On Almost-Sure Bounds for the LMS Algorithm

Michael A. Kouritzin

***Abstract*—Almost-sure (a.s.) bounds for linear, constant-gain, adaptive filtering algorithms are investigated. For instance, under general pseudo-stationarity and dependence conditions on the driving data $\{\psi_k, k = 1,2,3,\ldots\}$, $\{Y_k, k = 0,1,2,\ldots\}$ a.s. convergence and rates of a.s. convergence (as the algorithm gain $\epsilon \to 0$) are established for the LMS algorithm**

$$h_{k+1}^\epsilon = h_k^\epsilon + \epsilon Y_k(\psi_{k+1} - Y_k^T h_k^\epsilon)$$

subject to some nonrandom initial condition $h_0^\epsilon = h_0$. In particular, defining $\{g_k^\epsilon\}_{k=0}^\infty$ by $g_0^\epsilon = h_0$ and

$$g_{k+1}^\epsilon = g_k^\epsilon + \epsilon\Big(E[Y_k\psi_{k+1}] - E[Y_kY_k^T]g_k^\epsilon\Big) \quad \text{for } k = 0,1,2,\ldots,$$

we show that for any $\gamma > 0$

$$\max_{0\le k\le \gamma\epsilon^{-1}} |h_k^\epsilon - g_k^\epsilon| \to 0 \text{ as } \epsilon \to 0 \text{ a.s.}$$

and under a stronger dependency condition, we show that for any $0 < \zeta \le 1$ and $\gamma > 0$,

$$\max_{0\le k\le \gamma\epsilon^{-\zeta}} |h_k^\epsilon - g_k^\epsilon|$$

converges (as $\epsilon \to 0$) a.s. at a rate marginally slower than $O((\epsilon^{2-\zeta}\log\log(\epsilon^{-\zeta}))^{\frac{1}{2}})$. Then, under a stronger pseudo-stationarity assumption it is shown that similar results hold if the sequences $\{g_k^\epsilon\}_{k=0}^\infty, \epsilon > 0$ in the above results are replaced with the solution $g^0(\cdot)$ of a nonrandom linear ordinary differential equation, i.e. we have

$$\max_{0\le k\le \lfloor \gamma\epsilon^{-\zeta}\rfloor} |h_k^\epsilon - g^o(\epsilon k)| \to 0 \text{ as } \epsilon \to 0 \text{ a.s.},$$

where we can attach a rate to this convergence under the stronger dependency condition.

The almost-sure bounds contained in this paper complement previously developed weak convergence results in Kushner and Shwartz [IEEE Trans. Information Theory, IT-30(2), 177–182, 1984] and, as will be seen, are "near optimal". Moreover, the proofs used to establish these bounds are quite elementary.

***Index Terms*—Adaptive filtering, almost-sure bounds, method of averaging.**

I. Introduction

As mentioned in the editorial of Macchi [16], adaptive filtering, interpreted in the general sense of creating a stochastic approximation to estimate some nonrandom, optimal sequence of parameters, has wide-ranging and multifaceted applications in engineering. Suppose that $\{\psi_k,\ k = 1,2,3,\ldots\}$ and $\{Y_k,\ k = 0,1,2,\ldots\}$ are, respectively, $\Re$-and $\Re^d$-valued, second-order stochastic processes defined on a common probability space. A basic problem of adaptive filtering is to find some best mean-square linear approximation to ψ_{k+1} in terms of the components of Y_k, i.e. find a deterministic sequence, $\{f_k\}_{k=0}^\infty$, in $\Re^d$, which minimizes

$$E\{\psi_{k+1} - Y_k^T f_k\}^2 \quad \text{for } k = 0,1,2,\ldots \tag{1.1}$$

If $E(Y_kY_k^T)$ is nonsingular for each $k \ge 0$, then it is immediately apparent that $\{f_k\}_{k=0}^\infty$ is uniquely defined by

$$f_k \triangleq (E[Y_kY_k^T])^{-1}E[\psi_{k+1}Y_k] \quad \text{for } k = 0,1,2,\ldots \tag{1.2}$$

However, in practice $\{E[Y_kY_k^T]\}_{k=0}^\infty$ and $\{E[\psi_{k+1}Y_k]\}_{k=0}^\infty$ often are not readily discernible so (1.2) is of no direct use. Consequently, stochastic estimates of $\{f_k\}_{k=0}^\infty$ generated by adaptive algorithms, with either decreasing or constant gain, must suffice. The linear, decreasing-gain algorithm

$$h_{k+1} = h_k + \mu_k Y_k(\psi_{k+1} - Y_k^T h_k), \tag{1.3}$$

where $\{\mu_k\}_{k=0}^\infty$ is a sequence of real numbers converging to zero as $k \to \infty$ and $\{h_k(\omega),\ k = 0,1,2,\ldots\}$ forms our parameter estimates, is well-suited for ascertaining the minimum of (1.1) when f_k is independent of k. This algorithm has been studied extensively in the almost-sure case by, e.g., Eweda and Macchi [6] and Heunis [11]. On the other hand, it is well-noted that constant-gain adaptive-filtering algorithms provide estimates of $\{f_k\}_{k=0}^\infty$ with desirable tracking ability when $\{f_k\}_{k=0}^\infty$ fluctuates with time; however, few almost-sure results have been developed for the constant-gain version of (1.3).

In this paper we examine the linear stochastic recursion arising from basic linear constant-gain adaptive-filtering algorithms such as the following LMS algorithm:

$$h_{k+1}^\epsilon = h_k^\epsilon + \epsilon Y_k(\psi_{k+1} - Y_k^T h_k^\epsilon), \tag{1.4}$$

where the gain $\epsilon > 0$ is a small constant, and $\{\psi_k,\ k = 1,2,3,\ldots\}$ and $\{Y_k,\ k = 0,1,2,\ldots\}$ are processes defined as above which "drive" the algorithm. The limiting properties (as $\epsilon \to 0$) of the stochastic processes $\{h_k^\epsilon,\ k = 0,1,2,\ldots\}$, $\epsilon > 0$ have been investigated by Kushner and Shwartz in [12] using the theory of weak convergence of probability measures. Under reasonably general conditions on the driving processes $\{\psi_k,\ k = 1,2,3,\ldots\}$ and $\{Y_k,\ k = 0,1,2,\ldots\}$, we complement some of the asymptotic limits achieved in [12] with *almost-sure* bounds on the deviation between h_k^ϵ and a nonrandom recursion, g_k^ϵ, obtained from

$$g_{k+1}^\epsilon = g_k^\epsilon + \epsilon(E(Y_k\psi_{k+1}) - E(Y_kY_k^T)g_k^\epsilon)\ . \tag{1.5}$$

Manuscript received November 17, 1992; revised March 25, 1993.

The author was with the Institut für Mathematische Stochastik, Universität Freiburg, Hebelstraße 27, D-7800 Freiburg, Germany. He is now with the Department of Mathematics and Statistics, Carleton University, Ottawa, Ont. K1S 5B6, Canada.

IEEE Log Number 9400307.

Then, we assume (limiting the possible fluctuations in $\{f_k\}_{k=0}^{\infty}$) boundedness of $\{EY_l\psi_{l+1}\}_{l=0}^{\infty}$ and $\{EY_lY_l^T\}_{l=0}^{\infty}$ and existence of a matrix A and a vector b such that

$$A - \frac{1}{N}\sum_{l=0}^{N-1} E(Y_lY_l^T) = O(N^{-1}),$$
$$b - \frac{1}{N}\sum_{l=0}^{N-1} E(Y_l\psi_{l+1}) = O(N^{-1}), \qquad (1.6)$$

and we use the method of averaging to obtain a reasonably tight a.s. bound on the deviation between h_k^ϵ and $g^o(\epsilon k)$, where $g^o(\tau)$, $0 \le \tau \le 1$ is defined by the linear ordinary differential equation

$$\dot{g}^o(\tau) = -Ag^o(\tau) + b\,. \qquad (1.7)$$

Under mild stationary conditions, the nonrandom trajectory $\{g_k^\epsilon\}_{k=0}^{\infty}$ will tend towards and then track the trajectory $\{f_k\}_{k=0}^{\infty}$ provided $\epsilon > 0$ is chosen appropriately (see, for example, Solo [19] on the difficulty of choosing ϵ). However, unlike decreasing-gain adaptive algorithms, the effect of the driving stochastic processes on the right hand side of (1.4) does not diminish in time and, regardless of the value of $\epsilon > 0$, $\{h_k^\epsilon,\ k = 0,1,2,\ldots\}$ will eventually experience excursions away from the sequence $\{f_k\}_{k=0}^{\infty}$ with probability one. Still, large excursions should occur less frequently when ϵ is small and the almost-sure bounds between $\{h_k^\epsilon,\ k = 0,1,2,\ldots\}$ and $\{g_k^\epsilon\}_{k=0}^{\infty}$ mentioned above can be proven over time frames such as $0 \le k \le \gamma\epsilon^{-\zeta}$ for any $0 < \zeta \le 1, \gamma > 0$. Such results, established in this paper, might be used to show that $\{h_k^\epsilon,\ k = 0,1,2,\ldots\}$ must initially tend towards the optimal sequence $\{f_k\}_{k=0}^{\infty}$ with probability one and to provide further direction about the choice of ϵ.

In Section II we provide the regularity conditions which will be imposed, state the main results of this paper (Proposition 2.1 and Proposition 2.2) and provide some examples which indicate the scope of these regularity conditions. Section 3 contains the proofs of Propositions 2.1 and 2.2, and Section IV is a short discussion on the significance of these results. Finally, Appendix A is a collection of technical results required for the proofs in Section III. We mention that, in order to economize on notation and include algorithms other than the LMS algorithm given in (1.4), we generalize (1.4) and (1.5) slightly to the form

$$h_{k+1}^\epsilon = h_k^\epsilon + \epsilon(b_k - A_kh_k^\epsilon) \qquad (1.8)$$

and

$$g_{k+1}^\epsilon = g_k^\epsilon + \epsilon(Eb_k - E[A_k]\ g_k^\epsilon), \qquad (1.9)$$

where for some integer $M \ge 1$ and all $k = 0,1,2,\ldots$

$$A_k \triangleq \frac{1}{M}\sum_{l=\max\{k-M+1,0\}}^{k} Y_lY_l^T \qquad (1.10)$$

and

$$b_k \triangleq \frac{1}{M}\sum_{l=\max\{k-M+1,0\}}^{k} \psi_{l+1}Y_l\,. \qquad (1.11)$$

Clearly, the class of algorithms defined by (1.8), (1.10), and (1.11) includes the LMS algorithm, as well as other adaptive-filtering algorithms, discussed for example in [7]. All regularity conditions will be given with reference to the vector- and matrix-valued random processes $\{b_k,\ k = 0,1,2,\ldots\}$ and $\{A_k,\ k = 0,1,2,\ldots\}$ respectively so this whole class of adaptive filtering algorithms can be studied at once.

