Systems & Control Letters 11 (1988) 351–356
North-Holland

A method for adaptive estimation of ARMA processes

L. GUO

Department of Systems Engineering, R.S. Phys. S., Australian National University, GPO Box 4, Canberra, ACT 2601, Australia

H.F. CHEN

Institute of Systems Science, Academia Sinica, Beijing, People's Republic of China

Received 23 April 1988
Revised 29 June 1988

Abstract: When the noise process in adaptive identification of linear stochastic systems is correlated, and can be represented by a moving average model, extended least squares algorithms are commonly used, and converge under a strictly positive real (SPR) condition on the noise model. In this paper, we present an adaptive algorithm for the estimation of autoregressive moving average (ARMA) processes, and show that it is convergent without any SPR condition, and has a convergence rate of $O(\{\log\log t)/t\}^{1/2})$.

Keywords: Positive real, Adaptive algorithm, ARMA process, Spectral factorization, State space.

1. Introduction

Consider the following ARMA processes, described by

$$\sum_{j=0}^{p} a_j y(t-j) = \sum_{j=0}^{q} b_j \varepsilon(t-j), \tag{1a}$$

$$a_0 = b_0 = 1, \tag{1b}$$

where the innovation process $\{\varepsilon(t),\ t \geqslant 0\}$ is a stationary ergodic martingale difference sequence with respect to a family $\{\mathscr{F}_t\}$ of nondecreasing σ-algebras, satisfying

$$E\left[\varepsilon^2(t) \mid \mathscr{F}_{t-1}\right] = \sigma^2, \qquad E\,|\varepsilon(t)|^4 < \infty. \tag{2}$$

Set

$$A(z) = 1 + a_1 z^{-1} + \cdots + a_p z^{-p}, \tag{3a}$$

$$B(z) = 1 + b_1 z^{-1} + \cdots + b_q z^{-q}, \tag{3b}$$

and assume $A(z)$ and $B(z)$ are coprime polynomials in z^{-1} and

$$A(z) \neq 0, \quad B(z) \neq 0, \quad |z| \geqslant 1, \qquad a_p \cdot b_q \neq 0. \tag{4}$$

Estimation problems for both unknown parameters and orders of the above stationary ARMA processes have received extensive attention in the area of time series analysis. In this area, many approaches have been proposed, among which two kinds of methods are usually used. One is maximizing the likelihood function (e.g. [1]). This method cannot provide a closed form solution in general, and usually needs the use of iterative nonlinear optimization techniques. The other method involves first the use of increasing lag autoregressions to approximate the corresponding infinite order AR model, and then to do model reduction by invoking the BIC criterion (e.g. [2]) or via Hankel norm approximation or balanced truncation [3]. However, in the first step of this method, the least squares algorithm is used and the *dimension* of the corresponding design matrix *increases unboundedly* with time t (or data size).

Similar estimation problems have also extensively studied in the area of system identification. Commonly used algorithms for adaptive estimation of unknown parameters are *extended* least squares (ELS) and its variants (e.g. [4, 5]), where the orders are assumed to be known. These algorithms were shown to be convergent under a certain kind of strictly positive real (SPR) conditions imposed on the moving average part of the process. Specifically, for the ELS algorithm, it is required that

$$2\mathrm{Re}\{B^{-1}(\mathrm{e}^{\mathrm{i}\theta})\} - 1 > 0, \quad \forall\theta \in [0, 2\pi],$$

which qualitatively means that the system noise is not too 'colored'. This requirement is stronger than the usual minimum phase assumption on $B(z)$. It was also shown in [6] that for ARMA parameter estimation, if SPR condition is removed, counterexamples can be constructed such that the ELS algorithms does not converge. It is worth noting that the recursive prediction error methods (see e.g. [4]) avoid the SPR condition, but need an additional projection into an *a priori* known stability domain, and no convergence rate is guaranteed even in the case of global convergence.

Attempts to avoid SPR condition in adaptive estimation and control have been made in [7], where on-line spectral factorization ideas were used to estimate unknown parameters in moving average (MA) noise processes. Precise theoretical results have recently been established for the estimation of a class of nonstationary linear regression models in [8].

