Recursive identification for multidimensional ARMA processes with increasing variances

CHEN Hanfu

Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, China (email: hfchen@iss.ac.cn)

Received September 22, 2004; revised March 23, 2005

Abstract In time series analysis, almost all existing results are derived for the case where the driven noise $\{w_n\}$ in the MA part is with bounded variance (or conditional variance). In contrast to this, the paper discusses how to identify coefficients in a multidimensional ARMA process with fixed orders, but in its MA part the conditional moment $E(\|w_n\|^{\beta} \mid \mathcal{F}_{n-1})$, $\beta > 2$ is possible to grow up at a rate of a power of $\log n$. The well-known stochastic gradient (SG) algorithm is applied to estimating the matrix coefficients of the ARMA process, and the reasonable conditions are given to guarantee the estimate to be strongly consistent.

Keywords: multidimensional ARMA, increasing variance, recursive estimation, martingale difference sequence.

DOI: 10.1360/04yf0324

1 Introduction

Data coming from various systems such as social-economical, bio-medical, engineering, and ecological, are often modelled by an m-dimensional ARMA process with $m \geqslant 1$:

$$A(z)y_n = C(z)w_n, \qquad y_i = 0, \quad i < 0, \tag{1}$$

where

$$A(z) = I + A_1 z + \cdots + A_p z^p \quad \text{and} \quad C(z) = I + C_1 z + \cdots + C_r z^r \tag{2}$$

are matrix polynomials in backward-shift operator $z : zy_n = y_{n-1}$ with unknown coefficients

$$\theta^T \triangleq [-A_1, \cdots, -A_p, C_1, \cdots, C_r]. \tag{3}$$

The orders p and r of polynomials $A(z)$ and $C(z)$ are assumed to be available or set to be the upper bounds for the true ones. The problem discussed here is to recursively estimate θ on the basis of data $\{y_k,\ k = 0, 1, \cdots\}$.

By the classical approach in time series analysis, the second empirical moments for $\{y_k\}$ are first calculated, and the Yule-Walker equations are derived at the same time. The Yule-Walker equations in fact form a system of nonlinear algebraic equations with respect to θ, and the solution to the system leads to an estimate for θ. By this approach,

Copyright by Science in China Press 2005

it is necessary to assume that $\{y_k\}$ and $\{w_k\}$ both are stationary and ergodic with some additional assumptions on $\{w_k\}$ such as being iid or a martingale difference sequence (mds). There exists a vast of literature on this approach, e.g. refs. [1–5] among others. As a matter of fact, these conditions implicitly imply that $A(z)$ is stable, by which it is meant that $\det A(z) \neq 0,\ \forall z : |z| \leqslant 1$. To be precise, in ref. [6] it is shown that for a wide class of $\{w_n\}$, $\limsup\limits_{n\to\infty} \frac{1}{n} \sum\limits_{k=1}^{n} \|y_k\|^2 < \infty$ a.s. is equivalent to stability of $A(z)$, whatever $C(z)$ is. When $\det A(z)$ has no root in $|z| < 1$ but possibly has roots on the unit circle, the consistent estimate for θ is also possible to be obtained as shown in ref. [7].

It is clear that the estimation accuracy for all nonrecursive estimation methods depends on the sample size. For example, the estimate obtained by the Yule-Walker equation approach depends on the second empirical moments whose accuracy in turn depends on the sample size. Besides, the numerical solution to the nonlinear algebraic equations may give rise to additional errors.

The other approach to estimate θ is the recursive way by which the estimate for θ is updated after receiving a new y_n. For the one-dimensional AR process, i.e. $C(z) \equiv 1$, the least squares (LS) estimate gives consistent estimate for θ if $\{w_k\}$ is an mds satisfying the following conditions:

$$\liminf_{n\to\infty} E(w_{n+1}^2 \mid \mathcal{F}_n) > 0, \quad \text{and} \quad \sup_n E(|w_{n+1}|^\beta \mid \mathcal{F}_n) < \infty \quad \text{a.s.} \tag{4}$$

for some $\beta > 2$[8].

For multidimentional ARMAX systems, the consistent coefficient estimates can also be derived in a recursive way by using a diminishing excitation technique[9]. For this, among other conditions it should be assumed that

$$\liminf_{n\to\infty} \frac{1}{n}\lambda_{\min}\left(\sum_{k=1}^{n} w_k w_k^T\right) > 0, \quad \text{and} \quad \limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \|w_k\|^2 < \infty \quad \text{a.s.,} \tag{5}$$

where and hereafter $\lambda_{\min}(A)$ denotes the minimum eigenvalue of a matrix A. It is worth noting that for mds the condition $\sup\limits_n E(\|w_n\|^\beta \mid \mathcal{F}_{n-1}) < \infty$ with $\beta > 2$ implies that $\limsup\limits_{n\to\infty} \frac{1}{n} \sum\limits_{k=1}^{n} \|w_k\|^2 < \infty$.

In both recursive and nonrecursive approaches mentioned above, the variance in the time average sense $\frac{1}{n} \sum\limits_{k=1}^{n} \|w_k\|^2$ of $\{w_k\}$ is not allowed to grow up unboundedly.

When modelling a sequence of data received in real time, if it is observed that $\frac{1}{n} \sum\limits_{k=1}^{n} \|y_k\|^2$ grows up as fast as $(\log n)^s$, $s > 0$, then the Yule-Walker equation approach fails to work. One may still want to apply a recursive estimation method to fit the data into an ARMA process with the standard assumption on $\{w_k\}$ that $\{w_k\}$ is an mds satisfying (4). However, such a data sequence does not fit any ARMA process. This can be explained as follows. If $A(z)$ is stable, then as $n \to \infty$, $\frac{1}{n} \sum\limits_{1}^{n} \|y_k\|^2$ is bounded as shown in ref. [6], and hence a stable ARMA process does not fit the data. If $\det A(z)$ is unstable with some root inside the open unit disk, then $\frac{1}{n}\sum\limits_{1}^{n}\|y_k\|^2$ will exponentially diverge to infinity. If $\det A(z)$ has no explosive root but may have roots on the unit circle, then $\frac{1}{n}\sum\limits_{1}^{n}\|y_k\|^2$ may still grow up as fast as a polynomial as $n\to\infty$. Therefore, any unstable $A(z)$ describes a too fast growth rate of $\frac{1}{n}\sum\limits_{1}^{n}\|y_k\|^2$ compared with that for the data. Thus, we conclude that when modelling data with slowly increasing $\frac{1}{n}\sum\limits_{k=1}^{n}\|y_k\|^2$ by an ARMA process, the standard condition (4) is inappropriate.

In this paper, we discuss modelling by using ARMA processes with increasing $\frac{1}{n}\sum\limits_{k=1}^{n}\|w_k\|^2$ as $n\to\infty$. The rest of the paper is organized as follows. In section 2, some properties of ARMA processes with increasing variances are pointed out. The estimation algorithm and the behavior of the estimation error for w_k are presented in section 3. The convergence of the estimate is characterized in terms of transition matrices in section 4. The main result on strong consistency of θ_n is given in section 5. Some concluding remarks are given in the last section.

In ref. [9], a detailed analysis is given for the algorithm to be defined in section 3 for the bounded variance case. Techniques developed there are used here with modifications to cope with unbounded variances of $\{w_k\}$.

2 ARMA processes with increasing variances

As discussed above when $\frac{1}{n}\sum\limits_{k=1}^{n}\|y_k\|^2$ slowly grows up as $n\to\infty$, it is reasonable to model the data as an ARMA process with increasing variance, by which we mean $\frac{1}{n}\sum\limits_{k=1}^{n}\|w_k\|^2\to\infty$ as $n\to\infty$. Let us first characterize such kind of ARMA processes. For this we introduce the following conditions.

A1. $\{w_n,\mathcal{F}_n\}$ is an mds with

$$\liminf_{n}\frac{1}{n}\lambda_{\min}\left(\sum_{i=1}^{n}w_iw_i^T\right)\triangleq\lambda>0\quad\text{a.s.}\tag{6}$$

and

$$\sup_{n}(\log n)^{-\frac{\beta(1-\delta)}{2}}E\left(\|w_n\|^\beta\mid\mathcal{F}_{n-1}\right)\triangleq c_1<\infty\quad\text{a.s.}\tag{7}$$

for some $\beta>2$ and $\delta\in\left(\frac{4}{5},1\right]$.