II. Main Result

Suppose that $(\Omega, \mathcal{F}, P)$ is a probability space on which an $\Re^{d\times d}$-valued stochastic process $\{A_l(\omega),\ l = 0,1,2,\ldots\}$ and an $\Re^d$-valued stochastic process $\{b_l(\omega),\ l = 0,1,2,\ldots\}$ are defined. Furthermore, suppose that EA_l and Eb_l are well defined for each $l = 0,1,2,\ldots$ Then, the following conditions on $\{A_l,\ l = 0,1,2,\ldots\}$ and $\{b_l,\ l = 0,1,2,\ldots\}$ will be assumed for Proposition 2.1 (the first principal result of this paper):

(C0) $A_l(\omega)$ is symmetric and positive semi-definite for each $\omega \in \Omega$ and $l = 0,1,2,\ldots$

(C1) For each pair of integers $1 \le s,t \le d$, the $(s,t)^{\text{th}}$ component of $\tilde{A}_l \triangleq A_l - EA_l$ and the s^{th} component of $\tilde{b}_l \triangleq b_l - Eb_l$ satisfy the following moment bounds:

$$\text{(i)}\quad E\left|\sum_{l=p}^{q}\tilde{a}_l^{(s,t)}\right|^{2m} \le c_m(q-p+1)^m,$$
$$\text{(ii)}\quad E\left|\sum_{l=p}^{q}\tilde{b}_l^{(s)}\right|^{2m} \le c_m(q-p+1)^m,$$

for some real constants $m \ge 1, c_m > 0$ and all integers $0 \le p \le q < \infty$.

(C2) There is a constant $c' > 0$ such that

$$\text{(i)}\quad \limsup_{N\to\infty}\frac{1}{N}\sum_{l=0}^{N-1}|||EA_l||| < c',$$
$$\text{(ii)}\quad \limsup_{N\to\infty}\frac{1}{N}\sum_{l=0}^{N-1}|Eb_l| < c'$$

where $|x|$ denotes the Euclidean norm of d-dimensional vector x and $|||A||| \triangleq \sup_{y\in\Re^d}\frac{|Ay|}{|y|}$ is the corresponding matrix norm of d by d matrix A.

In preparation for the statement and development of the main results we define, for each $\epsilon > 0$, a stochastic process $\{h_k^\epsilon, k = 0,1,2,\ldots\}$ on $(\Omega, \mathcal{F}, P)$ and a nonrandom sequence $\{g_k^\epsilon\}_{k=0}^{\infty}$ by

$$h_{k+1}^\epsilon(\omega) = h_k^\epsilon(\omega) + \epsilon(b_k(\omega) - A_k(\omega)h_k^\epsilon(\omega))$$
$$\text{for}\quad k = 0,1,2,\cdots,\omega\in\Omega, \qquad (2.1)$$

and

$$g_{k+1}^\epsilon = g_k^\epsilon + \epsilon(Eb_k - E[A_k]\ g_k^\epsilon)$$
$$\text{for } k = 0,1,2,\ldots, \qquad (2.2)$$

subject to $g_0^\epsilon = h_0^\epsilon(\omega) = h_0$, a fixed nonrandom vector. Moreover, we use the notation that $\log_0(x) = x$ and $\log_k(x) =$

$\log(\log_{k-1}(x))$ for all large enough $x > 0$ and all integers $k \geq 1$ (for simplicity we take the log function to be such that $2^{\log(x)} = x$). Also, $\lfloor x \rfloor$ denotes the largest integer not larger than x for all $x > 0$.

Remark 2.1: By (2.2) and Condition (C2) it follows that for any $\gamma > 0$ and $0 < \zeta \leq 1$ there is a constant $c_{\gamma,\zeta} > 0$ such that

$$\begin{aligned}|g_k^\epsilon| &\leq \left|\left|\left|\prod_{l=0}^{k-1}(I - \epsilon EA_l)\right|\right|\right| |g_0^\epsilon| \\ &\quad + \sum_{j=0}^{k-1}\left|\left|\left|\prod_{l=j+1}^{k-1}(I - \epsilon EA_l)\right|\right|\right| \epsilon|Eb_j| \\ &\leq \prod_{l=0}^{k-1}(1 + \epsilon|||EA_l|||)\,|h_0| \\ &\quad + \sum_{j=0}^{k-1}\left(\prod_{l=j+1}^{k-1}(1 + \epsilon|||EA_l|||)\right)\epsilon|Eb_j| \\ &\leq \exp\left\{\epsilon\sum_{l=0}^{k-1}|||EA_l|||\right\}\left\{|h_0| + \epsilon\sum_{j=0}^{k-1}|Eb_j|\right\} < c_{\gamma,\zeta}\end{aligned} \tag{2.3}$$

for all $0 \leq k \leq \lfloor\gamma\epsilon^{-\zeta}\rfloor$, $0 < \epsilon \leq \gamma^{\frac{1}{\zeta}}$ so using (C2) again, we can define

$$D_{\gamma,\zeta} \triangleq \sup_{0<\epsilon\leq\gamma^{1/\zeta}} \max_{1\leq N\leq\lfloor\gamma\epsilon^{-\zeta}\rfloor} \frac{1}{N}\sum_{k=0}^{N-1}|Eb_k - EA_k\, g_k^\epsilon| < \infty. \tag{2.4}$$

We now give our first main result which is stated in terms of a nondecreasing sequence $\{\psi(l)\}_{l=2}^\infty$ to be explained following the statement of the proposition. The function $\omega \to L(\omega)$ in Proposition 2.1 will depend on $\psi(\cdot), m, \gamma$ and ζ but not ϵ. Thus, the following result is a rate of almost-sure convergence in terms of algorithm gain $\epsilon > 0$.

Proposition 2.1: Suppose $\{\psi(l)\}_{l=2}^\infty$ is a nondecreasing, positive sequence such that $\sum_{i=1}^\infty \frac{1}{\psi(2^i)} < \infty$. Then, under Conditions (C0), (C1) and (C2), given any $\gamma > 0$ and $0 < \zeta \leq 1$ there exists a function $\omega \to L(\omega)$ almost-surely finite such that for each $0 < \epsilon \leq \frac{1}{2}$:

(i) $\displaystyle\max_{0\leq k\leq\lfloor\gamma\epsilon^{-\zeta}\rfloor}|h_k^\epsilon - g_k^\epsilon| \leq L\epsilon^{1-\frac{\zeta}{2}}\log(\epsilon^{-\zeta})(\psi(\epsilon^{-\zeta}))^{\frac{1}{2}}$ if $m = 1$ or

(ii) $\displaystyle\max_{0\leq k\leq\lfloor\gamma\epsilon^{-\zeta}\rfloor}|h_k^\epsilon - g_k^\epsilon| \leq L\epsilon^{1-\frac{\zeta}{2}}(\psi(\epsilon^{-\zeta}))^{\frac{1}{2m}}$ if $m > 1$,

where $m \geq 1$ is the constant of (C1).

Notice that larger values of m in Condition (C1) diminish the effect of the asymptotically (as $\epsilon \to 0$) diverging term $(\psi(\epsilon^{-\zeta}))^{\frac{1}{2m}}$ and thereby produce sharper almost sure bounds on the right hand side of (ii) for small values of ϵ.

Example 2.1: Suppose κ is a large positive integer and σ is a small positive real number. Then, defining $\epsilon_0 \leq \frac{1}{2}$ to be a real number such that $\log_\kappa(\epsilon_0^{-\zeta}) \geq 1$ and

$$\psi(x) \triangleq \begin{cases}1 & x < \epsilon_0\\ \log(x)\,\log_2(x)\cdots\log_{\kappa-1}(x)(\log_\kappa(x))^{1+\sigma} & x \geq \epsilon_0,\end{cases} \tag{2.5}$$

we obtain the following bound from Proposition 2.1:

(i) $\displaystyle\max_{0\leq k\leq\lfloor\gamma\epsilon^{-\zeta}\rfloor}|h_k^\epsilon - g_k^\epsilon| \leq L\epsilon^{1-\frac{\zeta}{2}}((\log(\epsilon^{-\zeta}))^3\log_2(\epsilon^{-\zeta})\cdots\log_{\kappa-1}(\epsilon^{-\zeta}))^{\frac{1}{2}}(\log_\kappa\epsilon^{-\zeta})^{\frac{1+\sigma}{2}}$ if $m = 1$ or

(ii) $\displaystyle\max_{0\leq k\leq\lfloor\gamma\epsilon^{-\zeta}\rfloor}|h_k^\epsilon - g_k^\epsilon| \leq L\epsilon^{1-\frac{\zeta}{2}}(\log(\epsilon^{-\zeta})\log_2(\epsilon^{-\zeta})\cdots\log_{\kappa-1}(\epsilon^{-\zeta}))^{\frac{1}{2m}}(\log_\kappa(\epsilon^{-\zeta}))^{\frac{1+\sigma}{2m}}$ if $m > 1$,

for all $0 < \epsilon \leq \epsilon_0$. On the other hand, in the simple case where $d = 1, A_k(\omega) \equiv 0$ for all k and ω and $\{b_k, k = 0,1,2,\ldots\}$ is an i.i.d. sequence such that $Eb_1^2 < \infty$, one obtains from (2.1), (2.2) and Strassen's functional law of the iterated logarithm (see [21, Theorem 3]) that

$$\begin{aligned}\max_{0\leq k\leq\lfloor\gamma\epsilon^{-\zeta}\rfloor}|h_k^\epsilon - g_k^\epsilon| &= \epsilon\max_{0\leq k\leq\lfloor\gamma\epsilon^{-\zeta}\rfloor}\left|\sum_{l=0}^{k-1}(b_l - Eb_l)\right| \\ &= O(\epsilon^{1-\frac{\zeta}{2}}(\log_2(\epsilon^{-\zeta}))^{\frac{1}{2}})\end{aligned} \tag{2.6}$$

and no more-accelerated rate of convergence is possible. Hence in our more general setting with our modest conditions (see below) we obtain rates of convergence close to those known to be optimal in the simpler setting.

