INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING, VOL. 7, 525–537 (1993)

PERFORMANCE ANALYSIS OF THE FORGETTING FACTOR RLS ALGORITHM

LEI GUO
Institute of Systems Science, Academica Sinica, Beijing 100080, China

LENNART LJUNG
Department of Electrical Engineering, Linköping University, S-581 83 Linköping, Sweden

AND

PIERRE PRIOURET
Laboratoire de Probabilités, Université Pierre et Marie Curie, F-75252 Paris, France

SUMMARY

An analysis is given of the performance of the standard forgetting factor recursive least squares (RLS) algorithm when used for tracking time-varying linear regression models. Three basic results are obtained: (1) the '*P*-matrix' in the algorithm remains bounded if and only if the (time-varying) covariance matrix of the regressors is uniformly non-singular; (2) if so, the parameter tracking error covariance matrix is of the order $O(\mu + \gamma^2/\mu)$, where $\mu = 1 - \lambda$, λ is the forgetting factor and γ is a quantity reflecting the speed of the parameter variations; (3) this covariance matrix can be arbitrarily well approximated (for small enough μ) by an expression that is easy to compute.

KEY WORDS Adaptation Least squares Tracking Recursive identification

1. INTRODUCTION

Consider the linear time-varying regression

$$y(t) = \varphi^{\mathrm{T}}(t)\theta(t-1) + e(t), \quad t \geqslant 0 \tag{1}$$

where $y(t)$ and $e(t)$ are the observation and the noise respectively at time t, $\varphi(t) \in \mathbb{R}^d$ is the regressor and $\theta(t)$ is the unknown parameter vector to be estimated.

The well-known least squares estimator with constant forgetting factor $\lambda \in (0, 1)$ is defined by

$$\hat{\theta}(t) = \underset{\theta \in \mathbb{R}^d}{\operatorname{argmin}} \sum_{i=0}^{t} \lambda^{t-i}[y(i) - \theta^{\mathrm{T}}\varphi(i)]^2 \tag{2}$$

The quantity $\mu \triangleq 1 - \lambda$ is usually referred to as the speed of adaptation. Intuitively speaking, when the parameter process $\{\theta(t)\}$ is slowly time-varying, the adaptation speed should also be slow (i.e. μ small).

Since the function to be minimized in (2) is quadratic, the minimizer $\hat{\theta}(t)$ can be found

This paper was recommended for publication by editor C. F. N. Cowan

0890–6327/93/060525–13$11.50

Received 8 October 1992
Revised 30 June 1993

explicitly. Consequently, with some simple manipulations involving the use of the matrix inversion formula (as in e.g. Reference 1), it is easy to see that $\hat{\theta}(t)$ may be computed by the recursion (the RLS)

$$\hat{\theta}(t)=\hat{\theta}(t-1)+\mu P(t)\varphi(t)[y(t)-\varphi^{\mathrm{T}}(t)\hat{\theta}(t-1)] \tag{3}$$

$$P(t)=\frac{1}{1-\mu}\left(P(t-1)-\mu\frac{P(t-1)\varphi(t)\varphi^{\mathrm{T}}(t)P(t-1)}{1-\mu+\mu\varphi^{\mathrm{T}}(t)P(t-1)\varphi(t)}\right) \tag{4}$$

with deterministic initial conditions $\hat{\theta}(0)$ and $P(0)>0$.

Although this RLS algorithm is well motivated, a complete and rigorous analysis for its tracking performance does not seem to be given in the literature. Among the various previous studies on RLS we mention the works of Benveniste[2] for the case of $\mu\to 0$, Macchi and Eweda[3] for the case of constant parameters, Ljung and Priouret[4] for results obtained under a certain moment (stability) condition, Bittanti and Campi[5,6] for a special class of linear models, Niedźwiecki and Guo[7] for results over finite time horizons and Guo[8] for some stability studies.

In this paper we will first investigate the relationship between various excitation conditions and the L_p-uniform boundedness of the matrix $P(t)$, which is crucial to the performance analysis. Then we establish the tracking error bounds and give some simple approximations for the covariance matrix of this error. These results do not require that the time horizons are finite nor that the adaptation gain is decreasing (i.e. $\mu\to 0$).

Throughout the paper the minimum eigenvalue of a real matrix X is denoted by $\lambda_{\min}(X)$, the norm $\|X\|$ is defined as its maximum singular value, and by $\|X\|_p$ we mean the L_p-norm defined by $\|X\|_p=\{E\|X\|^p\}^{1/p}$.

2. THE UNIFORM BOUNDEDNESS OF $\|P(t)\|_p$

Let $\{\varphi(t),\mathscr{F}_t\}$ be an $\mathbb{R}^d$-valued adapted sequence with $\{\mathscr{F}_t\}$ being a non-decreasing sequence of σ-algebras defined on the basic probability space $(\Omega,\mathscr{F},P)$. For any integer $h>0$ let $A_h(t)$ be the matrix sequence defined by

$$A_h(t)=\sum_{i=t+1}^{t+h}\varphi(i)\varphi^{\mathrm{T}}(i),\quad t\geqslant 0 \tag{5}$$

Our first result concerns the minimum excitation condition needed for boundedness of $\|P(t)\|_p$, $p\geqslant 1$.

Proposition 1

Let $\sup_t E\|\varphi(t)\|^2<\infty$. If $\sup_t E\|P(t)\|<\infty$ holds for some $\mu\in(0,1)$, then there exists an integer $h>0$ such that

$$\inf_t E\lambda_{\min}(A_h(t))>0 \tag{6}$$

Proof. By (4) and the matrix inversion formula,

$$R(t)=(1-\mu)R(t-1)+\mu\varphi(t)\varphi^{\mathrm{T}}(t),\qquad R(t)\triangleq P^{-1}(t) \tag{7}$$

Consequently, for any $h>0$

$$\lambda_{\min}(R(t+h))\leqslant(1-\mu)^h\|R(t)\|+\lambda_{\min}(A_h(t)) \tag{8}$$

By the boundedness of $E\|\varphi(t)\|^2$ it follows from (7) that $\sup_t E\|R(t)\|<\infty$. Also note that $\|P(t)\|=\lambda_{\min}^{-1}(R(t))$. Then by Jensen's inequality $E\lambda_{\min}(R(t))\geqslant(E\|P(t)\|)^{-1}$. Therefore by the boundedness of $E\|P(t)\|$ we have $\inf_t E\lambda_{\min}(R(t)) > 0$. Hence it is easy to choose h large enough such that

$$\inf_t E[\lambda_{\min}(R(t+h)) - (1-\mu)^h \|R(t)\|] > 0$$

This in conjunction with (8) yields (6). □

Since (6) is the necessary excitation condition for boundedness of $\{\|P(t)\|_p\}$, $p \geqslant 1$, it is natural to expect that it is also sufficient. Unfortunately, this is not true in general (a more discussion will be given at the end of this section). To see this, we simply take $d=1$ and set $\varphi(t)=\varphi \neq 0$, $t \geqslant 0$, with $E\varphi^2 > 0$ and $E\varphi^{-2} = \infty$. Obviously (6) holds, but by (7) $EP(t) = ER^{-1}(t) \geqslant E[(1-\mu)^t R(0) + \varphi^2]^{-1} \xrightarrow[t\to\infty]{} \infty$ holds for all $\mu \in (0,1)$ by the monotonic convergence theorem.

