

SOME AVERAGING AND STABILITY RESULTS FOR RANDOM DIFFERENTIAL EQUATIONS*

STUART GEMAN†

Abstract. This paper concerns differential equations which contain strongly mixing random processes (processes for which the “past” and the “future” are asymptotically independent). When the “rate” of mixing is rapid relative to the rate of change of the solution process, information about the behavior of the solution is obtained. Roughly, the results fall into three categories:

1. Quite generally, the solution process is well approximated by a deterministic trajectory, over a finite time interval. 2. For more restricted systems, this approximation extends to the infinite interval $[0, \infty)$. 3. Conditions for the asymptotic stability of $\dot{x} = Ax$, where A is an $n \times n$ matrix-valued random process, are obtained.

1. Introduction. We are concerned with approximate solutions for a class of ordinary stochastic differential equations. These equations involve continuous and strongly mixing processes (stochastic processes for which the “past” and the “future” are asymptotically independent). The idea is to average (take the expected value of) the right hand side of the stochastic equation, and ask when does the solution of the resulting deterministic equation accurately describe the behavior of the original system. Roughly, the more rapid the mixing the more appropriate the approximation.

In the general, nonlinear case, we will establish conditions under which the random solution “stays close,” in a probabilistic sense, to the associated deterministic solution. In the linear homogeneous case, we will explore the relation between asymptotic stability in the averaged equation, and asymptotic stability in the random equation: specifically, when does the first imply the second?

Loosely speaking, let $\phi(t, \omega)$ be a continuous and mixing (though not necessarily stationary) R^m -valued random process, where ω is a sample point in a probability space. For every $\varepsilon > 0$, $\phi(t/\varepsilon, \omega)$ is also a mixing process, and, as $\varepsilon \rightarrow 0$, the mixing rate of $\phi(t/\varepsilon, \omega)$ becomes arbitrarily rapid. Let $H: R^n \times R^m \times R^1 \rightarrow R^n$, and, for each x, t , and $\varepsilon > 0$, define

$$G_\varepsilon(x, t) = E[H(x, \phi(t/\varepsilon, \omega), t)],$$

the expected value of $H(x, \phi(t/\varepsilon, \omega), t)$. Consider the following random differential equation, together with its associated “averaged equation”:

$$(1.1) \quad \begin{aligned} \dot{x}_\varepsilon(t, \omega) &= H(x_\varepsilon(t, \omega), \phi(t/\varepsilon, \omega), t), & x_\varepsilon(0, \omega) &= x_0 \in R^n, \\ \dot{y}_\varepsilon(t) &= G_\varepsilon(y_\varepsilon(t), t), & y_\varepsilon(0) &= x_0. \end{aligned}$$

We are interested in results such as

$$(1.2) \quad \lim_{\varepsilon \rightarrow 0} P\{ \sup_{t \in [0, T]} |x_\varepsilon(t) - y_\varepsilon(t)| > \eta \} = 0,$$

for some $T > 0$ and all $\eta > 0$, and

$$\lim_{\varepsilon \rightarrow 0} \sup_{t \geq 0} E|x_\varepsilon(t) - y_\varepsilon(t)|^2 = 0.$$

The following problem is closely related. For fixed ω , let $H(x, \omega, t)$ map $R^n \times R^1$ into R^n . Assume that for each x , $H(x, \omega, t)$ is a mixing process, and for each x and t

* Received by the editors December 19, 1977, and in revised form June 1, 1978. This work was supported by the National Science Foundation under Grant MPS75-15153.

† Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912.

define $G(x, t) = E[H(x, \omega, t)]$. Consider the random equation

$$\dot{x}_\varepsilon(t, \omega) = \varepsilon H(x_\varepsilon(t, \omega), \omega, t), \quad x_\varepsilon(0, \omega) = x_0,$$

with its averaged equation

$$\dot{y}_\varepsilon(t) = \varepsilon G(y_\varepsilon(t), t), \quad y_\varepsilon(0) = x_0.$$

We wish to approximate $x_\varepsilon(t, \omega)$ by $y_\varepsilon(t)$, for small ε . The change of variables $t \rightarrow t/\varepsilon$ makes the connection to the previous problem:

$$\begin{aligned} \dot{x}_\varepsilon(t, \omega) &= H(x_\varepsilon(t, \omega), \omega, t/\varepsilon), & x_\varepsilon(0, \omega) &= x_0, \\ \dot{y}_\varepsilon(t) &= G(y_\varepsilon(t), t/\varepsilon), & y_\varepsilon(0) &= x_0, \end{aligned}$$

and as $\varepsilon \rightarrow 0$, $H(x, \omega, t/\varepsilon)$ becomes an increasingly rapid mixing process.

Both problems will be treated together, and in slightly greater generality, by considering systems of the form

$$(1.3) \quad \begin{aligned} \dot{x}_\varepsilon(t, \omega) &= F(x_\varepsilon(t, \omega), \omega, t, t/\varepsilon), & x_\varepsilon(0, \omega) &= x_0, \\ \dot{y}_\varepsilon(t) &= G_\varepsilon(y_\varepsilon(t), t), & y_\varepsilon(0) &= x_0, \end{aligned}$$

where

(Ω, \mathcal{F}, P) is a probability space,
 $F: \mathbf{R}^n \times \Omega \times \mathbf{R}^1 \times \mathbf{R}^1 \rightarrow \mathbf{R}^n$,
 for fixed x and t , $F(x, \omega, t, \tau)$ is a mixing process, and
 $G_\varepsilon(x, t) = E[F(x, \omega, t, t/\varepsilon)]$.

With regards to stability, we will treat

$$(1.4) \quad \dot{x}(t, \omega) = A(t, \omega)x(t, \omega)$$

(where A is an $n \times n$ matrix-valued process), giving conditions on A under which asymptotic stability of

$$\dot{y}(t) = E[A(t)]y(t)$$

implies asymptotic stability for (1.4) (L^2 and almost sure).

Systems similar to (1.3) have already been studied by several authors (cf. Khasminskii [9], [10], Cogburn and Hersh [5], Papanicolaou and Kohler [14], White [19], and Blankenship and Papanicolaou [3]). Mostly, however, the results have been of a different nature, with particular attention given to conditions under which $x_\varepsilon(t)$ converges (as $\varepsilon \rightarrow 0$) to a diffusion process on finite intervals, $[0, T/\varepsilon]$, some $T > 0$. Averaging results, like (1.2), are usually discussed in conjunction with a central limit theorem for $(x_\varepsilon(t) - y_\varepsilon(t))/\sqrt{\varepsilon}$ on $[0, T]$ (cf. Khasminskii [9] and White [19]). The possibility of an extended averaging, to $[0, \infty)$, has not been explored (see, however, Blankenship and Papanicolaou [3] for some stability and diffusion approximation results on $[0, \infty)$). Finally, we should mention that results somewhat analogous to ours have been obtained in the context of Itô-like equations (see Vrkoc [17] and Lybrand [13]).

Section 2 is devoted to notation, definitions, and a preliminary lemma. In § 3, we will treat averaging in linear systems, on the interval $[0, \infty)$. As a by-product, the aforementioned stability result for the linear homogeneous random system is obtained. Sections 4 and 5 discuss averaging in nonlinear systems, on the intervals $[0, T]$ and $[0, \infty)$ respectively. Finally, § 6 explores the application of the theorems to several examples.