One sees from Conditions (C1) and Jensen's inequality that smaller values of $m \geq 1$ constitute a less stringent condition but from Proposition 2.1 small values of m also provide a looser a.s. bound for small enough ϵ. In fact, it is possible to obtain a.s. convergence of

$$\max_{0\leq k\leq\lfloor\gamma\epsilon^{-1}\rfloor}|h_k^\epsilon(\omega) - g_k^\epsilon|$$

without the rates of convergence in Proposition 2.1 under the milder version of (C1):

(C1′) For each set of integers $1 \leq s,\ t \leq d$ and $0 \leq p < q < \infty$, suppose

(i) $$E\left|\sum_{l=p}^q \tilde{a}_l^{(s,t)}\right|^2 \leq (q-p+1)f(q-p+1),$$

(ii) $$E\left|\sum_{l=p}^q \tilde{b}_l^{(s)}\right|^2 \leq (q-p+1)f(q-p+1),$$

where $\{f(l)\}_{l=1}^\infty$ is any positive-valued, nondecreasing sequence such that $\sum_{i=0}^\infty \frac{f(2^i)}{2^i} < \infty$ and $\liminf_{l\to\infty} f(Kl)/f(l) > 1$ for some integer $K \geq 2$.

Proposition 2.2: Under Conditions (C0), (C1′) and (C2), it follows that for any $\gamma > 0$

$$\max_{0\leq k\leq\lfloor\gamma\epsilon^{-1}\rfloor}|h_k^\epsilon(\omega) - g_k^\epsilon| \to 0 \text{ as } \epsilon \to 0 \text{ a.s.}$$

Example 2.2: Suppose for all $n \in \{1,2,3,\ldots\}$, some $K', \chi > 0$ and some integer $\beta \geq 1$

$$f(n) \triangleq \begin{cases}K' & 1\leq n<n_0\\ K'n/(\log(n)\log_2(n)\cdots\log_{\beta-1}(n)(\log_\beta(n))^{1+\chi}) & \text{otherwise}\end{cases}$$

and $n_0 = n_0(\beta, \chi, K')$ is chosen large enough that $f(\cdot)$ is nondecreasing. With this $f(\cdot)$ we obtain from Proposition 2.2

almost-sure convergence under dependency conditions similar to what Serfling (see Stout [20, Theorem 3.7.3]) used to obtain his strong law of large numbers.

Condition (C0) is a very mild condition which should be true in most applications including the ones mentioned in the introduction, Conditions (C1) and (C1′) define mild forms of dependence in the processes $\{A_l,\ l = 0,1,2,\ldots\}$ and $\{b_l,\ l=0,1,2,\ldots\}$, suitable for a wide range of applications, and Condition (C2) expresses a very weak type of pseudo-stationarity for $\{A_l,\ l=0,1,2,\ldots\}$ and $\{b_l,\ l=0,1,2,\ldots\}$. We now motivate Conditions (C1) and (C1′) by letting ξ_l represent some component, $a_l^{(s,t)}$, of A_l or some component, $b_l^{(s)}$, of b_l and amassing several examples where $\{\xi_l,\ l = 0,1,2,\ldots\}$ will yield a moment bound like those in (C1) or (C1′). For simplicity of notation let $\tilde{\xi}_l = \xi_l - E\xi_l$ for $l=0,1,2,\ldots$ in the following examples.

Example 2.3: As mentioned in Eweda and Macchi [6, p. 121], decaying-covariance assumptions suit data-transmission problems in which adaptive-filtering algorithms, such as the ones mentioned in the introduction, might be used. Suppose $\{\xi_l,\ l = 0,1,2,\ldots\}$ is a second-order process (i.e. $\{\tilde{\xi}_l,\ l = 0,1,2,\ldots\}$ is a zero-mean process with $E\left[\tilde{\xi}_l^2\right] < \infty$ for all $l=0,1,2,\ldots$) satisfying the following decaying-covariance condition:

$$\sup_{k\geq 0}\sum_{l=k}^{\infty}\left|E\left[\tilde{\xi}_l\tilde{\xi}_k\right]\right| < \infty. \tag{2.7}$$

Then there exists a constant $c>0$ such that for any $0\leq p\leq q<\infty$

$$E\left(\sum_{l=p}^{q}\tilde{\xi}_l\right)^2 \leq 2\sum_{k=p}^{q}\sum_{l=k}^{q}E\tilde{\xi}_k\tilde{\xi}_l \leq 2\sum_{k=p}^{q}\sum_{l=k}^{\infty}\left|E\tilde{\xi}_k\tilde{\xi}_l\right| \leq c\,(q-p+1)\,. \tag{2.8}$$

Thus a bound like (C1) is satisfied with $m=1$. Moreover, if there is a positive-valued, nondecreasing sequence $\{f(l)\}_{l=1}^{\infty}$ such that

$$\sum_{i=0}^{\infty}\frac{f(2^i)}{2^i} < \infty, \qquad \liminf_{l\to\infty} f(Kl)/f(l) > 1$$

(K an integer ≥ 2) and

$$\sup_{k\geq 0}\sum_{l=k}^{k+n-1}\left|E\left[\tilde{\xi}_j\tilde{\xi}_k\right]\right| \leq f(n) \text{ for all } n\geq 1,$$

then it follows that (C1′) is satisfied although (C1) may not be.

Before our next example (on strong mixing processes) we state without proof a third proposition which will only be used within the confines of Example 2.4.

Proposition 2.3: Let $\delta>0$ and $m>1$ be constants, let p and q be any positive integers such that $q\geq p$ and let $\{u_k,\ p\leq k\leq q\}$ be a $\Re$-valued, zero-mean stochastic process on $(\Omega,\mathcal{F},P)$ satisfying the following moment bound

$$M \triangleq \max_{p\leq k\leq q} E|u_k|^{m(2+\delta)} < \infty\ .$$

Moreover, let $\{\alpha(l)\}_{l=0}^{q-p}$ be a nonincreasing sequence such that

$$\max_{\substack{D\in\sigma\{u_k,\ p\leq k\leq r\}\\ E\in\sigma\{u_{r+l}\}}} |P(D\cap E)-P(D)P(E)| \leq \alpha(l)$$

for all integers $r,\ l$ such that $r,\ r+l\in\{p,p+1,\ldots,q\}$. Then, there exists a constant $c_{m,M,\delta}>0$ such that

$$E\left|\sum_{k=p}^{q}u_k\right|^{2m} \leq c_{m,M,\delta}\left(\sum_{k=0}^{q-p}[\alpha(k)]^{\frac{\delta}{m(2+\delta)}}\right)^m\cdot(q-p+1)^m.$$

Remark 2.2: Given the bounds on strong mixing processes (see for example Yokoyama [22, Theorems 1 and 2], Berbee [1, Lemma 3.2] and Doukhan and Portal [5, Theorems II.3 and II.4]), the above moment bound is not overly surprising. It can be proved in the continuous time setting by adapting arguments in Gerencsér [8, Theorem 1.1]. The discrete time version then follows via a construction similar to the one used in Example 2.5.

Example 2.4: Strong mixing conditions are widely used in the literature and appear to be satisfied by a fairly substantial class of processes including a wide variety of ARMA processes (see for example Mokkadem [17]). Suppose $\{\xi_l,\ l = 0,1,2,\ldots\}$ is a $\Re$-valued second-order process satisfying the following strong mixing condition: There exists a monotonically nonincreasing sequence $\{\alpha_\xi(l)\}_{l=0}^{\infty}$ and real constants $\delta>0,\ m\geq 1$ such that

$$\sup_{k\geq 0}\ \sup_{\substack{D\in\sigma\{\xi_j,\ j=k+l,k+l+1,\ldots\}\\ E\in\sigma\{\xi_j,\ j=0,1,\ldots,k\}}} |P(D\cap E)-P(D)P(E)| \leq \alpha_\xi(l), \tag{2.9}$$

for $l=0,1,2,\ldots$ and

$$\sum_{l=0}^{\infty}[\alpha_\xi(l)]^{\frac{\delta}{m(2+\delta)}} < \infty. \tag{2.10}$$

Moreover, suppose the process $\{\xi_l,\ l = 0,1,2,\ldots\}$ also satisfies the moment condition:

$$\sup_{k\geq 0} E\left|\tilde{\xi}_k\right|^{m(2+\delta)} < \infty \tag{2.11}$$

where $\delta>0$ and $m\geq 1$ are the same constants as in (2.10).

Then by Proposition 2.3, there exists a constant $c_m>0$ such that

$$E\left|\sum_{l=p}^{q}\tilde{\xi}_l\right|^{2m} \leq c_m(q-p+1)^m \tag{2.12}$$

for all integers $0\leq p\leq q<\infty$ which establishes Condition (C1). Alternatively, suppose (2.11) is satisfied with $m=1$ and suppose (2.10) is replaced with

$$\sum_{l=0}^{n-1} [\alpha_\xi(l)]^{\frac{\delta}{2+\delta}} \le f(n) \text{ for all } n \ge 1 \quad (2.13)$$

and some $f(\cdot)$ as in the previous example and Condition (C1′). Then using Proposition 2.3 (with $m = 1$), it follows that (C1′) is satisfied.

Example 2.5: In this example we consider the "stably generated" processes adopted by Davis and Vinter (see [3, Definition 5.1.1]) and the "L-mixing" processes of Gerencsér [8]. With this in mind, we assume that

1) for some $q \ge 1$, we have that

$$\sup_{k\ge 0} E|\xi_k|^{2q} < \infty,$$

2) there exists a family of σ-algebra pairs $\{(\mathcal{F}_k, \mathcal{F}_k^+),\ k \ge 0\}$ such that $\mathcal{F}_j \subset \mathcal{F}_k \subset \mathcal{F}$ and $\mathcal{F}_k^+ \subset \mathcal{F}_j^+ \subset \mathcal{F}$ for all integers $0 \le j \le k$, $\mathcal{F}_k$ is independent of $\mathcal{F}_k^+$ for all integers $k \ge 0$, ξ_k is $\mathcal{F}_k$-measurable for all integers $k \ge 0$ and

$$\sum_{k=0}^{\infty} \gamma_q(k) < \infty,$$

where $q \ge 1$ is the constant of 1 above and

$$\gamma_q(k) \triangleq \sup_{j\ge k} E^{\frac{1}{2q}} |\xi_j - E(\xi_j|\mathcal{F}_{j-k}^+)|^{2q} \text{ for } k = 0,1,2,\ldots$$

The processes described above are variations on the "exponentially stable" mixing processes introduced by Ljung (see S3 on page 772 of Ljung [14]) and have been found useful in studying inference in control theory. In particular, they have the useful property that a stable linear dynamical system driven by a process satisfying 1 and 2 generates an output which also satisfies 1 and 2 (see Lemma 2.4 on page 172 of [8] for a statement and proof of the continuous time version of this property and note that the discrete version is proved analogously). From Theorem 1.1 of [8] one sees that (C1) holds for $m = q$ when $\{\xi_l,\ l = 0,1,2,\ldots\}$ satisfies 1 and 2 above. Theorem 1.1 of [8] is actually stated in a continuous-time context for L-mixing processes, but the desired discrete-time version for processes satisfying 1 and 2 follows by converting the given discrete-time process, $\{\tilde{\xi}_k\}$, into a right-continuous, piecewise-constant, continuous-time process, i.e. letting $u_v = \tilde{\xi}_k$ for $v \in [k, k+1), k = 0,1,2,\ldots$ and applying Theorem 1.1 of [8] (which continues to hold under sufficient generality) to this continuous-time process. Finally, it is again possible to weaken 2 in a similar manner to what was done in Example 2.4 above and still satisfy (C1′).