Now we consider the following two types of strengthened excitation conditions.

$E_1(p)$. There exists an integer $h > 0$ such that

$$\sup_{t \geqslant 0} E[\lambda_{\min}^{-p}(A_h(t))] < \infty$$

E_2. There exist an integer $h > 0$ and constants $c > 0$ and $\delta > 0$ such that

$$P\{\lambda_{\min}(A_h(t)) \geqslant c \mid \mathscr{F}_t\} \geqslant \delta \quad \forall t \geqslant 0$$

We will prove that either of the above two conditions ensures uniform boundedness of $\|P(t)\|_p$. For this we need the following lemma which is proved in Appendix I.

Lemma 1

Let $\{w_n, \mathscr{F}_n\}$ and $\{v_n, \mathscr{F}_n\}$ be two non-negative adapted processes satisfying $P(w_{n+1} \geqslant c \mid \mathscr{F}_n) \geqslant \delta$ a.s. for some positive constants c and δ. Also let

$$v_{n+1} \leqslant \frac{v_n}{(1-\mu)^h(1+\mu v_n w_{n+1})}, \quad n \geqslant 0, \quad v_0 \neq 0$$

with constants $\mu \in (0,1)$ and $h > 0$. Then for any $p \geqslant 1$ and any $\mu_0 \in (0, 1-(1-\delta)^{1/ph})$

$$\sup_{\mu \in (0,\mu_0]} \sup_{n \geqslant 0} \|v_n\|_p < \infty$$

Theorem 1

Let $\{P(t)\}$ be defined by (4).

(i) If condition $E_1(p)$ holds for some $p \geqslant 1$, then for any $\mu_0 \in (0,1)$

$$\sup_{\mu \in (0,\mu_0]} \sup_{t \geqslant 0} \|P(t)\|_p < \infty$$

(ii) If condition E_2 holds, then for any $p \geqslant 1$ and any $\mu_0 \in (0, 1-(1-\delta)^{1/ph})$

$$\sup_{\mu \in (0,\mu_0]} \sup_{t \geqslant 0} \|P(t)\|_p < \infty$$

Proof. By (7) it is easy to see that we need only to prove the uniform boundedness of the subsequence $\|P(hn)\|_p$, $n \geqslant 1$. By (7)

$$R(nh+h) = (1-\mu)^h R(nh) + \mu \sum_{i=nh+1}^{nh+h} (1-\mu)^{nh+h-i}\varphi(i)\varphi^{\mathrm{T}}(i) \tag{9}$$

Hence by (5)

$$\lambda_{\min}\{R(nh+h)\} \geqslant (1-\mu)^h \lambda_{\min}\{R(nh)\} + \mu(1-\mu)^h \lambda_{\min}(A_h(nh))$$

or

$$u_{n+1} \geqslant \lambda^h (u_n + \mu w_{n+1}), \quad \lambda = 1-\mu, \quad n \geqslant 0 \tag{10}$$

where $u_n \triangleq \lambda_{\min}\{R(nh)\}$ and $w_{n+1} \triangleq \lambda_{\min}\{A_h(nh)\}$.

Now let $E_1(p)$ hold for some $p \geqslant 1$. By (10) and the Schwarz inequality

$$u_n \geqslant \mu\lambda^h \left(\sum_{k=1}^{n} (\lambda^h)^{n-k} w_k \right) \geqslant \mu\lambda^h \left(\sum_{k=1}^{n} (\lambda^h)^{n-k} \right)^2 \left(\sum_{k=1}^{n} (\lambda^h)^{n-k} w_k^{-1} \right)^{-1}$$

Consequently, by $E_1(p)$ and the Minkowski inequality there exists a constant K such that

$$\| u_n^{-1} \|^p \leqslant K\mu^{-1}\lambda^{-h} \left(\sum_{k=1}^{n} (\lambda^h)^{n-k} \right)^{-1} = K\mu^{-1}\lambda^{-h} \frac{1-\lambda^h}{1-\lambda^{nh}} \leqslant Kh(1-\mu_0)^{-h} \frac{1}{1-\lambda^{nh}}, \quad \mu \in (0, \mu_0]$$

where for the last inequality we have used the fact that $\mu^{-1}(1-\lambda^h) \leqslant h$, $\mu \in (0,1)$. Hence, by noticing that $(1-\mu)^{1/\mu}$, $\mu \in (0,1)$, is a decreasing function of μ and that $\lim_{\mu \to 0} (1-\mu)^{1/\mu} = \mathrm{e}^{-1}$, we have

$$\sup_{n \geqslant \mu^{-1}} \| u_n \|_p \leqslant Kh(1-\mu_0)^{-h} \frac{1}{1-(1-\mu)^{h/\mu}} \leqslant Kh(1-\mu_0)^{-h} \frac{1}{1-\mathrm{e}^{-h}} < \infty$$

On the other hand, by (10) $u_n^{-1} \leqslant \lambda^{-hn} u_0^{-1}$, hence for $\mu \in (0, \mu_0]$

$$\sup_{n \leqslant \mu^{-1}} \| u_n^{-1} \|_p \leqslant (1-\mu)^{-h/\mu} \| u_0^{-1} \|_p \leqslant (1-\mu_0)^{-h/\mu_0} \| u_0^{-1} \|_p$$

This completes the proof of assertion (i), while assertion (ii) is a direct consequence of Lemma 1, since by setting $v_n = u_n^{-1}$, we see from (10) that $v_{n+1} \leqslant \lambda^{-h}(1+\mu v_n w_{n+1})^{-1} v_n$. This completes the proof. □

