E − 2	1.20E − 1	2.40E − 2	1.20E − 1
150	4.30E − 2	4.80E − 2	1.50E − 2	4.70E − 3
200	3.10E − 2	1.60E − 3	1.10E − 2	3.50E − 3
250	1.30E − 2	1.11E − 14	3.65E − 3	1.90E − 15
500	1.20E − 2	1.10E − 14	3.60E − 3	1.90E − 15

2.5. *Comparison with the Durbin estimator*

How does the OS estimator compare to the conventional Durbin one [6]? Consider a second-order MA model, with parameters $b_0 = 1$, $b_1 = -0.532$ and $b_2 = 0.338$. In terms of bias and variance, the performance of these two estimators has been evaluated for different sample numbers with 1000 Monte Carlo runs and is shown in Tables 1 and 2. In the absence of noise, the OS estimator outperforms the Durbin estimator when $K \approx > 200$. Note that, as the number of samples increases (i.e. $K \approx > 250$, 500), the OS estimator has a zero error. The results are then more accurate than those of any other estimation procedure.

3. Noise effects

This section studies the effects of additive noise. The model is corrupted by a continuous additive noise $n(k)$, which is assumed to be i.i.d. (independent identically distributed) and statistically independent of the input $e(k)$. The signal to noise ratio is defined as

$$\mathrm{SNR} = 10\log_{10}\left[\frac{E[(z(k)-n(k))^2]}{E[n^2(k)]}\right].$$

The model is defined by the following equation:

$$z(k) = n(k) + \sum_{i=0}^{q} b_i e(k-i), \quad k \in \mathbb{Z}. \tag{16}$$

The 'analysis filter' outputs are shown in Fig. 3 (denoted by $y_j(k)$ for $j = 0, 1, \ldots, q$) and are given by

$$y_i(k) = b_0 e(k) + \sum_{i=1}^{q} (b_i + c_j b_{i-1}) e(k-i) + c_j b_q e(k-q-1) + n(k) + c_j n(k-1). \tag{17}$$

Thus, $y_j(k) = e_j(k) + n_j(k)$, where

$$e_j(k) = b_0 e(k) + \sum_{i=1}^{q} (b_i + c_j b_{i-1}) e(k-i) + c_j b_q e(k-q-1)$$

is a discrete variable with 2^{q+2} different values, denoted by v_i for $i = 1, \ldots, 2^{q+2}$, and where $n_j(k) = n(k) + c_j n(k-1)$ is a continuous variable. Let $f_{nj}(x)$ be the probability density function (p.d.f.) of $n_j(k)$. The p.d.f. of $y_j(k)$ is then given by

$$f_{y_j}(x) = \frac{1}{2^{q+2}} \sum_{i=1}^{2^{q+2}} f_{nj}(x - v_i). \tag{18}$$

Hence, the analysis filter output statistics can be determined. Assume that the $S_j(i)$ estimates obtained from Durbin estimates are correct, the matrix A from (11) is deterministic. The OS estimator bias and covariance matrix are thus given by (15a) and (15b). In the next part of the paper, $f_{y_j}(x)$ is used with (15a) and (15b) to determine the OS estimator bias and covariance matrix. The study is carried out for different additive noise models.

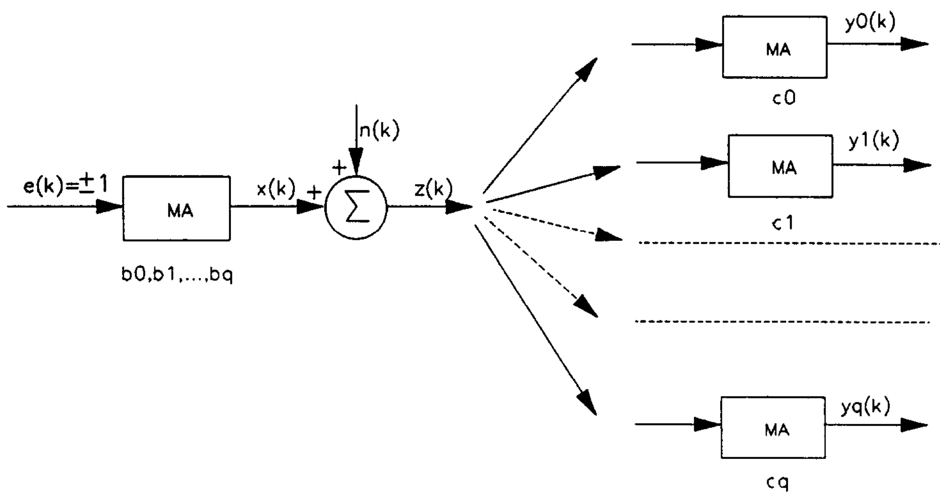

Fig. 3. Single-channel system with analysis filters.