Remark 2.3: In the above example, it is immediately obvious that L-mixing processes satisfy 1 and 2. Moreover, suppose that $\{\xi_k,\ k = 0,1,2,\ldots\}$ satisfies 1 above and (similar to Davis and Vinter [3, p. 217]) that $\{u_k,\ -\infty < k < \infty\}$ is an independent sequence of random variables such that ξ_k is $\sigma\{u_l,\ -\infty < l \le k\}$-measurable for all $k = 0,1,2,\ldots$. Moreover, suppose there exist constants $c > 0$ and $\lambda \in (0,1)$ such that for all integers k, j with $k \le j+1$, $k \ge 1$ there is a random variable $\xi_j[k]$ measurable with respect to $\sigma\{u_{j-k+1},\ldots,u_j\}$ and satisfying

$$E|\xi_j - \xi_j[k]|^{2q} \le c\,\lambda^k.$$

Then, it follows by Jensen's inequality for conditional expectations that

$$\begin{aligned} E|\xi_j - E(\xi_j|\sigma\{u_l,\ j-k < l < \infty\})|^{2q} \\ \le E|\xi_j - E(\xi_j|\sigma\{u_{j-k+1},\ldots,u_j\})|^{2q} \le 2^{2q}\,c\,\lambda^k \end{aligned}$$

for all integers $k \le j+1$, $k \ge 1$ and $\{\xi_l,\ l = 0,1,2,\ldots\}$ will also satisfy 2 of Example 2.3 with $\mathcal{F}_k \triangleq \sigma\{u_l,\ -\infty < l \le k\}$ and $\mathcal{F}_k^+ \triangleq \sigma\{u_l,\ k < l < \infty\}$.

Example 2.6: In [15], Longnecker and Serfling define a general class of mixing processes, the so-called "weak multiplicative" processes, which satisfy quasi-orthogonality dependence restrictions. Rather than repeat the definitions here we refer the reader to Section 2 of [15] where some half-dozen precise formulations of this concept are given. From Section 4 of [15], it is seen that if $\{\xi_l,\ l = 0,1,2,\ldots\}$ is weak multiplicative then condition (C1) holds for some positive integer m.

Example 2.7: Suppose that $\{\xi_l,\ l = 0,1,2,\ldots\}$ is a martingale difference sequence satisfying $\sup_{l\ge 0} E|\xi_l|^{2\nu} < \infty$ for some $\nu \ge 1$. Then $\{\tilde{\xi}_l,\ l = 0,1,2,\ldots\}$ satisfies (C1) with $m = \nu$. This follows directly from Burkholder's inequality (see Hall and Heyde [10, Theorem 2.10]) if $\nu = 1$ or from Theorem 3.7.8 of Stout [20] if $\nu > 1$.

Now, we motivate Condition (C2) via a simple example:

Example 2.8: Consider the dynamical system described by

$$y_{k+1}(\omega) = d_k y_k(\omega) + n_k u_k(\omega) \quad \text{for } k = 0,1,2,\ldots,$$

subject to $y_0 = 1$, where d_k, n_k may vary with time and $\{u_k,\ k = 0,1,2,\ldots\}$ is a sequence of independent $\mathcal{N}(0,1)$ random variables. Suppose we have access to corrupted versions of y_k and u_k defined by $\psi_k(\omega) = y_k(\omega) + \rho_k(\omega)$ and $e_k(\omega) = u_k(\omega) + \zeta_k(\omega)$ and where $\{\rho_k,\ k = 0,1,2,\ldots\}$ and $\{\zeta_k,\ k = 0,1,2,\ldots\}$ are sequences of zero-mean i.i.d. random variables mutually independent of each other and of $\{u_k, k = 0,1,2,\ldots\}$. Using the LMS algorithm to estimate d_k, n_k, we obtain the recursion

$$h_{k+1}^\epsilon = h_k^\epsilon + \epsilon(b_k(\omega) - A_k(\omega)\, h_k^\epsilon) \qquad \text{for } k = 0,1,2,\ldots,$$

subject to some initial guess h_0^ϵ at $[d_0\ n_0]^T$, where $b_k = [\psi_{k+1}\psi_k\ \ \psi_{k+1}e_k]^T$ and

$$A_k = \begin{bmatrix} \psi_k^2 & \psi_k e_k \\ \psi_k e_k & e_k^2 \end{bmatrix}.$$

(i) If d_k, n_k are constant for all k and $0 < |d_0| < 1$. Then it follows that

$$E\psi_k^2 = (d_0)^{2k} + \sum_{j=0}^{k-1}(d_0)^{2j}(n_0)^2 + E\rho_0^2$$
$$\text{for } k = 0,1,2,\ldots$$

so $E\psi_k^2$ varies with time; consequently, stationarity conditions such as those assumed in Eweda and Macchi [6] are not actually satisfied. However, it is easily seen that Condition (C2) is satisfied and thus (C2) is a more natural assumption for ARMA processes with transient behavior.

(ii) If d_k, n_k vary with time it seems unreasonable to insist that EA_k and Eb_k are time-invariant.

III. Proofs of Propositions 2.1 and 2.2

In this section we establish Proposition 2.1 and Proposition 2.2 as stated in Section 2. The proofs of Proposition 2.1(i), Proposition 2.1(ii) and Proposition 2.2 are all very similar so we will prove Proposition 2.1(ii) first and then describe the changes required for Proposition 2.1(i) and Proposition 2.2.

Proof of Proposition 2.1 (ii): We assume that $\zeta \in (0,1]$ has already been chosen and, to simplify notation, that $\gamma = 1$. Fix an $\omega \in \Omega$ and an $\epsilon \in (0, \frac{1}{2}]$. We have by (2.1) and (2.2) that

$$h_k^\epsilon - g_k^\epsilon = \epsilon\sum_{l=0}^{k-1}\tilde{b}_l - \epsilon\sum_{l=0}^{k-1}\tilde{A}_l g_l^\epsilon + \epsilon\sum_{l=0}^{k-1}A_l(g_l^\epsilon - h_l^\epsilon) \quad (3.1)$$

for any $k \in \{1,2,\ldots,\lfloor\epsilon^{-\zeta}\rfloor\}$, where $\tilde{A}_l \triangleq A_l - EA_l$ and $\tilde{b}_l \triangleq b_l - Eb_l$ for $l = 0,1,\ldots,\lfloor\epsilon^{-\zeta}\rfloor$. Hence, we have by (3.1)

$$|h_k^\epsilon - g_k^\epsilon| \le \epsilon\left\{\max_{1\le j\le\lfloor\epsilon^{-\zeta}\rfloor}\left|\sum_{l=0}^{j-1}\tilde{A}_l g_l^\epsilon\right| + \max_{1\le j\le\lfloor\epsilon^{-\zeta}\rfloor}\left|\sum_{l=0}^{j-1}\tilde{b}_l\right| + \sum_{l=0}^{k-1}|||A_l|||\,|g_l^\epsilon - h_l^\epsilon|\right\} \quad (3.2)$$

for $k = 0,1,\ldots,\lfloor\epsilon^{-\zeta}\rfloor$ and using the discrete Bellman-Gronwall inequality (see Desoer and Vidyasagar [4, page 254]) on (3.2) and the fact that $\zeta \le 1$

$$\begin{aligned}|h_k^\epsilon - g_k^\epsilon| &\le \left[\prod_{l=0}^{k-1}(1+\epsilon|||A_l|||)\right]\epsilon\left\{\max_{1\le j\le\lfloor\epsilon^{-\zeta}\rfloor}\left|\sum_{l=0}^{j-1}\tilde{A}_l g_l^\epsilon\right| + \max_{1\le j\le\lfloor\epsilon^{-\zeta}\rfloor}\left|\sum_{l=0}^{j-1}\tilde{b}_l\right|\right\}\\ &\le \exp\left\{\sum_{l=0}^{\lfloor\epsilon^{-1}\rfloor-1}\epsilon|||A_l|||\right\} \times \epsilon\left\{\max_{1\le j\le\lfloor\epsilon^{-\zeta}\rfloor}\left|\sum_{l=0}^{j-1}\tilde{A}_l g_l^\epsilon\right| + \max_{1\le j\le\lfloor\epsilon^{-\zeta}\rfloor}\left|\sum_{l=0}^{j-1}\tilde{b}_l\right|\right\}\end{aligned} \quad (3.3)$$

for $k = 0,1,\ldots,\lfloor\epsilon^{-\zeta}\rfloor$. Now, by (3.3) and Lemma A.5 there exists a function $\omega \to M(\omega)$ almost-surely finite and independent of ϵ such that for each $\omega \in \Omega$

$$\begin{aligned}&\max_{0\le k\le\lfloor\epsilon^{-\zeta}\rfloor}|h_k^\epsilon(\omega) - g_k^\epsilon|\\ &\le e^{M(\omega)}\epsilon\left\{\max_{1\le j\le\lfloor\epsilon^{-\zeta}\rfloor}\left|\sum_{l=0}^{j-1}\tilde{A}_l(\omega)g_l^\epsilon\right| + \max_{1\le j\le\lfloor\epsilon^{-\zeta}\rfloor}\left|\sum_{l=0}^{j-1}\tilde{b}_l(\omega)\right|\right\}\\ &\le e^{M(\omega)}\epsilon\left\{\sup_{2^{i_\epsilon}\le T\le 2^{i_\epsilon+1}}\max_{1\le j\le\lfloor T\rfloor}\left|\sum_{l=0}^{j-1}\tilde{A}_l(\omega)f_l^{T,\zeta}\right| + \max_{1\le j\le 2^{i_\epsilon+1}}\left|\sum_{l=0}^{j-1}\tilde{b}_l(\omega)\right|\right\}\end{aligned} \quad (3.4)$$

where i_ϵ is the integer such that $2^{-i_\epsilon-1} < \epsilon^\zeta \le 2^{-i_\epsilon}$ and $f_l^{T,\zeta} \triangleq g_l^{(1/T)^{1/\zeta}}$ for all $l = 0,1,2,\ldots$, and $T \in \Re$ such that $2^{i_\epsilon} \le T \le 2^{i_\epsilon+1}$.