A2. $A(z)$ is stable.

Remark 1. It is worth noting at once that the condition (7) means that $E(\|w_n\|^\beta\mid\mathcal{F}_{n-1})$ is allowed to diverge to infinity at a rate of $(\log n)^{\frac{\beta(1-\delta)}{2}}$. By the

Copyright by Science in China Press 2005

Lyapunov inequality it follows that

$$\sup_n \frac{1}{(\log n)^{1-\delta}} E(\|w_n\|^2 \mid \mathcal{F}_{n-1}) \leqslant \sup_n \frac{1}{(\log n)^{1-\delta}} \left[E(\|w_n\|^\beta \mid \mathcal{F}_{n-1})\right]^{\frac{2}{\beta}} \leqslant c_1^{\frac{2}{\beta}}.$$

Therefore, condition (7) implies that

$$E\big(\|w_n\|^2 \mid \mathcal{F}_{n-1}\big) \leqslant c_1^{\frac{2}{\beta}} (\log n)^{1-\delta}, \quad n > 1, \tag{8}$$

which means that the second conditional moments of $\{w_n\}$ are also allowed to diverge to infinity.

In what follows, by c_i, $i = 1, 2, \cdots$, we always denote positive values that are constants for any fixed sample (ω).

The following lemma shows that with A1 and A2 satisfied how fast $\frac{1}{n}\sum_{k=1}^{n} \|y_k\|^2$ and $\frac{1}{n}\sum_{k=1}^{n} \|w_k\|^2$ may diverge for an ARMA process with increasing variances.

Lemma 1. Assume A1 and A2 hold. Then

$$\sum_{k=1}^{n} \|w_k\|^2 = O(n(\log n)^{1-\delta}) \quad \text{a.s.} \tag{9}$$

and

$$\sum_{k=1}^{n} \|y_k\|^2 = O(n(\log n)^{1-\delta}) \quad \text{a.s.} \tag{10}$$

Proof. Since $\beta > 2$, by the C_r- and Lyapunov inequalities it follows that

$$E\left[\big(\|w_k\|^2 - E(\|w_k\|^2 \mid \mathcal{F}_{k-1})\big)^{\frac{\beta}{2}} \;\Big|\; \mathcal{F}_{k-1}\right] \leqslant 2^{\frac{\beta}{2}} E\big(\|w_k\|^\beta \mid \mathcal{F}_{k-1}\big). \tag{11}$$

Then by (7) and (11) we have

$$\sum_{k=2}^{\infty} E\left\{\left[\frac{\|w_k\|^2 - E(\|w_k\|^2 \mid \mathcal{F}_{k-1})}{k(\log k)^{1-\delta}}\right]^{\frac{\beta}{2}} \;\middle|\; \mathcal{F}_{k-1}\right\} \leqslant 2^{\frac{\beta}{2}} \sum_{k=2}^{\infty} \frac{c_1}{k^{\frac{\beta}{2}}} < \infty \quad \text{a.s.} \tag{12}$$

Since $\left(\frac{E(\|w_n\|^\beta \mid \mathcal{F}_{n-1})}{n^{\frac{\beta(1-\delta)}{2}}}\right)^{\frac{1}{\beta}}$ is nondecreasing with respect to β, without loss of generality we may assume $\frac{\beta}{2} \leqslant 2$ in ref. [7], and by the convergence theorem for mds[10], see also ref. [9]

$$\sum_{k=2}^{\infty} \frac{\|w_k\|^2 - E(\|w_k\|^2 \mid \mathcal{F}_{k-1})}{k(\log k)^{1-\delta}} < \infty \quad \text{a.s.}$$

From this by the Kronecker lemma it follows that

$$\frac{1}{n}\sum_{k=2}^{n} \frac{\|w_k\|^2 - E\big[\|w_k\|^2 \mid \mathcal{F}_{k-1}\big]}{(\log k)^{1-\delta}} \xrightarrow[n\to\infty]{} 0 \quad \text{a.s.},$$

which incorporating with (8) implies that

$$\sum_{k=2}^{n} \frac{\|w_k\|^2}{(\log k)^{1-\delta}} = O(n).$$

From here (9) follows, and by A2 (9) implies (10). Q.E.D.

Let

$$\varphi_n^0 \triangleq [y_n^T, \cdots, y_{n-p+1}^T, w_n^T, \cdots, w_{n-r+1}^T]^T,$$

which is called the stochastic regressor, since by (1) and (2) y_{n+1} can be written in a regression form:

$$y_{n+1} = \theta^T \varphi_n^0 + w_{n+1}. \tag{13}$$

Let

$$S_n^0 \triangleq \sum_{k=1}^{n} \varphi_k^0 \varphi_k^{0T} + \frac{1}{p+r} I. \tag{14}$$

The behavior of S_n^0 will play an important role in the coefficient estimation. However, because of involving $w_k, \cdots, w_{n-r+1}$, the regressor φ_k^0 and hence the matrix S_n^0 is unavailable. We need the following condition.

A3. $A(z)$ and $C(z)$ have no common left factor, and $[A_p \vdots C_r]$ is of row-full-rank.

Lemma 2. Assume A1 and A3 hold. Then

$$\liminf_{n\to\infty} \frac{1}{n} \lambda_{\min}(S_n^0) \triangleq c_2 > 0 \quad \text{a.s.} \tag{15}$$

Proof. The analysis method is similar to that for Theorem 6.3 in ref. [9], but here some modifications should be made to cope with the situation of increasing variance. We present the complete proof for readability.

Let

$$f_n \triangleq (\det A(z))\varphi_n^0, \quad \det A(z) = a_0 + a_1 z + \cdots + a_s z^s, \quad s \leqslant mp.$$

It is clear that

$$\lambda_{\min}\left(\sum_{k=1}^{n} f_k f_k^T\right) = \inf_{\|x\|=1} \sum_{k=1}^{n} (x^T f_k)^2 \leqslant (s+1) \sum_{k=0}^{s} a_k^2 \lambda_{\min}(S_n^0),$$

and for (15) it suffices to show

$$\liminf_{n\to\infty} \frac{1}{n} \lambda_{\min}\left(\sum_{k=1}^{n} f_k f_k^T\right) \neq 0. \tag{16}$$

Assume the converse: There exists a sequence $\{\eta_k\}$ of unit vectors $\|\eta_k\| = 1$ such that

$$\lim_{k\to\infty} n_k^{-1} \left(\sum_{i=1}^{n_k} \left(\eta_{n_k}^T f_i\right)^2\right) = 0. \tag{17}$$

Let

$$\eta_{n_k} \triangleq \left[\alpha_{n_k}^{(0)T}, \cdots, \alpha_{n_k}^{(p-1)T}, \beta_{n_k}^{(0)T}, \cdots, \beta_{n_k}^{(r-1)T}\right]^T,$$

where $\alpha_{n_k}^{(j)}$ and $\beta_{n_k}^{(j)}$ both are m-dimensional, and let

$$H_{n_k}(z) \triangleq \sum_{i=0}^{p-1} \alpha_{n_k}^{(i)T} z^i Adj A(z) C(z) + \sum_{i=0}^{r-1} \beta_{n_k}^{(i)T} z^i \det A(z) I_m \triangleq \sum_{i=0}^{\mu} h_{n_k}^{(i)T} z^i.$$

Copyright by Science in China Press 2005

From (1) we have

$$\det A(z)y_n = Adj A(z)C(z)w_n,$$

and hence

$$\begin{aligned} f_n &= \det A(z)\varphi_n^0 \\ &= \left[\det A(z)y_n^T, \cdots, \det A(z)y_{n-p+1}^T, \det A(z)w_n^T, \cdots, \det A(z)w_{n-r+1}^T\right]^T \\ &= \big[(Adj A(z)C(z)w_n)^T, \cdots, (Adj A(z)C(z)w_{n-p+1})^T, (\det A(z)w_n)^T, \cdots, \\ &\quad (\det A(z)w_{n-r+1})^T\big]^T. \end{aligned}$$