3.1. Additive Gaussian noise [11]

$n(k)$ is assumed to be a white Gaussian process with zero mean and variance σ^2. The variable $n_j(k) = n(k) + c_j n(k-1)$ is then Gaussian with zero mean and variance $\sigma_j^2 = (1 + c_j^2)\sigma^2$. From (18), the analysis filter output p.d.f. is given by

$$f_{y_j}(x) = \frac{1}{2^{q+2}} \sum_{i=1}^{2^{q+2}} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left[-\frac{(x - v_i)^2}{2\sigma_j^2} \right]. \quad (19)$$

The next calculation would be to determine the statistics of $\hat{M}$ but this is not an easy task. The determination of the exact statistics of each component $\hat{M}_j$ is even a difficult problem because the corresponding variables $y_j(k)$ are not independent. Much work has been done in the extreme value field. Gumbel [5] has shown the existence of three asymptotic distributions for the largest values. Each distribution assumes a specific behaviour for absolute large values of the variable. The asymptotic $\hat{M}_j$ distribution can be determined with the help of the following properties:

(1) The distribution of the largest normal values converges towards the Gumbel first asymptote [5].
(2) Given N p.d.f.'s $f_i(t)$ with $i = 1, \ldots, N$ of N variables whose largest value distributions converge towards the Gumbel first asymptote, it can be shown that the variable whose p.d.f. is $(1/N)\sum_{i=1}^{N} f_i(t)$ has a largest value distribution converging to the same asymptote.

Thus, for a high number of samples K, the $\hat{M}_j$ p.d.f. can be approximated by the first asymptote:

$$F(x) = P[X < x] = \exp[-\mathrm{e}^{-\alpha_K(x-u_K)}],$$
$$\alpha_K > 0, \quad x > 0.$$

The parameters α_K and u_K, which depend on the number of samples K, have then to be estimated. Numerous methods for estimating these two parameters are available in the literature [5]. They lead to a theoretical p.d.f. which can be compared with an $\hat{M}_j$ histogram. Figs. 4(a) and (b) present results for a second-order MA model with parameters $b_0 = 1$, $b_1 = -0.532$, $b_2 = 0.338$ and a signal to noise ratio (SNR) equal to 10 dB. For a $K = 10\,000$ number of samples, good agreement is shown between the theoretical and estimated $\hat{M}_j$ p.d.f.

The OS estimator bias can be determined from (15a) and from the $\hat{M}_j$ asymptotic distribution first moment given in [5]:

$$E[\hat{M}_j] = u_K + \frac{\gamma}{\alpha_K},$$

γ being the Euler constant ($\gamma \approx 0.577$). In the presence of an additive Gaussian noise, the main drawback of the OS estimator lies in its bias. When $n(k)$

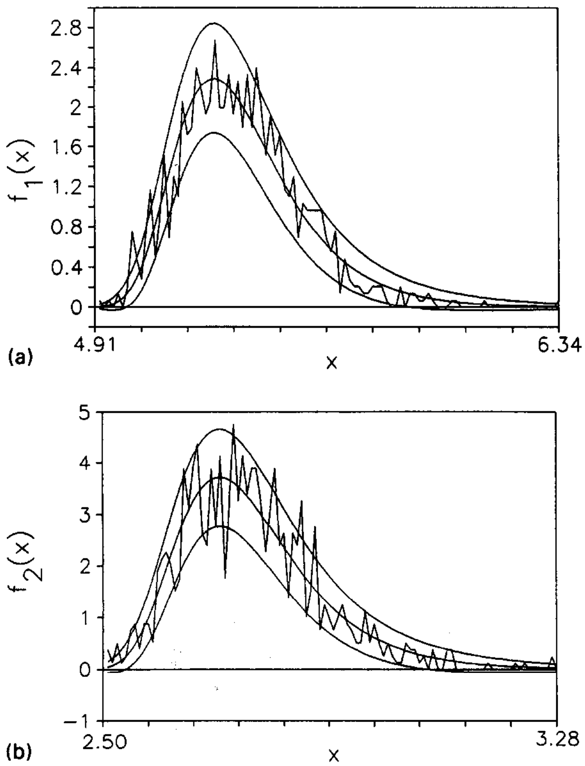

Fig. 4. Comparison between theoretical and estimated $\hat{M}_j$ p.d.f. with 95% confidence intervals: (a) $\hat{M}_1$ p.d.f.; (b) $\hat{M}_2$ p.d.f.