In this paper, we present an adaptive algorithm which requires no SPR condition for its convergence for the estimation of ARMA processes. The algorithm, which assumes the availability of the system order, is based on an adaptive spectral factorization procedure proposed in [8].

2. Main results

Let us describe the estimation algorithm first. Denote

$$m = \max(p, q), \qquad r_j = E[y(j)y(0)], \tag{5}$$

$$r_j(t) = \frac{1}{t}\sum_{i=0}^{t-j} y(i)y(j+i), \quad t \ge 1, \tag{6}$$

$$G_1(t) = \begin{bmatrix} r_1(t) & r_2(t) & \cdots & r_m(t) \\ r_2(t) & r_3(t) & & r_{m+1}(t) \\ \vdots & & \ddots & \vdots \\ r_m(t) & r_{m+1}(t) & \cdots & r_{2m-1}(t) \end{bmatrix}, \tag{7}$$

$$G_2(t) = \begin{bmatrix} r_2(t) & r_3(t) & \cdots & r_{m+1}(t) \\ r_3(t) & r_4(t) & & r_{m+2}(t) \\ \vdots & & \ddots & \vdots \\ r_{m+1}(t) & r_{m+2}(t) & \cdots & r_{2m}(t) \end{bmatrix}, \tag{8}$$

and set

$$H(t) = [1, 0, \ldots, 0]^{\mathrm{T}} \in \mathbb{R}^m, \tag{9}$$

$$M(t) = [r_1(t), r_2(t), \ldots, r_m(t)]^{\mathrm{T}} \in \mathbb{R}^m, \tag{10}$$

$$F(t) = G_1^+(t)G_2(t), \qquad L(t) = r_0(t), \tag{11}$$

where $^+$ denotes the Moore–Penrose pseudo-inverse.

Let $\hat{W}_t(z)$ and $\hat{\Omega}(t)$ be the estimates for the tranfer function $B(z)/A(z)$ and the noise variance σ^2 respectively. *The adaptive algorithm for computing* $\hat{W}_t(z)$ *and* $\hat{\Omega}(t)$ *is given as follows*:

$$\hat{W}_t(z) = I + H^{\mathrm{T}}(t)[zI - F^{\mathrm{T}}(t)]^{-1}\hat{K}(t), \tag{12}$$

$$\hat{\Omega}(t) = H^{\mathrm{T}}(t)\hat{\Sigma}(t)H(t) + L(t), \tag{13}$$

$$\hat{K}(t) = [F^{\mathrm{T}}\hat{\Sigma}(t)H(t) + M(t)]\hat{\Omega}^+(t), \tag{14}$$

$$\hat{\Sigma}(t) = \Sigma_t([\log^2 t]), \quad t \ge 1, \tag{15}$$

where $[\log^2 t]$ denotes the integer part of $\log^2 t$, and

$$\{\Sigma_t(s), 0 \le s \le [\log^2 t], t \ge 1\}$$

are recursively defined by

$$\begin{aligned} \Sigma_t(s+1) = {} & F^{\mathrm{T}}(t)\Sigma_t(s)F(t) \\ & -[F^{\mathrm{T}}(t)\Sigma_t(s)H(t) + M(t)] \\ & \cdot[H^{\mathrm{T}}(t)\Sigma_t(s)H(t) + L(t)]^+ \\ & \cdot[F^{\mathrm{T}}(t)\Sigma_t(s)H(t) + M(t)]^{\mathrm{T}}, \end{aligned} \tag{16a}$$

$$\Sigma_t(0) = 0. \tag{16b}$$

As usual, for a complex-valued function $f(z)$ which is defined and bounded on the unit circle $|z| = 1$, we denote its norm by

$$\| f(z) \|_\infty = \sup_{|z|=1} | f(z) |.$$

The main results of the paper are as follows.

