JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS **187**, 134–155 (1994)

Strong Convergence in the Stochastic Averaging Principle

A. J. HEUNIS AND M. A. KOURITZIN

Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

Communicated by S. M. Meerkov

Received October 15, 1990

In this note we consider the almost sure convergence (as $\varepsilon \to 0$) of solution $X^\varepsilon(\cdot)$, defined over the interval $0 \le \tau \le 1$, of the random ordinary differential equation

$$\dot{X}^\varepsilon(\tau) = F(X^\varepsilon(\tau), \tau/\varepsilon) \qquad \text{subject to } X^\varepsilon(0) = x_0.$$

Here $\{F(x, t, \omega), t \ge 0\}$ is a strong mixing process for each x and $(x, t) \to F(x, t, \omega)$ is subject to regularity conditions which ensure the existence of a unique solution over $0 \le \tau \le 1$ for all $\varepsilon > 0$. Under rather weak conditions it is shown that the function $X^\varepsilon(\cdot, \omega)$ converges a.s. to the solution $x^0(\cdot)$ of a non-random averaged differential equation

$$\dot{x}^0(\tau) = \overline{F}(x^0(\tau)) \qquad \text{subject to } x^0(0) = x_0,$$

the convergence being uniform over $0 \le \tau \le 1$.

1. INTRODUCTION

Consider the following ordinary differential equation in $\Re^d$.

$$\dot{Z}^\varepsilon(t) = \varepsilon F(Z^\varepsilon(t), \xi(t)) \qquad \text{subject to } Z^\varepsilon(0) = x_0, \tag{1.1}$$

where $\varepsilon > 0$ is a small parameter, $\{\xi(t), t \ge 0\}$ is a given function, and the right-hand side of (1.1) is regular enough to ensure that there is a unique solution $Z^\varepsilon(\cdot)$ defined on all intervals of the form $[0, I/\varepsilon]$ (I a

0022-247X/94 $6.00

fixed finite number). Differential equations with the structure of (1.1) are common in physics and engineering, and one is usually interested in the asymptotic limit, if any, of the solution $Z^\varepsilon(\cdot)$ over the interval $[0, l/\varepsilon]$ as $\varepsilon \to 0$. The study of this question is usually simplified by making the substitution

$$X^\varepsilon(\tau) \triangleq Z^\varepsilon(\tau/\varepsilon), \qquad 0 \le \tau \le l,$$

in which case (1.1) assumes the form

$$\dot{X}^\varepsilon(\tau) = F(X^\varepsilon(\tau), \xi(\tau/\varepsilon)) \qquad \text{subject to } X^\varepsilon(0) = x_0, \tag{1.2}$$

and the problem in question reduces to that of determining the asymptotic limit, if any, of the solutions $X^\varepsilon(\cdot)$, defined over the interval $[0, l]$, as $\varepsilon \to 0$. When the limit

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T F(x, \xi(t))\, dt = \bar{F}(x) \tag{1.3}$$

exists for each x in $\Re^d$ then it seems reasonable to expect that the solution $x^0(\cdot)$ of the ordinary differential equation

$$\dot{x}(\tau) = \bar{F}(x(\tau)) \qquad \text{subject to } x(0) = x_0 \tag{1.4}$$

(assumed, for the time being, to exist on the interval $[0, l]$ and to be unique) approximates $X^\varepsilon(\cdot)$ uniformly on $[0, l]$ for all small values of ε. This general intuitive idea is known as the *averaging principle* and (1.4) is sometimes called the *averaged differential equation*. As noted in Section 52 of Arnol'd [1], the idea of the averaging principle originates in celestial mechanics and has been in use for several centuries. Clarification of the precise conditions on (1.1) which ensure that the averaging principle is valid is comparatively recent and is largely due to Soviet mathematicians. The averaging principle is so commonly used in physics and engineering that finding conditions under which it holds is an important question; indeed, Sanders and Verhulst [19, page 33] note an instance ([9]) where a use of the averaging principle outside its domain of applicability leads to erroneous results. The first conditions on (1.1) which imply the validity of the averaging principle are due to Bogoliubov (see Chapter 4 in [3]), Gihman [8], and Krasnoselskii and Krein [11], who establish the following: If the function $x \to F(x, t)$ is Lipschitz continuous with a Lipschitz constant that is global with respect to (x, t), the function $(x, t) \to F(x, t)$ is uniformly bounded on $D \otimes [0, \infty)$ and the convergence in (1.3) holds for each $x \in D$, where D is any bounded subset of $\Re^d$ which contains the trajectory $\{x^0(\tau), 0 \leq \tau < l\}$ within its interior, then the solution $X^\varepsilon(\cdot)$ of (1.2) converges to the solution $x^0(\cdot)$ of (1.3) uniformly on $[0, l]$ as $\varepsilon \to 0$. A detailed survey of these results, along with many extensions, can be found in Volosov [22] and Sanders and Verhulst [19].

The question of conditions under which the averaging principle holds can also be raised in the case where $\xi(t)$ in (1.1) is not a single given function but a random process defined on some probability space. Now the ensemble of solutions $X^\varepsilon(\cdot, \omega)$ of (1.2), found pointwise for each ω, constitutes a random process and it is reasonable to expect that if a "law of large numbers" effect causes a convergence such as that in (1.3), and the limiting vector-field $\overline{F}(x)$ is independent of ω, then the solutions $X^\varepsilon(\cdot)$ of (1.2) should be approximated, in some sense, by the *non-random* solution $x^0(\cdot)$ of (1.4) for all small values of ε. Indeed, a modification of the arguments of Gihman [8] gives conditions on (1.2) which ensure that the quantity

$$\sup_{0 \leq \tau \leq l} |X^\varepsilon(\tau) - x^0(\tau)| \tag{1.5}$$

converges to zero in probability (for the details see Theorem 3.1 on page 217 of Freidlin and Wentzell [6]), a result which can be regarded as a type of weak law of large numbers. A corresponding functional central limit theorem has been obtained by Khas'minskii [10, Theorem 3.1] who formulates conditions on (1.2) which ensure that $\varepsilon^{-1/2}(X^\varepsilon(\cdot) - x^0(\cdot))$ converges weakly in $C[0, l]$ to a certain limiting Gauss–Markov process, while Freidlin [7, Theorem 2.1] associates a *rate* with the convergence (in probability) of (1.5) to zero in the form of a large deviations principle.

In this note we are interested in formulating fairly weak conditions on (1.2) such that (1.5) converges to zero almost surely as $\varepsilon \to 0$, that is, there holds a strong law of large numbers for $X^\varepsilon(\cdot)$ corresponding to the weak law of large numbers mentioned above. Almost-sure convergence in the averaging principle has received considerable attention in the guise of the theory of *recursive stochastic algorithms* (see [15, 12, 2, 17]) where, in place of (1.1), one considers a difference rather than a differential equation, and the step size used in solving this difference equation is allowed to converge to zero (see e.g. (4.1) on page 146 of [17]). Because the step size decreases to zero, results about a.s. convergence for recursive stochastic algorithms are usually accompanied by a restriction that, with probability one, the iterations (generated by the algorithm) are infinitely often in a certain compact neighbourhood (see e.g. Theorem C on page 146 of [17]), and the proofs of convergence are usually quite long and involved. Such a restriction is not relevant to our problem, and the proof of a.s. convergence given here parallels arguments used for establishing simple strong laws of large numbers (see, e.g., Section 7 in Lamperti [14]).

The assumed moment bounds are weak, being only slightly greater than second order, and the type of dependency structure needed for $\{\xi(t)\}$ is strong mixing in the sense of Rosenblatt, which is actually one of the weaker mixing restrictions for which meaningful results can be formulated. Moreover, stationarity of $\{\xi(t)\}$ is not assumed.