Thus,

$$\begin{aligned} \eta_{n_k}^T f_i &= \Big\{\alpha_{n_k}^{(0)} Adj A(z)C(z) + \cdots + \alpha_{n_k}^{(p-1)T} z^{p-1} Adj A(z)C(z) \\ &\quad + \beta_{n_k}^{(0)T} \det A(z)I + \cdots + \beta_{n_k}^{(r-1)T} z^{r-1} \det A(z)I\Big\} w_i \\ &= H_{n_k}(z)w_i = \sum_{j=0}^{\mu} h_{n_k}^{(j)T} w_{i-j}. \end{aligned}$$

By the converse assumption

$$\lim_{k\to\infty} n_k^{-1} \sum_{i=1}^{n_k} (\eta_{n_k}^T f_i)^2 = \lim_{k\to\infty} n_k^{-1} \sum_{i=1}^{n_k} \left(h_{n_k}^{(0)T} w_i + \cdots + h_{n_k}^{(\mu)T} w_{i-\mu}\right)^2 = 0. \qquad (18)$$

Paying attention to (8) we find that Theorem 2.8 in ref. [9] can be applied to the mds $\left\{\frac{w_n}{(\log n)^{\frac{1-\delta}{2}}}\right\}$, and hence for $j > s$ we have

$$\begin{aligned} &\left\| h_{n_k}^{(j)T} \sum_{i>j\vee s}^{n_k} w_{i-j} w_{i-s}^T h_{n_k}^{(s)} \right\| \\ &\leqslant \|h_{n_k}^{(j)}\| \|h_{n_k}^{(s)}\| \left\| \sum_{i>j\vee s}^{n_k} \big(\log(i-j)\big)^{\frac{1-\delta}{2}} \big(\log(i-s)\big)^{\frac{1-\delta}{2}} \frac{w_{i-j}}{\big(\log(i-j)\big)^{\frac{1-\delta}{2}}} \right. \\ &\quad \left. \cdot \frac{w_{i-s}}{\big(\log(i-s)\big)^{\frac{1-\delta}{2}}} \right\| \\ &\leqslant \|h_{n_k}^{(j)}\| \|h_{n_k}^{(s)}\| O\left(\left(\sum_{i=1}^{n_k} (\log i)^{1-\delta} \|w_i\|^2\right)^{\gamma}\right) \\ &\leqslant \|h_{n_k}^{(j)}\| \|h_{n_k}^{(s)}\| O\left((\log n_k)^{2\gamma(1-\delta)} n_k^{\gamma}\right), \end{aligned} \qquad (19)$$

where $\gamma \in (\frac{1}{2}, 1)$, and for the last inequality the estimate (9) of Lemma 1 is invoked.

Thus, by (6) and (19) we have

$$\sum_{i=1}^{n_k} (h_{n_k}^{(0)T} w_i + \cdots + h_{n_k}^{(\mu)T} w_{i-\mu})^2$$

$$\geqslant \sum_{i=1}^{n_k}\left[(h_{n_k}^{(0)T}w_i)^2+\cdots+(h_{n_k}^{(\mu)T}w_{i-\mu})^2\right]-\sum_{j=0}^{\mu}\sum_{\substack{s=0\\ s\neq j}}^{\mu}\|h_{n_k}^{(j)T}\sum_{i=1}^{n_k}w_{i-j}w_{i-s}^Th_{n_k}^{(s)}\|$$

$$\geqslant \lambda n_k\sum_{j=0}^{\mu}\|h_{n_k}^{(j)}\|^2-\sum_{j=0}^{\mu}\|h_{n_k}^{(j)}\|^2 O\left(n_k^{\gamma}(\log n_k)^{2\gamma(1-\delta)}\right)$$

$$\geqslant \lambda n_k(1+o(n_k))\sum_{j=0}^{\mu}\|h_{n_k}^{(j)}\|^2. \tag{20}$$

Combining (18) and (20) leads to

$$\|h_{n_k}^{(j)}\|\xrightarrow[k\to\infty]{}0,\quad j=0,1,\cdots,\mu,$$

and hence

$$H_{n_k}(z)\xrightarrow[k\to\infty]{}0.$$

Since $\{\eta_{n_k}\}$ is bounded, there is a convergent subsequence tending to a unit vector

$$\left[\alpha^{(0)T},\cdots,\alpha^{(p-1)T},\beta^{(0)T},\cdots,\beta^{(r-1)T}\right]^T.$$

Since $H_{n_k}(z)\xrightarrow[k\to\infty]{}0$, we have

$$\sum_{i=0}^{p-1}\alpha^{(i)T}z^i Adj A(z)C(z)+\sum_{i=0}^{r-1}\beta^{(i)T}z^i\det A(z)I=0. \tag{21}$$

Since $A(z)$ and $C(z)$ have no common left factor, there are $M(z)$ and $N(z)$ of compatible dimensions such that

$$A(z)M(z)+C(z)N(z)=I. \tag{22}$$

From (21) and (22) it follows that

$$\begin{aligned}\sum_{i=0}^{p-1}\alpha^{(i)T}z^i Adj A(z)&=\sum_{i=0}^{p-1}\alpha^{(i)T}z^i Adj A(z)(A(z)M(z)+C(z)N(z))\\&=\det A(z)\left(\sum_{i=0}^{p-1}\alpha^{(i)T}z^iM(z)-\sum_{i=0}^{r-1}\beta^{(i)T}z^iN(z)\right)\\&\triangleq\det A(z)\sum_{i=0}^{\nu}\mu^{(i)T}z^i.\end{aligned}$$

Multiplying the above expression respectively by $A(z)$ and $C(z)$ from the left, we derive

$$\sum_{i=0}^{p-1}\alpha^{(i)T}z^i=\sum_{i=0}^{\nu}\mu^{(i)T}z^iA(z) \tag{23}$$

and

$$-\sum_{i=0}^{r-1}\beta^{(i)T}z^i=\sum_{i=0}^{\nu}\mu^{(i)T}z^iC(z). \tag{24}$$

Since $[A_p \vdots C_r]$ is of row-full-rank, $\mu^{(i)T}[A_p \vdots C_r]\neq 0$ if $\mu^{(i)}\neq 0$. Assume $\mu^{(i)T}A_p\neq 0$. Then the right-hand side of (23) is with order greater than or equal to p, while the

Copyright by Science in China Press 2005

left-hand side of (23) is a polynomial of order less than or equal to $p-1$. This is impossible. Therefore, $\mu^{(i)T}A_p = 0$, and we must have $\mu^{(i)T}C_r \neq 0$. From (24) by the same argument we arrive at a contradiction. This means that $\mu^{(i)} = 0$, $i = 0, 1, \ldots, \nu$, which in turn imply that

$$\alpha^{(i)} = 0, \quad i = 0, \ldots, p-1, \quad \beta^{(j)} = 0, \quad j = 0, \ldots, r-1.$$

This contradicts with that $[\alpha^{(0)T}, \ldots, \alpha^{((p-1)T}, \beta^{(0)T}, \ldots, \beta^{(r-1)T}]^T$ is a unit vector. The contradiction proves (15).

3 Estimation algorithm

As mentioned in the last section, the stochastic regressor ϕ_n^0 in the regression model (13) is not available, and hence it cannot be used in the estimation algorithm.

Denote by $\hat{w}_k$ an estimate for w_k. Then ϕ_n^0 is estimated by

$$\phi_n \triangleq [y_n^T, y_{n-1}^T, \ldots, y_{n-p+1}^T, \hat{w}_n^T, \ldots, \hat{w}_{n-r+1}^T]^T. \tag{25}$$

We estimate θ by the following recursive algorithm:

$$\theta_{k+1} = \theta_k + \frac{\alpha\phi_k}{r_k}(y_{k+1} - \theta_k^T\phi_k)^T, \quad r_k = 1 + \sum_{i=1}^{k} \|\phi_i\|^2, \quad \alpha > 0. \tag{26}$$

From (13) it is seen that the estimate $\hat{w}_k$ for w_k is naturally to be defined as

$$\hat{w}_k = y_k - \theta_{k-1}^T\phi_{k-1}. \tag{27}$$

With an arbitrary θ_0 and initial values $\phi_i = 0$, $i < 0$, (25), (26) and (27) form a recursive algorithm estimating θ called as the stochastic gradient (SG) algorithm.