Theorem 1. *Assume that* $p \le q$ *in model* (1). *Then the estimation algorithm described by* (5)–(16) *has the following properties as* $t \to \infty$:

$$\left\| \hat{W}_t(z) - \frac{B(z)}{A(z)} \right\|_\infty = O\left(\left\{\frac{\log\log t}{t}\right\}^{1/2}\right) \quad \text{a.s.},$$

and

$$|\hat{\Omega}(t) - \sigma^2| = O\left(\left\{\frac{\log\log t}{t}\right\}^{1/2}\right) \quad \text{a.s.}$$

3. Auxiliary results

For the proof of Theorem 1 we need some results from linear system theory, autocorrelation approximation and adaptive spectral factorization.

We first state a result on the passage from Markov parameters to state-space equations (see [9]).

Lemma 1. *Suppose the rational* $l \times d$ *matrix* $G(z)$ *has* $G(\infty)=0$ *and is expanded as*

$$G(z)=\frac{A_1}{z}+\frac{A_2}{z^2}+\frac{A_3}{z^3}+\cdots.$$

The Markov parameters A_i *are arranged to form Hankel matrices* H_N *as follows*:

$$H_N=\begin{bmatrix} A_1 & A_2 & \dots & A_N \\ A_2 & A_3 & & A_{N+1} \\ \vdots & & \ddots & \vdots \\ A_N & A_{N+1} & \dots & A_{2N-1} \end{bmatrix}.$$

Since $G(z)$ *is rational, there always exists the first integer* r *such that*

$$\text{rank } H_r = \text{rank } H_{r+1} = \text{rank } H_{r+2} = \cdots.$$

Denote $n = \text{rank } H_r$, *and let* P *and* Q *be nonsingular matrices such that*

$$PH_rQ=\begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}.$$

Then the following matrices 'realize' $G(z)$, *in the sense that*

$G(z)=M^{\mathrm{T}}(zI-F)^{-1}H$:

$H = n \times d$ *top left corner of* PH_r,

$M^{\mathrm{T}} = l \times n$ *top left corner of* H_rQ,

$F = n \times n$ *top left corner of* $P(\sigma H_r)Q$,

where

$$\sigma H_r=\begin{bmatrix} A_2 & A_3 & \dots & A_{r+1} \\ A_3 & A_4 & & A_{r+2} \\ \vdots & & \ddots & \vdots \\ A_{r+1} & A_{r+2} & \dots & A_{2r} \end{bmatrix}.$$

Moreover, $[F, H]$ *is completely reachable and* $[F, M]$ *is completely observable.*

Next, we present a result concerning asymptotic properties of autocorrelation approximation (see [10]).

Lemma 2. *Let* r_j *and* $r_j(t)$ *be defined as in* (5) *and* (6) *respectively. Then as* $t \to \infty$,

$$\max_{0\le j\le P(t)} |r_j(t)-r_j| = \mathrm{O}\left(\left\{\frac{\log\log t}{t}\right\}^{1/2}\right) \quad \text{a.s.,}$$

where $P(t)$ *is any sequence of integers such that* $P(t)=\mathrm{O}(\log^a t)$ *for some* $a>0$.

Finally, we will need the following fact about the convergence of an adaptive spectral factorization algorithm, which is established in [8].

Lemma 3. *Let* $\Phi(z)$ *be a power spectrum function represented by*

$$\Phi(z)=L+H^{\mathrm{T}}(zI-F^{\mathrm{T}})^{-1}M + M^{\mathrm{T}}(z^{-1}I-F)^{-1}H$$

and satisfying the following conditions:

A1. $\Phi(z)$ *is positive on the unit circle.*

A2. *The matrix quadruple* $\{F, H, M, L/2\}$ *is minimal, i.e.,* $[F, H]$ *and* $[F, M]$ *are respectively reachable and observable.*

A3. *All eigenvalues of* F *lie in the unit circle.*

A4. *The quantity* $(\det F)L - M^{\mathrm{T}}[\operatorname{Adj} F]H$ *is non-zero.*