2. Preliminaries. First, let us define the types of mixing to be used. Given the probability space (Ω, \mathcal{F}, P) , let $\mathcal{F}_0^t \subset \mathcal{F}$ and $\mathcal{F}_t^\infty \subset \mathcal{F}$ be two families of σ -fields indexed by $t \geq 0$. Mixing will be defined in terms of these families, and, in each application, these families will be defined in terms of the right-hand side of the stochastic system. (For now, think of \mathcal{F}_a^b as the sigma field generated by some stochastic process between the times a and b , any $a \leq b$.)

In formulating and utilizing the various strong mixing conditions, it is convenient to introduce a certain signed measure: for every $t \geq 0$ and $\delta \geq 0$, let $v_{t,\delta}(\cdot)$ be the measure, on $(\Omega \times \Omega, \mathcal{F}_0^t \times \mathcal{F}_{t+\delta}^\infty)$, defined by

$$v_{t,\delta}(B) = P(\omega : (\omega, \omega) \in B) - P \times P(B)$$

$B \in \mathcal{F}_0^t \times \mathcal{F}_{t+\delta}^\infty$. Since $\{B \in \mathcal{F}_0^t \times \mathcal{F}_{t+\delta}^\infty : (\omega : (\omega, \omega) \in B) \in \mathcal{F}\}$ is a monotone class which contains the elementary sets, $(\omega : (\omega, \omega) \in B)$ is in \mathcal{F} for all $B \in \mathcal{F}_0^t \times \mathcal{F}_{t+\delta}^\infty$, so the definition makes sense. Notice that $v_{t,\delta} = 0$ whenever \mathcal{F}_0^t and $\mathcal{F}_{t+\delta}^\infty$ are independent with respect to P (extend from the rectangles). Loosely speaking, the total variation of $v_{t,\delta}$ will be small whenever \mathcal{F}_0^t and $\mathcal{F}_{t+\delta}^\infty$ are “nearly” independent.

Here are three versions of “strong mixing”:

$$\text{Type I.} \quad \rho(\delta) \equiv \sup_{t \geq 0} \sup_{\substack{A \in \mathcal{F}_0^t \\ B \in \mathcal{F}_{t+\delta}^\infty \\ P(A) > 0}} \frac{1}{P(A)} |v_{t,\delta}(A \times B)| \rightarrow 0$$

as $\delta \rightarrow \infty$,

$$\text{Type II.} \quad \rho(\delta) \equiv \sup_{t \geq 0} \sup_{A \in \mathcal{F}_0^t \times \mathcal{F}_{t+\delta}^\infty} |v_{t,\delta}(A)| \rightarrow 0$$

as $\delta \rightarrow \infty$,

$$\text{Type III.} \quad \rho(\delta) \equiv \sup_{t \geq 0} \sup_{\substack{A \in \mathcal{F}_0^\infty \\ B \in \mathcal{F}_{t+\delta}^\infty}} |v_{t,\delta}(A \times B)| \rightarrow 0$$

as $\delta \rightarrow \infty$.

The first two were introduced by Kolmogorov (see Volkonskii and Rozanov [16]), and the last by Rosenblatt [15]. Type I has been frequently used in the context of stochastic differential equations (cf. Khasminskii [10], Cogburn and Hersh [5], and Papanicolaou and Kohler [14]). It is not hard to show that $\text{I} \Rightarrow \text{II} \Rightarrow \text{III}$, and further implications can be disproved by counterexample. If $\{\mathcal{F}_a^b\}$ is generated by a stationary process, then any of the strong mixing conditions implies ergodicity.

$|v|_{t,\delta}(\cdot)$ will refer to the total variation measure. Since $v_{t,\delta}$ is the difference of two probability measures:

$$|v|_{t,\delta}(\Omega \times \Omega) \leq 2, \quad \text{and} \quad |v|_{t,\delta}(\Omega \times \Omega) = 2 \sup_{A \in \mathcal{F}_0^t \times \mathcal{F}_{t+\delta}^\infty} |v_{t,\delta}(A)|.$$

So, with Type II (or I) mixing:

$$|v|_{t,\delta}(\Omega \times \Omega) \leq 2\rho(\delta).$$

The following lemma plays a central role in obtaining each of our results. After its proof, we will show, heuristically, how it is applied to the averaging problem. The procedure outlined here can also be used to establish the consistency of some continuous time stochastic approximation algorithms, as will be demonstrated in a forthcoming article.

For convenience, we will not distinguish between “for almost every ω ” and “for every ω ”. Look at the systems

$$\begin{aligned} \dot{x}(t, \omega) &= H(x(t, \omega), \omega, t), & x(0, \omega) &= x_0 \in R^n, \\ \dot{y}(t) &= G(y(t), t), & y(0) &= x_0, \end{aligned}$$

where

1. H is jointly measurable in its three arguments.
2. $G(x, t) = E[H(x, \omega, t)]$, and for all i and j

$$\frac{\partial}{\partial x_j} G_i(x, t) \text{ exists, and } G(x, t) \text{ and } \frac{\partial}{\partial x_j} G_i(x, t) \text{ are continuous (in } (x, t)).$$

3. For some $T > 0$:

- a. There exists a unique solution, $x(t, \omega)$, on $[0, T]$ for every ω ,
- b. A solution to

$$\frac{\partial}{\partial t} g(t, s, x) = G(g(t, s, x), t), \quad g(s, s, x) = x,$$

exists on $[0, T] \times [0, T] \times R^n$.

We will use the following notation:

1. $g_s(t, s, x) = (\partial/\partial s)g(t, s, x)$.
2. $g_x(t, s, x) =$ the $n \times n$ matrix with (i, j) component

$$\frac{\partial}{\partial x_j} g_i(t, s, x).$$

3. If $K: R^n \rightarrow R^1$ and $K \in C^1$,

$$K'(x) = \left(\frac{\partial}{\partial x_1} K(x), \dots, \frac{\partial}{\partial x_n} K(x) \right), \quad \text{and}$$

$$\frac{\partial}{\partial x} K(g(t, s, x(s, \omega))) = \left(\frac{\partial}{\partial x_1} K(g(t, s, x)), \dots, \frac{\partial}{\partial x_n} K(g(t, s, x)) \right)_{x=x(s, \omega)}.$$

4. Define the families of σ -fields \mathcal{F}_0^t and \mathcal{F}_t^∞ such that, for each $t \geq 0$,

\mathcal{F}_0^t contains the σ -field generated by

$$\{H(x, \omega, s): 0 \leq s \leq t, -\infty < x < \infty\}, \quad \text{and}$$

\mathcal{F}_t^∞ contains the σ -field generated by

$$\{H(x, \omega, s): t \leq s < \infty, -\infty < x < \infty\}.$$

Two facts that will be needed are:

1. $g_s(t, s, x) = -g_x(t, s, x)G(x, s)$

for all $t \in [0, T]$, $s \in [0, T]$, and $x \in R^n$ (cf. Hartman [7, Chap. 5]), and

2. For any $f: \Omega \times \Omega \rightarrow R^1$, $\mathcal{F}_0^t \times \mathcal{F}_t^\infty$ measurable,

$$\int_{\Omega \times \Omega} \{f(\omega, \omega) - f(\omega, \eta)\} dP(\omega) dP(\eta) = \int_{\Omega \times \Omega} f(\omega, \eta) dv_{t,0}$$

(follows easily from the definition of $v_{t,0}$, and a monotone class argument for functions on $\Omega \times \Omega$).