is not bounded, the following property is verified:

$$\lim_{K \to +\infty} E[\hat{M}_j] = \lim_{K \to +\infty} u_K + \frac{\gamma}{\alpha_K} = +\infty.$$

In the presence of an additive Gaussian noise, the OS estimator is asymptotically biased. It can be shown that the Durbin estimator performs better than the OS estimator [11]. The bias of the OS estimator, for models corrupted by an additive Gaussian noise, leads to an unsatisfactory behaviour. The next step is to design an unbiased version of the OS estimator for models corrupted by an additive bounded noise.

3.2. *Bounded additive noise* [12]

3.2.1. *Modeling*

The usual way of modelling system perturbations with an additive Gaussian noise can be explained as follows:

- The Gaussian assumption can sometimes be justified by means of theoretical works such as those on central limit theorems.
- The estimation of the noise statistics usually leads to a Gaussian-shaped p.d.f.
- In most applications, conventional tests (such as Kolmogorov or Chi2 ones), allow us to assume that real noise data are Gaussian.

However, the perturbations of physical systems are always bounded, because of sensor finite dynamics for instance. Under these conditions, perturbation model can be made with a Gaussian-shaped bounded noise. A zero mean bounded noise $n(k)$ can be generated with a Gaussian process $g(k)$, a one-to-one application f from the real set to a bounded interval $[-N, +N]$ and the relation $n(k) = f[g(k)]$. For example, the smooth clipping function $f(x) = \beta \operatorname{Arctg}(\alpha x)$ (which is conventional in electronics) preserves the main lobe shape of the Gaussian p.d.f. (due to its approximate linearity) and changes only its tails. This function leads to the p.d.f. of $n(k)$:

$$f_n(x) = \frac{1 + \operatorname{tg}^2(x/\beta)}{\alpha\beta\sqrt{2\pi}} \exp\left[-\frac{\operatorname{tg}^2(x/\beta)}{2\alpha^2} \right],$$

$$x \in \left] -\tfrac{1}{2}\pi\beta, \tfrac{1}{2}\pi\beta \right[. \tag{20}$$

The two parameters α and β can be determined from noise observations. Fig. 5 gives an insight of how this bounded noise compares with a physical one. In this figure, the p.d.f. of the bounded noise is compared with a histogram from real noise data (interferences for an Instrument Landing System [1]). Good agreement is shown between the theoretical noise p.d.f. and the histogram of the actual noise. Thus, the interferences in an Instrument Landing System can be modeled as a bounded noise obtained with a Gaussian process transformation. Consider now the study of the OS estimator for the identification of models corrupted by such an additive bounded noise.

3.2.2. *Modified OS estimator*

For an additive noise $n(k)$ bounded by its maximum N, the maximum of the analysis filter output

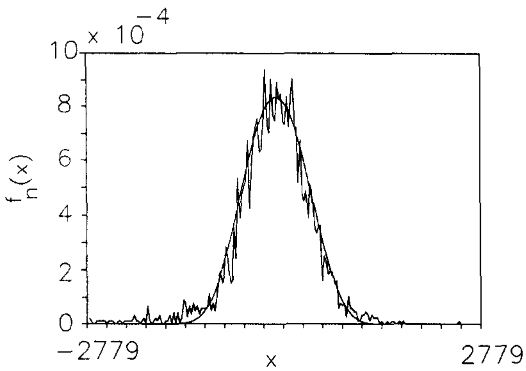

Fig. 5. Comparison between the bounded noise p.d.f. and the ILS interference histogram.

is given by

$$M_j = |b_0| + \sum_{i=1}^{q} |b_i + c_j b_{i-1}| + |c_j b_q| + (1 + |c_j|)N. \quad (21)$$

The concatenation of this equation for $j = 0, \ldots, q+1$ leads to a matrix system very similar to (7) with $q+2$ equations and $q+2$ unknowns $b_0, b_1, \ldots, b_q$ and N:

$$Ab = M, \quad (22)$$

with $M = [M_0, \ldots, M_{q+1}]^{\mathrm{T}}$, $b = [b_0, \ldots, b_q, b_{q+1}]^{\mathrm{T}}$ and $b_{q+1} = N$. This system of equations allows simultaneous estimation of the model parameters and the noise maximum. The properties of this *modified OS estimator* are now examined.