Remarks and discussion

(a) If under condition E_2 δ can be taken arbitrarily close to unity and the horizon h (which may depend on δ) satisfies $h = o(|\log(1-\delta)|)$ as $\delta \to 1^-$, then in Theorem 1 (ii) μ_0 can be taken as any value in (0, 1). However, in general, under condition E_2 μ_0 cannot be arbitrarily close to unity. In fact, we have the following example where $1-(1-\delta)^{1/h}$ is actually a critical point for boundedness of $\| P(t) \|_1$. Let $\{\varphi(t)\}$ be an i.i.d sequence with $P(\varphi(1)=0) = q = 1 - P(\varphi(1)=1)$. Then E_2 holds with $h=1$, $c=1$ and $\delta = 1-q$ and hence $1-(1-\delta)^{t/h} = 1-q$. However, for $\mu > 1-q$ it is easy to see that $EP(t) \geqslant E[P(t)I(\varphi(i)=0,\ 1 \leqslant i \leqslant t)] = (1-\mu)^{-t}R^{-1}(0)q^t \xrightarrow[t \to \infty]{} \infty$.

(b) If $\{\varphi(t)\}$ is a (strictly) stationary sequence, then condition $E_1(p)$ really is that $E[\lambda_{\min}^{-p}(\sum_{i=1}^{h} \varphi(i)\varphi^{\mathrm{T}}(i))] < \infty$ for some integer $h > 0$. This condition was previously used by Macchi and Eweda[3] and studied by Niedźwiecki and Guo.[7] If $\{\varphi(t)\}$ is an i.i.d. sequence of d-dimensional vectors with suitably high moment, then Niedźwiecki and Guo (Reference 7, pp. 200–202) showed that condition $E_1(p)$ holds for some $p \geqslant 1$ if and only if there exist constants $K > 0$, $\gamma > 0$ and $x_0 > 0$ such that for all $\beta \in \mathbb{R}^d$, $|\beta| = 1$, $P(|\beta^{\mathrm{T}}\varphi(1)| \leqslant x) \leqslant Kx^{\gamma}$, $0 \leqslant x \leqslant x_0$. In a similar way it is easy to show that condition E_2 holds if and only if $P(\beta^{\mathrm{T}}\varphi(1) \neq 0) > 0$ $\forall \beta \in \mathbb{R}^d$. Furthermore, if $P(\beta^{\mathrm{T}}\varphi(1) \neq 0) = 1$ $\forall \beta \in \mathbb{R}^d$, then E_2 holds for $h = d$ and for all $\delta \in (0,1)$. In many cases, as noted in Reference 7 the verification of the condition E_1 (or E_2) can be transferred to the i.i.d situation. Typical examples are the M-dependent case and the case where $\{\varphi(t)\}$ is generated by a state space model

$$x(t+1) = Ax(t) + B\eta(t), \qquad \varphi(t) = Cx(t) + D\eta(t) \tag{11}$$

where the driving signal $\{\eta(t)\}$is an i.i.d sequence. The first case is trivial to handle. For the second case we need only to note that for output-reachable model (11) there exists a constant $c > 0$ such that (see e.g. Reference 9, p. 353)

$$\lambda_{\min}(A_h(t)) \geqslant c\lambda_{\min}\left(\sum_{t+1+\nu}^{t+h} \bar{\eta}(i)\bar{\eta}^{\mathrm{T}}(i)\right), \quad t \geqslant 0$$

holds for all $h > \nu$, where ν is the McMillan degree of the system (11) and $\bar{\eta}(i) = [\eta^{\mathrm{T}}(i), \ldots, \eta^{\mathrm{T}}(i-\nu)]^{\mathrm{T}}$.

(c) Conditions E_1 and E_2 are not equivalent in general. In the i.i.d case, as implicitly mentioned in (b), E_1 is stronger than E_2; however, when the sequence $\{\varphi(t)\}$ is strongly correlated (especially when the predictable part is not bounded from below), E_1 is likely to be weaker than E_2. This can be easily seen by simply taking $d = 1$, $\varphi(t) = \varphi$ $\forall t$, with $E|\varphi|^{-2p} < \infty$, $P(|\varphi| < x) > 0$ $\forall x > 0$.

(d) If we are only concerned with the boundedness of $\|P(t)\|_p$ rather than the uniform (with respect to μ) boundedness as in Theorem 1, then condition E_2 may be further weakened so that the case where $\{\varphi(t)\}$ is generated by a class of time-varying models can be included.[8] However, without the uniform boundedness of $\|P(t)\|_p$ it is not known how to give a meaningful performance analysis for the tracking algorithm (see the next two sections).

Next we proceed to show that for a large class of weakly dependent sequences $\{\varphi(t)\}$ condition E_2 is actually equivalent to the minimum condition (6).

Recall that a random process $\{x(i),\ i \geqslant 1\}$ is called ϕ-mixing or α-mixing if $\phi(n) \to 0$ or $\alpha(n) \to 0$ as $n \to \infty$, where $\phi(n)$ and $\alpha(n)$ are the mixing rates defined by

$$\phi(n) = \sup_{k \geqslant 1} \sup_{A \in \mathscr{F}_1^k, B \in \mathscr{F}_{k+n}^\infty} |P(B \mid A) - P(B)|$$

$$\alpha(n) = \sup_{k \geqslant 1} \sup_{A \in \mathscr{F}_1^k, B \in \mathscr{F}_{k+n}^\infty} |P(AB) - P(A)P(B)|$$

with $\mathscr{F}_i^j \triangleq \sigma\{x_k,\ i \leqslant k \leqslant j\}$, $1 \leqslant i \leqslant j \leqslant \infty$. Obviously $\alpha(n) \leqslant \phi(n)$ and $\phi(n) \to 0$ is true if $\{\varphi(t)\}$ is M-dependent or is the output of a stable finite-dimensional linear system driven by bounded white noise.

Lemma 2

Let $\{\varphi(t)\}$ be ϕ-mixing with $\sup_t \|\varphi(t)\|_p < \infty$ for some $p > 2$. Then the following conditions are equivalent:

(i) $E\lambda_{\min}(A_h(t)) \geqslant \delta$ $\forall t \geqslant 0$
(ii) $P\{\lambda_{\min}(A_h(t)) \geqslant c\} \geqslant \delta$ $\forall t \geqslant 0$
(iii) $P\{\lambda_{\min}(A_h(t)) \geqslant c \mid \mathscr{F}_t\} \geqslant \delta$ $\forall t \geqslant 0$, $\mathscr{F}_t = \sigma\{\varphi(i),\ i \leqslant t\}$
(iv) $\lambda_{\min}(EA_h(t)) \geqslant \delta$ $\forall t \geqslant 0$

where h, δ and c are positive constants which may differ from condition to condition.