In Section 2 we state the regularity conditions which are assumed throughout this note and make some remarks on these conditions. In Section 3 the proof of the main convergence result is given, while in Appendices 1–3 useful technical results are collected; these results are stated in a self-contained manner and are used in the proofs in Section 3. The main tools of proof are an adaptation to the continuous-parameter setting of moment bounds for strong mixing processes due to Serfling [20] and Sotres and Ghosh [21] as well as a maximal inequality due to Longnecker and Serfling [16].

2. Conditions

Suppose that $(\Omega, \mathcal{F}, P)$ is a probability space on which is defined a system of $\Re^d$-valued processes $\{F(x, s, \omega), s \geq 0\}$ indexed by $x \in \Re^d$ and jointly measurable in (s, ω) for each x. The following conditions are assumed throughout this note:

(C0) There exists a set $\Lambda_1 \in \mathcal{F}$, such that $P(\Lambda_1) = 0$ and, if $\omega \notin \Lambda_1$, then

$$\int_0^t |F(0, s)|\, ds < \infty \qquad \text{for all } t \geq 0.$$

Remark 2.1. Throughout this note $|\cdot|$ is used to denote the Euclidean norm in $\Re^d$ as well as the absolute value of a real number.

(C1) There exist non-negative functions $L(t, \omega)$ and $L_0(\omega)$ such that, for almost all ω;

(i) $\limsup_{T\to\infty} \frac{1}{T}\int_0^T L(t, \omega)\, dt \leq L_0(\omega) < \infty$

(ii) $|F(x, t) - F(x', t)| \leq L(t, \omega)|x - x'|$, for all $x, x' \in \Re^d$ and $t \geq 0$.

In view of (C0) and (C1) there exists some P-null set $\Lambda \in \mathcal{F}$ (not dependent on ε) such that, for each $\omega \notin \Lambda$ and $\varepsilon > 0$, the ordinary differential equation

$$\dot{x}(\tau) = F(x(\tau), \tau/\varepsilon), \qquad \text{subject to } x(0) = x_0, \tag{2.1}$$

has a unique solution, denoted by $X^\epsilon(\tau, \omega v)$, defined on the interval $0 \leq \tau \leq 1$. This follows from, for example, the standard theory in Chapter II of Reid [18].

(C2) There exist σ-algebras $\{\mathcal{F}_s^t, 0 \leq s \leq t \leq \infty\}$ such that for each x and $t \geq 0$, $F(x, t)$ is $\mathcal{F}_t^t$-measurable with respect to ω where

(i) $\mathcal{F}_s^t \subset \mathcal{F}$ for $0 \leq s \leq t \leq \infty$

(ii) $\mathcal{F}_s^t \subset \mathcal{F}_u^v$ for $0 \leq u \leq s \leq t \leq v \leq \infty$

(iii) There is some non-increasing function $\alpha(\cdot)$ such that

$$\sup_{t\geq 0} \sup_{\substack{A\in\mathcal{F}_0^t \\ B\in\mathcal{F}_{t+\tau}^\infty}} |P(A \cap B) - P(A)P(B)| \leq \alpha(\tau) \qquad \text{for } 0 \leq \tau \leq \infty. \tag{2.2}$$

Without loss of generality, it is assumed that $0 \leq \alpha(\tau) \leq 1$ for all $\tau \geq 0$.

(C3) There is some $\delta > 0$ such that

$$M \triangleq \sup_{t\geq 0} \| F(0, t) \|_{2+\delta} < \infty \qquad \text{and} \qquad N \triangleq \sup_{t\geq 0} \| L(t) \|_{2+\delta} < \infty,$$

where, for any $r > 0$ and random vector $X = (X^1 \cdots X^d)$, $\|X\|_r$ denotes the Euclidean length of the vector $(\|X^1\|_r \cdots \|X^d\|_r)$ and $\|Z\|_r$ is the L_r-norm of a random variable Z.

(C4) The Rosenblatt mixing coefficient, α, defined in (C2) satisfies

$$\alpha(\tau) \leq \eta\tau^{-\theta}$$

for some constants $\theta > 1 + 2\delta^{-1}$, $\eta > 0$, and all $\tau \geq 1$, where δ is the constant of (C3).

(C5) For each $x \in \Re^d$, the limit

$$\bar{F}(x) \triangleq \lim_{T\to\infty} \frac{1}{T} \int_0^T EF(x, t)\, dt$$

exists. Note that by (C1, ii) and (C3) the above integral is well defined for $T \geq 0$.

Now, by condition (C1) for $x, x' \in \Re^d$;

$$|EF(x', t) - EF(x, t)| \leq \sup_{t\geq 0} EL(t) \cdot |x' - x| \leq N \cdot |x' - x|. \tag{2.3}$$

Clearly, in view of (C5) and (2.3), for $x, x' \in \Re^d$;

$$|\overline{F}(x) - \overline{F}(x')| \leq \limsup_{T\to\infty} \frac{1}{T}\int_0^T |EF(x,t) - EF(x',t)|\,dt \leq N|x - x'|, \quad (2.4)$$

so $\overline{F}$ has a global Lipschitz constant N and the differential equation

$$\dot{x}(\tau) = \overline{F}(x(\tau)), \qquad \text{subject to } x(0) = x_0, \quad (2.5)$$

has a unique solution, $x^0(\tau)$, on $0 \leq \tau \leq 1$. Moreover, for any $\tau \in [0, 1]$,

$$|x^0(\tau) - x_0| \leq \left|\int_0^\tau \overline{F}(x^0(s)) - \overline{F}(x_0)\,ds\right| + \tau|\overline{F}(x_0)|$$

and so by the Gronwall inequality (see for example Lemma 1.1 in Chapter 2 of Freidlin and Wentzell [6]) and (2.4),

$$\sup_{0\leq\tau\leq 1} |x^0(\tau)| \leq e^N|\overline{F}(x_0)| + |x_0| \triangleq D. \quad (2.6)$$

Moreover, from condition (C3), and Eqs. (2.4) and (2.6);

$$\begin{aligned} |x^0(\tau) - x^0(\tau')| &\leq \int_{\tau'}^{\tau} |\overline{F}(x^0(s)) - \overline{F}(0)|\,ds + \int_{\tau'}^{\tau} |\overline{F}(0)|\,ds \\ &\leq (ND + M)|\tau - \tau'| \end{aligned} \quad (2.7)$$

whenever $0 \leq \tau' \leq \tau \leq 1$. The bounds (2.6) and (2.7) are useful for proofs in later sections.

The following definitions are made for ease of notation:

$$\tilde{F}(x, t) \triangleq F(x, t) - EF(x, t) \quad (2.8)$$

and

$$\|\Psi\|_C \triangleq \max_{0\leq t\leq 1} |\Psi(t)| \qquad \text{for all } \Psi \in C[0, 1]. \quad (2.9)$$

Remark 2.2. Conditions (C2) and (C4) ensure that there is enough mixing for a law of large numbers effect to average out the right-hand side of (1.2). Note the trade-off involved between (C3) and (C4): weak moment restrictions (corresponding to small values of $\delta > 0$ in (C3)) require a fast rate of mixing in (C4), while a slow rate of mixing is permissible when strong moment bounds (i.e., large values of δ) are postulated in (C3).

Remark 2.3. The strong mixing in the sense of Rosenblatt assumed in (C2) is among the least restrictive of a variety of mixing hypotheses. A comparison of mixing conditions can be found in Bradley [4, pages 165–192].