LEMMA 1. For any C^1 function $K: \mathbf{R}^n \rightarrow \mathbf{R}^1$ and $t \in [0, T]$:

$$E[K(x(t))] = K(y(t)) + \int_0^t \int_{\Omega \times \Omega} \left(\frac{\partial}{\partial x} K(g(t, s, x(s, \omega))) \right) \cdot H(x(s, \omega), \eta, s) dv_{s,0} ds,$$

provided that

$$\left(\frac{\partial}{\partial x} K(g(t, s, x(s, \omega))) \right) \cdot H(x(s, \omega), \eta, s), \quad \text{and}$$

$$\left(\frac{\partial}{\partial x} K(g(t, s, x(s, \omega))) \right) \cdot H(x(s, \omega), \omega, s)$$

are absolutely integrable on $\Omega \times \Omega \times [0, T]$, with respect to $dP(\omega) dP(\eta) ds$.

Proof.

$$\begin{aligned} & K(x(t, \omega)) - K(y(t)) \\ &= \int_0^t \frac{\partial}{\partial s} K(g(t, s, x(s, \omega))) ds \\ &= \int_0^t K'(g(t, s, x(s, \omega))) \cdot \{g_x(t, s, x(s, \omega))\dot{x}(s, \omega) + g_s(t, s, x(s, \omega))\} ds \\ &= \int_0^t K'(g(t, s, x(s, \omega))) \cdot \{g_x(t, s, x(s, \omega))H(x(s, \omega), \omega, s) \\ &\quad - g_x(t, s, x(s, \omega))G(x(s, \omega), s)\} ds \\ &= \int_0^t \int_{\Omega} \left(\frac{\partial}{\partial x} K(g(t, s, x(s, \omega))) \right) \cdot \{H(x(s, \omega), \omega, s) - H(x(s, \omega), \eta, s)\} dP(\eta) ds. \end{aligned}$$

Hence

$$\begin{aligned} E[K(x(t))] &= K(y(t)) + \int_0^t \int_{\Omega \times \Omega} \left(\frac{\partial}{\partial x} K(g(t, s, x(s, \omega))) \right) \\ &\quad \cdot \{H(x(s, \omega), \omega, s) - H(x(s, \omega), \eta, s)\} dP(\eta) dP(\omega) ds \\ &= K(y(t)) + \int_0^t \int_{\Omega \times \Omega} \left(\frac{\partial}{\partial x} K(g(t, s, x(s, \omega))) \right) \cdot H(x(s, \omega), \eta, s) dv_{s,0} ds. \quad \text{Q.E.D.} \end{aligned}$$

Here is an example of how the lemma may be used. Return to (1.1), and let \mathcal{F}_a^b be the sigma field generated by

$$\{\phi(t, \omega): a \leq t \leq b\},$$

using strict inequality for a or b infinite. Assume Type II mixing, and observe that

$$|v_{|t, \delta}(\Omega \times \Omega)| \leq 2\rho(\delta/\varepsilon)$$

for the process $\phi(t/\varepsilon, \omega)$ (i.e. now defining \mathcal{F}_a^b by the process $\phi(t/\varepsilon)$ instead of $\phi(t)$).

Letting $K(x) = |x - y_\varepsilon(t)|^2$ in Lemma 1:

$$E|x_\varepsilon(t) - y_\varepsilon(t)|^2 = \int_0^t \int_{\Omega \times \Omega} \left(\frac{\partial}{\partial x} K(g_\varepsilon(t, s, x_\varepsilon(s, \omega))) \right) \cdot H(x_\varepsilon(s, \omega), \phi(s/\varepsilon, \eta), s) dv_{s,0} ds.$$

Then, when we can, we write (for any small $\delta > 0$)

$$\begin{aligned}
 E|x_\epsilon(t) - y_\epsilon(t)|^2 &= O(\delta) + \int_\delta^t \int_{\Omega \times \Omega} \left(\frac{\partial}{\partial x} K(g_\epsilon(t, s, x_\epsilon(s - \delta, \omega))) \right) \\
 &\quad \cdot H(x_\epsilon(s - \delta, \omega), \phi(s/\epsilon, \eta), s) dv_{s,0} ds \\
 &= O(\delta) + \int_\delta^t \int_{\Omega \times \Omega} \left(\frac{\partial}{\partial x} K(g_\epsilon(t, s, x_\epsilon(s - \delta, \omega))) \right) \\
 &\quad \cdot H(x_\epsilon(s - \delta, \omega), \phi(s/\epsilon, \eta), s) dv_{s-\delta,\delta} ds \\
 &= O(\delta) + O(\rho(\delta/\epsilon)). \qquad \qquad \qquad \text{(since } x_\epsilon(s - \delta) \text{ is } \mathcal{F}_0^{s-\delta} \text{ measurable)}
 \end{aligned}$$

Finally then,

$$\lim_{\epsilon \rightarrow 0} E|x_\epsilon(t) - y_\epsilon(t)|^2 = O(\delta) \Rightarrow \lim_{\epsilon \rightarrow 0} E|x_\epsilon(t) - y_\epsilon(t)|^2 = 0,$$

and, hopefully, uniformly on some t interval.

3. Linear systems.

A. Stability for the homogeneous equation. The question of averaging relates naturally to the question of stability. If $A(t, \omega)$ is an $n \times n$ matrix of random processes, and

$$B = E[A(t)]$$

is constant, when does asymptotic stability of the averaged system

$$(3.1) \qquad \qquad \qquad \dot{y}(t) = By(t)$$

imply some manner of stability for

$$(3.2) \qquad \qquad \qquad \dot{x}(t, \omega) = A(t, \omega)x(t, \omega)?$$

In this direction, Infante [8] showed that if $A(t, \omega)$ is ergodic, and if for some symmetric positive definite matrix P ,

$$(3.3) \qquad \qquad \qquad E[\max \text{eigenvalue}(A^T(t) + PA(t)P^{-1})] < 0,$$

then

$$x(t) \rightarrow 0 \quad \text{a.s.}$$

Now suppose (3.3) is satisfied, and let R be the symmetric positive definite square root of P . Then (3.3) \Rightarrow

$$\begin{aligned}
 E[\max_{\substack{x \in R^n \\ |x|=1}} x^T RA(t)R^{-1}x] &< 0 \\
 \Rightarrow \max_{\substack{x \in R^n \\ |x|=1}} x^T RBR^{-1}x &< 0
 \end{aligned}$$

\Rightarrow (3.1) is asymptotically stable.

However, the converse is not true, i.e. the asymptotic stability of (3.1) does not imply (3.3). In fact, it is easy to find systems in which (3.1) and (3.2) are stable, but (3.3) is violated. (Recently, Blankenship [2] considerably generalized the condition (3.3). But still, the above discussion holds, i.e. stability in (3.1) is not sufficient for the generalized version of (3.3).)

In a similar vein, Blankenship and Papanicolaou [3] discuss (3.2) when $A(t)$ is ergodic Markov with compact state space, and the equation is “close” to a related Ito equation, in a suitable sense. In this case, (3.2) inherits the stability properties of the Ito equation, and the latter are well understood. The point, then, is to utilize the available machinery for Ito equations in analyzing the stability of (3.2).