Noticing that $\phi_k(y_{k+1} - \theta^T\phi_k)^T$ is the gradient of $\|y_{k+1} - \theta^T\phi_k\|^2$, we find that the SG algorithm, roughly speaking, is a stochastic approximation algorithm[11] for minimizing

$$\min_\theta E\|y_{k+1} - \theta^T\phi_k\|^2,$$

if $\sum\limits_{k=1}^{\infty} \frac{1}{r_k} = \infty$.

Let ζ_k and ϕ_k^ζ denote the estimation errors for w_{k+1} and ϕ_k, respectively, i.e.

$$\zeta_k \triangleq \hat{w}_{k+1} - w_{k+1} = y_{k+1} - \theta_k^T\phi_k - w_{k+1}, \tag{28}$$

$$\phi_k^\zeta = \phi_k - \phi_k^0. \tag{29}$$

For the convergence analysis of the algorithm (25)–(27), the behavior of ζ_k is of crucial importance, because the estimation error for the regressor ϕ_k^0 is totally determined by $\{\zeta_k\}$.

The property formulated in the following lemma was established in ref. [12] for the case where $\{w_k\}$ is of bounded variance. We now show that this property remains true for $\{w_k\}$ being of increasing variance.

We need the following condition.

A4. $C(z)$ is strictly positive real (SPR), i.e. $C(e^{i\lambda}) + C^T(e^{-i\lambda}) > 0$, $\forall\lambda \in [0, 2\pi]$.

Remark 2. From A4 it is clear that there is a small number $\alpha > 0$ such that $C(e^{i\lambda}) + C^T(e^{-i\lambda}) > \alpha I$, $\forall\lambda \in [0, 2\pi]$. Thus, A4 implies that there exists $\alpha > 0$ such that $C(z) - \frac{\alpha}{2}I$ is SPR. From now on, it is assumed that α in (26) is selected in such a way.

Lemma 3. Assume A1, A3, and A4 hold. Then for $\{\zeta_k\}$ given by (28)

$$\sum_{k=0}^{\infty} \frac{\|\zeta_k\|^2}{r_k} < \infty \quad \text{a.s.} \tag{30}$$

Proof. Let the estimation error produced by (25)–(27) be denoted by

$$\tilde{\theta}_k = \theta - \theta_k.$$

From (26), (28) it follows that

$$\tilde{\theta}_{k+1} = \tilde{\theta}_k - \frac{\alpha\phi_k}{r_k}(\zeta_k^T + w_{k+1}^T). \tag{31}$$

Notice that

$$\begin{aligned} C(z)\zeta_k &= C(z)(y_{k+1} - \theta_k^T\phi_k - w_{k+1}) \\ &= y_{k+1} - C(z)w_{k+1} + (C(z) - I)(y_{k+1} - \theta_k^T\phi_k) - \theta_k^T\phi_k \\ &= \theta^T\phi_k - \theta_k^T\phi_k = \tilde{\theta}_k^T\phi_k. \end{aligned} \tag{32}$$

By Remark 2, α in (26) is selected such that $C(z) - \frac{\alpha}{2}I$ is SPR, by which there is a constant $\epsilon_1 > 0$ such that

$$t_n \triangleq 2\alpha\sum_{i=1}^{n}\zeta_i^T\left(\tilde{\theta}_i\phi_i - \frac{\alpha(1+\epsilon_1)}{2}\zeta_i\right) \geqslant 0, \quad t_0 = 0. \tag{33}$$

By (31) it follows that

$$\begin{aligned} \mathrm{tr}\tilde{\theta}_{k+1}^T\tilde{\theta}_{k+1} + \frac{t_{k+1}}{r_k} &= \mathrm{tr}\tilde{\theta}_k^T\tilde{\theta}_k - 2\alpha(\zeta_k^T + w_{k+1}^T)\frac{\tilde{\theta}_k^T\phi_k}{r_k} \\ &\quad + \frac{\alpha^2\|\phi_k\|^2}{r_k^2}\|\zeta_k + w_{k+1}\|^2 + \frac{t_{k+1}}{r_k} \\ &= \mathrm{tr}\tilde{\theta}_k^T\tilde{\theta}_k - \frac{2\alpha[\zeta_k^T(\tilde{\theta}_k\phi_k - \frac{\alpha}{2}(1+\epsilon_1)\zeta_k]}{r_k} \\ &\quad - \frac{\alpha^2(1+\epsilon_1)\|\zeta_k\|^2}{r_k} - \frac{2\alpha w_{k+1}^T\tilde{\theta}_k^T\phi_k}{r_k} + \frac{\alpha^2\|\phi_k\|^2\|\zeta_k\|^2}{r_k^2} \\ &\quad + \frac{\alpha^2\|\phi_k\|^2}{r_k^2}\|w_{k+1}\|^2 + \frac{2\alpha^2\|\phi_k\|^2}{r_k^2}\zeta_k^T w_{k+1} + \frac{t_{k+1}}{r_k} \\ &\leqslant \mathrm{tr}\tilde{\theta}_k^T\tilde{\theta}_k + \frac{t_k}{r_{k-1}} - \frac{\alpha^2\epsilon_1\|\zeta_k\|^2}{r_k} - \frac{2\alpha w_{k+1}^T\tilde{\theta}_k^T\phi_k}{r_k} \\ &\quad + \frac{\alpha^2\|\phi_k\|^2}{r_k^2}\|w_{k+1}\|^2 + \frac{2\alpha^2\|\phi_k\|^2}{r_k^2}\zeta_k^T w_{k+1}. \end{aligned}$$

Copyright by Science in China Press 2005

Summing up both sides of the above inequality from 0 to $n+1$ and assuming $r_{-1}=1$, we arrive at

$$\begin{aligned}&\operatorname{tr}\tilde{\theta}_{n+1}^T\tilde{\theta}_{n+1}+\frac{t_{n+1}}{r_n}+\frac{\alpha^2\epsilon_1}{2}\sum_{i=0}^{n}\frac{\|\zeta_i\|^2}{r_i}\\&\leqslant \operatorname{tr}\tilde{\theta}_0^T\tilde{\theta}_0+t_0-2\alpha\sum_{i=0}^{n}\frac{w_{i+1}^T\tilde{\theta}_i^T\phi_i}{r_i}+\alpha^2\sum_{i=0}^{n}\frac{\|\phi_i\|^2}{r_i^2}\|w_{i+1}\|^2\\&+2\alpha^2\sum_{i=0}^{n}\frac{\|\phi_i\|^2}{r_i^2}\zeta_i^Tw_{i+1}-\frac{\alpha^2\epsilon_1}{2}\sum_{i=1}^{n}\frac{\|\zeta_i\|^2}{r_i}.\end{aligned}\tag{34}$$

By Lemma 2 it follows that there is $c_3>0$ a.s. such that

$$\sum_{i=1}^{n}\|y_i\|^2\geqslant c_3n\qquad\forall n.$$

Then we have

$$r_n\geqslant 1+\sum_{1}^{n}\|y_i\|^2\geqslant c_3n.\tag{35}$$

We now estimate the sums at the right-hand side of (34).

By (7) and Theorem 2.8 in ref. [9] we have

$$\begin{aligned}&\sum_{i=1}^{n}\frac{w_{i+1}^T}{\log(i+1))^{\frac{1-\delta}{2}}}\cdot\frac{\tilde{\theta}_i^T\phi_i(\log(i+1))^{\frac{1-\delta}{2}}}{r_i}\\&=O\left(\left(\sum_{i=1}^{n}\frac{\|\tilde{\theta}_i^T\phi_i\|^2\log(i+1)}{r_i^2}\right)^{\gamma}\right)\\&=O\left(\left(\sum_{i=1}^{n}\frac{\|\zeta_i\|^2}{r_i}\right)^{\gamma}\right)\quad\text{for }\gamma\in\left(\frac{1}{2},1\right),\end{aligned}\tag{36}$$

where (32), (35) are invoked for the last equality.