If $L(t)$, $H(t)$, $F(t)$ *and* $M(t)$ (*not necessarily defined by* (9)–(11)) *are consistent estimates of* L, H, F, *and* M *respectively, then* $\hat{W}_t(z)$ *and* $\hat{\Omega}(t)$ *defined by* (12)–(16) *are convergent, with convergence rates*

$$\|\hat{W}_t(z)-W(z)\|_\infty=\mathrm{O}(\Delta(t))+\mathrm{O}\left(\frac{1}{t^\alpha}\right),$$

$$\|\hat{\Omega}(t)-\Omega\|=\mathrm{O}(\Delta(t))+\mathrm{O}\left(\frac{1}{t^\alpha}\right), \quad \forall\alpha>1,$$

where $\{W(z), \Omega\}$ *with* $W(\infty)=1$ *is the unique stable and minimum phase spectral factor of* $\Phi(z)$, *i.e.,*

$$\Phi(z)=W(z)\Omega W^{\mathrm{T}}(z^{-1}),$$

and where $\Delta(t)$ *is the estimation error at time* t:

$$\Delta(t)=\|F(t)-F\|+\|H(t)-H\| + \|M(t)-M\|+\|L(t)-L\|. \tag{17}$$

Remark 1. Actually, the quantity $\hat{\Sigma}(t)$ used in (14) can be defined in many ways, for example, if instead of (15) $\hat{\Sigma}(t)$ is defined as $\Sigma_t([d(t)])$, $t \ge 0$, where $d(t)>0$ is any non-decreasing sequence such that $d(t) \to \infty$, as $t \to \infty$, then the resulting convergence rate in Lemma 3 is now

$$O(\Delta(t)) + O(\exp\{-\delta\, d(t)\}),$$

for some $\delta > 0$. The choice of $d(t) = \log^2 t$ in (15) enables us to obtain the convergence rate $o(\{(\log\log t)/t\}^{1/2})$ in Theorem 1. This rate is the same as that in the well known laws of the iterated logarithm, and is the best possible.

4. Proof of Theorem 1

If we can express the power spectrum $\Phi(z)$ of $\{y(t)\}$ in the form required in Lemma 3 with Conditions A1–A4 satisfied and if $[F, H, M]$ and L in this representation are estimated by

$$[F(t), H(t), M(t)] \quad \text{and} \quad L(t)$$

defined in (9)–(11) with estimation error of order $o(\{(\log\log)/t\}^{1/2})$, then applying Lemma 3 leads to the desired results.

It is well known that the power spectrum of the stationary process $\{y(t)\}$ defined by (1) is

$$\Phi(z) = \sigma^2 \frac{B(z)}{A(z)} \frac{B(z^{-1})}{A(z^{-1})}, \tag{18}$$

and it is related to the covariance in the standard way

$$\Phi(z) = \sum_{i=-\infty}^{\infty} [Ey(i)y(0)] z^{-i}. \tag{19}$$

Now, let A, B and C be matrices such that all eigenvalues of A lie in the unit circle and

$$\frac{B(z)}{A(z)} = C^{\mathrm{T}}(I - z^{-1}A)^{-1}B. \tag{20}$$

This is possible because $A(z)$ is stable. Then let S be the solution of the Lyapunov equation $S = ASA^{\mathrm{T}} + BB^{\mathrm{T}}$. We have

$$\begin{aligned}
&(I - z^{-1}A)^{-1}BB^{\mathrm{T}}(I - zA^{\mathrm{T}})^{-1} \\
&\quad = (zI - A)^{-1}(S - ASA^{\mathrm{T}})(z^{-1}I - A^{\mathrm{T}})^{-1} \\
&\quad = (zI - A)^{-1}\{(zI - A)S(z^{-1}I - A^{\mathrm{T}}) \\
&\qquad\qquad + AS(z^{-1}I - A^{\mathrm{T}}) \\
&\qquad\qquad + (zI - A)SA^{\mathrm{T}}\}(z^{-1}I - A^{\mathrm{T}})^{-1} \\
&\quad = (zI - A)^{-1}AS + SA^{\mathrm{T}}(z^{-1}I - A^{\mathrm{T}})^{-1} + S;
\end{aligned}$$

consequently, by (18)–(20),

$$\begin{aligned}
\Phi(z) = {} & \sigma^2 C^{\mathrm{T}}(zI - A)^{-1}ASC \\
& + \sigma^2 C^{\mathrm{T}}SA^{\mathrm{T}}(z^{-1}I - A^{\mathrm{T}})^{-1}C + \sigma^2 C^{\mathrm{T}}SC,
\end{aligned}$$