The modified OS estimator bias can be determined from $\hat{M}_j$ statistics as before. $y_j(k)$ is a bounded random variable. Using Gumbel's results [5], the distribution of its largest value converges towards the third asymptote given by

$$F(x) = \exp\left[-\left(\frac{\omega_K - x}{\omega_K - v_K}\right)^{k_K}\right], \quad x \leqslant \omega_K, \quad (23)$$

with $v_k < \omega_K$ and $k_K > 0$. The parameters k_K, ω_K and v_K (which depend on the number of samples K) can be estimated with the different methods that can be found in [5]. The first moment corresponding to this asymptote is

$$E[\hat{M}_j] = \omega_K + (\omega_K - v_K)\Gamma(1 + 1/k_K). \quad (24)$$

It can be shown that (see [5])

$$\lim_{K\to+\infty} v_K = \lim_{K\to+\infty} \omega_K = M_j. \quad (25)$$

Thus, *the modified OS estimator is asymptotically unbiased.*

The determination of the modified OS estimator covariance matrix is a difficult problem because variables $\hat{M}_j$ are not independent (see (15b)). However, the following results can be obtained.

The jth diagonal term of $C_{\hat{M}}$ is equal to the $\hat{M}_j$ variance (i.e. the first asymptote order-two moment given in [5])

$$\mathrm{Var}[\hat{M}_j] = (v_K - \omega_K)^2 [\Gamma(1 + 2/(k_K) - \Gamma(1 + 1/k_K)^2].$$

Eq. (25) then gives

$$\lim_{K\to+\infty} \mathrm{Var}[\hat{M}_j] = 0. \quad (26)$$

The diagonal terms of $C_{\hat{M}}$ tend to zero asymptotically.

Using the Schwarz inequality

$$0 \leqslant |\mathrm{cov}(\hat{M}_i, \hat{M}_j)|^2 \leqslant \mathrm{Var}[\hat{M}_i]\,\mathrm{Var}[\hat{M}_j] \quad (27)$$

Eqs. (26) and (27) then give

$$\lim_{K\to+\infty} \mathrm{cov}(\hat{M}_i, \hat{M}_j) = 0,$$

with $\mathrm{cov}(\hat{M}_i, \hat{M}_j) = E[(\hat{M}_i - E(\hat{M}_i))(\hat{M}_j - E(\hat{M}_j))]$. The off-diagonal terms of $C_{\hat{M}}$ tend to zero asymptotically.

Eq. (15b) then shows that, *in the presence of an additive bounded noise, the modified OS estimator is consistent.*

In terms of bias and variance (estimated from 1000 Monte Carlo runs), a comparison between the modified OS and Durbin estimator performance is presented in Figs. 6 and 7, for a finite number of samples. Despite being asymptotically unbiased, the modified OS estimator gives large bias for low signal to noise ratio. The results for the variance are more satisfactory for models with low perturbations.

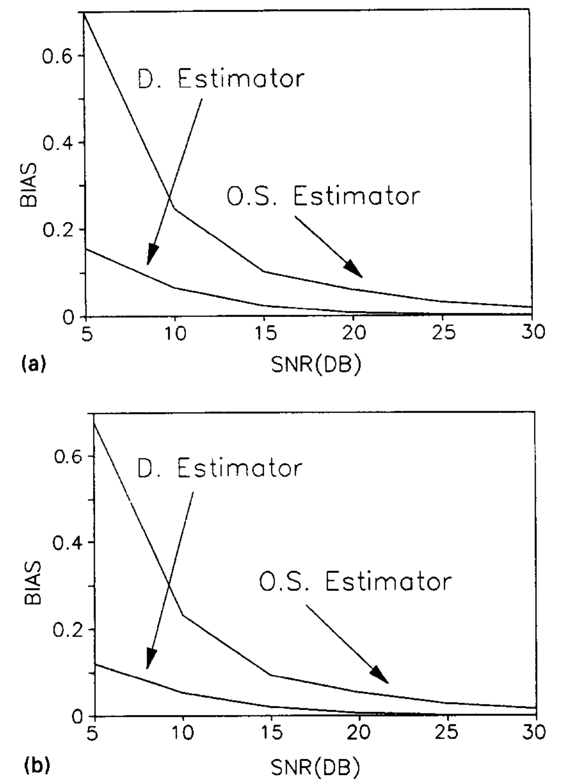

Fig. 6. Comparison between the modified OS and the Durbin estimator bias for different signal to noise ratios: (a) $\hat{b}_1$ bias, (b) $\hat{b}_2$ bias.