By (7) and the convergence theorem for the following martingale difference sequence

$$\frac{\|w_k\|^2}{(\log k)^{1-\delta}}-\frac{E(\|w_k\|^2|\mathcal{F}_{k-1})}{(\log k)^{1-\delta}},$$

we see that

$$\sum_{i=1}^{\infty}\frac{\|\phi_i\|^2}{r_i^2}(\|w_{i+1}\|^2-E(\|w_{i+1}\|^2|\mathcal{F}_i))\tag{37}$$

is convergent on the set where

$$\sum_{i=2}^{\infty}\frac{\|\phi_i\|^2}{r_i^2}(\log i)^{1-\delta}<\infty.\tag{38}$$

Notice that for any $\mu\in(0,1)$

$$\sum_{i=2}^{\infty}\frac{\|\phi_i\|^2}{r_i^{1+\mu}}=\sum_{i=2}^{\infty}\int_{r_{i-1}}^{r_i}\frac{dx}{r_i^{1+\mu}}\leqslant\sum_{i=2}^{\infty}\int_{r_{i-1}}^{r_i}\frac{dx}{x^{1+\mu}}=\int_2^{\infty}\frac{dx}{x^{1+\mu}}<\infty.\tag{39}$$

Further, by (8) and (38) it follows that

$$\sum_{i=0}^{\infty} \frac{\|\phi_i\|^2}{r_i^2} E(\|w_{i+1}\|^2|\mathcal{F}_i) < \infty \quad \text{a.s.}$$

This together with convergence of (37) implies

$$\sum_{i=0}^{\infty} \frac{\|\phi_i\|^2}{r_i^2} \|w_{i+1}\|^2 < \infty \quad \text{a.s.} \tag{40}$$

Again, applying Theorem 2.8 of ref. [9] we see that by (35) for $\gamma \in (\frac{1}{2}, 1)$

$$\begin{aligned}\sum_{i=0}^{n} \frac{\|\phi_i\|^2}{r_i^2} \zeta_i^T w_{i+1} &= O\bigg(\Big(\sum_{i=0}^{n} \frac{\|\phi_i\|^4}{r_i^4} \|\zeta_i\|^2 \cdot \log(i+1)\Big)^{\gamma}\bigg) \\ &= O\bigg(\Big(\sum_{i=0}^{n} \frac{\|\zeta_i\|^2}{r_i}\Big)^{\gamma}\bigg).\end{aligned} \tag{41}$$

Since $\gamma \in (\frac{1}{2}, 1)$, the terms on the right-hand sides of (36) and (41) are either bounded or dominated by a quantity of $o\Big(\sum_{i=0}^{n} \frac{\|\zeta_i\|^2}{r_i}\Big)$. Therefore, by (36), (40), and (41) it follows that the right-hand side of (34) is bounded.

Thus, (34) yields that $\{\|\theta_k\|\}$ is bounded and (30) is correct.

4 Auxiliary results

We are planning to prove strong consistency of θ_k defined in the last section. For this we first establish some auxiliary results.

From (13), (26), and (29) it follows that

$$\begin{aligned}\tilde{\theta}_{k+1} &= \tilde{\theta}_k - \frac{\alpha\phi_k}{r_k}\Big(\phi_k^{0T}\theta + w_{k+1}^T - \phi_k^T\theta_k\Big) \\ &= \tilde{\theta}_k - \frac{\alpha\phi_k}{r_k}\Big((\phi_k^T - \phi_k^{\zeta T})\theta - \phi_k^T\theta_k + w_{k+1}^T\Big) \\ &= (I - \frac{\alpha\phi_k\phi_k^T}{r_k})\tilde{\theta}_k + \frac{\alpha\phi_k\phi_k^{\zeta T}\theta}{r_k} - \frac{\alpha\phi_k}{r_k} w_{k+1}^T.\end{aligned} \tag{42}$$

From here it is seen that $(I - \frac{\alpha\phi_k\phi_k^T}{r_k})$ serving as the transition matrix plays an important role in the behavior of $\{\tilde{\theta}_k\}$.

Let us recursively define $\Phi(n+1, i)$ and $\Phi^0(n+1, i)$ for $n+1 \geqslant i, i = 0, 1, \dots$ as follows:

$$\Phi(n+1, i) = \left(I - \frac{\alpha\phi_n\phi_n^T}{r_n}\right)\Phi(n, i), \quad \Phi(i, i) = I, \tag{43}$$

$$\Phi^0(n+1, i) = \left(I - \frac{\alpha\phi_n^0\phi_n^{0T}}{r_n^0}\right)\Phi^0(n, i), \quad \Phi^0(i, i) = I, \tag{44}$$

Copyright by Science in China Press 2005

where

$$r_n^0 = 1 + \sum_{i=1}^{n} \|\phi_i^0\|^2. \tag{45}$$

The following lemma shows the properties of $\Phi(n,i)$ and its relationship with $\Phi^0(n,i)$.

Lemma 4. i) For the algebraic recurrence (43) the following inequalities take place:

$$\sum_{j=0}^{\infty} \frac{\|\Phi(j,0)\phi_j\|^2}{r_j} \leqslant \frac{m(p+r)}{\alpha}, \tag{46}$$

$$\sum_{i=0}^{n-1} \frac{\|\Phi(n,i+1)\phi_i\|^2}{r_i} \leqslant \frac{m(p+r)}{\alpha}. \tag{47}$$

ii) Assume A1, A3, and A4 hold. Then

$$S^0 \triangleq \{\omega : \Phi^0(n,0) \xrightarrow[n\to\infty]{} 0\} \subset S \triangleq \{\omega : \Phi(n,0) \xrightarrow[n\to\infty]{} 0\} \tag{48}$$

with possible exception on a set of probability zero.

Proof. The assertion i) is derived by a purely algebraic manipulation without using any assumption on the system. For its proof we refer to Lemma 4.2 of ref. [9].

In ii) all conditions A1, A3, and A4 are used for deriving (30). With (30) having been established, the proof can be carried out along the lines of the proof for Theorem 4.4 in ref. [9].

Theorem 1. Assume A1–A4 hold. Then the estimate θ_k given by the SG algorithm with an arbitrary initial value converges to θ on S defined by (48) with possible exception on a set of probability zero.

Proof. The proof is modified from that for Theorem 4.3 of ref. [9] by taking notice of increasing variances of $\{w_k\}$.

By (42) and (43) it follows that

$$\tilde{\theta}_{n+1} = \Phi(n+1,0)\tilde{\theta}_0 + \alpha \sum_{j=0}^{n} \Phi(n+1,j+1)\frac{\phi_j \phi_j^{\zeta T}}{r_j}\theta - \alpha \sum_{j=0}^{n} \Phi(n+1,j+1)\frac{\phi_j}{r_j} w_{j+1}^T. \tag{49}$$

For proving the theorem it suffices to show that the last two terms of (49) a.s. converge to zero on S.

For the second term on the right-hand side of (49) we have

$$\left\| \sum_{j=0}^{n} \Phi(n+1,j+1)\frac{\phi_j \phi_j^{\zeta T}}{r_j} \right\| \leqslant \left\| \sum_{j=0}^{N} \Phi(n+1,j+1)\frac{\phi_j \phi_j^{\zeta T}}{r_j} \right\| + \left(\sum_{j=N+1}^{n} \frac{\|\Phi(n+1,j+1)\phi_j\|^2}{r_j} \right)^{\frac{1}{2}} \left(\sum_{j=N+1}^{\infty} \frac{\|\phi_j^{\zeta}\|^2}{r_j} \right)^{\frac{1}{2}}. \tag{50}$$

For any fixed N, the first term on the right-hand side of (50) tends to zero on S as $n \to \infty$, while the last term may be arbitrarily small if N is sufficiently large by (47) and Lemma 3. Therefore, the second term on the right-hand side of (49) tends to zero a.s. on S.

We now consider the last term of (49).