Remark 2.4. In the great majority of applications condition (C1) holds with $L(t, \omega) \equiv L_0$ for all (t, ω) where L_0 is a constant. This corresponds to $F(x, t, \omega)$ having a global Lipschitz constant L_0 with respect to x in (C1). However, there are right-hand sides in (2.1), often suggested by problems in data-communication and adaptive filtering, where the extra generality provided by condition (C1) proves useful. For example, if $\{Y(t)\}$ and $\{\psi(t)\}$ are given random processes defined on a common probability space such that $Y(t, \omega)$ is a d by 1 column vector and $\psi(t, \omega)$ is real-valued, then the continuous-parameter version of a commonly used adaptive filter (see e.g. (1.1) in Kushner and Schwartz [1]) is given by

$$\dot{Z}^{\varepsilon} = \varepsilon F(Z^{\varepsilon}(t), t), \qquad t \geq 0,$$

where

$$F(x, t, \omega) \triangleq -(Y(t) \cdot Y^T(t))x + Y(t)\psi(t).$$

Here there is generally no global Lipschitz constant for $F(x, t, \omega)$ with respect to x (unless we assume the presence of a uniform bound for the function $(t, \omega) \to Y(t, \omega)$, which is generally an unacceptable restriction in adaptive filtering), but, as is seen later, it is easily verified that (C1) holds if the processes $\{Y(t)\}$ and $\{\psi(t)\}$ are strong mixing with a suitable mixing rate. Condition (C1) has been motivated by condition (E4) on page 146 of Metivier and Prioret [17].

Remark 2.5. Our goal is to establish a.s. convergence of $X^{\varepsilon}(\cdot)$ to $x^0(\cdot)$ uniformly on the unit interval [0, 1]. One possible approach is to show that the conditions of the deterministic averaging principle (see, for example, page 429 of Bogoliubov and Mitropolskii [3]) are satisfied "pointwise" for almost all ω. However, proofs of the deterministic averaging principle depend quite strongly on *uniform boundedness* of the right-hand side with respect to (x, t) taken over $D \otimes [0, \infty)$ where D is some (usually bounded) subset of $\Re^d$. This uniform boundedness condition may fail to hold almost surely for random right-hand sides $F(x, t, \omega)$, as can be seen by considering e.g. the very common right-hand side

$$F(x, t, \omega) \triangleq Ax + \xi(t),$$

where $\{\xi(t)\}$ is a strong mixing $\mathfrak{R}^d$-valued Gaussian process and A is a fixed square matrix. Here, for arbitrary ω, there generally fails to exist a finite number $M(\omega)$ such that

$$\sup_{\substack{x \in D \\ t \geq 0}} |F(x, t, \omega)| \leq M(\omega)$$

since one cannot ensure that $\sup_{t \geq 0} |\xi(t)| < \infty$ a.s. for a Gaussian process. Actually, one can adapt existing proofs of the deterministic averaging principle to overcome this problem and to fit conditions (C0) to (C5), but this requires showing that, for almost all ω, one has

$$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} \varepsilon \left[\sup_{\substack{0 \leq t < \varepsilon^{-1} \\ x \in D}} \left| \int_0^t \{F(x, v, \omega) - \overline{F}(x)\} \, dv \right| \right] = 0$$

for some bounded set $D \subset \mathfrak{R}^d$ which contains the trajectory $\{x^0(\tau), 0 \leq \tau \leq 1\}$ in its interior (see the definition of $F(\varepsilon)$ after (26.16) on page 432 of Bogoliubov and Mitropol'skii [3] along with the assertion that $F(\varepsilon) \to 0$ as $\varepsilon \to 0$). Rather than adopt this approach it is easier to give a direct proof of convergence.

Remark 2.6. In this note convergence of $X^\varepsilon(\cdot)$ to $x^0(\cdot)$ uniformly on the interval $[0, 1]$ is considered. The results generalise in a trivial manner to uniform convergence on any finite interval $[0, l]$.

3. Main Result

The main result of this note is the following strong convergence result. It is a functional strong law of large numbers for the stochastic averaging principle when the "driving" random process is Rosenblatt mixing.

Proposition 1. *Under the conditions* (C0)–(C5) *of Section* 2;

$$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} \|X^\varepsilon - x^0\|_C = 0 \qquad \text{a.s.,}$$

where $\{X^\varepsilon(\cdot), \varepsilon > 0\}$ and $x^0(\cdot)$ are as defined in Section 2.

Proof. Without loss of generality the P-null set Λ originating in the conditions (C0) and (C1) of Section 1 can be assumed empty since Λ can always be subtracted from the underlying probability space $(\Omega, \mathcal{F}, P)$ to yield a new probability space. Thus we suppose that $X^\varepsilon(\tau, \omega)$ is defined on $0 \le \tau \le 1$ for each $\varepsilon > 0$ and $\omega \in \Omega$.

Fix an $\omega \in \Omega$. For any $0 \le \tau \le 1$, $0 < \varepsilon \le 1$,

$$\begin{aligned} |X^\varepsilon(\tau) - x^0(\tau)| \le & \left| \int_0^\tau F(X^\varepsilon(s), s/\varepsilon) - F(x^0(s), s/\varepsilon)\, ds \right| \\ & + \left| \int_0^\tau F(x^0(s), s/\varepsilon) - EF(x^0(s), s/\varepsilon)\, ds \right| \\ & + \left| \int_0^\tau EF(x^0(s), s/\varepsilon) - \bar{F}(x^0(s))\, ds \right| \end{aligned} \tag{3.1}$$

and by condition (C1), Eq. (2.8), and Lemma A.2 of Appendix 3 we have

$$\begin{aligned} |X^\varepsilon(\tau) - x^0(\tau)| \le & \int_0^\tau L(s/\varepsilon, \omega)|X^\varepsilon(s) - x^0(s)|\, ds \\ & + \max_{0 \le \tau \le 1} \left| \int_0^\tau \tilde{F}(x^0(s), s/\varepsilon)\, ds \right| + \rho(\varepsilon) \end{aligned} \tag{3.2}$$

for some function $\rho(\varepsilon) \to 0$ as $\varepsilon \to 0$. By (3.2) and the Bellman–Gronwall inequality (see [5, page 252])

$$\begin{aligned} & |X^\varepsilon(\tau) - x^0(\tau)| \\ & \le \exp\left(\int_0^1 L(s/\varepsilon, \omega)\, ds \right) \left\{ \max_{0 \le \tau \le 1} \left| \int_0^\tau \tilde{F}(x^0(s), s/\varepsilon)\, ds \right| + \rho(\varepsilon) \right\} \end{aligned} \tag{3.3}$$

for all $0 \le \tau \le 1$. But by conditions (C1.i) there is some $A(\omega) < \infty$ such that

$$\int_0^1 L(s/\varepsilon, \omega)\, ds = \varepsilon \int_0^{\varepsilon^{-1}} L(s, \omega)\, ds \le A(\omega) \qquad \text{for all } 0 < \varepsilon \le 1$$

and so there exists constant $B(\omega) < \infty$ such that

$$|X^\varepsilon(\tau) - x^0(\tau)| \le B(\omega) \left\{ \max_{0 \le \tau \le 1} \left| \int_0^\tau \tilde{F}(x^0(s), s/\varepsilon)\, ds \right| + \rho(\varepsilon) \right\} \quad \text{for all } 0 \le \tau \le 1. \tag{3.4}$$