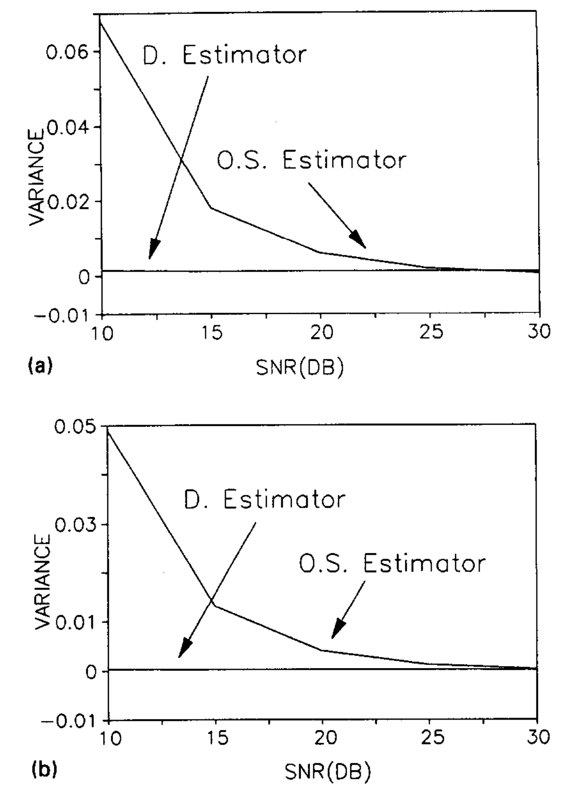

Fig. 7. Comparison between the modified OS and the Durbin estimator variance for different signal to noise ratios: (a) $\hat{b}_2$ variance, (b) $\hat{b}_2$ variance.

3.3. *Comparison with the Yellin and Porat estimator* [15]

There are some similarities between the OS estimator and the Yellin and Porat estimator:

- both methods are restricted to discrete-alphabet inputs and provide, in the noiseless case, the exact MA parameters.
- Both methods are sensitive to the noise level but do not require assumptions on the nature of the noise (the noise can be Gaussian, non-Gaussian, independent or correlated with the input, etc.)

A comparison between the two estimators for the first test case developed in [15] leads to the following results:

Yellin and Porat algorithm
Bias (absolute value)
$= \{0.0071, 0.0050, 0.0045, 0.0052, 0.0074\}$,
Standard deviation
$= \{0.0069, 0.0044, 0.0038, 0.0040, 0.0054\}$.

Our algorithm (modified OS estimator)
Bias (absolute value)
$= \{0.0081, 0.0014, 0.0012, 0.0031, 0.013\}$,
Standard deviation
$= \{0.012, 0.0057, 0.024, 0.0033, 0.047\}$.

It can be seen that the two approaches give very similar results. However, our estimator can be preferred because of its lower computational cost. The method of Yellin and Porat consists in determining equivalent measurements (measurements that correspond to the same input sequence). The algorithm for finding the equivalent measurements is not simple. A set S, which will appear in clusters, has to be constructed. These clusters have then to be identified and to be separated. The set S contains outliers that have to be removed with another algorithm. Moreover, the method of Yellin and Porat is recursive. As it is discussed in their paper, the estimation of the first parameter b_0 is crucial. Our algorithm is not recursive and does not suffer from the problem induced by the estimation of b_0.

3.4. Model order mismatch

When the model order is underestimated, the performance of our algorithm is very similar to the performance of the Yellin and Porat algorithm. The unmodeled part of the MA model output can be viewed as an additive zero mean discrete noise. Our algorithm is sensitive to the noise level but not to the nature of the additive noise. It will perform well as long as the energy of the unmodeled part is negligible with respect to the energy of the modeled part. For instance, consider the case of a fifth-order MA model with parameters

$$b_1 = -0.532, \quad b_2 = 0.338, \quad b_3 = 0.061,$$
$$b_4 = -0.0525, \quad b_5 = 0.012.$$

The model order used by the algorithm is $q = 2$. For an SNR = 40 dB, the following results are obtained:

Bias (absolute value) = $\{0.031, 0.0195\}$,

Standard deviation = $\{0.0092, 0.0085\}$.

These results are very similar to the results obtained with the Yellin and Porat algorithm in the case of underestimation.