The proof is given in Appendix II. Now, combining Proposition 1, Theorem 1 and Lemma 2, we immediately get the following theorem.

Theorem 2

Let $\{\varphi(t)\}$ be a ϕ-mixing sequence with $\sup_t E\|\varphi(t)\|^{2+\varepsilon}<\infty$ for some $\varepsilon>0$. Then for $p\geqslant 1$ there exists a constant $\mu_0\in(0,1)$ such that $\sup_{\mu\in(0,\mu_0)}\sup_t\|P(t)\|_p<\infty$ if and only if there exist constants $h>0$ and $\delta>0$ such that

$$\sum_{i=t+1}^{t+h} E[\varphi(i)\varphi^{\mathrm{T}}(i)]\geqslant\delta I\quad\forall t\geqslant 0$$

3. TRACKING ERROR BOUNDS

Let us assume that the parameter process $\{\theta(t)\}$ can be modelled by

$$\theta(t)=\theta(t-1)+\gamma w(t),\quad t\geqslant 0,\quad \gamma>0 \tag{12}$$

The purpose of this section is to obtain an upper bound for the tracking error $\|\tilde{\theta}(t)\|_p$, $\tilde{\theta}(t)\triangleq\theta(t)-\hat{\theta}(t)$, which is expected to be of the order $O(\sqrt{\mu}+\gamma/\sqrt{\mu})$ under some standard assumptions. The idea in the analysis is that instead of directly analysing $\|\tilde{\theta}(t)\|_p$, we first obtain upper bounds for $\|R(t)\tilde{\theta}(t)\|_p$, then a combined use of the Holder inequality and Theorem 1 will yield the desired result. For simplicity of discussion we introduce the following definition.

Definition 1

A random process $\{x(t)\}$ is said to belong to the class $\mathcal{M}_p$, $p\geqslant 1$, if there exists a constant $c_p(x)$ depending only on p and $\{x(t)\}$ such that

$$\left\|\sum_{i=m+1}^{m+n} x(i)\right\|_p\leqslant c_p(x)n^{1/2}\quad\forall n\geqslant 1,\quad\forall m\geqslant 0$$

As the next lemma shows, $\mathcal{M}_p$ includes a large class of random processes of interest. Throughout the sequel x_p^* denotes $\sup_{t\geqslant 0}\|x(t)\|_p$.

Lemma 3

Let $p\geqslant 2$ and $x_p^*<\infty$, $Ex(t)=0$. Then $\{x(t)\}\in\mathcal{M}_p$ if either

(i) $\{x(t)\}$ is a martingale difference sequence
(ii) $\{x(t)\}$ is ϕ-mixing with $\Sigma_{t=1}^{\infty}\phi^{1/2}(t)<\infty$
(iii) $\{x(t)\}$ is stationary α-mixing with $x_{p+\delta}^*<\infty$ and $\Sigma_{t=1}^{\infty} t^{p/2-1}[\alpha(t)]^{\delta/(p+\delta)}<\infty$ for some $\delta>0$
(iv) $\{x(t)\}$ is a stationary aperiodic Markov sequence which is Markov ergodic and satisfies Doeblin's condition.

The proof is essentially a collection of known results (see Appendix III). A useful result for processes in $\mathcal{M}_p$ is the following lemma.

Lemma 4

Let $x(t)\in\mathbb{R}^d$ and $Z(t)\in\mathbb{R}^{d\times d}$, $t\geqslant 1$, be two random processes and let $Y(t)=(1-\mu)Y(t-1)+\mu Z(t)$, $t\geqslant 1$. Then for $S(t,k)\triangleq\Sigma_{i=k}^{t} x(i)$ and $\mu\in(0,1)$ the following properties hold:

(i) $\sum_{k=1}^{t}(1-\mu)^{t-k}Y(k)x(k)=(1-\mu)^{t}Y(0)S(t,1)$

$$+\sum_{k=1}^{t}\mu(1-\mu)^{k-1}Z(t-k+1)S(t,t-k+1)$$

(ii) $\left\|\sum_{k=1}^{t}(1-\mu)^{t-k}x(k)\right\|_p\leqslant 2c_p(x)\mu^{-1/2}$, $t\geqslant 1$, provided that $\{x(k)\}\in\mathcal{M}_p$

(iii) $\left\|\sum_{k=1}^{t}(1-\mu)^{t-k}Y(k)x(k)\right\|_p\leqslant c_{2p}(x)\mu^{-1/2}[\|Y(0)\|+Z_{2p}^{*}]$, $t\geqslant 1$, provided that $\{x(k)\}\in\mathcal{M}_{2p}$.

Proof. Assertion (i) is easily verified by summation by parts. Taking $Z(t)\equiv I$, $Y(0)=I$ and applying Minkowski's inequality to (i), we obtain

$$\left\|\sum_{k=1}^{t}(1-\mu)^{t-k}x(k)\right\|_p\leqslant c_p(x)\left((1-\mu)^{t}t^{1/2}+\sum_{k=1}^{t}\mu(1-\mu)^{k-1}k^{1/2}\right)$$

Hence (ii) follows immediately by noticing that $\max_{0\leqslant\mu\leqslant 1}[\sqrt{\mu}(1-\mu)^{t}t^{1/2}]\leqslant 1$ $\forall t\geqslant 1$ and $\Sigma_{k=1}^{\infty}(1-\mu)^{k-1}k=1/\mu^{2}$. Similarly, (iii) follows from (i) by the Minkowski and Schwarz inequalities. □

Theorem 3

For $q>1$, $r>1$ and $1/q+1/r\leqslant 1$ let the following conditions be satisfied:

(i) there exists a constant $\mu_0\in(0,1)$ such that $K_q\triangleq\sup_{\mu\in(0,\mu_0]}\sup_t\|P(t)\|_q<\infty$
(ii) $\{\varphi(t)e(t)\}\in\mathcal{M}_r$
(iii) $\{w(t)-Ew(t)\}\in\mathcal{M}_{2r}$ and $\|\Sigma_{i=k+1}^{n}Ew(i)\|\leqslant\delta(n-k)$ for all $n\geqslant k\geqslant 0$ and for some $\delta>0$.