For ease of notation define $T_\varepsilon \triangleq \varepsilon^{-1}$, $n^\varepsilon \triangleq [T_\varepsilon]$, and

$$\tilde{F}_U(s) \triangleq \begin{cases} \tilde{F}(x^0(s/U), s) & \text{for } 0 \le s \le U \\ 0 & \text{for } s > U, \end{cases} \tag{3.5}$$

where $[a]$ denotes the largest integer not greater than a. We now reduce convergence (as $\varepsilon \to 0$) of the *set* of functions $\{X^\varepsilon(\cdot)\}$ to convergence of some *sequence* of functions, to which a.s. convergence arguments using the Borel–Cantelli lemma can be applied. By (3.4), (3.5), the change of variables formula, and the fact $T_\varepsilon \geq 1$:

$$\begin{aligned}\|X^\varepsilon - x^0\|_C &\leq B(\omega)\left\{\frac{1}{T_\varepsilon}\max_{0\leq t\leq T_\varepsilon}\left|\int_0^t \bar{F}_{T_\varepsilon}(s)\,ds\right| + \rho(\varepsilon)\right\} \\ &\leq B(\omega)\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}|\bar{F}_{T_\varepsilon}(s) - \bar{F}_{n_\varepsilon+1}(s)|\,ds \\ &\quad + B(\omega)\frac{1}{T_\varepsilon}\max_{0\leq t\leq n_\varepsilon+1}\left|\int_0^t \bar{F}_{n_\varepsilon+1}(s)\,ds\right| + B(\omega)\rho(\varepsilon) \qquad (3.6) \\ &\leq B(\omega)\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}|\bar{F}_{T_\varepsilon}(s) - \bar{F}_{n_\varepsilon+1}(s)|\,ds \\ &\quad + B(\omega)\frac{2}{n_\varepsilon+1}\max_{0\leq t\leq n_\varepsilon+1}\left|\int_0^t \bar{F}_{n_\varepsilon+1}(s)\,ds\right| + B(\omega)\rho(\varepsilon)\end{aligned}$$

(since $1 + n_\varepsilon \leq 2T_\varepsilon$ for all $0 < \varepsilon \leq 1$). It is now shown that each of the terms on the far right of (3.6) converges a.s. to zero as $\varepsilon \to 0$. By (3.5), (2.3), (2.7), and condition (C1, ii and i) there exists $H(\omega) < \infty$ such that

$$\begin{aligned}&\frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}|\bar{F}_{T_\varepsilon}(s) - \bar{F}_{n_\varepsilon+1}(s)|\,ds \\ &\quad \leq \frac{1}{T_\varepsilon}\int_0^{T_\varepsilon}\{L(s,\omega) + N\}[ND + M]\left(\frac{s}{T_\varepsilon} - \frac{s}{n_\varepsilon+1}\right)ds \qquad (3.7) \\ &\quad \leq (ND + M)\frac{n_\varepsilon + 1 - T_\varepsilon}{T_\varepsilon^2}\int_0^{T_\varepsilon} L(s) + N\,ds \leq \frac{ND + M}{T_\varepsilon}H(\omega),\end{aligned}$$

since by (2.8) and (2.3), $\bar{F}(x, t)$ has a Lipschitz constant of $L(t, \omega) + N$ with respect to x. Now consider the second term on the right of (3.6). By (3.5) and Lemma A1 of Appendix 3;

$$\|\bar{F}_n(s)\|_{2+\delta} \leq 2M + 2ND < \infty, \qquad \text{for all } s \geq 0, n = 1, 2, \ldots \qquad (3.8)$$

Thus, in view of conditions (C2) and (C4) and Corollary A1 of Appendix 1, there exists some $\beta > 0$ and $\Gamma > 0$ such that

$$E\left|\int_t^u \bar{F}_n(s)\,ds\right|^{2+\beta} \leq \Gamma\cdot(u - t)^{1+\beta/2} \leq [h(t, u)]^{1+\beta/2} \qquad (3.9)$$

for all $n = 1, 2, 3, \ldots$ and $0 \le t \le u \le n$ where

$$h(t, u) \triangleq \Gamma^{2/(2+\beta)} \cdot (u - t).$$

Thus by (3.9) and Corollary A2 of Appendix 2 (with $\gamma \triangleq 1 + \beta/2$, $\nu \triangleq 2 + \beta$), there exists some $A_\beta > 0$ depending only on β such that

$$E\left\{\max_{0\le t\le n}\left|\int_0^t \bar{F}_n(s)\,ds\right|^{2+\beta}\right\} \le A_\beta \Gamma\, n^{1+\beta/2}, \qquad n = 1, 2, 3, \ldots \tag{3.10}$$

By (3.10) and the Chebyshev inequality;

$$\begin{aligned} P\left\{\frac{1}{n}\max_{0\le t\le n}\left|\int_0^t \bar{F}_n(s)\,ds\right| \ge n^{-\beta/(s+4\beta)}\right\} &\le \frac{A_\beta \Gamma n^{1+\beta/2}}{n^{2+(3/4)\beta}} \\ &\le A_\beta \Gamma n^{-(1+\beta/4)}, \end{aligned} \tag{3.11}$$

so by the Borel Cantelli lemma;

$$\frac{1}{n}\max_{0\le t\le n}\left|\int_0^t \bar{F}_n(s)\,ds\right| \to 0 \qquad \text{as } n\to\infty \text{ a.s.} \tag{3.12}$$

Letting $n \triangleq n_\varepsilon + 1$ in (3.12) for the second term on the right of (3.6), we have

$$\|X^\varepsilon - x^0\|_C \to 0 \qquad \text{as } \varepsilon \to 0 \text{ a.s.} \tag{3.13}$$

by (3.6), (3.7), (3.12), and the fact that $\rho(\varepsilon) \to$ as $\varepsilon \to 0$ (see Lemma A2). ∎

Remark 3.1. As an example consider the continuous-parameter version of the adaptive filtering algorithm mentioned in Remark 2.4 Let

$$F(x, t, \omega) \triangleq -(Y(t, \omega) Y^T(t, \omega))x + \psi(t) Y(t) \tag{3.14}$$

for all $x \in \Re^d$, $t \ge 0$, $\omega \in \Omega$, where $\{\psi(t), t \ge 0\}$ and $\{Y(t), t \ge 0\}$ are, respectively, real and $\Re^d$ valued jointly measurable stochastic processes defined on a common probability space (Ω, F, P). We assume that $\{\psi(t)\}$ and $\{Y(t)\}$ satisfy the following conditions.

(A1) There is some $\delta > 0$ such that, for all $i, j = 1, 2, \ldots, d$,

$$\sup_{t\ge 0} \| Y^i(t) Y^j(t)\|_{2+\delta} < \infty \qquad \text{and} \qquad \sup_{t\ge 0} \|\psi(t) Y^j(t)\|_{2+\delta} < \infty$$

(where $Y^i(t)$ is the ith element of the vector $Y(t)$).

(A2) There is a $d \times d$ matrix R, along with a d-vector b, such that

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T E(Y(t)Y^T(d))\,dt = R \qquad \text{and} \qquad \lim_{T\to\infty}\frac{1}{T}\int_0^T E(Y(t)\psi(t))\,dt = b.$$

(A3) There are σ-algebras $\{\mathscr{F}_s^t,\ 0 \le s \le t \le \infty\}$ which satisfy (i–iii) of condition (C2) of Section 2, such that $Y(t, \omega)$ and $\psi(t, \omega)$ are $\mathscr{F}_t^t$-measurable in ω for all $t \ge 0$. Moreover, the Rosenblatt mixing coefficient $\alpha(\cdot)$ for these σ-algebras satisfies

$$\alpha(\tau) \le \eta\tau^{-\theta}, \qquad \text{for all } \tau \ge 1,$$

where $\eta > 0$ and $\theta > 1 + 2\delta^{-1}$ are constants (δ being given by (A1)).