Here, we define sufficient conditions on the process $A(t)$ under which the stability of (3.1) implies that of (3.2). The main requirements are that: (1) $A(t)$ be bounded, and (2) roughly, $A(t)$ be nearly independent of itself at sufficiently small separations of t . For (2), Type I mixing with “rapidly” decreasing $\rho(\delta)$ will suffice (for example see remark 2 below). Notice that $A(t)$ need not be stationary (in fact, B may depend on t).

Specifically, let

$$\dot{x}(t, \omega) = A(t, \omega)x(t, \omega), \quad x(0, \omega) = x_0 \in R^n,$$

where

1. $A(t, \omega)$ is an $n \times n$ matrix of real valued random processes satisfying:
 - a. $A(t, \omega)$ is piecewise continuous for each ω , and
 - b. $|A(t, \omega)| \leq c_1$ for all t and ω , and some constant c_1 .¹
2. The equation

$$(3.4) \quad \dot{y}(t) = B(t)y(t),$$

where

$$B(t) = E[A(t)],$$

is exponentially stable, i.e. if $\Phi(t, s)$ is the transition matrix for (3.4), then

$$(3.5) \quad |\Phi(t, s)| \leq \gamma e^{-\lambda(t-s)}$$

for all $t \geq s \geq 0$, and some positive constants γ and λ .

For each $t \geq 0$, define \mathcal{F}_0^t to be the σ -field generated by $\{A(s) : 0 \leq s \leq t\}$, and \mathcal{F}_t^∞ to be the σ -field generated by $\{A(s) : t \leq s \leq \infty\}$. For each $\delta > 0$ let

$$\rho(\delta) = \sup_{t \geq 0} \sup_{\substack{A \in \mathcal{F}_0^t \\ B \in \mathcal{F}_{t+\delta}^\infty \\ P(A) > \delta}} \frac{1}{P(A)} |v_{t,\delta}(A \times B)|.$$

THEOREM 1. *There exists an $r_0 > 0$, depending only on n, c_1, γ , and λ , such that*

$$\min_{\delta \geq 0} (\rho(\delta) + \delta) < r_0 \Rightarrow x(t) \rightarrow 0$$

in mean square and almost surely.

(Theorems 1 and 2 will be proved together.)

Remarks: 1. Since $|B(t)| \leq c_1$, (3.5) is equivalent to a variety of seemingly weaker statements, such as

$$\int_{t_0}^{t_1} |\Phi(t, t_0)| dt \leq M$$

for all $t_1 \geq t_0 \geq 0$ and some constant M (see Brockett [4]).

¹For any $m \times n$ matrix A , $|A|$ refers to the “induced norm”, i.e.

$$|A| = \max_{\substack{x \in R^n \\ |x|=1}} |Ax|,$$

using the Euclidean norm in R^n and R^m .

2. As an example, consider the system

$$\dot{x}_\varepsilon(t, \omega) = A(t/\varepsilon, \omega)x_\varepsilon(t, \omega),$$

where $E[A(t, \omega)]$ is a stability matrix, and A is bounded. If the components of A form a (vector valued) ergodic Markov process, satisfying Doeblin's condition, then A is Type I mixing. The theorem says that

$$x_\varepsilon(t, \omega) \rightarrow 0$$

in mean square and almost surely, for all ε sufficiently small.

3. For situations like the one discussed in remark 2, the question of whether a.s. convergence holds for *every* $\varepsilon > 0$ is unresolved. (As far as mean square convergence goes, counterexamples, when ε is too large, are easily constructed.)

Here is a particularly simple example. Let A_1 and A_2 be constant matrices with the property that

$$pA_1 + (1-p)A_2$$

is negative definite, for some $1 > p > 0$. Choose an i.i.d. sequence of A_1 's and A_2 's with probabilities p and $1-p$ respectively. For $i = 0, 1, \dots$, define $A(t, \omega)$ to be the i th member of the sequence, on the interval $t \in [i, i+1)$. Does $x(t, \omega)$, defined by

$$(3.6) \quad \dot{x}(t, \omega) = A(t, \omega)x(t, \omega),$$

converge to 0 almost surely? Theorem 1 says that if, instead, we choose $A(t, \omega)$ piecewise constant on the intervals $[i\varepsilon, i\varepsilon + \varepsilon)$, then $x(t, \omega) \rightarrow 0$ a.s. for all ε sufficiently small.

Notice that the Kolomogorov zero-one law applies, with the following implication: if $\phi(t, s, \omega)$ is the transition matrix for (3.6), then either $|\phi(t, 0, \omega)| \rightarrow 0$ a.s., or

$$p\{|\phi(t, 0, \omega)| \rightarrow 0\} = 0.$$

B. Averaging in inhomogenous systems. For each ε , $1 \geq \varepsilon > 0$, define

$$(3.7) \quad \dot{x}_\varepsilon(t, \omega) = A(t, t/\varepsilon, \omega)x_\varepsilon(t, \omega) + d(t, t/\varepsilon, \omega),$$

and

$$\dot{y}_\varepsilon(t) = B_\varepsilon(t)y_\varepsilon(t) + e_\varepsilon(t)$$

with

$$x_\varepsilon(0, \omega) = y_\varepsilon(0) = x_0 \in \mathbf{R}^n,$$

where

1. For each t, τ , and ω , $A(t, \tau, \omega)$ is an $n \times n$ matrix $\{a_{ij}(t, \tau, \omega)\}$, and $d(t, \tau, \omega)$ is an n -vector $\{d_i(t, \tau, \omega)\}$. For each i and j

- a. a_{ij} and d_i are jointly measurable in their 3 arguments, and
- b. for each ω and $\varepsilon > 0$

$$a_{ij}(t, t/\varepsilon, \omega) \quad \text{and} \quad d_i(t, t/\varepsilon, \omega)$$

are piecewise continuous.

2. There exist constants c_1 and c_2 such that

$$|A(t, \tau, \omega)| \leq c_1 \quad \text{and} \quad |d(t, \tau, \omega)| \leq c_2$$

for all t, τ , and ω .

3. The σ -fields \mathcal{F}_0^t and \mathcal{F}_t^∞ satisfy the Type I mixing condition, where \mathcal{F}_0^t is the σ -field generated by

$$\{A(s, \tau, \omega): 0 \leq \tau \leq t, 0 \leq s < \infty\} \cup \{d(s, \tau, \omega): 0 \leq \tau \leq t, 0 \leq s < \infty\},$$

and \mathcal{F}_t^∞ is the σ -field generated by

$$\{A(s, \tau, \omega): t \leq \tau < \infty, 0 \leq s < \infty\} \cup \{d(s, \tau, \omega): t \leq \tau < \infty, 0 \leq s < \infty\}.$$

4. $B_\varepsilon(t) = E[A(t, t/\varepsilon)]$ and $e_\varepsilon(t) = E[d(t, t/\varepsilon)]$. If $\Phi_\varepsilon(t, s)$ is the transition matrix for

$$\dot{y}_\varepsilon(t) = B_\varepsilon(t)y_\varepsilon(t),$$

then there exist positive constants γ and λ such that

$$|\Phi_\varepsilon(t, s)| \leq \gamma e^{-\lambda(t-s)}$$

for all $t \geq s \geq 0$ and $1 \geq \varepsilon > 0$.