Then for $p\triangleq(1/q+1/r)^{-1}$ and for all $\mu\in(0,\mu_0]$, $t\geqslant 1$,

$$\|\tilde{\theta}(t)\|_p\leqslant K_q\{(1-\mu)^{t}\|R_0\|\,\|\tilde{\theta}(0)\|_r+2c_r(\varphi,e)\mu^{1/2}$$
$$+\gamma\mu^{-1/2}c_{2r}(w)(\|R_0\|+\varphi_{4r}^{*2})+\delta\mu^{-1}(\|R_0\|+\varphi_{2r}^{*2})\}$$

where the constants $c_r(\varphi,e)$ and $c_{2r}(w)$ are given by conditions (ii) and (iii) (see Definition 1). Moreover, if $\{e(t),\mathscr{F}_t\}$ and $\{w(t),\mathscr{F}_t\}$ are martingale difference sequences with $\mathscr{F}_t\triangleq\sigma\{e(s), w(s), \varphi(s+1), s\leqslant t\}$, then there exists a constant c_r depending only on r such that for all $\mu\in(0,\mu_0]$, $t\geqslant 1$,

$$\|\tilde{\theta}(t)\|_p\leqslant c_rK_q\{(1-\mu)^{t}\|R_0\|\,\|\tilde{\theta}(0)\|_r+\mu^{1/2}e_{2r}^{*}\varphi_{2r}^{*}+\gamma\mu^{-1/2}w_{2r}^{*}(\|R_0\|+\varphi_{4r}^{*2})\}$$

Proof. We need only to prove the first assertion, since from this the second one follows easily. As we noted earlier, the Hölder inequality gives

$$\|\tilde{\theta}(t)\|_p\leqslant\|P(t)\|_q\|R(t)\tilde{\theta}(t)\|_r\leqslant K_q\|R(t)\tilde{\theta}(t)\|_r \tag{13}$$

so we need only to consider $\|R(t)\tilde{\theta}(t)\|_r$. By (1) and (12), from (3) we get the error equation

$$\tilde{\theta}(t)=[I-\mu P(t)\varphi(t)\varphi^{\mathrm{T}}(t)]\tilde{\theta}(t-1)+\mu P(t)\varphi(t)e(t)-\gamma w(t) \tag{14}$$

From this and (7) it follows that

$$R(t)\tilde{\theta}(t)=(1-\mu)R(t-1)\tilde{\theta}(t-1)+\mu\varphi(t)e(t)-\gamma R(t)w(t) \tag{15}$$

or

$$R(t)\tilde{\theta}(t)=(1-\mu)^tR(0)\tilde{\theta}(0)+\sum_{k=1}^{t}(1-\mu)^{t-k}[\mu\varphi(k)e(k)-\gamma R(k)w(k)] \tag{16}$$

By Lemma 4 (ii) and condition (ii)

$$\mu\left\|\sum_{k=1}^{t}(1-\mu)^{t-k}\varphi(k)e(k)\right\|_r\leqslant 2c_r(\varphi,e)\mu^{1/2},\quad t\geqslant 1 \tag{17}$$

By Lemma 4 (iii) and condition (iii)

$$\gamma\left\|\sum_{k=1}^{t}(1-\mu)^{t-k}R(k)[w(k)-Ew(k)]\right\|_r\leqslant\gamma c_{2r}(w)\mu^{-1/2}[\,\|R(0)\|+\varphi_{4r}^{*2}],\quad t\geqslant 1 \tag{18}$$

By Lemma 4 (i) and condition (iii) again

$$\begin{aligned}\gamma\left\|\sum_{k=1}^{t}(1-\mu)^{t-k}R(k)Ew(k)\right\|_r&\leqslant\delta\gamma\left\{(1-\mu)^t t\,\|R(0)\|+\varphi_{2r}^{*2}\sum_{k=1}^{t}\mu(1-\mu)^{k-1}k\right\}\\&\leqslant\delta\gamma\mu^{-1}\{\,\|R(0)\|+\varphi_{2r}^{*2}\},\quad t\geqslant 1\end{aligned} \tag{19}$$

Finally, combining (16)–(19), we get the desired estimate for $\|R(t)\tilde{\theta}(t)\|_r$ and hence the proof is complete.

4. APPROXIMATION OF THE MEAN TRACKING ERRORS

Throughout this section we assume the following condition:

(H) $$E[e(t)\mid\mathscr{F}_{t-1}]=0\quad\text{and}\quad E[w(t)\mid\mathscr{F}_{t-1}]=E[e(t)w(t)\mid\mathscr{F}_{t-1}]=0\quad\text{a.s.} \tag{20}$$

$$E[e^2(t)\mid\mathscr{F}_{t-1}]=R_e(t)\quad\text{and}\quad E[w(t)w^{\mathrm{T}}(t)\mid\mathscr{F}_{t-1}]=R_w(t)\quad\text{a.s.} \tag{21}$$

where $\{w(t)\}$ is defined by (12), $\mathscr{F}_t=\sigma\{w(s),\,e(s),\,\varphi(s+1),\,s\leqslant t\}$ and $R_e(t)$ and $R_w(t)$ are deterministic sequences.

Remark. It may be remarked that (21) is more general than assuming e and w to be independent sequences. Take e.g. $e(t)=\varepsilon(t)\mathrm{sign}\{\varepsilon(t-1)\}$, where $\{\varepsilon(t)\}$ is i.i.d with $P(\varepsilon(t)=2)=\frac{1}{3}$ and $P(\varepsilon(t)=-1)=\frac{2}{3}$. This makes $\{e(t)\}$ a dependent sequence but it is still a martingale difference with deterministic conditional variance.

Set

$$\Pi(t)=E[\tilde{\theta}(t)\tilde{\theta}^{\mathrm{T}}(t)],\qquad S(t)=E[\varphi(t)\varphi^{\mathrm{T}}(t)] \tag{22}$$

Following Ljung and Priouret,[4] we define $\hat{\Pi}(t)$ recursively by $\hat{\Pi}(0)=\Pi(0)$,

$$\hat{\Pi}(t)=\bar{A}(t)\hat{\Pi}(t-1)\bar{A}^{\mathrm{T}}(t)+\mu^2\bar{P}(t)S(t)\bar{P}(t)R_e(t)+\gamma^2R_w(t) \tag{23}$$

where $\bar{A}(t)=I-\mu\bar{P}(t)S(t)$, $\bar{P}(t)=[\bar{R}(t)]^{-1}$ and

$$\bar{R}(t)=(1-\mu)\bar{R}(t-1)+\mu E[\varphi(t)\varphi^{\mathrm{T}}(t)],\qquad \bar{R}(0)=R(0) \tag{24}$$

We will prove that $\Pi(t)$ can be well approximated by $\hat{\Pi}(t)$ for small μ.