It then follows at once that (C0) and (C2)–(C5) in Section 2 hold, where

$$\bar{F}(x) \triangleq -Rx + b. \tag{3.15}$$

As for condition (C1), if we define

$$L(t, \omega) \triangleq |||\, Y(t, \omega) Y^T(t, \omega)\, |||$$

and

$$L_0(\omega) \triangleq |||\, R\, ||| \qquad \text{for all } t \ge 0,\ \omega \in \Omega$$

(here $|||\cdot|||$ denotes the Euclidean norm of a $d \times d$ matrix), then it is a consequence of the law of large numbers for non-stationary strong mixing processes (see Theorem 2.2 in Sotres and Ghosh [21]) that

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T L(t, \omega)\,dt = L_0 \qquad \text{a.s.}$$

so condition (C1) holds as well. Applying Proposition 1, we see that

$$\sup_{0\le\tau\le 1} |X^\varepsilon(\tau) - x^0(\tau)| \tag{3.16}$$

converges to zero a.s. as $\varepsilon \to 0$, where $X^\varepsilon(\cdot)$ and $x^0(\cdot)$ are the solutions of (2.1) and (2.5) with right-hand sides given by (3.14) and (3.15), respectively. This strong convergence result complements the convergence *in probability* to zero of (3.16) which was established by Kushner and Shwartz [13, page 178] (in the discrete parameter setting) using a method based on the Stroock–Varadhan martingale problem and weak convergence of probability measures. Note that conditions (A1) to (A3) imply only very weak stationarity restrictions on the given processes $\{\psi(t)\}$ and $\{Y(t)\}$, which is in contrast to many studies of this adaptive filtering algorithm where rather strong stationarity is often assumed. For example, if (A1) and (A3) hold, and $t \to E(Y(t)\ Y^T(t))$ and $t \to E(Y(t)\psi(t))$ are periodic with period λ, then (A2) holds as well with

$$R \triangleq \frac{1}{\lambda}\int_0^\lambda E(Y(t)Y^T(t))\,dt, \qquad b \triangleq \frac{1}{\lambda}\int_0^\lambda E(Y(t)\psi(t))\,dt.$$

APPENDIX 1: Moment Bounds

In the development of Proposition 1 we required moment bounds on row-wise integrals of an array of continuous-parameter strong mixing stochastic processes $\{\tilde{F}_n(t),\ t \geq 0\}$, for $n = 1, 2, \ldots$ (see Eq. (3.9)). The following result, which is a consequence of bounds obtained by Sotres and Ghosh [21] and Serfling [20], is the main tool for establishing these bounds.

Theorem A1. *Let $\{X_i,\ i = 1, 2, \ldots\}$ be a zero mean $\Re^d$-valued process on $(\Omega, \mathcal{F}, P)$ such that*

(a) $E|X_i|^{2+\delta} \leq M < \infty$ *for all* $1 \leq i < \infty$ *and some* $\delta > 0$ *and*

(b) *there exists some non-increasing sequence of numbers* $\{\tilde{\alpha}(j),\ j = 1, 2, \ldots\}$ *such that*

$$\sup_{k \geq 1} \sup_{\substack{A \in \mathcal{M}_1^k \\ B \in \mathcal{M}_{k+j}^\infty}} |P(A \cap B) - P(A)P(B)| \leq \tilde{\alpha}(j)$$

(where $\mathcal{M}_a^b \triangleq \sigma\{X_i,\ a \leq i \leq b\}$ for $1 \leq a \leq b < \infty$, $\mathcal{M}_a^\infty \triangleq \sigma\{X_i,\ a \leq i < \infty\}$ for $1 \leq a < \infty$) and

$$\tilde{\alpha}(\tau) \leq \tilde{\eta} j^{-\theta} \qquad \textit{for all } j = 1, 2, 3, \ldots,$$

where $\tilde{\eta} > 0$ and $\theta > 1 + 2\delta^{-1}$ are constants. Then there exist constants $\tilde{\Gamma} > 0$ and $0 < \beta < \delta$ such that

$$E\left|\sum_{i=a+1}^{a+m} X_i\right|^{2+\beta} \leq \tilde{\Gamma} \cdot m^{1+\beta/2} \qquad \text{(A1.1)}$$

for all integers $a \geq 0$ and $m \geq 1$, where β and $\tilde{\Gamma}$ depend only on M, θ, $\bar{\eta}$, and δ.

Proof. Under conditions (a) and (b), Sotres and Ghosh show (see pages 3 and 4 in [21]) that there exist finite positive constants c_1, c_2, s, depending only on M, δ, $\bar{\eta}$, and θ, such that

$$E\left[m^{-1}\left(\sum_{i=a+1}^{a+m} X_i\right)^2\right] \leq c_1 \tag{A1.2}$$

and

$$E\left|E\left(m^{-1}\left(\sum_{i=a+1}^{a+m} X_i\right)^2 \middle| \mathcal{M}_1^a\right) - E\left[m^{-1}\left(\sum_{i=a+1}^{a+m} X_i\right)^2\right]\right| \leq c_2 m^{-s} \tag{A1.3}$$

for all integers $a \geq 0$ and $m \geq 1$. Now Theorem A1 follows from these bounds and the following theorem which is due to Serfling [20].

THEOREM. *Under the conditions* (a) *of Theorem* A1 *and Eqs.* (A1.2) *and* (A1.3) *there exists some $\tilde{\Gamma} < \infty$ and $\beta > 0$ depending only on M, δ, c_1, c_2, and s such that*

$$E\left|m^{-1/2} \sum_{i=a+1}^{a+m} X_i\right|^{2+\beta} \leq \tilde{\Gamma}$$

for all integers $a \geq 0$ and $m \geq 1$. ∎

Remark A1.1. Lines (A1.2) and (A1.3) above correspond to bounds (2.9) and (2.10) respectively on page 3 of [21]. The above theorem of Serfling is a direct consequence of Lemma 2.1 of [20] and the sufficiency part of Theorem 3.1 of [20]. It should be noted that the proof of the sufficiency part of Theorem 3.1 continues to hold with no changes at all when condition (A2) on page 1158 in [20] is replaced with the slightly weaker condition that (A1.2) hold for all integers $a \geq 0$, $m \geq 1$.

The next corollary adapts Theorem A1 to continuous parameter processes and is used in line (3.9) of Section 3.

COROLLARY A1. *Suppose that $\{\xi(s), s \geq 0\}$ is a zero mean, $\Re^d$-values, jointly measurable process on $(\Omega, \mathcal{F}, P)$ such that*

(a) $E|\xi(s)|^{2+\delta} \leq M < \infty$ *for all $s \geq 0$ and some $\delta > 0$.*

(b) *There exists σ-algebras $\{\mathcal{F}_s^t, 0 \leq s \leq t \leq \infty\}$ which satisfy* (i), (ii), *and* (iii) *of condition* (C2), *Section* 1, *such that $\xi(t)$ is $\mathcal{F}_t^t$-measurable in ω for all $t \geq 0$, and let $\alpha(\cdot)$ be as in* (2.2).

(c) *There are constants $\theta > 1 + 2\delta^{-1}$ (where $\delta > 0$ is the constant in* (a)) *and $\eta \geq 0$ such that the mixing coefficient $\alpha(\cdot)$ in* (b) *satisfies*

$$\alpha(\tau) \leq \eta\tau^{-\theta} \qquad \text{for all } \tau \geq 1.$$

Then there exist constants $\Gamma > 0$ and $0 < \beta < \delta$ such that

$$E\left|\int_t^u \xi(s)\,ds\right|^{2+\beta} \leq \Gamma \cdot (u-t)^{1+\beta/2} \qquad \text{for all } 0 \leq t \leq u. \tag{A1.4}$$

where Γ and β depend only on M, δ, η, and θ.

Proof. Without loss of generality, we assume that $\eta \geq 2^{-\theta}$.