Let $\gamma > \frac{1}{2}$. Then by (35) there is $\mu \in (0,1)$ such that

$$\frac{(\log i)^{1-\delta}}{r_i^{2\gamma}} < \frac{c_4}{r_i^{1+\mu}} \tag{51}$$

for some $c_4 > 0$ a.s., where δ is the one figured in A1. By (39), (51) it follows that

$$\sum_{i=1}^{\infty} \frac{\|\phi_i\|^2}{r_i^{2\gamma}} (\log i)^{1-\delta} < \infty \quad \text{a.s.} \tag{52}$$

Then, by the convergence theorem for martingale difference sequences[9,10], (8) and (52) yield that

$$\sum_{i=1}^{\infty} \frac{\phi_i w_{i+1}^T}{r_i^{\gamma}} < \infty \quad \text{a.s.} \quad \forall \gamma > \frac{1}{2}. \tag{53}$$

This implies

$$Q_n \triangleq \sum_{i=1}^{n} \frac{\phi_i w_{i+1}^T}{r_i} \xrightarrow[n\to\infty]{} \sum_{i=1}^{\infty} \frac{\phi_i w_{i+1}^T}{r_i} \triangleq Q < \infty \quad \text{a.s.} \tag{54}$$

Defining $Q_{-1} \triangleq 0$, we find that

$$\left\| \sum_{i=1}^{n} \Phi(n+1, j+1) \frac{\phi_j}{r_j} w_{j+1}^T \right\| = \left\| \sum_{j=0}^{n} \Phi(n+1, j+1)(Q_j - Q_{j-1}) \right\|$$

$$= \left\| Q_n - \sum_{j=0}^{n} [\Phi(n+1, j+1) - \Phi(n+1, j)] Q_{j-1} \right\|$$

$$= \Bigg\| Q_n - \sum_{j=0}^{n} [\Phi(n+1, j+1) - \Phi(n+1, j)] Q$$

$$+ \sum_{j=0}^{n} [\Phi(n+1, j+1) - \Phi(n+1, j)] \sum_{i=j}^{\infty} \frac{\phi_i w_{i+1}^T}{r_i} \Bigg\|$$

$$= \left\| Q_n - Q + \Phi(n+1, 0) Q + \alpha \sum_{j=0}^{n} \Phi(n+1, j+1) \frac{\phi_j \phi_j^T}{r_j} \sum_{i=j}^{\infty} \frac{\phi_i w_{i+1}^T}{r_i} \right\|,$$

and hence on S

$$\lim_{n\to\infty} \Bigg\{ \left\| \sum_{j=0}^{n} \Phi(n+1, j+1) \frac{\phi_j}{r_j} w_{j+1}^T \right\|$$
$$- \alpha \left\| \sum_{j=0}^{n} \Phi(n+1, j+1) \frac{\phi_j \phi_j^T}{r_j} \sum_{i=j}^{\infty} \frac{\phi_i w_{i+1}^T}{r_i} \right\| \Bigg\} = 0. \tag{55}$$

Copyright by Science in China Press 2005

If we can show that the second term on the left-hand side of (55) tends to zero as $n \to \infty$, then so does its first term. This combining with convergence of (50) to zero leads to that all terms on the right-hand side of (49) tend to zero a.s. on S and thus completes the proof of the theorem.

In order to show

$$\left\| \sum_{j=0}^{n} \Phi(n+1, j+1) \frac{\phi_j \phi_j^T}{r_j} \sum_{i=j}^{\infty} \frac{\phi_i w_{i+1}^T}{r_i} \right\| \xrightarrow[n\to\infty]{} 0 \quad \text{a.s. on } S, \tag{56}$$

we first prove that for any given $\epsilon_1 > 0$, N can be selected large enough such that

$$\left\| r_j^{\nu} \sum_{i=j}^{\infty} \frac{\phi_i w_{i+1}^T}{r_i} \right\| < \epsilon_1, \quad \forall j \geqslant N, \tag{57}$$

where $\nu \in (0, \frac{1}{2})$ is arbitrarily fixed. Set

$$\sigma_n = \sum_{i=n}^{\infty} \frac{\phi_i w_{i+1}^T}{r_i^{1-\nu}},$$

which is finite by (53).

We find that

$$\begin{aligned} r_N^{\nu} \left\| \sum_{i=N}^{\infty} \frac{\phi_i w_{i+1}^T}{r_i} \right\| &= r_N^{\nu} \left\| \sum_{i=N}^{\infty} (\sigma_i - \sigma_{i+1}) \frac{1}{r_i^{\nu}} \right\| \\ = r_N^{\nu} \left\| \frac{\sigma_N}{r_N^{\nu}} - \sum_{i=N}^{\infty} \sigma_{i+1} \left(\frac{1}{r_i^{\nu}} - \frac{1}{r_{i+1}^{\nu}} \right) \right\| \\ \leqslant \|\sigma_N\| + r_N^{\nu} \sum_{i=N}^{\infty} \|\sigma_{i+1}\| \left(\frac{1}{r_i^{\nu}} - \frac{1}{r_{i+1}^{\nu}} \right) \xrightarrow[N\to\infty]{} 0 \quad \text{a.s.} \end{aligned}$$

By (57) we now have

$$\begin{aligned} &\left\| \sum_{j=N}^{n} \Phi(n+1, j+1) \frac{\phi_j \phi_j^T}{r_j} \sum_{i=j}^{\infty} \frac{\phi_i w_{i+1}^T}{r_i} \right\| \\ = &\left\| \sum_{j=N}^{n} \Phi(n+1, j+1) \frac{\phi_j \phi_j^T}{r_j^{1+\nu}} r_j^{\nu} \sum_{i=j}^{\infty} \frac{\phi_i w_{i+1}^T}{r_i} \right\| \\ \leqslant &\, \epsilon_1 \left(\sum_{j=N}^{n} \frac{\|\Phi(n+1, j+1)\phi_j\|}{r_j} \right)^{\frac{1}{2}} \left(\sum_{j=N}^{n} \frac{\|\phi_j\|^2}{r_j^{1+2\nu}} \right)^{\frac{1}{2}}, \end{aligned}$$

which tends to zero as $n \to \infty$ then $N \to \infty$ by (47) and (39). From this, the convergence (56) is derived and at the same time the proof is completed.

5 Main result

We now in a position to prove our main result.

Theorem 2. Assume A1–A4 hold. Let θ_k be given by the SG algorithm (25)–(27). Then $\theta_k \xrightarrow[n\to\infty]{} \theta$ a.s.

Proof. By Lemma 4 and Theorem 1 it suffices to show that

$$\Phi^0(n,0) \xrightarrow[n\to\infty]{} 0 \quad \text{a.s.},$$

where $\Phi^0(n,0)$ is recursively given by (44), (45).

Step 1. Let us first explain the idea of the proof. For this we define

$$s_n \triangleq \sum_{i=n_0}^{n-1} \frac{\|\phi_i^0\|^2}{r_i^0(\log r_{i-1}^0)^{1-\epsilon}}, \tag{58}$$

where ϵ is taken such that: $\frac{4}{5} \leqslant \epsilon < \delta$ and n_0 possibly depending on sample n_0 is large enough so that $\log r_{n_0}^0 > 1$. The selection of ϵ is possible since $\delta \in (\frac{4}{5}, 1]$ as required in A1. We now show that $s_n \xrightarrow[n\to\infty]{} \infty$ a.s. By Lemma 1

$$r_n^0 = O\big(n(\log n)^{1-\delta}\big). \tag{59}$$

By (15) there is $c_5 > 0$ a.s. such that

$$r_{n-1}^0 \geqslant c_5 n \tag{60}$$

for all sufficiently large n.

Combining (59), (60) leads to

$$r_n^0 \leqslant c_6 r_{n-1}^0 (\log r_{n-1}^0)^{1-\delta} \tag{61}$$

for some $c_6 > 0$ a.s. and all large enough n. Assume n_0 in (58) is large enough so that both (60), (61) hold for $n \geqslant n_0$. Then

$$\begin{aligned} s_{n+1} &\geqslant \frac{1}{c_6}\sum_{i=n_0}^{n} \frac{\|\phi_i^0\|^2}{r_{i-1}^0(\log r_{i-1}^0)^{2-\epsilon-\delta}} = \frac{1}{c_6}\sum_{i=n_0}^{n}\int_{r_{i-1}^0}^{r_i^0} \frac{dx}{r_{i-1}^0(\log r_{i-1}^0)^{2-\epsilon-\delta}} \\ &\geqslant \frac{1}{c_6}\sum_{i=n_0}^{n}\int_{r_{i-1}^0}^{r_i^0} \frac{dx}{x(\log x)^{2-\epsilon-\delta}} \\ &= \frac{1}{(\epsilon+\delta-1)c_6}[(\log r_n^0)^{\epsilon+\delta-1} - (\log r_{n_0}^0)^{\epsilon+\delta-1}] \xrightarrow[n\to\infty]{} \infty \quad \text{a.s.}, \end{aligned}$$

since $\epsilon + \delta > \frac{8}{5} > 1$. This means that there is $c_7 > 0$ a.s. such that

$$s_{n+1} \geqslant c_7 (\log r_n^0)^{\epsilon+\delta-1} \to \infty \quad \text{a.s.} \tag{62}$$

From $s_n \xrightarrow[n\to\infty]{} \infty$ it is clear that

$$m(t) \triangleq \max\{m : s_m \leqslant t\} \tag{63}$$

is finite for any $t \in (0,\infty)$ and diverges to infinity as $t \to \infty$.