Theorem 4

Let, in addition to condition (H), the following conditions hold:

(i) there exists some $\mu_0 \in (0, 1)$ such that $\sup_{\mu \in (0,\mu_0]} \sup_t \| P(t) \|_8 < \infty$
(ii) $(\varphi_{2p}^* + w_p^* + e_p^*) < \infty$ for $p = 16$
(iii) $\{\varphi(t)\varphi^{\mathrm{T}}(t) - E[\varphi(t)\varphi^{\mathrm{T}}(t)\} \in \mathcal{M}_4$

Then there exists a constant $c > 0$ such that for all $\mu \in (0, \mu_0]$ and all $t \geqslant 1$

$$\| \Pi(t) - \hat{\Pi}(t) \| \leqslant c\mu^{1/2} [(1-\mu)^{2t} \| \tilde{\theta}(0) \|_8^2 + \mu + \gamma^2 \mu^{-1}]$$

where $\hat{\Pi}(t)$ is defined by (23).

Remark. Note that condition (i) is implied by more explicit conditions in Theorems 1 and 2. Recall also the definition of $\mathcal{M}_4$ in Definition 1.

Proof. Recursively define

$$\bar{\theta}(t) = [I - \mu\bar{P}(t)S(t)]\bar{\theta}(t-1) + \mu\bar{P}(t)\varphi(t)e(t) - \gamma w(t), \qquad \bar{\theta}(0) = \tilde{\theta}(0) \tag{25}$$

It is immediately verified that $\hat{\Pi}(t) = E[\bar{\theta}(t)\bar{\theta}^{\mathrm{T}}(t)]$ and

$$\bar{R}(t)\bar{\theta}(t) = (1-\mu[\bar{R}(t-1)\bar{\theta}(t-1)] + \mu\varphi(t)e(t) - \gamma\bar{R}(t)w(t) \tag{26}$$

Note that

$$\begin{aligned} \| \Pi(t) - \hat{\Pi}(t) \| &= \| E\{ [\tilde{\theta}(t) - \bar{\theta}(t)]\tilde{\theta}^{\mathrm{T}}(t) + \bar{\theta}(t)[\tilde{\theta}(t) - \bar{\theta}(t)]^{\mathrm{T}} \} \| \\ &\leqslant \| \tilde{\theta}(t) - \bar{\theta}(t) \|_2 [\| \tilde{\theta}(t) \|_2 + \| \bar{\theta}(t) \|_2] \\ &\leqslant [\| \tilde{\theta}(t) - \bar{P}(t)R(t)\tilde{\theta}(t) \|_2 + \| \bar{P}(t)R(t)\tilde{\theta}(t) - \bar{\theta}(t) \|_2] [\| \tilde{\theta}(t) \|_2 + \| \bar{\theta}(t) \|_2] \end{aligned} \tag{27}$$

We proceed to estimate the right-hand side of (27) term-by-term. However, first of all we prove that for $\mu \in (0, 1)$

$$\| \bar{P}(t) \| \leqslant c, \qquad \| \bar{R}(t) - R(t) \|_4 \leqslant c\sqrt{\mu} \tag{28}$$

where and hereafter c denotes a generic constant independent of μ and t. By Proposition 1 condition (i) implies (6), so taking expectations on both sides of (9) and using a similar (but simpler) argument to that in the proof of Theorem 1(i), we conclude the first assertion in (28). By (7) and (24)

$$R(t) - \bar{R}(t) = \mu \sum_{k=1}^{t} (1-\mu)^{t-k} [\varphi(t)\varphi^{\mathrm{T}}(t) - E\varphi(t)\varphi^{\mathrm{T}}(t)]$$

Hence by condition (iii) and Lemma 4 (ii) the second assertion in (28) is also true. We now turn to consider the terms in (27). Note that

$$\| \bar{\theta}(t) \|_2 \leqslant \| \bar{P}(t) \| \, \| \bar{R}(t)\bar{\theta}(t) \|_2$$

Noting also the similarity between equations (26) and (15), we can use a similar method to that for Theorem 3 to get upper bounds for $\| \bar{R}(t)\bar{\theta}(t) \|_2$ and then, by noting (28),

$$\| \bar{\theta}(t) \|_2 \leqslant c\{ (1-\mu)^t \| \tilde{\theta}(0) \|_4 + \sqrt{\mu} + \gamma/\sqrt{\mu} \} \tag{29}$$

By (28) again

$$\|\tilde{\theta}(t)-\bar{P}(t)R(t)\tilde{\theta}(t)\|_2 \leqslant \|\bar{P}(t)\|\,\|[\bar{R}(t)-R(t)]\tilde{\theta}(t)\|_2 \leqslant c\|\bar{R}(t)-R(t)\|_4\|\tilde{\theta}(t)\|_4 \leqslant c\sqrt{\mu}\,\|\tilde{\theta}(t)\|_4 \quad (30)$$

By (15), (26) and Lemma 4 (ii)

$$\begin{aligned}\|\bar{P}(t)R(t)\bar{\theta}(t)-\bar{\theta}(t)\|_2 &\leqslant \|\bar{P}(t)\|\,\|R(t)\bar{\theta}(t)-\bar{R}(t)\bar{\theta}(t)\|_2\\ &\leqslant c\left\|\sum_{k=1}^{t}(1-\mu)^{t-k}\gamma[R(t)-\bar{R}(t)]\,w(t)\right\|_2\\ &\leqslant c\gamma\mu^{-1/2}\sup_{t\geqslant 0}\|R(t)-\bar{R}(t)\|_4 \leqslant c\gamma \end{aligned} \quad (31)$$

Substituting (29)–(31) into (27) and applying Theorem 3, we finally get the desired results. □

Next we show that if $S(t)$, $R_e(t)$ and $R_w(t)$ are time-invariant, then $\hat{\Pi}(t)$ can be simplified so that $\Pi(t)$ can be approximated by a simple function of μ.