THEOREM 2. For all ε sufficiently small, $\sup_{t \geq 0} E x_\varepsilon^2(t) < \infty$, and

$$\lim_{\varepsilon \rightarrow 0} \sup_{t \geq 0} E |x_\varepsilon(t) - y_\varepsilon(t)|^2 = 0.$$

Remark. For systems of the form

$$\dot{x}_\varepsilon(t, \omega) = \varepsilon A(t, \omega)x_\varepsilon(t, \omega) + \varepsilon d(t, \omega),$$

the change of variables $t \rightarrow t/\varepsilon$ shows that, whenever $E[A(t)]$ is constant, exponential stability of

$$\dot{y}(t) = E[A(t)]y(t)$$

is sufficient for (4).

Proofs of Theorems 1 and 2. (The proof of Theorem 1 by itself is quite simple. Nevertheless, it is more efficient to combine the two.)

For now, consider the equations

$$\begin{aligned} \dot{x}(t, \omega) &= A(t, \omega)x(t, \omega) + d(t, \omega), \\ (3.8) \quad \dot{y}(t) &= B(t)y(t) + e(t), \\ x(0, \omega) &= y(0) = x_0. \end{aligned}$$

The conditions on A , d , B , and e are those of Theorem 2, taking $A(t, \tau, \omega)$ and $d(t, \tau, \omega)$ independent of their first arguments, and looking only at $\varepsilon = 1$. However, do not yet assume that $\rho(\delta) \rightarrow 0$.

We will require some bounds on $x(t)$ and $y(t)$. These follow easily from the bounds on A and d . For all $t \geq 0$, $s \geq 0$, and ω :

$$(3.9) \quad |x(t) - x(s)| \leq |t - s| e^{c_1(t-s)} (c_1 |x(s)| + c_2),$$

$$(3.10) \quad |x(t) - x_0| \leq t e^{c_1 t} (c_1 |x_0| + c_2),$$

and

$$(3.11) \quad |y(t) - x_0| \leq t e^{c_1 t} (c_1 |x_0| + c_2).$$

Below, we will introduce constants k_1, k_2, \dots, k_{10} . It is important to note the dependence of these constants: they are (only) functions of $c_1, c_2, \gamma, \lambda, |x_0|$, and n .

Let $g(t, s, x)$ be the solution to (3.8) satisfying

$$y(s) = x.$$

Then

$$g(t, s, x) = \phi(t, s)x + \int_s^t \phi(t, \sigma)e(\sigma) d\sigma$$

where $\phi(t, s)$ is the transition matrix for

$$\dot{y}(t) = B(t)y(t).$$

With $y(0) = x_0$, the solution to (3.8) satisfies

$$|y(t)| = |g(t, 0, x_0)| \leq k_1$$

for some constant k_1 (use condition 4 of Theorem 2).

Put $K(x) = |x - y(t)|^2$ in Lemma 1 (for linear systems, Lemma 1 extends easily to piecewise continuous right-hand sides):

$$E|x(t) - y(t)|^2 = 2 \int_0^t \int_{\Omega \times \Omega} \left(\phi(t, s)x(s, \omega) + \int_s^t \phi(t, \sigma)e(\sigma) d\sigma - y(t) \right) \phi(t, s) \cdot (A(s, \eta)x(s, \omega) + d(s, \eta)) dv_{s,0} ds.$$

For any δ such that $0 < \delta \leq t$ and $\delta \leq 1$:

$$E|x(t) - y(t)|^2 = 2 \int_0^\delta \int_{\Omega \times \Omega} \left(\phi(t, s)x(s, \omega) + \int_s^t \phi(t, \sigma)e(\sigma) d\sigma - y(t) \right) \phi(t, s) \cdot (A(s, \eta)x(s, \omega) + d(s, \eta)) dv_{s,0} ds$$

(I)

$$+ 2 \int_\delta^t \int_{\Omega \times \Omega} \left(\phi(t, s)x(s - \delta, \omega) + \int_s^t \phi(t, \sigma)e(\sigma) d\sigma - y(t) \right) \phi(t, s)$$

(II)

$$\cdot (A(s, \eta)x(s - \delta, \omega) + d(s, \eta)) dv_{s,0} ds$$

$$+ 2 \int_\delta^t \int_{\Omega \times \Omega} \left\{ \phi(t, s) \right.$$

$$\cdot (x(s, \omega) - x(s - \delta, \omega)) \phi(t, s) (A(s, \eta)x(s, \omega) + d(s, \eta))$$

$$+ \left(\phi(t, s)x(s - \delta, \omega) + \int_s^t \phi(t, \sigma)e(\sigma) d\sigma - y(t) \right)$$

(III)

$$\cdot \phi(t, s) A(s, \eta) (x(s, \omega) - x(s - \delta, \omega)) \left. \right\} dv_{s,0} ds.$$

Treat each term (I, II, and III) separately.

For a bound on I we have, for some constant k_2 ,

$$“I” \leq k_2 \delta e^{-\lambda t} \quad \text{for all } 0 < \delta \leq t, \delta \leq 1.$$

In II, $v_{s,0}$ can be replaced by $v_{s-\delta, \delta}$ (these measures agree on $(\Omega \times \Omega, \mathcal{F}_0^{s-\delta} \times \mathcal{F}_s^\infty)$, $s \geq \delta$, and, of course, $x(s - \delta, \omega)$ is $\mathcal{F}_0^{s-\delta}$ -measurable). Expand the matrix multiplications (because they involve only η dependence, two of the terms vanish when the v

integration is done):

$$\begin{aligned}
\text{“II”} &= 2 \sum_{l,m,n,0} \int_{\delta}^t \int_{\Omega \times \Omega} \phi_{ln}(t,s) \phi_{lm}(t,s) x_n(s-\delta, \omega) x_0(s-\delta, \omega) a_{m0}(s, \eta) dv_{s-\delta, \delta} ds \\
&+ 2 \sum_{l,m,n} \int_{\delta}^t \int_{\Omega \times \Omega} \phi_{ln}(t,s) \phi_{lm}(t,s) x_n(s-\delta, \omega) d_m(s, \eta) dv_{s-\delta, \delta} ds \\
&+ 2 \sum_{l,m,n,0} \int_{\delta}^t \int_{\Omega \times \Omega} \int_s^t \phi_{ln}(t, \sigma) \phi_{lm}(t, s) e_n(\sigma) x_0(s-\delta, \omega) a_{m0}(s, \eta) d\sigma dv_{s-\delta, \delta} ds \\
&- 2 \sum_{l,m,n} \int_{\delta}^t \int_{\Omega \times \Omega} y_l(t) \phi_{lm}(t, s) x_n(s-\delta, \omega) a_{mn}(s, \eta) dv_{s-\delta, \delta} ds.
\end{aligned}$$

If two functions, f and g , are $\mathcal{F}_0^{s-\delta}$ and \mathcal{F}_s^{∞} -measurable, respectively, and if

$$E|f| < \infty$$

and

$$|g| \leq c \quad \text{almost surely,}$$

then (see Billingsley [1], Chapter 4)

$$\left| \int_{\Omega \times \Omega} f(\omega) g(\eta) dv_{s-\delta, \delta} \right| \leq 2\rho(\delta) E|f|c.$$