When the model order is overestimated, the algorithm breaks down. The choice of the analysis filter parameters, for having a full rank matrix A, becomes impossible because some r_i are equal (see Eq. (8)). This property could induce a model order selection procedure. In practice, our algorithm is limited to high signal to noise ratio (> 30 dB). Under these conditions, the model order selection can be performed with the usual techniques developed in [6].

4. Non-minimum phase MA systems

Recent years have shown an increasing interest in non-minimum phase MA systems. Using higher-order statistics, many techniques have been developed for the identification of these systems driven by non-Gaussian inputs [4, 18, 14]. One of these techniques, proposed by Giannakis and Mendel [4], is based on the following equations (denoted by GM equations):

$$\sum_{k=0}^{q} b_k^2 C_z(m-k) = \left(\frac{\sigma_e^2}{\gamma_{3e}}\right) \sum_{k=0}^{q} b_k C_{3z}(m-k, m-k), \tag{28}$$

$C_z(k)$ and $C_{3z}(k, k)$ being, respectively, the autocorrelation and diagonal third-order cumulants of the process $z(n)$, and σ_e^2, γ_{3e} the input variance and third-order cumulant. In the noise-free case, the concatenation of Eqs. (28) for $m = -q, \ldots, +2q$ (q being the MA model order) leads to a linear overdetermined system with $2q + 1$ unknowns b_k, b_k^2 (for $k = 1, 2, \ldots, q$) and $\varepsilon = \sigma_e^2/\gamma_{3e}$. The solution of this system can be determined using different least-squares techniques. For our simulations, the use of singular value decomposition (SVD) and total least squares have been used as was suggested in [8]. For models embedded in additive Gaussian noise, the range of values taken by m cannot include $0, 1, \ldots, q$. Thus, the following equations (developed by Tugnait [14]) have to be considered:

$$\sum_{k=1}^{q} b_k C_{3z}(m-k, m-k+q) - \left(\frac{\gamma_{3e} b_q}{\sigma_e^2}\right) C_z(m) = -C_{3z}(m, m). \tag{29}$$

The combination of Eqs. (28) and (29), respectively, for $m = -1, \ldots, -q, q+1, \ldots, 2q$ and $m = -q, \ldots, +q$ leads to an overdetermined system that can be solved with least-squares techniques. In the case of a symmetrically distributed input, the third-order cumulants $C_{3y}(m, m)$ are equal to zero. Consequently, Eqs. (28) and (29) have to be extended to order four. For instance,

(28) becomes

$$\sum_{k=0}^{q} b_k^3 C_z(m-k) = \left(\frac{\sigma_v^2}{\gamma_{4v}}\right) \sum_{k=0}^{q} b_k C_{4z}(m-k, m-k, m-k). \quad (30)$$

The modified OS estimator performance is now compared with that obtained with the GM estimator for the identification of non-minimum phase systems driven by binary inputs. The following non-minimum phase MA model is considered [4]:

$$z(k) = e(k) - 2.0833e(k-1) + e(k-2) + n(k). \quad (31)$$

This model has two reciprocal zeros at $\frac{4}{3}$ and $\frac{3}{4}$ and MA parameters corresponding to the minimum phase system are

$$b_{MP}(0) = 1, \quad b_{MP}(1) = -1.5, \quad b_{MP}(2) = 0.5625. \quad (32)$$

Input samples $e(k)$ are binary, with equally likely values $+1$ and -1. The additive noise $n(k)$ is white Gaussian and statistically independent of $e(k)$. MA parameters are estimated using GM and modified OS estimators from $K = 10\,000$ output measurements. In terms of bias and variance, a comparison between the performance of these two estimators is presented in Figs. 8 and 9. The variance performance of the modified OS estimator is better for high signal to noise ratio than those obtained with the GM algorithm. However, for a finite number of samples, a large bias may remain.