Theorem 5

If, in addition to the conditions of Theorem 4, $S(t)=S$, $R_e(t)=R_e$ and $R_w(t)=R_w$ $\forall t\geqslant 0$, then there exists a constant $c>0$ such that for all $t\geqslant 1$ and all $\mu\in(0,\mu_0]$

$$\|\Pi(t)-\tfrac{1}{2}(\mu S^{-1}R_e+\mu^{-1}\gamma^2R_w)\| \leqslant c\{(1-\mu)^{2t}\|\tilde{\theta}(0)\|_8^2+[(1-\mu)^t+\sqrt{\mu}](\mu+\gamma^2\mu^{-1})\}$$

Proof. First of all, condition (i) of Theorem 4 implies (6) by Proposition 1, which in turn implies that $\inf_t \lambda_{\min}(EA_h(t))>0$ or $\inf_t \lambda_{\min}(hS)>0$ and hence $S>0$. Note that $I-\mu\bar{P}(t)S(t)=(1-\mu)[\bar{R}(t)]^{-1}\bar{R}(t-1)$, so setting $\hat{\Pi}_1(t)=\bar{R}(t)\hat{\Pi}(t)\bar{R}(t)$, we see from (23) that

$$\hat{\Pi}_1(t)=(1-\mu)^2\hat{\Pi}_1(t-1)+\mu^2SR_e+\gamma^2\bar{R}(t)R_w\bar{R}(t) \quad (32)$$

By (24) it is easy to see that

$$\|\bar{R}(t)-S\| \leqslant (1-\mu)^t\|R_0-S\| \quad (33)$$

and so by (28)

$$\|\bar{P}(t)-S^{-1}\| \leqslant c(1-\mu)^t \quad (34)$$

Hence (32) can be rewritten as

$$\hat{\Pi}_1(t)=(1-\mu)^2\hat{\Pi}_1(t-1)+\mu^2SR_e+\gamma^2SR_wS+O(\gamma^2(1-\mu)^t)$$

Consequently,

$$\hat{\Pi}_1(t)=\mu^{-1}(2-\mu)^{-1}(\mu^2SR_e+\gamma^2SR_wS) + O((1-\mu)^{2t}[\,\|\tilde{\theta}(0)\|_2^2+\mu+\gamma^2\mu^{-1}])+O((1-\mu)^t\gamma^2\mu^{-1}) \quad (35)$$

Note that

$$\mu^{-1}(2-\mu)^{-1}(\mu^2SR_e+\gamma^2SR_wS)=\tfrac{1}{2}(\mu SR_e+\gamma^2\mu^{-1}SR_wS)+O(\mu^2+\gamma^2) \quad (36)$$

Substituting this into (35) and using (33) and (34) again, we have

$$\hat{\Pi}_1(t) = \bar{R}(t)\,[\tfrac{1}{2}\,(\mu S^{-1}R_e + \mu^{-1}\gamma^2 R_w)]\,\bar{R}(t) + O(\mu^2+\gamma^2) + O((1-\mu)^{2t}\,\|\,\tilde{\theta}(0)\,\|_2^2) + O((1-\mu)^t(\mu+\gamma^2\mu^{-1}))$$

Hence by the boundedness of $\|\,\bar{P}(t)\,\|$

$$\begin{aligned} &\|\,\hat{\Pi}(t) - \tfrac{1}{2}\,(\mu S^{-1}R_e + \mu^{-1}\gamma^2 R_w)\,\| \\ &\leqslant \|\,\bar{P}(t)\,\|^2\,\|\,\hat{\Pi}_1(t) - \tfrac{1}{2}\,\bar{R}(t)(\mu S^{-1}R_e + \mu^{-1}\gamma^2 R_w)\bar{R}(t)\,\| \\ &\leqslant O((1-\mu)^{2t}\,\|\,\tilde{\theta}(0)\,\|_2^2) + O((1-\mu)^t(\mu+\gamma^2\mu^{-1})) + O(\mu^2+\gamma^2) \end{aligned}$$

Hence Theorem 5 follows from this and Theorem 4 immediately. □

5. CONCLUSIONS

The main result of this contribution is Theorem 4 (in conjunction with Theorem 1 for condition (i)).

The fact that the approximating expression $\hat{\Pi}(t)$ obeys a simple equation gives us a good handle on the true error covariance matrix $\Pi(t)$. A special feature of the result is that it is generally valid over unlimited time intervals, including the transient. Also, it is valid for all γ and for all μ in an interval and the constant c can be computed from the properties of e, φ and w.

It is of course true that the interest of the result relates to small values of μ (but not necessarily of γ, since the relative error decays as $\sqrt{\mu}$).

The result thus parallels and extends those of Reference 4. For non-gradient methods Reference 4 has an assumption that $\|\,\tilde{\theta}\,\|_4$ be bounded by $C\|\,\tilde{\theta}\,\|_2$. In this paper we have removed that annoying assumption for the RLS algorithm. It would be highly desirable to be able to do the same for general adaptation schemes.

APPENDIX I: PROOF OF LEMMA 1

Take a constant $A > h/\delta c$ to define a function $f(\mu)$ as $f(\mu) = (1-\mu)^{-hp}[1-\delta+\delta(1+\mu Ac)^{-p}]$. It is easy to verify that $f(0) = 1$ and $f'(0) = p(h-\delta Ac) < 0$ and hence there exists a constant $\mu_1 \in (0, \mu_0)$ small enough such that $f(\mu) \in (0,1)\ \forall \mu \in (0, \mu_1]$. Now set $z_{k+1} = (1+\mu v_n w_{n+1})^{-p}$. Then on the set $\{v_n \geqslant A\}$ we have

$$\begin{aligned} z_{n+1} &\leqslant (1+\mu A w_{n+1})^{-p} \leqslant (1+\mu Ac)^{-p} I(w_{n+1} \geqslant c) + I(w_{n+1} < c) \\ &\leqslant (1+\mu Ac)^{-p} + [1-(1+\mu Ac)^{-p}]\, I(w_{n+1} < c) \end{aligned}$$

Hence by the assumption we have

$$E[z_{n+1} \mid \mathscr{F}_n] \leqslant 1-\delta+\delta(1+\mu Ac)^{-p} \quad \text{on } \{v_n \geqslant A\} \tag{37}$$

Consequently, we have

$$\begin{aligned} E[v_{n+1}^p \mid \mathscr{F}_n] &= E[v_{n+1}^p \mid \mathscr{F}_n]\,[I(v_n \geqslant A) + I(v_n < A)] \\ &\leqslant f(\mu) v_n^p I(v_n \geqslant A) + (1-\mu)^{-hp} v_n^p I(v_n < A) \\ &\leqslant f(\mu) v_n^p + [(1-\mu)^{-hp} - f(\mu)]\, v_n^p I(v_n < A) \end{aligned} \tag{38}$$