Now, using (2.2), (C2), (2.3), and the discrete Bellman-Gronwall inequality, we discover that $T \to f_l^{T,\rho}$ whence

$$T \to \max_{0\le j\le\lfloor T\rfloor}\left|\sum_{l=0}^{j-1}\tilde{A}_l f_l^{T,\zeta}\right|$$

is right continuous and the supremum term in (3.4) is F-measurable. Hence, we have by the monotone convergence theorem and Lemma A.1 (ii) that there is a constant $\alpha_m^\zeta > 0$ such that

$$E\left\{\sum_{i=1}^\infty\frac{\sup_{2^i\le T\le 2^{i+1}}\max_{1\le j\le\lfloor T\rfloor}\left|\sum_{l=0}^{j-1}\tilde{A}_l f_l^{T,\zeta}\right|^{2m}}{2^{im}\,\psi(2^i)}\right\} \le \sum_{i=1}^\infty\frac{\alpha_m^\zeta 2^m}{\psi(2^i)} \quad (3.5)$$

and since $\sum_{i=1}^\infty\frac{1}{\psi(2^i)} < \infty$ we have a function $\omega \to N(\omega)$ almost-surely finite such that

$$\sup_{2^i\le T\le 2^{i+1}}\max_{1\le j\le\lfloor T\rfloor}\left|\sum_{l=0}^{j-1}\tilde{A}_l(\omega)f_l^{T,\zeta}\right| \le N(\omega)2^{\frac{i}{2}}(\psi(2^i))^{\frac{1}{2m}} \quad (3.6)$$

for all $\omega \in \Omega$ and $i = 1,2,3,\ldots$. Similarly, using the monotone convergence theorem, and Lemma A.2(ii) there exists a function $\omega \to K(\omega)$ almost-surely finite such that

$$\max_{1\le j\le 2^{i+1}}\left|\sum_{l=0}^{j-1}\tilde{b}_l(\omega)\right| \le K(\omega)2^{\frac{i}{2}}(\psi(2^i))^{\frac{1}{2m}} \quad (3.7)$$

for all $\omega \in \Omega$ and $i = 1,2,3,\cdots$. Substituting (3.6) and (3.7) into (3.4) yields the existence of some $H(\omega)$ almost surely finite such that:

$$\begin{aligned}\max_{0\le k\le\lfloor\epsilon^{-\zeta}\rfloor}|h_k^\epsilon(\omega) - g_k^\epsilon| &\le e^{M(\omega)}\epsilon(K(\omega) + N(\omega))2^{\frac{i_\epsilon}{2}}(\psi(2^{i_\epsilon}))^{\frac{1}{2m}}\\ &\le H(\omega)\epsilon^{1-\frac{\zeta}{2}}(\psi(\epsilon^{-\zeta}))^{\frac{1}{2m}}\end{aligned} \quad (3.8)$$

for all $\omega \in \Omega$ and $0 < \epsilon \le \frac{1}{2}$ and Proposition 2.1 (ii) follows when $\gamma \le 1$; the case where $\gamma > 1$ follows similarly defining i_ϵ to be the integer such that $2^{-i_\epsilon-1} < \epsilon^\zeta/\gamma \le 2^{-i_\epsilon}$ and replacing $f_l^{T,\zeta}$ with $f_l^{T,\zeta,\gamma} \triangleq g_l^{(\gamma/T)^{\frac{1}{\zeta}}}$.

Proof of Proposition 2.1 (i): We follow the same path as for Proposition 2.1(ii) but use Lemma A.1(i) and Lemma A.2(i) in place of Lemma A.1(ii) and Lemma A.2(ii) and replace (3.5) with

$$E\left\{\sum_{i=1}^{\infty}\frac{\sup_{2^i\le T\le 2^{i+1}}\max_{1\le j\le\lfloor T\rfloor}\left|\sum_{l=0}^{j-1}\tilde{A}_l f_l^{T,\zeta}\right|^2}{2^i\, i^2\, \psi(2^i)}\right\} \le \sum_{i=1}^{\infty}\frac{\alpha_1^\zeta\, 2(i+2)^2}{i^2\psi(2^i)} < \infty.$$

After repeating the work in (3.6) and (3.7), (3.8) becomes:

$$\max_{0\le k\le\lfloor\epsilon^{-\zeta}\rfloor}|h_k^\epsilon(\omega)-g_k^\epsilon| \le H(\omega)\epsilon^{1-\frac{\zeta}{2}}(\log\epsilon^{-\zeta})(\psi(\epsilon^{-\zeta}))^{\frac{1}{2}} \tag{3.9}$$

for all $\omega\in\Omega$ and $0<\epsilon\le\frac{1}{2}$ so we have the case where $\gamma\le 1$ and the case where $\gamma>1$ follows similarly. □

Finally, we prove Proposition 2.2 which assumes only Condition (C1′) rather than (C1). However, its proof requires only a trivial modification of the proof of Proposition 2.1.

Proof of Proposition 2.2: For ease of notation we will assume that $\gamma=1$. Now it follows from (3.4) of the proof of Proposition 2.1 (ii) that

$$\max_{0\le k\le\lfloor\epsilon^{-\zeta}\rfloor}|h_k^\epsilon(\omega)-g_k^\epsilon| \le e^{M(\omega)} \times\epsilon\left\{\sup_{2^{i_\epsilon}\le T\le 2^{i_\epsilon+1}}\max_{1\le j\le\lfloor T\rfloor}\left|\sum_{l=0}^{j-1}\tilde{A}_l(\omega)f_l^T\right| + \max_{1\le j\le 2^{i_\epsilon+1}}\left|\sum_{l=0}^{j-1}\tilde{b}_l(\omega)\right|\right\} \tag{3.10}$$

for each $\omega\in\Omega$ and $\epsilon\in(0,1]$, where i_ϵ is the integer such that $2^{-i_\epsilon+1}<\epsilon\le e^{-i_\epsilon}$ and $f_l^T \triangleq g_l^{1/T}$. Moreover, it follows from the monotone convergence theorem, Lemma A.1 (iii) and Condition (C1′) that there is a constant $c_1>0$ such that

$$E\left\{\sum_{i=0}^{\infty}\frac{\sup_{2^i\le T\le 2^{i+1}}\max_{1\le j\le\lfloor T\rfloor}\left|\sum_{l=0}^{j-1}\tilde{A}_l f_l^T\right|^2}{2^{2i}}\right\} \le \sum_{i=0}^{\infty}\frac{c_1 f(2^{i+1})}{2^i}<\infty, \tag{3.11}$$

where $f(\cdot)$ is the function defined in Condition (C1′). Hence, we have that

$$\frac{\sup_{2^i\le T\le 2^{i+1}}\max_{1\le j\le\lfloor T\rfloor}\left|\sum_{l=0}^{j-1}\tilde{A}_l f_l^T\right|}{2^i}\to 0 \text{ as } i\to\infty \text{ a.s.} \tag{3.12}$$

Similarly, using the monotone convergence theorem and Lemma A.2 (iii), we have that

$$\frac{\max_{1\le j\le 2^{i+1}}\left|\sum_{l=0}^{j-1}\tilde{b}_l\right|}{2^i}\to 0 \text{ as } i\to\infty \text{ a.s.} \tag{3.13}$$

Proposition 2.2 follows by substituting (3.12) and (3.13) into (3.10). □

IV. Discussion

In the preceding sections, we have shown that under mild stationarity and dependency conditions one can bound the difference between the random recursion (2.1) and the nonrandom recursion (2.2). Under additional stationarity conditions on the processes $\{b_k,\ k=0,1,2,\ldots\}$ and $\{A_k,\ k=0,1,2,\ldots\}$ one can associate a "limiting ordinary differential equation" with the recursion (2.1), whose (nonrandom) solution defines the limiting behaviour of $\{h_k^\epsilon,\ k=0,1,2,\ldots\}$ as $\epsilon\to 0$. Such a result is an immediate consequence of Proposition 2.2 if there is sufficient regularity in the nonrandom sequences $\{Eb_k\}$ and $\{EA_k\}$ to allow one to apply (a discrete-time version of) the classical averaging principle of Bogoliubov-Krylov-Mitropolskii [2] to the nonrandom recursion (2.2). For example, if

$$\sup_{l\ge 0}|Eb_l|<\infty,\quad \sup_{l\ge 0}|||EA_l|||<\infty, \tag{4.1}$$

and the following limits exist

$$\lim_{N\to\infty}\frac{1}{N}\sum_{l=0}^{N-1}EA_l=A,\quad \lim_{N\to\infty}\frac{1}{N}\sum_{l=0}^{N-1}Eb_l=b \tag{4.2}$$

for some d-vector b and d by d matrix A, then it follows from Lemma A.7 that for any $\gamma>0$

$$\max_{0\le k\le\lfloor\gamma\epsilon^{-1}\rfloor}|g_k^\epsilon-g^0(k\epsilon)|\to 0 \tag{4.3}$$

as $\epsilon\to 0$, where $g^0(\cdot)$ is the solution of the differential equation

$$\dot{g}^0(\tau)=-Ag^0(\tau)+b \text{ subject to } g^0(0)\triangleq h_0. \tag{4.4}$$

Combining (4.3) with Proposition 2.2 one sees that if (4.1) and (4.2) hold, in addition to Conditions (C0) and (C1′) of Section 2, then

$$\lim_{\substack{\epsilon\to 0\\ \epsilon>0}}\max_{0\le k\le\lfloor\gamma\epsilon^{-1}\rfloor}|h_k^\epsilon(\omega)-g^0(k\epsilon)|=0 \text{ a.s..} \tag{4.5}$$

This result complements Theorem 1 of [12] which establishes the convergence *in probability* of the quantity on the left of (4.5) to zero as $\epsilon\to 0$, under conditions somewhat related to those above (see the Remark on page 179 of [12]). Moreover, one can also get almost-sure rate bounds for the convergence in (4.5) merely by assuming enough regularity for the sequences $\{EA_k\}$ and $\{Eb_k\}$. Indeed if, instead of (4.2), one has

$$|||\frac{1}{N}\sum_{l=0}^{N-1}EA_l-A|||=O(N^{-1}) \tag{4.6}$$

with a similar bound for the $\{Eb_k\}$ sequence, then it follows by Lemmas A.8 and A.9 that for any $\gamma>0$, $0<\zeta\le 1$

$$\max_{0\le k\le\lfloor\gamma\epsilon^{-\zeta}\rfloor}|g_k^\epsilon-g^0(\epsilon k)|=O(\epsilon^{1-\zeta/2}), \tag{4.7}$$

and hence from Proposition 2.1 one sees that

$$\max_{0\le k\le \lfloor \gamma\epsilon^{-\zeta}\rfloor} |h_k^\epsilon(\omega) - g^0(k\epsilon)| \tag{4.8}$$

is, for almost all ω, either $O(\epsilon^{1-\frac{\zeta}{2}}(\log\epsilon^{-\zeta})(\psi(\epsilon^{-\zeta}))^{\frac{1}{2}})$ or $O(\epsilon^{1-\frac{\zeta}{2}}(\psi(\epsilon^{-\zeta}))^{\frac{1}{2m}})$ depending on whether $m=1$ or $m>1$ in (C1).