which is similar to the expression given in Lemma 3, but for which it is not clear whether Conditions A1–A4 are satisfied. Nevertheless, from this and (19) we find that

$$\sum_{i=1}^{\infty} r_i z^{-i} = \sigma^2 C^{\mathrm{T}}(zI - A)^{-1}ASC.$$

which is obviously rational, analytic in $|z| \geq 1$, and vanishes at $z = \infty$. Hence we can apply Lemma 1. For this we denote

$$G_{1m} = \begin{bmatrix} r_1 & r_2 & \cdots & r_m \\ r_2 & r_3 & & r_{m+1} \\ \vdots & & \ddots & \vdots \\ r_m & r_{m+1} & \cdots & r_{2m-1} \end{bmatrix},$$

$$G_{2m} = \begin{bmatrix} r_2 & r_3 & \cdots & r_{m+1} \\ r_3 & r_4 & \cdots & r_{m+2} \\ \vdots & & \ddots & \vdots \\ r_{m+1} & r_{m+2} & \cdots & r_{2m} \end{bmatrix},$$

where m and r_j are defined in (5).

It is easy to see that

$$\operatorname{rank} G_{1m} = \operatorname{rank} G_{1(m+1)} = \operatorname{rank} G_{1(m+2)} = \cdots$$

since by (1), the correlation function $\{r_j\}$ satisfies the following equations:

$$r_{m+j} + a_1 r_{m+j-1} + \cdots + a_m r_j = 0, \quad \forall j \geq 1,$$

where $a_i \triangleq 0$, for $i > p$.

Furthermore, it is known that (see e.g. [11], Lemma 5.6),

$$\operatorname{rank} G_{lm} = m.$$

Thus, by setting

$$H = [1, 0, \ldots, 0]^{\mathrm{T}} \in \mathbb{R}^m,$$

$$M = [r_1, r_2, \ldots, r_m]^{\mathrm{T}} \in \mathbb{R}^m,$$

$$F = (G_{1m})^{-1}G_{2m}, \qquad L = r_0,$$

and taking $P = (G_{1m})^{-1}$, $Q = I$ in Lemma 1, we have

$$\sum_{j=1}^{\infty} r_j z^{-j} = M^{\mathrm{T}}(zI - F)^{-1}H. \tag{21}$$

From Lemma 2 and (9)–(11) we see that the estimation error defined by (17) has the convergence rate

$$\Delta(t) = \mathrm{O}\left(\left\{\frac{\operatorname{loglog} t}{t}\right\}^{1/2}\right).$$

Hence to complete the proof we need only to check Conditions A1–A4 of Lemma 3. First of all,

$$\Phi(e^{i\theta}) = \sigma^2 | B(e^{i\theta})/A(e^{i\theta}) |^2 \geq 0$$

which, in fact, is strictly positive because of (4). Thus, Condition A1 is satisfied.

Noticing that $\sum_{j=1}^{\infty} r_j z^{-j}$ is analytic in $|z| \geq 1$ and $\{F, H, M\}$ is minimal, from (21) we know that all eigenvalues of F lie in the unit circle. Thus, Condition A3 is verified, while condition A2 is one of the conclusions of Lemma 1.

Therefore, it remains to verify Condition A4 in Lemma 3, i.e.,

$$(\det F)L - M^{\mathrm{T}}[\operatorname{Adj} F]H \neq 0. \tag{22}$$

Let us write

$$\det(zI - F) = z^m + \alpha_1 z^{(m-1)} + \cdots + \alpha_{m-1} z + \alpha_m \tag{23}$$

and

$$M^{\mathrm{T}}[\operatorname{Adj}(zI - F)]H = \beta_1 z^{m-1} + \cdots + \beta_{m-1} z + \beta_m \tag{24}$$

then we know that

$$\det F = (-1)^m \alpha_m, \quad |\alpha_m| + |\beta_m| \neq 0, \tag{25}$$

since $\{F, H, M\}$ is a minimal realization of the transfer function $M^{\mathrm{T}}(zI - F)^{-1}H$.