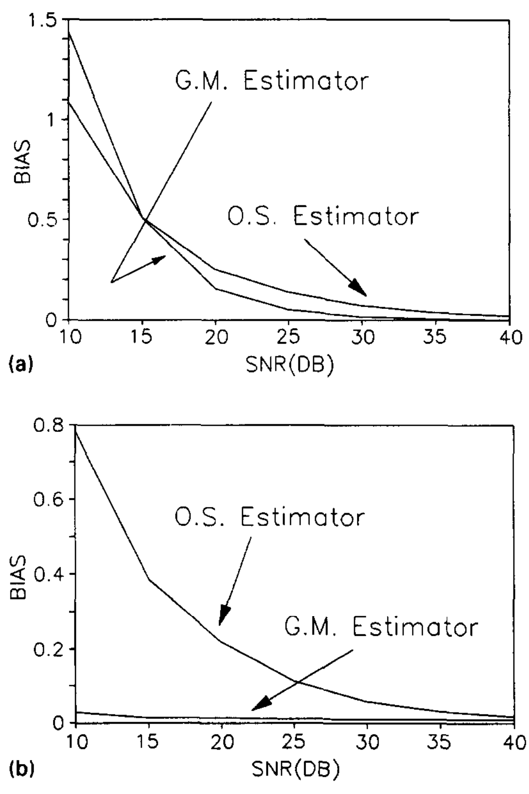

Fig. 8. Comparison between the modified OS and the GM estimator bias for different signal to noise ratios: (a) $\hat{b}_1$ bias, (b) $\hat{b}_2$ bias.

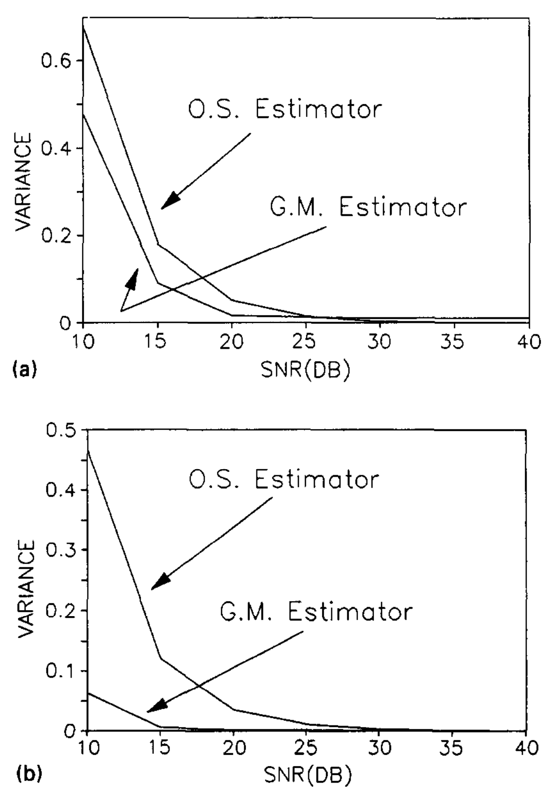

Fig. 9. Comparison between the modified OS and the GM estimator variance for different signal to noise ratios: (a) $\hat{b}_1$ variance, (b) $\hat{b}_2$ variance.

5. Conclusion

The maximum and the order statistics of a moving average model driven by a binary input have been used to estimate its parameters. In the absence of noise, the main property of the estimator (denoted OS estimator) is a zero error for a finite number of samples with a high known probability.

The OS estimator performance has then been studied for the identification of models embedded in additive Gaussian or bounded noise. In terms of bias and variance, a comparison between the OS estimator and the Durbin estimator has been proposed.

The estimation algorithm is not restricted to non-minimum phase systems. A comparison has been presented with a higher-order estimator developed by Giannakis and Mendel. For high signal to noise ratio, the variance of the OS estimator outperforms conventional estimators. For a finite number of samples, its bias may remain large, even in the case of an additive bounded noise.

Appendix A. Choice of analysis filter parameters

Let $r_i = -b_i/b_{i-1}$ for $i = 1, \dots, q$ and $r_{q+1} = 0$. Arranging the r_i ratios in increasing order give $q+1$ parameters denoted by s_i such that

$$s_1 < s_2 < \cdots < s_{q+1}.$$

The vector $[s_1, \dots, s_{q+1}]^{\mathrm{T}}$ is the *order statistics* of the vector $[r_1, \dots, r_{q+1}]^{\mathrm{T}}$. This appendix shows that analysis filter parameters defined by

$$c_0 < s_j^* < c_1 < s_2^* < \cdots < s_q^* < c_q < s_{q+1}^* \qquad \text{(A.1)}$$

lead to $q+1$ *linearly independent vectors* $\underline{V}_j$.

Consider the simple case where all MA parameters are positive (the other cases could be studied in similar way). We then obtain

$$S(0) = 1,$$

$$S_j(i) = 1 \Leftrightarrow b_i + c_i b_{i-1} \geqslant 0 \Leftrightarrow c_j \geqslant -b_i/$$

$$b_{i-1}, S_j(q+1) = 1 \Leftrightarrow c_j > 0.$$

When the parameters b_i are all positive, all ratios s_i with $i = 1, \dots, q+1$ are negative and then $s_{q+1} = 0$. We then get the following results:

For $c_j > s_{q+1} = 0$:

$$S(0) = S_j(1) = \cdots = S_j(q+1) = 1,$$

for $c_j \in]s_q, s_{q+1}[$:

$$S(0) = S_j(1) = \cdots = S_j(q) = 1, \quad S_j(q+1) = -1,$$

etc.