Taking expectations, we see that

$$E(v_{n+1}^p) \leqslant f(\mu) E(v_n^p) + A_p[(1-\mu)^{-hp} - f(\mu)]$$

and consequently, since $f(\mu) < 1$, $\mu \in (0, \mu_1]$,

$$E(v_n^p) \leqslant [f(\mu)]^n E(\| R_0 \|^{-p}) + A \frac{(1-\mu)^{-hp} - f(\mu)}{1 - f(\mu)}$$

The last term is bounded for $\mu \in (0, \mu_1]$, since by L'Hospital's rule

$$\lim_{\mu \to 0} \frac{(1-\mu)^{-hp} - f(\mu)}{1 - f(\mu)} = \frac{hp - f'(0)}{-f'(0)}$$

Hence we have proved that

$$\sup_{\mu \in (0, \mu_1]} \sup_{n \geqslant 0} \| v_n \|_p < \infty \tag{39}$$

Next we consider the case where $\mu \in (\mu_1, \mu_0]$. Note that

$$v_{n+1} \leqslant \frac{v_n}{(1-\mu)^h} I(w_{n+1} < c) + \frac{v_n}{(1-\mu)^h (1 + \mu c v_n)} I(w_{n+1} \geqslant c)$$

from which it follows that

$$\| v_{n+1} \|_p \leqslant \frac{(1-\delta)^{1/p}}{(1-\mu)^h} \| v_n \|_p + \frac{1}{(1-\mu)^h \mu c}$$

Hence it is obvious that

$$\sup_{\mu \in (\mu_1, \mu_0]} \sup_{n \geqslant 0} \| v_n \|_p < \infty$$

This in conjunction with (39) completes the proof of Lemma 1. □

APPENDIX II: PROOF OF LEMMA 2

(i) ⇒ (ii). Denote $A_t = \{\lambda_{\min}(A_h(t)) \geqslant \delta/2\}$. Then (ii) follows by observing that

$$\begin{aligned} \delta \leqslant E\lambda_{\min}(A_h(t)) &\leqslant \delta/2 + E[\lambda_{\min}(A_h(t)) I(A_t)] \\ &\leqslant \delta/2 + \| A_h(t) \|_{p/2} [P(A_t)]^{(p-2)/p} \quad \text{(by the Holder inequality)} \end{aligned}$$

(ii) ⇒ (iii). The proof is similar to the proof of Lemma 2.3 in Reference 8.

(iii) ⇒ (iv). The proof is trivial, since (iii) ⇒ (ii) ⇒ (i) ⇒ (iv) is obvious.

To complete the proof, we have to show that (iv) ⇒ (i). By the Holder-type inequality for a ϕ-mixing sequence (e.g. Reference 10, p. 278) it is easy to show (via a direct calculation) that $\sup_{t \geqslant 0} \| A_m(t) - EA_m(t) \|_2 = o(m)$ as $m \to \infty$. However, by (iv) it is easy to convince oneself that $\inf_{t \geqslant 0} \lambda_{\min}(EA_m(t)) \geqslant cm$ for all $m > h$ and some $c > 0$. Consequently, for suitably large m we have

$$\inf_{t \geqslant 0} E\lambda_{\min}(A_m(t)) \geqslant \inf_{t \geqslant 0} \{\lambda_{\min}(EA_m(t)) - \| A_m(t) - EA_m(t) \|_2\} > 0$$

Hence (i) holds and the proof is complete.

APPENDIX III: PROOF OF LEMMA 3

(i) By first applying Burkholder's inequality (e.g. Reference 10, p. 23) and then the C_r-inequality, we know that $\{x(i)\} \in \mathcal{M}_p$ with $c_p(x) = c_p x_p^*$ for some constant c_p depending only on p. (ii) By the Holder-type inequality (e.g. Reference 10, p. 278) it is directly verified that $\{x(i)\} \in \mathcal{M}_2$ with $c_2(x) = c_2 x_2^*$ for some constant $c_2 > 0$. Hence from Lemma 3.2 of Reference 11 the assertion (ii) follows. (iii) According to Yoko Yama,[12] this holds with $c_p(x) = c_p x_{p+\delta}^*$. Finally, (iv) can be found in Reference 13, p. 225.

REFERENCES

1. Ljung, L., and T. Söderström, *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, MA, 1983.

2. Benveniste, A., 'Design of adaptive algorithms for the tracking of time-varying systems', *Int. j. adapt. control signal process.*, **1**, 1–29 (1987).
3. Macchi, O., and E. Eweda, 'Compared speed and accuracy of RLS and LMS algorithms with constant forgetting factors', *APII*, **22**, 255–267 (1988).
4. Ljung, L., and P. Priouret, 'A result on the mean square error obtained using general tracking algorithms', *Int. j. adapt control signal process.*, **5**, 231–250 (1991).
5. Bittanti, S., and M. Campi, 'Adaptive RLS algorithms under stochastic excitation—L^2 convergence analysis', *IEEE Trans. Automatic Control*, **AC-36**, 963–967 (1991).
6. Bittanti, S., and M. Campi, 'Tuning the forgetting factor in RLS identification algorithms', *Proc. 30th IEEE Conf. on Decision and Control*, Brighton, 1991, IEEE, New York, 199X, pp. 1688–1689.
7. Niedźwiecki, M., and L. Guo, 'Nonasymptotic results for finite-memory WLS filters', *IEEE Trans. Automatic Control*, **AC-36**, 198–206 (1991).
8. Guo, L., 'Stability of recursive stochastic tracking algorithms', *SIAM J. Control Optim.*, (1993), in press.
9. Green, M., and J. B. Moore, 'Persistency of excitation in linear systems', *Syst. Control Lett.*, **7**, 351–360 (1986)
10. Hall, P., and C. C. Heyde, *Martingale Limit Theory and Its Applications*, Academic, New York, 1980.
11. Shao, Q. M., and C. L. Lu, 'Strong approximation for partial sums of weakly dependent random variables', *Sci. Sinica (Ser. A)*, **XXX**, (6), 575–587 (1987)
12. Yoko Yama, R., 'Moment bounds for stationary mixing sequences', *Z. Wahrsch. Verw. Gebiete*, **52**, 45–57 (1980).
13. Doob, J. L., *Stochastic Processes*, Wiley, New York, 1953.