Apply this to II and get:

$$\begin{aligned}
\text{“II”} &\leq 4n^4 c_1 \gamma^2 \rho(\delta) \int_{\delta}^t e^{-2\lambda(t-s)} Ex^2(s-\delta) ds \\
&+ 4n^3 c_2 \gamma^2 \rho(\delta) \int_{\delta}^t e^{-2\lambda(t-s)} \sqrt{Ex^2(s-\delta)} ds \\
&+ 4n^4 c_1 c_2 \frac{\gamma^2}{\lambda} \rho(\delta) \int_{\delta}^t e^{-\lambda(t-s)} \sqrt{Ex^2(s-\delta)} ds \\
&+ 4n^3 c_1 y(t) \gamma \rho(\delta) \int_{\delta}^t e^{-\lambda(t-s)} \sqrt{Ex^2(s-\delta)} ds \\
&\leq k_3 \rho(\delta) \int_0^t e^{-\lambda(t-s)} Ex^2(s) ds \\
&+ k_4 (c_2 + y(t)) \rho(\delta) \int_0^t e^{-\lambda(t-s)} \sqrt{Ex^2(s)} ds, \quad \text{for all } 0 < \delta \leq t, \\
&\delta \leq 1.
\end{aligned}$$

For III, use (3.9):

$$\begin{aligned}
\text{“III”} &\leq 2 \int_{\delta}^t \int_{\Omega \times \Omega} \gamma^2 e^{-2\lambda(t-s)} \delta e^{c_1 \delta} (c_1 |x(s, \omega)| + c_2)^2 d|v|_{s,0} ds \\
&+ 2 \int_{\delta}^t \int_{\Omega \times \Omega} \left\{ \gamma e^{-\lambda(t-s)} |x(s, \omega)| + \gamma e^{-\lambda(t-s)} \delta e^{c_1 \delta} (c_1 |x(s, \omega)| + c_2) \right. \\
&\left. + \int_s^t \gamma e^{-\lambda(t-\sigma)} c_2 d\sigma + y(t) \right\} \gamma e^{-\lambda(t-s)} c_1 \delta e^{c_1 \delta} (c_1 |x(s, \omega)| + c_2) d|v|_{s,0} ds.
\end{aligned}$$

There are three types of terms, those involving Ex^2 , those involving $E|x|$, and those not involving x :

$$\begin{aligned} \text{“III”} &\leq k_5 \delta \int_0^t e^{-\lambda(t-s)} Ex^2(s) ds \\ &\quad + k_6 (c_2 + y(t)) \delta \int_0^t e^{-\lambda(t-s)} \sqrt{Ex^2(s)} ds \\ &\quad + k_7 c_2 \delta, \quad \text{for all } 0 < \delta \leq t, \delta \leq 1. \end{aligned}$$

When $0 \leq t < \delta \leq 1$, (3.10) and (3.11) imply

$$(3.12) \quad E|x(t) - y(t)|^2 \leq k_8 \delta e^{-\lambda t}.$$

Combine (3.12) with the bounds on I, II, and III: for all $0 < \delta \leq 1$ and $t \geq 0$

$$(3.13) \quad \begin{aligned} E|x(t) - y(t)|^2 &\leq k_9 \left\{ \delta e^{-\lambda t} + (\rho(\delta) + \delta) \int_0^t e^{-\lambda(t-s)} Ex^2(s) ds \right. \\ &\quad \left. + (c_2 + y(t))(\rho(\delta) + \delta) \int_0^t e^{-\lambda(t-s)} \sqrt{Ex^2(s)} ds + c_2 \delta \right\}. \end{aligned}$$

For $T > 0$, let $z(T) = \sup_{s \in [0, T]} \sum Ex^2(s)$. If $t \in [0, T]$ then

$$(3.14) \quad \begin{aligned} Ex^2(t) &\leq 2y(t) \sqrt{Ex^2(t)} + E|x(t) - y(t)|^2 \\ &\leq \frac{k_9}{\lambda} (\rho(\delta) + \delta) z^2(T) + \left\{ 2k_1 + \frac{3k_9}{\lambda} (c_2 + k_1) \right\} z(T) + k_9 + k_9 c_2 \delta. \end{aligned}$$

Choose $0 < \alpha_1 < 1$, and then choose r_0 such that

$$\frac{k_9}{\lambda} r_0 = \alpha_1.$$

Note that r_0 depends only on $c_1, c_2, \gamma, \lambda, |x_0|$, and n . For any δ such that

$$(3.15) \quad \rho(\delta) + \delta < r_0$$

(3.14) implies

$$Ex^2(t) \leq \alpha_1 z^2(T) + \alpha_2 z(T) + \alpha_3$$

for all $t \in [0, T]$, where α_2 and α_3 depend only on $c_1, c_2, \gamma, \lambda, |x_0|$, and n , and where $\alpha_1 < 1$. Hence

$$\begin{aligned} z^2(T) &\leq \alpha_1 z^2(T) + \alpha_2 z(T) + \alpha_3 \\ &\Rightarrow z^2(T) < \alpha < \infty \end{aligned}$$

for some α depending only on $c_1, c_2, \gamma, \lambda, |x_0|$, and n . Finally then,

$$\sup_{t \in [0, \infty)} Ex^2(t) < \alpha < \infty.$$

Now look at (3.7). $A(t, t/\varepsilon, \omega)$ and $d(t, t/\varepsilon, \omega)$ have joint mixing rate

$$\rho_\varepsilon(\delta) = \rho(\delta/\varepsilon).$$

Hence (3.15) becomes

$$(3.16) \quad \rho(\delta/\varepsilon) + \delta < r_0.$$

Since $\rho(\delta) \rightarrow 0$, for all ε sufficiently small there will exist a δ satisfying (3.16). Therefore

$$(3.17) \quad \sup_{t \in [0, \infty)} Ex_\varepsilon^2(t) < \alpha < \infty$$

for all ε sufficiently small.

Return to (3.13). Put $\delta = \varepsilon^{1/2}$ and choose $\varepsilon < 1$ small enough so that (3.17) holds:

$$\begin{aligned} E|x_\varepsilon(t) - y_\varepsilon(t)|^2 &\leq \varepsilon^{1/2} k_9 + k_9(\rho(\varepsilon^{-1/2}) + \varepsilon^{1/2}) \frac{\alpha}{\lambda} \\ &\quad + k_9(c_2 + k_1)(\rho(\varepsilon^{-1/2}) + \varepsilon^{1/2}) \frac{\alpha^{1/2}}{\lambda} + k_9 c_2 \varepsilon^{1/2} \end{aligned}$$

for all $t \geq 0$, which gives Theorem 2.