(I) Consider the case $0 \leq u - t \leq 1$, where $t, u \geq 0$. The case $t = u$ is trivial, so assume that $0 < u - t \leq 1$. For any $0 < \beta < \delta$, we have, by Jensen's inequality, Fubini's theorem, and Holder's inequality,

$$\begin{aligned} E\left|\int_t^u \xi(s)\,ds\right|^{2+\beta} &= (u-t)^{2+\beta} E\left|\int_{t/(u-t)}^{u/(u-t)} \xi((u-t)s')\,ds'\right|^{2+\beta} \\ &\leq (u-t)^{2+\beta} E\int_{t/(u-t)}^{u/(u-t)} |\xi((u-t)s')|^{2+\beta}\,ds' \qquad \text{(A1.5)} \\ &\leq (u-t)^{2+\beta}\int_{t/(u-t)}^{u/(u-t)} (E|\xi((u-t)s')|^{2+\delta})^{(2+\beta)/(2+\delta)}\,ds'. \end{aligned}$$

Therefore by (A1.5), hypothesis (a), and the fact that $0 \leq u - t \leq 1$,

$$\begin{aligned} E\left|\int_t^u \xi(s)\,ds\right|^{2+\beta} &\leq (u-t)^{2+\beta} M^{(2+\beta)/(2+\delta)} \\ &\leq (u-t)^{1+\beta/2} M^{(2+\beta)/(2+\delta)} \end{aligned} \tag{A1.6}$$

(II) Consider the case $1 < u - t$, where $t, u \geq 0$. Then

$$\int_t^u \xi(s)\,ds = \int_t^{1+[t]} \xi(s)\,ds + \sum_{i=2+[t]}^{[u]} X_i + \int_{[u]}^u \xi(s)\,ds \qquad \text{a.s.,} \tag{A1.7}$$

where

$$X_i \triangleq \int_{i-1}^i \xi(s)\,ds \qquad \text{for all } i = 1, 2, \ldots \qquad \text{a.s.} \tag{A1.8}$$

For any $0 < \beta \le \delta$ we have, by (A1.7), Part (I), and the fact that $u - t > 1$,

$$\begin{aligned} E\left|\int_t^u \xi(s)\,ds\right|^{2+\beta} &\le 3^{2+\beta}\left\{E\left|\int_t^{1+[t]} \xi(s)\,ds\right|^{2+\beta} + E\left|\sum_{i=2+[t]}^{[u]} X_i\right|^{2+\beta} + E\left|\int_{[u]}^u \xi(s)\,ds\right|^{2+\beta}\right\} \\ &\le 3^{2+\beta}M^{(2+\beta)/(2+\delta)}\{([t]+1-t)^{1+\beta/2} \qquad \text{(A1.9)} \\ &\quad + (u-[u])^{1+\beta/2}\} + 3^{2+\beta}E\left|\sum_{i=2+[t]}^{[u]} X_i\right|^{2+\beta} \\ &\le 2\cdot 3^{2+\beta}M^{(2+\beta)/(2+\delta)}(u-t)^{1+\beta/2} + 3^{2+\beta}E\left|\sum_{i=2+[t]}^{[u]} X_i\right|^{2+\beta}. \end{aligned}$$

By Jensen's inequality, Fubini's theorem, and (a),

$$E|X_i|^{2+\delta} \le E\int_{i-1}^{i} |\xi(s)|^{2+\delta}\,ds \le M, \qquad \text{for all } i = 1, 2, \ldots, \qquad \text{(A.10)}$$

and

$$EX_i = \int_{i-1}^{i} E\xi(s)\,ds = 0. \qquad \text{(A.11)}$$

Define

$$\mathcal{M}_a^b \triangleq \sigma\{X_i, a \le i \le b\} \quad \text{for all integers } 1 \le a \le b < \infty \qquad \text{(A.12)}$$

$$\mathcal{M}_a^\infty \triangleq \sigma\{X_i, a \le i < \infty\} \qquad \text{for all integers } 1 \le a < \infty$$

$$\tilde{\alpha}(j) \triangleq \begin{cases} \alpha(j-1) & \text{for } j = 2, 3, \ldots \\ \eta 2^\theta & \text{for } j = 1 \end{cases} \qquad \text{(A1.13)}$$

Then $\mathcal{M}_1^k \subset \mathcal{F}_0^k$, $\mathcal{M}_{k+j}^\infty \subset \mathcal{F}_{k+j-1}^\infty$, and $\tilde{\alpha}(j) \le \eta 2^\theta j^{-\theta}$ for all integers $j, k = 1, 2, 3, \ldots$, so by condition (b),

$$\sup_{k\ge 1}\ \sup_{\substack{A\in\mathcal{M}_1^k \\ B\in\mathcal{M}_{k+j}^\infty}} |P(A\cap B) - P(A)P(B)| \le \tilde{\alpha}(j) \qquad \text{(A1.14)}$$

Also, $\tilde{\alpha}(0) \geq \tilde{\alpha}(1) \geq \ldots$. Therefore, by Theorem A1, there are $0 < \beta < \delta$ and $\tilde{\Gamma} > 0$ depending only on M, θ, η, and δ such that

$$E\left|\sum_{i=2+[t]}^{[u]} X_i\right|^{2+\beta} \leq \tilde{\Gamma}([u] - [t] - 1)^{1+\beta/2} \leq \tilde{\Gamma}(u - t)^{1+\beta/2}. \quad \text{(A1.15)}$$

By (A1.9) and (A1.15), there exists $\Gamma > 0$ depending only on M, θ, η, δ and β such that

$$E\left|\int_t^u \xi(s)\, ds\right|^{2+\beta} \leq \Gamma(u - t)^{1+\beta/2} \quad \text{(A1.16)}$$

and the corollary follows from (A1.6) and (A1.16). ∎

Remark A1.2. The essence of the above-quoted bounds of Serfling [20] and Sotres and Ghosh [21] is that if $\{X_i\}$ is a zero mean sequence, suitably bounded and mixing, such that

$$E\left[\left(m^{-1/2}\sum_{i=a+1}^{a+m} X_i\right)^2\right] = O(1) \quad \text{(A1.17)}$$

for all integers $a \geq 0$, $m \geq 1$, then (A1.17) continues to hold when the exponent 2 is increased to $2 + \beta$. It is this constant $\beta > 0$ which allows use of the Borel–Cantelli lemma in (3.11).

APPENDIX 2: A MAXIMAL INEQUALITY

The following maximal inequality, established in Theorem 1 of Longnecker and Serfling [16], is the essential tool in establishing Corollary A2 (which follows), which is the maximal inequality used in line (3.10) of Section 3.

THEOREM A2. *Let $X_1, \ldots, X_n$ be $\mathfrak{R}^d$-valued random vectors. Suppose there are constants $\gamma > 1$, $\nu > 0$ such that*

(i) $E\left|\sum_{k=i}^{j} X_k\right|^{\nu} \leq [g(i,j)]^{\gamma}$ for all $1 \leq i \leq j \leq n$

and the function $g(\cdot, \cdot)$ satisfies

(ii) $g(i,j) + g(j+1, k) \leq g(i,k)$ for all $1 \leq i \leq j < k \leq n$.

Then there exists a constant $\tilde{A}_{\nu,\gamma} > 0$ (depending only on γ, ν) such that

$$E\left[\max_{1\le s\le n}\left|\sum_{k=1}^{s} X_k\right|^{\nu}\right] \le \tilde{A}_{\nu,\gamma}[g(1,n)]^{\gamma}.$$

Remark A2.1. The proof of Theorem A2 given in [16] extends trivially to $\mathfrak{R}^d$-valued random vectors with Euclidean norm. $\tilde{A}_{\nu,\gamma}$ is defined in terms of ν and γ by (2.2) and (2.3) of [16] and it is important to note that $\tilde{A}_{\nu,\gamma}$ does not depend on n.