Therefore, for $\Phi^0(n,0) \xrightarrow[n\to\infty]{} 0$ it is equivalent to show $\Phi^0(m(t),0) \xrightarrow[t\to\infty]{} 0$.

The idea of proving $\Phi^0(n,0) \xrightarrow[n\to\infty]{} 0$ consists in showing

$$\|\Phi^0(m(n_0+kd)), m(n_0+(k-1)d))\| \leqslant 1 - ck^{-\alpha_1} \tag{64}$$

with $c > 0$ and $0 < \alpha_1 \leqslant 1$, where $d > 1$ is fixed.

If this is done, then

$$\lim_{n\to\infty} \|\Phi^0(n,0)\| = \lim_{n\to\infty} \|\Phi^0(m(n_0+nd),0)\|$$

Copyright by Science in China Press 2005

$$\leqslant \lim_{n\to\infty} \|\Phi^0(m(n_0,0))\| \prod_{k=1}^{n} \|\Phi^0(m(n_0+kd), m(n_0+(k-1)d))\|$$

$$\leqslant \lim_{n\to\infty} \prod_{k=1}^{n}(1-ck^{-\alpha_1}) = 0,$$

and the proof of the theorem will be completed.

Thus, our goal is to prove (64).

Step 2. We now show that there is $c_8 > 0$ a.s. such that

$$\sum_{i=m(n_0+(k-1)d)}^{m(n_0+kd)-1} \frac{\phi_i^0 \phi_i^{0T}}{r_i^0} > c_8 k^{\frac{\delta-\epsilon}{\epsilon}} I, \forall k = 1, 2, \ldots \tag{65}$$

From the definition (63), for $m(t)$ it is seen that

$$t < s_{m(t)+1} < s_{m(t)} + 1 \leqslant t + 1. \tag{66}$$

Noticing that by (61) for large n_0

$$\log r_n^0 \leqslant 2 \log r_{n-1}^0 \quad \forall n \geqslant n_0,$$

and hence

$$s_n \leqslant 2 \sum_{i=n_0}^{n-1} \frac{\|\phi_i^0\|^2}{r_i(\log r_i)^{1-\epsilon}} \leqslant 2 \sum_{i=n_0}^{n-1} \int_{r_{i-1}^0}^{r_i^0} \frac{dx}{x(\log x)^{1-\epsilon}} \leqslant \frac{2}{\epsilon}(\log r_{n-1}^0)^{\epsilon},$$

we find that

$$\log r_{m(n_0+(k-1)d)}^0 \geqslant \left(\frac{\epsilon}{2} s_{m(n_0+(k-1)d)+1}\right)^{\frac{1}{\epsilon}} \geqslant \left(\frac{\epsilon}{2}(n_0+(k-1)d)\right)^{\frac{1}{\epsilon}}, \tag{67}$$

where for the last inequality (66) is invoked.

For large enough n_0 from (15) we see that $S_i^0 \geqslant \frac{c_2 i}{2} I$, $\forall i \geqslant n_0$, and from (59), (60) $r_{i-1}^0 \leqslant c_9 i(\log r_{i-1}^0)^{1-\delta}$, $\forall i \geqslant n_0$. Consequently, for $i-1 \geqslant n_0$ we have

$$\frac{S_{i-1}^0}{r_{i-1}^0} \geqslant \frac{c_2}{2c_9(\log r_{i-1}^0)^{1-\delta}} I. \tag{68}$$

Thus, summing by parts and taking notice of (68) we have

$$\sum_{i=m(n_0+(k-1)d)}^{m(n_0+kd)-1} \frac{\phi_i^0 \phi_i^{0T}}{r_i^0} = \sum_{i=m(n_0+(k-1)d)}^{m(n_0+kd)-1} \frac{S_i^0 - S_{i-1}^0}{r_i^0}$$

$$= \frac{S_{m(n_0+kd)-1}^0}{r_{m(n_0+kd)-1}^0} - \frac{S_{m(n_0+(k-1)d)-1}^0}{r_{m(n_0+(k-1)d)}^0} + \sum_{i=m(n_0+(k-1)d)+1}^{m(n_0+kd)-1} S_{i-1}^0 \frac{\|\phi_i^0\|^2}{r_i^0 r_{i-1}^0}$$

$$\geqslant -I + \frac{c_2}{2c_9} \sum_{i=m(n_0+(k-1)d)+1}^{m(n_0+kd)-1} \frac{\|\phi_i^0\|^2}{r_i^0(\log r_{i-1}^0)^{1-\delta}} I$$

$$\geqslant -I + \frac{c_2}{2c_9}(\log r_{m(n_0+(k-1)d)}^0)^{\delta-\epsilon} \sum_{i=m(n_0+(k-1)d)+1}^{m(n_0+kd)-1} \frac{\|\phi_i^0\|^2}{r_i^0(\log r_{i-1}^0)^{1-\epsilon}} I$$

$$\geqslant -I+\frac{c_2}{2c_9}\left[\frac{\epsilon}{2}(n_0+(k-1)d)\right]^{\frac{\delta-\epsilon}{\epsilon}}(d-1)I \geqslant c_8 k^{\frac{\delta-\epsilon}{\epsilon}} I \tag{69}$$

with $c_8 > 0$ a.s., where for the last but one inequality, (67) and the definitions (58), (63) are used.

Step 3. We now prove (64). Let ρ_k be the largest eigenvalue of $\Phi^{0T}(m(n_0+kd), m(n_0+(k-1)d))\Phi^0(m(n_0+kd), m(n_0+(k-1)d))$ and $x_{m(n_0+(k-1)d)}$ the corresponding unit eigenvector.

For $i \in [m(n_0+(k-1)d), m(n_0+kd)-1]$ recursively define x_i by

$$x_{i+1}=\left(I-\frac{\alpha\phi_i^0\phi_i^{0T}}{r_i^0}\right)x_i. \tag{70}$$

Then, we have

$$\begin{aligned} x^T_{m(n_0+kd)}x_{m(n_0+kd)} &= x^T_{m(n_0+(k-1)d)}\Phi^{0T}(m(n_0+kd), m(n_0+(k-1)d))\cdot \\ &\quad \cdot \Phi^0(m(n_0+kd), m(n_0+(k-1)d))x_{m(n_0+(k-1)d)} \\ &= \rho_k x^T_{m(n_0+(k-1)d)}x_{m(n_0+(k-1)d)} = \rho_k. \end{aligned} \tag{71}$$

Further, from (70) we have

$$x^T_{i+1}x_{i+1} \leqslant x_i^T x_i - \alpha x_i^T \frac{\phi_i^0\phi_i^{0T}}{r_i^0}x_i,$$

which, by summing up both sides, leads to

$$\alpha \sum_{i=m(n_0+(k-1)d)}^{m(n_0+kd)-1} \frac{\|\phi_i^{0T}x_i\|^2}{r_i^0} \leqslant \|x_{m(n_0+(k-1)d)}\|^2 - \|x_{m(n_0+kd)}\|^2 = 1-\rho_k, \tag{72}$$

where for the last equality, (71) is used.