The above a.s. rate bounds are all slightly slower than $O(\epsilon^{1-\frac{\zeta}{2}})$, which is very slow convergence indeed. The question then arises as to what extent it is possible to improve these bounds under perhaps more stringent conditions. We note first of all that if one defines $\{X^\epsilon(\tau), 0\le\tau\le 1\}$ by

$$X^\epsilon(\tau) \triangleq \epsilon^{-\frac{1}{2}}(h_k^\epsilon(\omega) - g^0(k\epsilon)) \quad \text{for } \tau\in[k\epsilon, k\epsilon+\epsilon), k=0,1,\ldots,\lfloor\epsilon^{-1}\rfloor, \tag{4.9}$$

then one expects from the weak convergence analysis in [12, Section V] that, under suitable strengthening of the regularity conditions in Section 2, the family of processes $\{X^\epsilon(\cdot)\}$ converges weakly to some limiting Gauss-Markov process, as $\epsilon\to 0$. This at once implies that for a.a. ω the rate of convergence in (4.5) cannot be faster than $O(\epsilon^{\frac{1}{2}})$. Actually, based on the functional law of the iterated logarithm for sums of random variables, we believe that the quantity in (4.8) can be shown to be of the form $O((\epsilon^{2-\zeta}\log\log\epsilon^{-\zeta})^{\frac{1}{2}})$ for a.a. ω and that no further improvement in this rate bound is possible. However, this will likely require an involved proof as well as regularity conditions much more stringent than those of Section II. As illustrated in Example 2.1, this paper establishes an a.s. convergence rate almost as good as this best bound under very general conditions and by a very simple proof.

APPENDIX

Technical Results

This appendix contains various technical results used to support the proofs in Section III and substantiate the claims made in Section IV. The first two results, Lemma A.1 and Lemma A.2, are used directly in the proofs of Propositions 2.1 and 2.2.

Lemma A.1: Under Condition (C2) and either Condition (C1) or (C1′) of Section 2 there exists a constant $\alpha_m^{\zeta,\gamma}>0$ such that

$$\text{(i)}\quad E\left\{\sup_{U\le T\le V}\max_{1\le j\le\lfloor T\rfloor}\left|\sum_{l=0}^{j-1}\tilde A_l f_l^{T,\zeta,\gamma}\right|^2\right\} \le \alpha_1^{\zeta,\gamma} V[log(2V)]^2$$

for all $1\le U\le V\le 2U$ if (C1) is satisfied with $m=1$,

$$\text{(ii)}\quad E\left\{\sup_{U\le T\le V}\max_{1\le j\le\lfloor T\rfloor}\left|\sum_{l=0}^{j-1}\tilde A_l f_l^{T,\zeta,\gamma}\right|^{2m}\right\} \le \alpha_m^{\zeta,\gamma} V^m$$

for all $1\le U\le V\le 2U$ if (C1) is satisfied with some $m>1$, or

$$\text{(iii)}\quad E\left\{\sup_{U\le T\le V}\max_{1\le j\le\lfloor T\rfloor}\left|\sum_{l=0}^{j-1}\tilde A_l f_l^{T,1,\gamma}\right|^2\right\} \le \alpha_1^{1,\gamma}\lfloor V\rfloor f(\lfloor V\rfloor)$$

for all $1\le U\le V\le 2U$ if Condition (C1′) is satisfied, where $\tilde A_l \triangleq A_l - EA_l$ and $f_l^{T,\zeta,\gamma}\triangleq g_l^{(\gamma/T)^{1/\zeta}}$ (g_l^ϵ being defined in (2.2) for each $\epsilon>0$) for $l=0,1,2,\ldots$ and $T\ge 1$. Here, $0<\zeta\le 1$, $\gamma\ge 1$ are constants, and $f(\cdot)$ is the function of Condition (C1′).

Proof: Consider (i), (ii) and (iii) simultaneously and fix a U and V such that $1\le U\le V\le 2U$. Then it follows by (2.2) of Section 2 that

$$\begin{aligned}
&E\left\{\sup_{U\le T\le V}\max_{1\le j\le\lfloor T\rfloor}\left|\sum_{l=0}^{j-1}\tilde A_l f_l^{T,\zeta,\gamma}\right|^{2m}\right\}\\
&\le 2^{2m}\left[E\left\{\max_{1\le j\le\lfloor V\rfloor}\left\|\left|\sum_{l=0}^{j-1}\tilde A_l\right|\right\| |h_0|\right\}^{2m} + \left(\frac{\gamma}{U}\right)^{\frac{2m}{\zeta}}\right.\\
&\times E\left\{\sup_{U\le T\le V}\max_{2\le j\le\lfloor T\rfloor}\left|\sum_{l=0}^{j-1}\tilde A_l\right.\right.\\
&\left.\left.\left.\sum_{r=0}^{l-1}(Eb_r - [EA_r]f_r^{T,\zeta,\gamma})\right|^{2m}\right\}\right].
\end{aligned}\tag{A.1}$$

Now by (2.4) of Section 2

$$\begin{aligned}
&\sup_{U\le T\le V}\max_{2\le j\le\lfloor T\rfloor}\left|\sum_{l=0}^{j-1}\tilde A_l\sum_{r=0}^{l-1}(Eb_r-[EA_r]f_r^{T,\zeta,\gamma})\right|^{2m}\\
&=\sup_{U\le T\le V}\max_{2\le j\le\lfloor T\rfloor}\left|\sum_{r=0}^{j-2}\sum_{l=r+1}^{j-1}\tilde A_l(Eb_r-[EA_r]f_r^{T,\zeta,\gamma})\right|^{2m}\\
&\le\sup_{U\le T\le V}\max_{2\le j\le\lfloor T\rfloor}\left\{\sum_{r=0}^{j-2}\max_{0\le p\le q<\lfloor V\rfloor}\left\|\left|\sum_{l=p}^{q}\tilde A_l\right|\right\|\left|Eb_r-[EA_r]f_r^{T,\zeta,\gamma}\right|\right\}^{2m}\\
&\le V^{2m}\left\{\sup_{0<\epsilon\le\gamma^{1/\zeta}}\max_{2\le j\le\lfloor\gamma\epsilon^{-\zeta}\rfloor}\frac{1}{j-1}\sum_{r=0}^{j-2}|Eb_r-[EA_r]g_r^\epsilon|\right\}^{2m}\\
&\times\max_{0\le p\le q<\lfloor V\rfloor}\left\|\left|\sum_{l=p}^{q}\tilde A_l\right|\right\|^{2m}\\
&\le D_{\gamma,\zeta}^{2m}V^{2m}\max_{0\le p\le q<\lfloor V\rfloor}\left\|\left|\sum_{l=p}^{q}\tilde A_l\right|\right\|^{2m}\\
&\le D_{\gamma,\zeta}^{2m}2^{2m}U^{2m}\max_{0\le p\le q<\lfloor V\rfloor}\left\|\left|\sum_{l=p}^{q}\tilde A_l\right|\right\|^{2m}
\end{aligned}\tag{A.2}$$

for each $\omega \in \Omega$. Hence by (A.1) and (A.2) there exists a number $c_1 > 0$ independent of U and V such that

$$E\left\{\sup_{U\le T\le V}\max_{1\le j\le \lfloor T\rfloor}\left|\sum_{l=0}^{j-1}\tilde{A}_l f_l^{T,\zeta,\gamma}\right|^{2m}\right\} \le c_1 E\left\{\max_{0\le p\le q<\lfloor V\rfloor}\left|\left|\left|\sum_{l=p}^{q}\tilde{A}_l\right|\right|\right|^{2m}\right\} \le d^{2m}c_1 E\left\{\max_{0\le p\le q<\lfloor V\rfloor}\max_{1\le s,t\le d}\left|\sum_{l=p}^{q}\tilde{a}_l^{(s,t)}\right|^{2m}\right\} \le d^{2m}c_1\sum_{s,t=1}^{d}E\left\{\max_{0\le p\le q<\lfloor V\rfloor}\left|\sum_{l=p}^{q}\tilde{a}_l^{(s,t)}\right|^{2m}\right\} \quad (A.3)$$

where $\tilde{a}_l^{(s,t)}$ denotes the $(s,t)^{\text{th}}$ component of $\tilde{A}_l$. However, by (C1)

$$E\left|\sum_{l=p}^{q}\tilde{a}_l^{(s,t)}\right|^{2m} \le c_m(q-p+1)^m \quad (A.4)$$

or by (C1′)

$$E\left|\sum_{l=p}^{q}\tilde{a}_l^{(s,t)}\right|^{2} \le (q-p+1)f(q-p+1) \quad (A.5)$$

for all integers p,q,s,t such that $0 \le p \le q < \lfloor V\rfloor$ and $1 \le s,t \le d$. (i) follows from (A.3), (A.4) and Theorem A.3, where we use $\{X_l,\ l=0,1,2,\ldots\} = \{\tilde{a}_l^{(s,t)},\ l=0,1,2,\ldots\}$ for each $1\le s,\ t\le d$ and $g(i,j) = c_m(j-i+1)$ for all integers $0\le i\le j$. Similarly, (ii) and (iii) follow from (A.3), (A.4), (A.5) and Theorem A.4, where $g(n) = c_m n^m$ for all integers $n \ge 1$ when we are proving (ii) and $g(n) = n\,f(n)$ for all integers $n\ge 1$ when we are proving (iii). □

Lemma A.2: Under Condition (C2) and either Condition (C1) or Condition (C1′) of Section 2 there exists a constant $\beta_m > 0$ such that

$$\text{(i)}\quad E\left\{\max_{1\le j\le\lfloor V\rfloor}\left|\sum_{l=0}^{j-1}\tilde{b}_l\right|\right\}^2 \le \beta_1 V[\log(2V)]^2$$

for all $V \ge 1$ if (C1) is satisfied with $m = 1$,

$$\text{(ii)}\quad E\left\{\max_{1\le j\le\lfloor V\rfloor}\left|\sum_{l=0}^{j-1}\tilde{b}_l\right|\right\}^{2m} \le \beta_m V^m$$

for all $V \ge 1$ if (C1) is satisfied with some $m > 1$,

$$\text{(iii)}\quad E\left\{\max_{1\le j\le\lfloor V\rfloor}\left|\sum_{l=0}^{j-1}\tilde{b}_l\right|\right\}^2 \le \beta_1\lfloor V\rfloor f(\lfloor V\rfloor)$$

for all $V\ge 1$ if (C1′) is satisfied, where $\tilde{b}_l \triangleq b_l - Eb_l$ for $l = 0,1,2,\ldots$ and $f(\cdot)$ is the function of Condition (C1′).