By (18), (19), (21), (23) and (24) it follows that

$$\sigma^2 \left\{\frac{1 + b_1 z^{-1} + b_q z^{-q}}{1 + a_1 z^{-1} + \cdots + a_p z^{-p}}\right\} \cdot \left\{\frac{1 + b_1 z + \cdots + b_q z^q}{1 + a_1 z + \cdots + a_p z^p}\right\} = \frac{\beta_1 z^{m-1} + \cdots + \beta_m}{z^m + \alpha_1 z^{m-1} + \cdots + \alpha_m} + \frac{\beta_1 z^{-(m-1)} + \cdots + \beta_m}{z^{-m} + \alpha_1 z^{-(m-1)} + \cdots + \alpha_m} + r(0). \tag{26}$$

We now prove (22) by considering the following two cases.

(i) $p = q$. Letting $z \to \infty$ (26), we see that

$$\frac{\beta_m}{\alpha_m} + r(0) = \sigma^2 \frac{b_q}{a_p}.$$

Thus, $\alpha_m \neq 0$ and

$$\begin{aligned}(\det F)L - M^{\mathrm{T}}[\operatorname{Adj} F]H &= (-1)^m \alpha_m [r(0) - M^{\mathrm{T}} F^{-1} H] \\ &= (-1)^m \alpha_m \left[r(0) + \frac{\beta_m}{\alpha_m}\right] \\ &= (-1)^m \alpha_m \sigma^2 \frac{b_q}{a_p} \neq 0.\end{aligned}$$

(ii) $p < q$. By letting $z \to \infty$ in (26) we see that

$$\frac{\beta_m}{\alpha_m} + r(0) = \infty.$$

Therefore $\alpha_m = 0$, and we need only verify $M^{\mathrm{T}}[\operatorname{Adj} F]H \neq 0$. Note that by (24),

$$M^{\mathrm{T}}[\operatorname{Adj} F]H = (-1)^{m+1} \beta_m$$

but by (25), $\beta_m \neq 0$. Hence the proof of Theorem 1 is complete.

5. Conclusion

We have presented an adaptive algorithm which does not require any SPR condition for its convergence for the parameter estimation of a class of ARMA processes. Currently under investigation are the more general problems of adaptive estimation of both unknown parameters and orders of nonstationary ARMAX processes when the noise model satisfies no SPR condition.

Acknowledgement

The authors would like to thank Dr. D.W. Huang for his valuable discussions.

References

[1] G.E.P. Box and G.M. Jenkins, *Time Series Analysis, Forecasting and Control* (Holden-Day, San Francisco, CA, 1976).

[2] E.J. Hannan and L. Kavalieris, A method for autoregressive moving-average estimation, *Biometrica* **71** (1984) 281.

[3] E.J. Hannan, Rational transfer function approximation, *Statist. Sci.* **2** (1987) 135–161.

[4] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification* (M.I.T. Press, Cambridge, MA, 1983).

[5] H.F. Chen, *Recursive Estimation and Control for Stochastic Systems* (John Wiley, New York, 1985).

[6] L. Ljung, T. Söderström and I. Gustavsson, Counterexamples to general convergence of a commonly used recursive identification method, *IEEE Trans. Automat. Control* **20** (1975) 643.

[7] J.B. Moore, Side-stepping the positive real restriction for stochastic adaptive schemes, *Ricerche Automat.* **8** (1982) 501–523.

[8] L. Guo and J.B. Moore, Stochastic system identification via adaptive spectral factorization, Technical Report, Dept. Systems Engineering, Australian National University (1987).

[9] P.L. Ho and R.E. Kalman, Effective construction of linear state-variable model from input/output functions, *Regelungstechnik* **14** (1966) 545–548.

[10] H.Z. An, Z.G. Chen and E.J. Hannan, Autocorrelation, autoregression and autoregressive approximation, *Ann. Statist.* **10** (1982) 926–936.

[11] R.S. Tsay and G.C. Tiao, Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models, *J. Amer. Statist. Assoc.* **79** (1984) 84.