To get Theorem 1, rewrite (3.13) using $c_2 = 0$ and the bound $|y(t)| = |\phi(t, o)x_0| \leq |x_0|\gamma e^{-\lambda t}$:

$$\begin{aligned} E|x(t) - y(t)|^2 &\leq k_9 e^{-\lambda t} + k_9(\rho(\delta) + \delta) \int_0^t e^{-\lambda(t-s)} Ex^2(s) ds \\ &\quad + 3k_9|x_0|\gamma e^{-\lambda t} \int_0^t e^{-\lambda(t-s)} \sqrt{Ex^2(s)} ds \end{aligned}$$

for all $\delta \leq 1$ and $t \geq 0$. Now suppose (3.15) holds for some $\delta \leq 1$. Then

$$\begin{aligned} Ex^2(t) &\leq 2y(t)\sqrt{Ex^2(t)} + E|x(t) - y(t)|^2 \\ &\leq 2|x_0|\gamma e^{-\lambda t} \alpha^{1/2} + k_9 e^{-\lambda t} + \lambda \alpha_1 \int_0^t e^{-\lambda(t-s)} Ex^2(s) ds + 3k_9|x_0|\gamma e^{-\lambda t} \frac{\alpha^{1/2}}{\lambda} \\ &= k_{10} e^{-\lambda t} + \lambda \alpha_1 \int_0^t e^{-\lambda(t-s)} Ex^2(s) ds \end{aligned}$$

for all $t \geq 0$, and some constant k_{10} . Multiply by $e^{\lambda t}$ and apply the Gronwall inequality (use (3.9) and (3.10) for the required continuity of $Ex^2(t)$) to obtain

$$Ex^2(t) \leq k_{10} e^{-\lambda t(1-\alpha_1)}.$$

Since $\alpha_1 < 1$,

$$(3.18) \quad Ex^2(t) \rightarrow 0,$$

which is one part of Theorem 1 (in a linear system, (3.18) cannot depend on $|x_0|$).

As it turns out, $Ex^2(t) \rightarrow 0$ fast enough to insure almost sure convergence as well.

Fix $\varepsilon > 0$ and choose $\eta > 0$ such that

$$\eta e^{c_1 \eta} \left(c_1 \frac{\varepsilon}{2} + c_2 \right) < \frac{\varepsilon}{2}.$$

If

$$A_n = \left\{ \omega : |x(n\eta)| \geq \frac{\varepsilon}{2} \right\}$$

then

$$P(A_n) \leq \frac{4}{\varepsilon} Ex^2(n\eta) \leq \frac{4}{\varepsilon} k_{10} e^{-\lambda n \eta (1-\alpha_1)}$$

$$\Rightarrow P(A_n \text{ infinitely often}) = 0.$$

Fix $\omega \in (A_n \text{ infinitely often})^c$ ($(\cdot)^c$ refers to set complement). For some N ,

$$n > N \Rightarrow |x(n\eta, \omega)| < \frac{\varepsilon}{2}.$$

For any $t > N\eta$, $|t - n_0\eta| \leq \eta$ for some $n_0 > N$. Hence, by (3.9), $|x(t, \omega) - x(n_0\eta, \omega)| \leq \eta e^{c_1\eta}(c_1(\varepsilon/2) + c_2) < \varepsilon/2 \Rightarrow |x(t, \omega)| < \varepsilon$. Q.E.D.

4. Averaging on finite intervals. Now look at some more general systems. Theorems 3 and 4 concern:

$$(4.1) \quad \begin{aligned} \dot{x}_\varepsilon(t, \omega) &= F(x_\varepsilon(t, \omega), \omega, t, t/\varepsilon), \\ \dot{y}_\varepsilon(t) &= G_\varepsilon(y_\varepsilon(t), t), \\ x_\varepsilon(0, \omega) &= y_\varepsilon(0) = x_0 \in R^n \end{aligned}$$

where

1. F is jointly measurable with respect to its four arguments, and for all i, j, k , and ω ,

$$F_i(x, \omega, t, \tau), \quad \frac{\partial}{\partial x_j} F_i(x, \omega, t, \tau), \quad \frac{\partial^2}{\partial x_j \partial x_k} F_i(x, \omega, t, \tau)$$

are continuous (in (x, t, τ)).

2. F is Type II mixing, with \mathcal{F}_0^t the σ -field generated by

$$\{F(x, \omega, s, \tau): 0 \leq \tau \leq t, -\infty < x < \infty, 0 \leq s < \infty\},$$

and \mathcal{F}_t^∞ the σ -field generated by

$$\{F(x, \omega, s, \tau): t \leq \tau < \infty, -\infty < x < \infty, 0 \leq s < \infty\}.$$

3. $G_\varepsilon(x, t) = E[F(x, \omega, t, t/\varepsilon)]$.

THEOREM 3. Assume also that:

4. There exist continuous functions $B_1(r, t)$, $B_2(r, t)$, and $B_3(r, t)$, such that for all $i, j, k, \tau \geq 0$, and ω :

- $|F_i(x, \omega, t, \tau)| \leq B_1(|x|, t)$;
- $|(\partial/\partial x_j)F_i(x, \omega, t, \tau)| \leq B_2(|x|, t)$;
- $|(\partial^2/\partial x_j \partial x_k)F_i(x, \omega, t, \tau)| \leq B_3(|x|, t)$.

5. $\sup_{\varepsilon > 0, t \in [0, T]} |y_\varepsilon(t)| \leq K$, for some constants T and K .

Then

$$\sup_{t \in [0, T]} |x_\varepsilon(t) - y_\varepsilon(t)| \rightarrow 0$$

in probability as $\varepsilon \rightarrow 0$.

Remarks: 1. 5 will be satisfied (for every $T > 0$) if, for example,

$$B_1(r, t) \leq B(t)(r+1)$$

for some continuous function $B(t)$.

2. 1 through 5 are not sufficient to guarantee the existence of a solution, $x_\varepsilon(t)$, on $[0, T]$ for every ω . Some realizations of $x_\varepsilon(t)$ may be singular, but this does not affect the conclusion of the theorem (the definition of $x_\varepsilon(t)$ after such a singularity is arbitrary). See § 6, Example C for a demonstration of this behavior.

3. For linear and certain related systems the theorem remains true using Type III in place of Type II mixing.

Proof. Fix η such that $1 > \eta > 0$. Let $R(r): R^1 \rightarrow R^1$ such that

$$R(r) = \begin{cases} 1, & r \leq (K+1)^2, \\ 0, & r \geq (K+2)^2 \end{cases}$$

and $R \in C^2$. Define

$$\hat{H}_\varepsilon(x, \omega, t) = R(|x|^2)F(x, \omega, t, t/\varepsilon)$$

and let

$$\hat{G}_\varepsilon(x, t) = E[\hat{H}_\varepsilon(x, \omega, t)].$$

Define $\hat{x}_\varepsilon(t, \omega)$ and $\hat{y}_\varepsilon(t)$ by

$$(4.2) \quad \begin{aligned} \hat{x}_\varepsilon(t, \omega) &= \hat{H}_\varepsilon(\hat{x}_\varepsilon(t, \omega), \omega, t), & \hat{x}(0, \omega) &= x_0, \quad \text{and} \\ \hat{y}_\varepsilon(t) &= \hat{G}_\varepsilon(\hat{y}_\varepsilon(t), t), & \hat{y}_\varepsilon(0) &= x_0. \end{aligned}$$

Condition 5 implies that

$$y_\varepsilon(t) = \hat{y}_\varepsilon(t)$$

for all $\varepsilon > 0$ and $t \in [0, T]$. Consequently, if for some ω_0 and $\varepsilon_0 > 0$,

$$\sup_{t \in [0, T]} |\hat{x}_{\varepsilon_0}(t, \omega_0) - \hat{y}_{\varepsilon_0}(t)| \leq \eta,$$

then

$$x_{\varepsilon_0}(t, \omega_0) = \hat{x}_{\varepsilon_0}(t, \omega_0)$$

for all $t \in [0, T]$. Hence

$$P\left\{ \sup_{t \in [0, T]} |x_\varepsilon(t) - y_\varepsilon(t)| > \eta \right\} \leq P\left\{ \sup_{t \in [0, T]} |\hat{x}_\varepsilon(t) - \hat{y}_\varepsilon(t)| > \eta \right\}$$

for all $\varepsilon > 0$. First we will prove

$$(4.3) \quad \lim_{\varepsilon \rightarrow 0} \sup_{t \in [0, T]} E|\hat{x}_\varepsilon(t) - \hat{y}_\varepsilon(t)|^2 = 0,$$

and then we will show that

$$(4.4) \quad \lim_{\varepsilon \rightarrow 0} P\left\{ \sup_{t \in [0, T]} |\hat{x}_\varepsilon(t) - \hat{y}_\varepsilon(t)| > \eta \right\} = 0$$

is a consequence of (4.3).