Proof: Similar to the proof of Lemma A.1. □

The following theorem is a trivial extension to Serfling's maximal inequality (see [20, Theorem 2.4.1]). It is used to establish Lemma A.1 (i) and Lemma A.2 (i) above.

Theorem A.3: Let $X_0, X_1, \ldots, X_{n-1}$ $(n \ge 1)$ be real-valued random variables such that

$$E\left|\sum_{k=i}^{j}X_k\right|^2 \le g(i,j) \text{ for all integers } 0\le i\le j<n$$

where $g(\cdot,\cdot)$ is a nonnegative function satisfying

$$g(i,j) + g(j+1,k) \le g(i,k) \text{ for all } 0\le i\le j<k<n.$$

Then

$$E\left\{\max_{0\le i\le j<n}\left|\sum_{k=i}^{j}X_k\right|^2\right\} \le 2[\log_2(2n)]^2 g(0,n).$$

The following maximal inequality is an immediate consequence of Lai and Stout's maximal inequality (see [13, Theorem 5]). It is used to establish Lemma A.1 (ii), (iii) and Lemma A.2 (ii), (iii) above.

Theorem A.4: Let $X_0, X_1, \ldots, X_{n-1}$ $(n\ge 1)$ be real-valued random variables. Suppose there is a constant $\nu > 0$ and a positive, nondecreasing function $g(l)$, $l = 1,2,3,\ldots$ satisfying

$$\liminf_{l\to\infty} g(Kl)/g(l) > K$$

for some integer $K \ge 2$ and

$$E\left|\sum_{k=i}^{j}X_k\right|^{\nu} \le g(j-i+1) \text{ for all integers } 0\le i\le j<n.$$

Then there exists a constant A (independent of n) such that

$$E\left\{\max_{0\le i\le j<n}\left|\sum_{k=i}^{j}X_k\right|^{\nu}\right\} \le Ag(n) \text{ for all } n\ge 1.$$

Next, in Lemma A.5, we establish the uniform (with respect to ϵ) bound which is required in the proofs of Propositions 2.1 and 2.2. Since Condition (C1) clearly implies Condition (C1′), Lemma A.5 holds under the hypothesis of Proposition 2.1 as well as those of Proposition 2.2.

Lemma A.5: Under Conditions (C0), (C1′) and (C2) of Section 2 there exists a function $M : \Omega \to (0,\infty]$ almost surely finite such that

$$\epsilon\sum_{l=0}^{\lfloor\gamma\epsilon^{-1}\rfloor-1}|||A_l(\omega)||| \le M(\omega)$$

for all $\omega\in\Omega$ and $0<\epsilon\le\gamma$, where $\gamma > 0$ is some constant.

Proof: We will just prove the result in the case where $\gamma = 1$ since the case $\gamma \neq 1$ is virtually identical. Fix an $\omega \in \Omega$, an $\epsilon \in (0,1]$ and a pair of integers $1 \leq s,t \leq d$. Then it follows by Condition (C2) (i) that there is some $c_1 > 0$ such that

$$\left|\epsilon \sum_{l=0}^{\lfloor \epsilon^{-1} \rfloor -1} a_l^{(s,t)}\right| \leq \left|\frac{1}{N_\epsilon} \sum_{l=0}^{N_\epsilon -1} \tilde{a}_l^{(s,t)}\right| + c_1, \tag{A.6}$$

where $N_\epsilon \triangleq \lfloor \epsilon^{-1} \rfloor$ and $a_l^{(s,t)}$ respectively $\tilde{a}_l^{(s,t)}$ is the $(s,t)^{\text{th}}$ component of A_l respectively $\tilde{A}_l \triangleq A_l - EA_l$. Now we have by Condition (C1′) (i) that

$$E\left|\sum_{l=j}^{j+n-1} \tilde{a}_l^{(s,t)}\right|^2 \leq n\, f(n) \tag{A.7}$$

for all integers $j \geq 0$ and $n \geq 1$, where $f(\cdot)$ is the function of Condition (C1′). Hence, it follows by Theorem A.6 (to follow) that

$$\frac{1}{n}\sum_{l=0}^{n-1} \tilde{a}_l^{(s,t)} \to 0 \text{ as } n \to \infty \text{ a.s.} \tag{A.8}$$

From (A.8) and (A.6) we obtain a function $\omega \to L_{s,t}(\omega)$ almost surely finite and independent of ϵ such that

$$\left|\epsilon \sum_{l=0}^{\lfloor \epsilon^{-1} \rfloor -1} a_l^{(s,t)}(\omega)\right| \leq L_{s,t}(\omega) \text{ for all } \omega \in \Omega \tag{A.9}$$

so letting $L(\omega) = d \cdot \max_{1 \leq s,t \leq d} L_{s,t}(\omega)$ for all $\omega \in \Omega$ it must follow that

$$|||\epsilon \sum_{l=0}^{\lfloor \epsilon^{-1} \rfloor -1} A_l(\omega)||| \leq L(\omega) \text{ for all } \omega \in \Omega. \tag{A.10}$$

Now since each $A_l(\omega)$ is symmetric and positive semidefinite, there exists a d by d matrix $P_l(\omega)$ such that $A_l(\omega) = P_l(\omega)P_l^T(\omega)$ for all $l = 0, \ldots, \lfloor \epsilon^{-1} \rfloor - 1$ and $\omega \in \Omega$ (see for example Proposition D.1.2 of Davis and Vinter [3]). Hence fixing $\omega \in \Omega$

$$||| \sum_{l=0}^{\lfloor \epsilon^{-1} \rfloor -1} A_l||| \geq \sum_{l=0}^{\lfloor \epsilon^{-1} \rfloor -1} a_l^{(i,i)} = \sum_{l=0}^{\lfloor \epsilon^{-1} \rfloor -1} |P_l^T e_i|^2 \tag{A.11}$$

for $i = 1, 2, \ldots, d$, where e_i is the i^{th} unit vector and by basic properties of induced matrix norms (see for example pages 14 and 15 of [9]) we have by letting $p_l^{(i,j)}$ denote the $(i,j)^{\text{th}}$ component of P_l that

$$\begin{aligned} d||| \sum_{l=0}^{\lfloor \epsilon^{-1} \rfloor -1} A_l||| &\geq \sum_{l=0}^{\lfloor \epsilon^{-1} \rfloor -1} \sum_{i=1}^{d}\sum_{j=1}^{d} (p_l^{(i,j)})^2 \\ &\geq \sum_{l=0}^{\lfloor \epsilon^{-1} \rfloor -1} |||P_l||| \cdot |||P_l^T||| \geq \sum_{l=0}^{\lfloor \epsilon^{-1} \rfloor -1} |||A_l|||. \end{aligned} \tag{A.12}$$

Substituting (A.10) into (A.12) we find

$$\epsilon \sum_{l=0}^{\lfloor \epsilon^{-1} \rfloor -1} |||A_l(\omega)||| \leq M(\omega), \tag{A.13}$$

where $M(\omega) \triangleq d \cdot L(\omega)$ is almost surely finite. □

The following strong law of large numbers, Theorem A.6, is a slight modification of Lai and Stout's law of large numbers [13, Theorem 7] and is proved in exactly the same manner as their result (let $p = 2$, $g(n) = nf(n)$ and replace their log terms with $n/f(n)$). We use Theorem A.6 in line (A.8) of Lemma A.5.

Theorem A.6: Suppose that $\{X_i, i = 0,1,2,\ldots\}$ is a sequence of random variables such that

$$E\left[\sum_{i=a}^{a+n-1} X_i\right]^2 \leq nf(n) \text{ for all } a \geq 0,\ n \geq 1,$$

where $\{f(n)\}_{n=1}^{\infty}$ is a nonnegative, nondecreasing sequence satisfying the constraints given in Condition (C1′) of Section II. Then $\frac{1}{n}\sum_{i=0}^{n-1} X_i \to 0$ almost surely.

We now commence establishing the as yet unproven assertions of Section IV. The first result, Lemma A.7, is used in Section IV in conjunction with Proposition 2.2 to prove that

$$\lim_{\substack{\epsilon \to 0 \\ \epsilon > 0}} \max_{0 \leq k \leq \lfloor \gamma \epsilon^{-\zeta} \rfloor} |h_k^\epsilon(\omega) - g^0(k\epsilon)| = 0 \text{ a.s.},$$

where $g^0(\cdot)$ is the solution of the differential equation given in (4.4).

Lemma A.7: Suppose $\{b_l,\ l = 0,1,2,\ldots\}$ and $\{A_l,\ l = 0,1,2,\ldots\}$ are respectively $\Re^d$- and $\Re^{d \times d}$-valued (nonrandom) sequences, b is a d-vector and A is a d by d-matrix such that

$$M \triangleq \sup_{l \geq 0}|b_l| < \infty, \quad P \triangleq \sup_{l \geq 0}|||A_l||| < \infty,$$

$$\limsup_{N \to \infty}\left|\frac{1}{N}\sum_{l=0}^{N-1} b_l - b\right| = 0,$$

$$\text{and} \quad \limsup_{N \to \infty}|||\frac{1}{N}\sum_{l=0}^{N-1} A_l - A||| = 0.$$

Then, it follows that for any constant $\gamma > 0$

$$\max_{0 \leq k \leq \lfloor \gamma \epsilon^{-1} \rfloor} |g_k^\epsilon - g^0(\epsilon k)| \to 0 \text{ as } \epsilon \to 0, \tag{A.14}$$

where $\{g_k^\epsilon,\ k = 0,1,2,\ldots\}$ and $g^0(\cdot)$ are defined by

$$g_{k+1}^\epsilon = g_k^\epsilon + \epsilon(b_k - A_k g_k^\epsilon) \text{ for all } k \geq 0, \epsilon > 0, \tag{A.15}$$

and

$$\dot{g}^0(\tau) = b - Ag^0(\tau) \text{ for all } 0 \leq \tau \leq 1 \tag{A.16}$$

subject to $g_0^\epsilon = g^0(0)$ is some fixed (independent of ϵ) vector.