Two different c_j values belonging respectively to intervals $]s, s_{i+1}[$ and $]s_{i+1}, s_{i+2}[$ yield two vectors $\underline{V}_j$ which differ with respect to only one component. With the previous choice of parameters c_j (see (A.1)), the vectors $\underline{V}_j$ for $j = 0, \dots, q$ are all different but only one component of two consecutive vectors is different. An example of such vectors $\underline{V}_j$ is given below:

$$\underline{V}_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix}, \quad \underline{V}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ -1 \end{bmatrix}, \quad \underline{V}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ -1 \\ -1 \end{bmatrix},$$

$$\dots, \quad \underline{V}_q = \begin{bmatrix} 1 \\ -1 \\ -1 \\ \vdots \\ -1 \\ -1 \end{bmatrix}.$$

These vectors $\underline{V}_j$ for $j = 0, \dots, q$ are linearly independent.

For practical implementation, analysis filter parameters are determined from parameters s_j^* (and not from s_j). When the rough estimations of parameters b_i are sufficiently accurate, the following result can be obtained:

$$c_i = \frac{s_i^* + s_{i+1}^*}{2} \Rightarrow c_i \in]s_i, s_{i+1}[,$$

which leads to $q+1$ linearly independent vectors $\underline{V}_j$.

Acknowledgements

The authors would like to acknowledge C. Mailhes and F. Castanié for their helpful comments.

References

[1] F. Castanié and C. Lambert-Nebout, "High performance periodic signal rejection using improved Prony modeling", *Proc. 4th Conf. on Sig. Process. (EUSIPCO'88)*, Grenoble, September 1988, pp. 403–406.

[2] R.O. Duda and P. E. Hart, *Pattern Recognition and Scene Analysis*, Wiley, New York, 1973.

[3] B. Friedlander, "System identification techniques for adaptive signal processing", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-30, April 1982, pp. 240–246.

[4] G.B. Giannakis and J. M. Mendel, "Identification of non-minimum phase systems using higher-order statistics", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-37, March 1989, pp. 360–377.

[5] E.J. Gumbel, *Statistics of Extremes*, Columbia University Press, New York, 1958.

[6] S.M. Kay, *Modern Spectral Estimation*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[7] M.G. Kendall and A. Stuart, *The Advanced Theory of Statistics*, Griffin, London, 1961.

[8] J.M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications", *Proc. IEEE*, Vol. 79, March 1991, pp. 277–305.

[9] J.Y. Tourneret, Contribution à l'étude de modèles ARMA non gaussiens, PhD Dissertation, Enseeiht/Gapse, Toulouse, France, Order No. 605, 1992.

[10] J.Y. Tourneret and B. Lacaze, "A new estimator of moving average model parameters based on order statistics", *Proc. 26th Conf. on Inf. Sciences and Syst.* (*CISS'92*), Princeton University, 18–20 March 1992, pp. 608–613.

[11] J. Y. Tourneret and B. Lacaze, "Study of MA model identification using order statistics in the presence of noise", *Proc. 6th Conf. on Sig. Process.* (*EUSIPCO'92*), Brussels, 24–27 August 1992, pp. 827–830.

[12] J.Y. Tourneret and B. Lacaze, "Identification d'un modèle MA perturbé par un bruit additif à support borné", *Proc. 14th GRETSI Symp. on Sig. and Image Proc.*, Juans-les-Pins, 13–16 September 1993, pp. 77–80.

[13] J. Y. Tourneret and B. Lacaze, "Comparison between two nonminimum phase system estimators", *Proc. 7th IEEE Mediterranean Electrotechnical Conf.* (*melecon'94*), Antalya, 12–14 April 1994, pp. 117–120.

[14] J. Tugnait, "Approaches to FIR system identification with noisy data using higher-order statistics", *IEEE Trans. Acoust. Speech Signal Process.*, Vol. ASSP-38, July 1990, pp. 1307–1317.

[15] D. Yellin and B. Porat, "Blind identification of FIR systems excited by discrete-alphabet inputs", *IEEE Trans. Signal Process.*, Vol. SP-41, March 1993, pp.1331–1339.