From (70) by (72) it follows that for $i \in [m(n_0+(k-1)d), m(n_0+kd)-1]$ we have

$$\begin{aligned} \|x_i - x_{m(n_0+(k-1)d)}\| &= \alpha\left\| \sum_{j=m(n_0+(k-1)d)}^{i-1} \frac{\phi_j^0\phi_j^{0T}x_j}{r_j^0}\right\| \\ &\leqslant \alpha(\log r^0_{m(n_0+kd)-1})^{\frac{1-\epsilon}{2}} \sum_{i=m(n_0+(k-1)d)}^{m(n_0+kd)-1} \\ &\quad \frac{\|\phi_j^0\|}{(r_j^0)^{\frac12}(\log r^0_{j-1})^{\frac{1-\epsilon}{2}}}\cdot\frac{\|\phi_j^{0T}x_j\|}{(r_j^0)^{\frac12}} \\ &\leqslant \sqrt{\alpha}(\log r^0_{m(n_0+kd)-1})^{\frac{1-\epsilon}{2}}(1+d)^{\frac12}(1-\rho_k)^{\frac12}. \end{aligned} \tag{73}$$

By (62) and (66) we see

$$\log r^0_{m(n_0+kd)-1} \leqslant \left(\frac{1}{c_7}s_{m(n_0+kd)}\right)^{\frac{1}{\epsilon+\delta-1}} \leqslant \left(\frac{n_0+kd}{c_7}\right)^{\frac{1}{\epsilon+\delta-1}}. \tag{74}$$

Combining (73) and (74) yields

$$\|x_i - x_{m(n_0+(k-1)d)}\| \leqslant \sqrt{\alpha}\left(\frac{n_0+kd}{c_7}\right)^{\frac{1-\epsilon}{2(\epsilon+\delta-1)}}(1+d)^{\frac12}(1-\rho_k)^{\frac12}$$

Copyright by Science in China Press 2005

$$\leqslant c_{10} k^{\frac{1-\epsilon}{2(\epsilon+\delta-1)}} (1-\rho_k)^{\frac{1}{2}} \quad \text{for some } c_{10} > 0 \text{ a.s.} \tag{75}$$

By (65) incorporating with (75) we arrive at

$$\begin{aligned} c_8 k^{\frac{\delta-\epsilon}{\epsilon}} &\leqslant x^T_{m(n_0+(k-1)d)} \sum_{j=m(n_0+(k-1)d)}^{m(n_0+kd)-1} \frac{\phi^0_j \phi^{0T}_j}{r^0_j} (x_{m(n_0+(k-1)d)} - x_j + x_j) \\ &\leqslant (\log r^0_{m(n_0+kd)-1})^{1-\epsilon} \sum_{j=m(n_0+(k-1)d)}^{m(n_0+kd)-1} \frac{\|\phi^0_j\|^2}{r^0_j (\log r^0_j)^{1-\epsilon}} \cdot c_{10} k^{\frac{1-\epsilon}{2(\epsilon+\delta-1)}} (1-\rho_k)^{\frac{1}{2}} \\ &\quad + (\log r^0_{m(n_0+kd)-1})^{\frac{1-\epsilon}{2}} \sum_{j=m(n_0+(k-1)d)}^{m(n_0+kd)-1} \frac{\|\phi^0_j\|}{(r^0_j)^{\frac{1}{2}} (\log r^0_{j-1})^{\frac{1-\epsilon}{2}}} \cdot \frac{\|\phi^{0T}_j x_j\|}{(r^0_j)^{\frac{1}{2}}}, \end{aligned}$$

which, by (72), (74), and the Schwarz inequality, leads to

$$\begin{aligned} c_8 k^{\frac{\delta-\epsilon}{\epsilon}} &\leqslant c_{11} k^{\frac{1-\epsilon}{\epsilon+\delta-1} + \frac{1-\epsilon}{2(\epsilon+\delta-1)}} (1-\rho_k)^{1/2} + c_{12} k^{\frac{1-\epsilon}{2(\epsilon+\delta-1)}} (1-\rho_k)^{1/2} \\ &\leqslant c_{13} k^{\frac{3}{2}(\frac{1-\epsilon}{\epsilon+\delta-1})} (1-\rho_k)^{1/2} \end{aligned} \tag{76}$$

for some $c_{11} > 0$, $c_{12} > 0$, $c_{13} > 0$ a.s.

From (76) it follows that

$$\rho_k \leqslant 1 - c_{14} k^{-\alpha_1}, \quad \alpha_1 \triangleq \frac{3(1-\epsilon)}{\epsilon+\delta-1} - \frac{2(\delta-\epsilon)}{\epsilon}, \quad c_{14} > 0, \quad \text{a.s.} \tag{77}$$

Since $\delta \in (\frac{4}{5}, 1]$ and $\epsilon \in [\frac{4}{5}, \delta)$, we have $2\delta(1-\delta) + \epsilon(1-\epsilon) > 0$, which implies that $\alpha_1 > 0$. The inequality $\alpha_1 \leqslant 1$ is equivalent to

$$\frac{1-\epsilon}{\delta} + \frac{1-\delta}{\epsilon} \leqslant \frac{1}{2},$$

which is clearly held for $\delta > \frac{4}{5}$ and $\epsilon \geqslant \frac{4}{5}$. Consequently, we have

$$0 < \alpha_1 \leqslant 1. \tag{78}$$

Since $\|\Phi^0(m(n_0+kd), m(n_0+(k-1)d)\| = \rho_k^{\frac{1}{2}}$, we have

$$\|\Phi^0(m(n_0+kd), m(n_0+(k-1)d)\| \leqslant (1 - c_{14} k^{-\alpha_1})^{\frac{1}{2}} \leqslant 1 - \frac{c_{14}}{2} k^{-\alpha_1}.$$

This verifies (64) with $c = \frac{c_{14}}{2}$ and at the same time completes the proof of the theorem.

6 Concluding remarks

The strong consistency of the coefficient estimate given by the SG algorithm is proved for multidimensional ARMA processes with increasing variances. We now give a few comments on the reasonability of the imposed conditions and on the possible further research.

Condition A1 describes what kind of increasing variances is considered in the paper. If we find that the variance of a time series slowly increases, then probably to impose A2 is necessary, because any instability of $A(z)$ will give rise to a quick (at least at a rate of polynomial) divergence of the second moment of the data.

In A3 the condition on having no common left factor for $A(z)$ and $C(z)$ is natural for identifiability of coefficients, but the row-full-rank condition of $[A_p \vdots C_r]$ is rather technical, because the orders p and r are assumed to be the upper bounds of the true ones. Concerning A4, it is a strong condition on $C(z)$ and may be weakened, e.g. to a stability condition. This is for further research. Besides, it is of interest to consider other estimation methods, e.g. the extended least squares (ELS) algorithm. It is not clear if the ELS works for the case considered in the paper.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant Nos. G0221301, 60334040 and 60474004).

References

1. Anderson, T. W., The statistical Analysis of Time Series, New York: Wiley, 1971.
2. Brockwell, P. T., Davis, R. A., Time Series: Theory and Methods, New York: Springer, 1987.
3. Caines, P. E., Linear Stochastic Systems, New York: Wiley, 1988.
4. Hanann, E. J., Diestler, M., The Statistical Theory of Linear Systems, New York: Wiley, 1988.
5. Ljung, L., Söderström, T. S., Theory and Practice of Recursive Identification, Boston: The MIT Press, 1983.
6. Chen, H. F., On stability and trajectory boundedness in mean-square sense for ARMA processes, Acta Mathematicae Applicatae Sinica, English Series, 2003, 19: 573–580.
7. Zhang, H. M., A loglog law for unstable ARMA processes with applications to time series analysis, J. Multivariate Analysis, 1992, 40: 173–204.
8. Lai, T. L., Wei, C. Z., Asymptotic properties of general autoregressive models and strong consistency of least squares of their parameters, J. Multivariate Analysis, 1992, 13: 1–23.
9. Chen, H. F., Guo, L., Identification and Stochastic Adaptive Control, Boston: Birkhäuser, 1991.
10. Chow, Y. S., Local convergence of martingales and the law of large numbers, Ann. Math. Stat., 1965, 36: 552–558.
11. Chen, H. F., Stochastic Approximation and Its Applications, Dordrecht: Kluwer, 2002.
12. Goodwin, G. C., Ramadge, P. J., Caines, P. E., Discrete-time stochastic adaptive control, SIAM J. Control and Optimization, 1981, 19: 829–853.

Copyright by Science in China Press 2005