COROLLARY A2. *Let $\{\xi(s),\ 0 \le s \le T\}$ be an arbitrary continuous parameter, $\mathfrak{R}^d$-valued, jointly measureable process such that $\int_0^T |\xi(s)|\, ds < \infty$ a.s. Suppose that for some constants $\nu > 0$ and $\gamma > 1$,*

$$\text{(a)}\quad E\left|\int_t^u \xi(s)\, ds\right|^{\nu} \le [h(t,u)]^{\gamma} \qquad \text{for all } 0 \le t \le u \le T,$$

where the function h satisfies

$$\text{(b)}\quad h(s,t) + h(t,u) \le h(s,u), \qquad \text{for all } 0 \le s \le t \le u \le T.$$

Then there exists some constant $A_{\nu,\gamma}$ (depending only on γ, ν) such that

$$E\left[\max_{0\le t\le T}\left|\int_0^t \xi(s)\, ds\right|^{\nu}\right] \le A_{\nu,\gamma}[h(0,T)]^{\gamma}.$$

Proof. Fix an integer $m \ge 1$ and define for almost all ω

$$\xi_k^m \triangleq \int_{(k-1)2^{-m}}^{k2^{-m}} \xi(s)\, ds \qquad \text{for all } k = 1, 2, \ldots, [T2^m]. \tag{A2.1}$$

Then by (A2.1) and hypothesis (a), for integers $1 \le i \le j \le [T2^m]$,

$$\begin{aligned} E\left[\left|\sum_{k=i}^{j} \xi_k^m\right|^{\nu}\right] &= E\left[\left|\int_{(i-1)2^{-m}}^{j2^{-m}} \xi(s)\, ds\right|^{\nu}\right] \\ &\le \left[h\left(\frac{i-1}{2^m}, \frac{j}{2^m}\right)\right]^{\gamma} = [g_m(i,j)]^{\gamma}, \end{aligned} \tag{A2.2}$$

where

$$g_m(i,j) \triangleq h\left(\frac{i-1}{2^m}, \frac{j}{2^m}\right) \qquad \text{for all } 1 \le i \le j \le [T2^m]. \tag{A2.3}$$

Moreover, by hypothesis (b), for all $1 \le i \le j < k \le [T2^m]$,

$$g_m(i,j) + g_m(j+1,k) \le h\left(\frac{i-1}{2^m}, \frac{k}{2^m}\right) = g_m(i,k). \tag{A2.4}$$

In view of (A2.2), (A2.3), and the Longnecker–Serfling inequality (Theorem A2), there exists some $\tilde{A}_{\nu,\gamma}$ such that

$$E\left[\max_{0\le j\le [T2^m]} \left|\int_0^{j2^{-m}} \xi(s)\,ds\right|^{\nu}\right] = E\left[\max_{1\le j\le [T2^m]} \left|\sum_{k=1}^{j} \xi_k^m\right|^{\nu}\right] \tag{A2.5}$$
$$\le \tilde{A}_{\nu,\gamma}[g_m(1,[T2^m])]^{\gamma}.$$

Now, let T_m be the largest number of the form $l2^{-m}$ such that $T \ge l2^{-m}$ where l is an integer. Then, by (A2.5) and (A2.3),

$$E\left[\max_{\substack{0\le t\le T_m\\ t=l2^{-m}}} \left|\int_0^{t} \xi(s)\,ds\right|^{\nu}\right] \le \tilde{A}_{\nu,\gamma}[g_m(1,T_m2^m)]^{\gamma} = \tilde{A}_{\nu,\gamma}[h(0,T_m)]^{\gamma} \tag{A2.6}$$
$$\le \tilde{A}_{\nu,\gamma}[h(0,T)]^{\gamma} \qquad \text{for all } m = 1,2,\ldots,$$

since by hypothesis (a) $h(T_m, T) \ge 0$ and so by hypothesis (b)

$$h(0,T_m) \le h(0,T). \tag{A2.7}$$

But for almost all ω,

$$\max_{\substack{0\le t\le T_m\\ t=l2^{-m}}} \left|\int_0^{t} \xi(s)\,ds\right|^{\nu} \nearrow \max_{0\le t\le T} \left|\int_0^{t} \xi(s)\,ds\right|^{\nu} \qquad \text{as } m \nearrow \infty. \tag{A2.8}$$

The corollary follows from (A2.6), (A2.8), and the Monotone Convergence theorem. ∎

APPENDIX 3: REQUIRED BOUNDS

LEMMA A1. *Assume conditions* (C0), (C1), (C3), *and* (C5) *in Section* 2. *Then, for each* $1 \le \lambda \le 2 + \delta$,

$$\sup_{T\ge 0}\sup_{0\le t\le T} \|\tilde{F}(x^0(t/T), t)\|_{\lambda} \le 2M + 2ND,$$

where $\tilde{F}(x,t)$ *is given by* (2.8) *and* $x^0(\cdot)$ *is the unique solution of* (2.5).

Proof. Fix $T > 0$ and $0 \le t \le T$, then by (2.8), condition (C1.ii), and (2.3),

$$|\bar{F}(x^0(t/T), t)| \le \{L(t, \omega) + N\} |x^0(t/T)| + |\bar{F}(0, t)| \tag{A3.1}$$

and so by (2.6), Minkowski's inequality, (C1), and (C3),

$$\|\bar{F}(x^0(t/T), t)\|_\lambda \le (\|L(t)\|_\lambda + N)D + \|\bar{F}(0, t)\|_\lambda \le 2ND + 2M \tag{A3.2}$$

for all $0 \le t \le T$. ▮

The following result is used in line (3.2) of Section 3.

LEMMA A2. *Under conditions* (C0), (C1), (C3), *and* (C5) *of Section* 2, *there exists some function* $\rho: (0, 1] \to [0, \infty)$ *such that*

$$\sup_{0 \le \tau \le 1} \left| \int_0^\tau EF(x^0(s), s/\varepsilon) - \bar{F}(x^0(s))\, ds \right| \le \rho(\varepsilon) \tag{A3.3}$$

for all $0 < \varepsilon \le 1$ *and*

$$\lim_{\substack{\varepsilon \to 0 \\ \varepsilon > 0}} \rho(\varepsilon) = 0. \tag{A3.4}$$

Proof. This proof uses an argument due to Gihman [8, page 217]. For each $\tau \in [0, 1]$ and $\varepsilon \in (0, 1]$, define

$$\Theta^\varepsilon(\tau) \triangleq \int_0^\tau \{EF(x^0(s), s/\varepsilon) - \bar{F}(x^0(s))\}\, ds \tag{A3.5}$$

$$= \varepsilon \int_0^{\tau/\varepsilon} \{EF(x^0(u\varepsilon), u) - \bar{F}(x^0(u\varepsilon))\}\, du.$$

In a manner similar to that used in Lemma A1, we find for all $0 \le u \le \varepsilon^{-1}$, $\varepsilon > 0$,

$$|EF(x^0(u\varepsilon), u) - \bar{F}(x^0(u\varepsilon))| \le 2M + 2ND \tag{A3.6}$$

and hence by (6.5),

$$|\Theta^\varepsilon(\tau) - \Theta^\varepsilon(\tau')| \le (2M + 2ND)|\tau - \tau'| \qquad \text{for all } 0 \le \tau, \tau' \le 1. \tag{A3.7}$$