Proof: Fix a $\delta > 0$, an $\epsilon > 0$ and an integer k such that $1 \le k \le \lfloor \gamma\epsilon^{-1} \rfloor$ and define $E \triangleq \sup_{0\le\tau\le\gamma} |g^0(\tau)|$. Then,

$$|g_k^\epsilon - g^0(k\epsilon)| \le \epsilon \left| \sum_{l=0}^{k-1} (b_l - A_l g^0(\epsilon l)) - \int_0^k (b - A g^0(\epsilon s)) ds \right| + \epsilon \sum_{l=0}^{k-1} |||A_l||| \cdot |g^0(\epsilon l) - g_l^\epsilon| \quad (A.17)$$

and defining $\bar{A}_l \triangleq A_l - A$ and $\bar{b}_l \triangleq b_l - b$ for all $l = 0, 1, 2, \ldots,$ we have for any $n > 0$ that

$$\begin{aligned} &\left| \sum_{l=0}^{k-1} (b_l - A_l g^0(\epsilon l)) - \int_0^k (b - A g^0(\epsilon s)) ds \right| \\ &\le \sum_{l=0}^{\lfloor \gamma\epsilon^{-1} \rfloor - 1} |||A||| \cdot |g^0(\epsilon l) - \int_l^{l+1} g^0(\epsilon s)\, ds| \\ &+ \sum_{l=0}^{\lfloor \gamma\epsilon^{-1} \rfloor - 1} |||\bar{A}_l||| \cdot \left| g^0(\epsilon l) - g^0\left(\frac{\lfloor l\epsilon n \rfloor}{n} \right) \right| \\ &+ \sum_{i=0}^{\lfloor k\epsilon n \rfloor - 1} \left| \sum_{l \in I_i^{\epsilon,n}} \left(\bar{b}_l - \bar{A}_l g^0(i/n) \right) \right| \\ &+ \left| \sum_{\substack{l \in I_{\lfloor k\epsilon n \rfloor} \\ l<k}} \left(\bar{b}_l - \bar{A}_l\, g^0\left(\frac{\lfloor k\epsilon n \rfloor}{n} \right) \right) \right|, \end{aligned} \quad (A.18)$$

where for each $i \in \{0, 1, 2, \ldots \lfloor k\epsilon n \rfloor\}$

$$I_i^{\epsilon,n} \triangleq \{ \lceil i\epsilon^{-1}n^{-1} \rceil, \lceil i\epsilon^{-1}n^{-1} \rceil + 1, \ldots, \lceil (i+1)\epsilon^{-1}n^{-1} \rceil - 1 \}. \quad (A.19)$$

(In the above line and in the remainder of this proof we use $\lceil x \rceil$ to represent the smallest integer not smaller than x where x is any real number.) Hence, if we make n large enough that

$$\max_{0\le s\le\gamma} |g^0(s) - g^0\left(\frac{\lfloor sn \rfloor}{n} \right)| < \frac{\delta}{\exp(P\gamma) \cdot P\gamma \cdot 8} \quad (A.20)$$

and

$$(M + PE)/n < \frac{\delta}{\exp(P\gamma) \cdot 8} \quad (A.21)$$

then we have by (A.18), (A.20) and (A.21) an $\epsilon_0(\delta) > 0$ such that

$$\begin{aligned} &\epsilon \left| \sum_{l=0}^{k-1} (b_l - A_l g^0(\epsilon l)) - \int_0^k (b - A g^0(\epsilon s)) ds \right| \\ &\le \gamma\epsilon P \cdot \max_{0\le l<\lfloor \gamma\epsilon^{-1} \rfloor} \int_l^{l+1} \int_l^s |b - A g^0(\epsilon t)| dt ds + \frac{\delta}{4} \exp(-P\gamma) \\ &+ 2\gamma \sum_{i=0}^{\lfloor \gamma n \rfloor - 1} \left| \frac{\epsilon n}{i+1} \sum_{l=0}^{\lceil (i+1)\epsilon^{-1}n^{-1} \rceil - 1} \left(\bar{b}_l - \bar{A}_l g^0(i/n) \right) \right| \\ &+ \epsilon \sum_{l=\lceil \lfloor k\epsilon n \rfloor \epsilon^{-1}n^{-1} \rceil}^{\lceil (\lfloor k\epsilon n \rfloor + 1)\epsilon^{-1}n^{-1} \rceil - 1} \left| \bar{b}_l - \bar{A}_l g^0\left(\frac{\lfloor k\epsilon n \rfloor}{n} \right) \right| \\ &\le \epsilon \frac{\gamma P}{2} (M + PE) + \frac{\delta}{2} \exp(-P\gamma) + 2(M + PE)(n^{-1} + \epsilon) \\ &\le \delta \exp(-P\gamma) \text{ for all } \epsilon < \epsilon_0(\delta). \end{aligned} \quad (A.22)$$

The lemma follows from (A.17), (A.22) and the discrete Bellman-Gronwall inequality. □

Next, under more stringent conditions than Lemma A.7, we obtain (by combining Lemmas A.8 and A.9) a rate of convergence in (A.14). In preparation for the statement of Lemma A.8 and Lemma A.9, we presuppose a (nonrandom) sequence, $\{A_l,\ l = 0, 1, 2, \ldots\}$, of d by d matrices, a (nonrandom) sequence, $\{b_l,\ l = 0, 1, 2, \ldots\}$, of d-vectors, a d by d matrix, A, and a d-vector, b. Finally, we define $\{g_k^\epsilon,\ k = 0, 1, 2, \ldots\}$ and $g^0(\cdot)$ as in (A.15) and (A.16) above and an additional sequence $\{v_k^\epsilon,\ k = 0, 1, 2, \ldots\}$ by

$$v_{k+1}^\epsilon = v_k^\epsilon + \epsilon \left(\frac{1}{N_\epsilon} \sum_{l=0}^{N_\epsilon - 1} b_{k+l} - \frac{1}{N_\epsilon} \sum_{l=0}^{N_\epsilon - 1} A_{k+l} v_k^\epsilon \right) \quad \text{for all } k \ge 0, \epsilon > 0, \quad (A.23)$$

subject to $v_0^\epsilon = g_0^\epsilon = g^0(0)$ for all $\epsilon > 0$, where $N_\epsilon \triangleq \epsilon^{-\zeta/2}$ for some constant $0 < \zeta \le 1$.

Lemma A.8: Suppose

$$\sup_{l\ge0} |||A_l||| < \infty \text{ and } \sup_{l\ge0} |b_l| < \infty. \quad (A.24)$$

Then, there exists a $c > 0$ independent of ϵ such that

$$\max_{0\le k\le\lfloor \gamma\epsilon^{-\zeta} \rfloor} |g_k^\epsilon - v_k^\epsilon| \le c\epsilon^{1-\zeta/2} \text{ for all } \epsilon > 0. \quad (A.25)$$

Proof: Lemma A.8 follows by an adaptation to the discrete-time setting of the arguments used to establish Lemma 3.2.9 of Sanders and Verhulst [18]

Lemma A.9: Suppose (A.24) is satisfied and

$$\sup_{N\ge0} \left|\left|\left| \sum_{l=0}^{N} (A_l - A) \right|\right|\right| < \infty \text{ and } \sup_{N\ge0} \left| \sum_{l=0}^{N} (b_l - b) \right| < \infty. \quad (A.26)$$

Then, there is a $c > 0$ independent of ϵ such that

$$\max_{0\le k\le\lfloor \gamma\epsilon^{-\zeta} \rfloor} |v_k^\epsilon - g^0(\epsilon k)| \le c\epsilon^{1-\zeta/2} \text{ for all } \epsilon > 0. \quad (A.27)$$

Proof: Lemma A.9 follows by an adaptation to the discrete-time setting of the arguments used to establish Lemma 3.3.2 of Sanders and Verhulst [18].

ACKNOWLEDGMENT

The author thanks Professor A. J. Heunis for reviewing a rough draft of this note and Professors A. J. Heunis and D. E. Miller for valuable suggestions related to this work. The author also thanks a reviewer for suggesting the current compact way of stating Proposition 2.1.

REFERENCES

[1] H. Berbee, "Convergence rates in the strong law for bounded mixing sequences," *Prob. Theory Rel. Fields*, vol. 74, pp. 255–270, 1987.

[2] N. N. Bogoliubov and Yu. A. Mitropol'skii, *Asymptotic Methods in the Theory of Non-Linear Oscillations.* Delhi, India: Hindustan, 1961. *International Monographs on Advanced Mathematics and Physics.*

[3] M. H. A. Davis and R. B. Vinter, *Stochastic Modelling and Control.* New York: Chapman and Hall, 1985.

[4] C. A. Desoer and M. Vidyasagar, *Feedback Systems: Input-Output Properties.* New York: Academic, 1975.

[5] P. Doukhan and F. Portal, "Principle d'invariance faible pour la fonction de repartition empirique dans un cadre multidimensionnel et melangeant," *Prob. Math. Stat.*, vol. 8, pp. 117–132, 1987.
[6] E. Eweda and O. Macchi, "Convergence of an adaptive linear estimation algorithm," *IEEE Trans. Automat. Contr.*, vol. AC-29, pp. 119–127, Feb. 1984.
[7] D. C. Farden, "Stochastic approximation with correlated data," *IEEE Trans. Inform. Theory*, IT-27, pp. 105–113, Jan. 1981.
[8] L. Gerencsér, "On a class of mixing processes," *Stochastics*, vol. 26, pp. 165–191, 1989.
[9] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: The Johns Hopkins University Press, 1983.
[10] P. Hall and C. C. Heyde, *Martigale Limit Theory and Its Application*. New York: Academic, 1980.
[11] A. J. Heunis, "Rates of convergence for an adaptive filtering algorithm driven by stationary dependent data," *SIAM J. Contr. Optimization*, to be published.
[12] H. J. Kushner and A. Shwartz, "Weak convergence and asymptotic properties of adaptive filters with constant gain," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 177–182, Mar. 1984.
[13] T. L. Lai and W. Stout, "Limit theorems for sums of dependent random variables," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 51, pp. 1–14, 1980.
[14] L. Ljung, "Convergence analysis of parametric identification methods," *IEEE Trans. Automat. Contr.*, AC-23(5), pp. 770–782, 1978.
[15] M. Longnecker and R. J. Serfling, "Moment inequalities for S_n under general dependence restrictions, with applications," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 43, pp. 1–21, 1978.
[16] O. M. Macchi, "Guest editorial," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 131–133, Mar. 1984.
[17] A. Mokkadem, "Mixing properties of ARMA processes," *Stochast. Process. Appl.*, vol. 29, pp. 309–315, 1988.
[18] J. A. Sanders and F. Verhulst, *Averaging Methods in Nonlinear Dynamical Systems*. New York:Springer-Verlag, 1985.
[19] V. Solo, "The error variance of LMS with time-varying weights," *IEEE Trans. Signal Processing*, vol. 40, pp. 803–813, Apr. 1992.
[20] W. F. Stout, *Almost Sure Convergence*. New York: Academic, 1974.
[21] V. Strassen, "An invariance principle for the law of the iterated logarithm," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 3, pp. 211–226, 1964.
[22] R. Yokoyama, "Moment bounds for stationary mixing sequences," *Z. Wahrscheinlichkeitstheorie Verw. Gebiete*, vol. 52, pp. 45–57, 1980.