Thus, $\{\Theta^\varepsilon, \varepsilon > 0\}$ is equicontinuous and it remains to show $\lim_{\varepsilon \to 0} \Theta^\varepsilon(\tau) = 0$ for each $0 \le \tau \le 1$. Fix such a τ, a $\delta > 0$, define

$$G^\varepsilon(x, s) \triangleq EF(x, s/\varepsilon) - \bar{F}(x) \qquad \text{for all } x \in \Re^d, 0 \le s \le 1, \tag{A3.8}$$

and find an integer n large enough that

$$\max_{0\le s\le 1}\left|x^0(s) - x^0\left(\frac{[ns]}{n}\right)\right| < \frac{\delta}{4N}. \tag{A3.9}$$

Then, by (2.3), (2.4) and (A3.9):

$$\begin{aligned}
|\Theta^\varepsilon(\tau)| &\le \left|\int_0^\tau G^\varepsilon(x^0(s), s) - G^\varepsilon\left(x^0\left(\frac{[ns]}{n}\right), s\right) ds\right| \\
&\quad + \left|\int_0^\tau G^\varepsilon\left(x^0\left(\frac{[ns]}{n}\right), s\right) ds\right| \\
&\le \tau \cdot 2N \cdot \max_{0\le s\le 1}\left|x^0(s) - x^0\left(\frac{[ns]}{n}\right)\right| + \left|\int_0^\tau G^\varepsilon\left(x^0\left(\frac{[ns]}{n}\right), s\right) ds\right| \\
&\le \frac{\delta}{2} + \left|\int_0^\tau G^\varepsilon\left(x^0\left(\frac{[ns]}{n}\right), s\right) ds\right|.
\end{aligned}$$

Now, by (A3.8) and condition (C5), there exists an $\varepsilon_0(\delta, n)$ such that

$$\begin{aligned}
&\left|\int_0^\tau G^\varepsilon\left(x^0\left(\frac{[ns]}{n}\right), s\right) ds\right| \\
&\le \sum_{i=0}^{[\tau n]-1}\left|\int_{i/n}^{(i+1)/n} G^\varepsilon\left(x^0\left(\frac{i}{n}\right), s\right) ds\right| + \left|\int_{[\tau n]/n}^{\tau} G^\varepsilon\left(x^0\left(\frac{[\tau n]}{n}\right), s\right) ds\right| \\
&\le 2\sum_{i=1}^{[\tau n]}\left|\int_0^{i/n} G^\varepsilon\left(x^0\left(\frac{i}{n}\right), s\right) ds\right| + 2\left|\int_0^\tau G^\varepsilon\left(x^0\left(\frac{[\tau n]}{n}\right), s\right) ds\right| \\
&\le 2\sum_{i=1}^{[\tau n]}\left|\varepsilon\int_0^{i/n\varepsilon} G^\varepsilon\left(x^0\left(\frac{i}{n}\right), s\varepsilon\right) ds\right| \\
&\quad + 2\left|\varepsilon\int_0^{\tau/\varepsilon} G^\varepsilon\left(x^0\left(\frac{[\tau n]}{n}\right), s\varepsilon\right) ds\right| \\
&\le 2\sum_{i=1}^{[\tau n]}\left|\frac{n\varepsilon}{i}\int_0^{i/n\varepsilon} EF\left(x^0\left(\frac{i}{n}\right), s\right) - \bar{F}\left(x^0\left(\frac{i}{n}\right)\right) ds\right| \\
&\quad + 2\left|\frac{\varepsilon}{\tau}\int_0^{\tau/\varepsilon} EF\left(x^0\left(\frac{[\tau n]}{n}\right), s\right) - \bar{F}\left(x^0\left(\frac{[\tau n]}{n}\right)\right) ds\right| \\
&< \frac{\delta}{2} \qquad \text{for all } \varepsilon \le \varepsilon_0(\delta, n).
\end{aligned} \tag{A3.11}$$

The lemma follows from the equicontinuity of $\{\Theta^{\varepsilon}, \varepsilon > 0\}$, (A3.10), and (A3.11). ▌

REFERENCES

1. V. I. ARNOL'D, "Mathematical Methods of Classical Mechanics," Graduate Texts in Mathematics Series, Vol. 60, Springer Verlag, New York, 1978.
2. A. BENVENISTE, M. METIVIER, AND P. PRIOURET, "Algorithmes Adaptifs et Approximations Stochastiques: Theorie et Applications," Masson, Paris, 1987.
3. N. N. BOGOLIUBOV AND YU. A. MITROPOL'SKĬI, "Asymptotic Methods in the Theory of Non-Linear Oscillations," International Monographs on Advanced Mathematics and Physics, Hindustan Publ., Delhi, 1961.
4. R. C. BRADLEY, Basic properties of strong mixing conditions, *in* "Dependence in Probability and Statistics, A Survey of Recent Results," Birkhauser, Basel, 1986.
5. C. A. DESOER AND M. VIDYASAGAR, "Feedback Systems; Input-Output Properties," Academic Press, New York, 1975.
6. M. I. FREIDLIN AND A. D. WENTZELL, "Random Perturbations of Dynamical Systems," Grundlehren Series, Vol. 260, Springer Verlag, New York, 1984.
7. M. I. FREIDLIN, The averaging principle and theorems on large deviations, *Russian Math. Surveys* **33,** No. 5 (1978), 117–176.
8. I. I. GIHMAN, On a theorem of N. N. Bogoliubov (in Russian), *Ukrain. Math. Zh.*, **4** (1952), 215–219.
9. J. A. JEANS, "Astronomy and Cosmogony," Cambridge Univ. Press, Cambridge, 1928.
10. R. Z. KHAS'MINSKII, On stochastic processes defined by differential equations with a small parameter, *Theory Probab. Appl.* **2** (1966), 211–228.
11. M. A. KRASNOSEL'SKII AND S. G. KREIN, On the averaging principle in nonlinear mechanics, *Uspekhi Mat. Nauk.* **10,** No. 3 (1955), 147–152.
12. H. J. KUSHNER AND D. S. CLARK, "Stochastic Approximation Methods for Constrained and Unconstrained Systems," Applied Mathematical Sciences Series, Vol. 26, Springer-Verlag, New York, 1978.
13. H. J. KUSHNER AND A. SHWARTZ, Weak convergence and asymptotic properties of adaptive filters with constant gains, *IEEE Trans. Inform. Theory,* **30,** No. 2 (1984), 177–182.
14. J. LAMPERTI, "Probability," Benjamin-Cummings, Reading, MA, 1966.
15. L. LJUNG, Analysis of recursive stochastic algorithms, *IEEE Trans. Automat. Control* **AC-22,** (1977), 551–575.
16. M. LONGNECKER AND R. J. SERFLING, General moment and probability inequalities for the maximum partial sum, *Acta Math. Acad. Sci. Hungar.* **30** (1977), 129–133.
17. M. METIVIER AND P. PRIOURET, Applications of a Kushner and Clark lemma to general classes of stochastic algorithms, *IEEE Trans. Inform. Theory,* **IT-20,** No. 2 (1984), 140–151.
18. W. T. REID, "Ordinary Differential Equations," Wiley, New York, 1970.
19. J. A. SANDERS AND F. VERHULST, "Averaging Methods in Nonlinear Dynamical Systems," Applied Mathematical Sciences Series, Vol. 59, Springer-Verlag, New York, 1985.
20. R. J. SERFLING, Contributions to central limit theory for dependent variables, *Ann. Math. Statist.* **39** (1968), 1158–1175.
21. D. A. SOTRES AND M. GHOSH, Strong convergence of linear rank statistics for mixing processes, *Sankhyā Ser. B* **39** (1977), 1–11.
22. V. M. VOLOSOV, Averaging in systems of ordinary differential equations, *Russian Math. Surveys* **17** No. 6, (1962), 1–126.