Let $\hat{g}_\varepsilon(t, s, x)$ be the solution to

$$(4.5) \quad \begin{aligned} \frac{\partial}{\partial t} \hat{g}_\varepsilon(t, s, x) &= \hat{G}_\varepsilon(\hat{g}_\varepsilon(t, s, x), t), \\ \hat{g}_\varepsilon(s, s, x) &= x. \end{aligned}$$

Here are some bounds for the $\hat{\cdot}$ system:

(i) because $R(|x|^2) = 0$ whenever $|x| \geq K+2$,

$$|\hat{x}_\varepsilon| \leq K+2$$

for all $\varepsilon > 0$, $t \in [0, T]$, and ω ;

ii) because of condition 4, there is a constant which bounds

$$|\hat{H}_{\varepsilon i}|, \quad \left| \frac{\partial}{\partial x_j} \hat{H}_{\varepsilon i} \right|, \quad \left| \frac{\partial^2}{\partial x_j \partial x_k} \hat{H}_{\varepsilon i} \right|, \quad |\hat{G}_{\varepsilon i}|, \quad \left| \frac{\partial}{\partial x_j} \hat{G}_{\varepsilon i} \right|, \quad \text{and} \quad \left| \frac{\partial^2}{\partial x_j \partial x_k} \hat{G}_{\varepsilon i} \right|$$

for all $i, j, k, t \in [0, T], \varepsilon > 0, \omega$, and x ;

iii) there is a constant which bounds

$$|\hat{g}_{\varepsilon i}|, \quad \left| \frac{\partial}{\partial x_j} \hat{g}_{\varepsilon i} \right|, \quad \text{and} \quad \left| \frac{\partial^2}{\partial x_j \partial x_k} \hat{g}_{\varepsilon i} \right|$$

for all $T \geq t > s \geq 0, \varepsilon > 0$, and x (differentiate (4.5) once, and then twice, with respect to x and use (ii)).

Choose δ and t such that $0 \leq \delta \leq t \leq T$, and apply Lemma 1 to (4.2) using $K(x) = |x - \hat{y}_\varepsilon(t)|^2$:

$$\begin{aligned} & E|\hat{x}_\varepsilon(t) - \hat{y}_\varepsilon(t)|^2 \\ &= \int_0^t \int_{\Omega \times \Omega} \left(\frac{\partial}{\partial x} K(\hat{g}_\varepsilon(t, s, \hat{x}_\varepsilon(s, \omega))) \right) \cdot \hat{H}_\varepsilon(\hat{x}_\varepsilon(s, \omega), \eta, s) dv_{s,0} ds \\ \text{(I)} \quad &= \int_0^\delta \int_{\Omega \times \Omega} \left(\frac{\partial}{\partial x} K(\hat{g}_\varepsilon(t, s, \hat{x}_\varepsilon(s, \omega))) \right) \cdot \hat{H}_\varepsilon(\hat{x}_\varepsilon(s, \omega), \eta, s) dv_{s,0} ds \\ \text{(II)} \quad &+ \int_\delta^t \int_{\Omega \times \Omega} \left(\frac{\partial}{\partial x} K(\hat{g}_\varepsilon(t, s, \hat{x}_\varepsilon(s - \delta, \omega))) \right) \cdot \hat{H}_\varepsilon(\hat{x}_\varepsilon(s - \delta, \omega), \eta, s) dv_{s,0} ds \\ &+ \int_\delta^t \int_{\Omega \times \Omega} \left\{ \left(\frac{\partial}{\partial x} K(\hat{g}_\varepsilon(t, s, \hat{x}_\varepsilon(s, \omega))) \right) \cdot \hat{H}_\varepsilon(\hat{x}_\varepsilon(s, \omega), \eta, s) \right. \\ \text{(III)} \quad &\left. - \left(\frac{\partial}{\partial x} K(\hat{g}_\varepsilon(t, s, \hat{x}_\varepsilon(s - \delta, \omega))) \right) \cdot \hat{H}_\varepsilon(\hat{x}_\varepsilon(s - \delta, \omega), \eta, s) \right\} dv_{s,0} ds. \end{aligned}$$

Now bound each term separately. (i), (ii), and (iii) imply

$$\text{“I”} = O(\delta).$$

In II, replace $v_{s,0}$ by $v_{s-\delta,\delta}$ and again use (i), (ii) and (iii):

$$\text{“II”} = O(\rho(\delta/\varepsilon))$$

(the mixing rate of F , and therefore \hat{H}_ε , is $\rho(\delta/\varepsilon)$). Finally,

$$\begin{aligned} \text{“III”} &\leq \int_\delta^t \int_{\Omega \times \Omega} |\hat{x}_\varepsilon(s, \omega) - \hat{x}_\varepsilon(s - \delta, \omega)| \\ &\quad \sup_{|x| \leq K+2} \left| \nabla \left\{ \left(\frac{\partial}{\partial x} K(\hat{g}_\varepsilon(t, s, x)) \right) \cdot \hat{H}_\varepsilon(x, \eta, s) \right\} \right| d|v|_{s,0} ds. \end{aligned}$$

From (i), (ii), and (iii):

$$|\hat{x}_\varepsilon(s, \omega) - \hat{x}_\varepsilon(s - \delta, \omega)| = O(\delta),$$

and

$$\sup_{|x| \leq K+2} \left| \nabla \left\{ \left(\frac{\partial}{\partial x} K(\hat{g}_\varepsilon(t, s, x)) \right) \cdot \hat{H}_\varepsilon(x, \eta, s) \right\} \right| = O(1),$$

and consequently

$$\text{“III”} = O(\delta) \quad (\text{for } t \in [\delta, T]).$$

Hence, for any $\delta \in [0, T]$,

$$\sup_{t \in [\delta, T]} E|\hat{x}_\varepsilon(t) - \hat{y}_\varepsilon(t)|^2 = O(\delta) + O(\rho(\delta/\varepsilon)).$$

Since

$$\sup_{t \in [0, \delta]} E|\hat{x}_\varepsilon(t) - \hat{y}_\varepsilon(t)|^2 = O(\delta),$$

$$\lim_{\varepsilon \rightarrow 0} \sup_{t \in [0, T]} E|\hat{x}_\varepsilon(t) - \hat{y}_\varepsilon(t)|^2 = O(\delta),$$

which proves (4.3).

Because of (ii), the derivatives

$$\dot{\hat{x}}_\varepsilon(t) \quad \text{and} \quad \dot{\hat{y}}_\varepsilon(t)$$

are bounded for all $\varepsilon > 0$ and $t \in [0, T]$. Under these circumstances, (4.4) follows easily from (4.3). Q.E.D.

5. Averaging on $[0, \infty)$. Here, the conditions are more severe. Look again at (4.1) with assumptions 1, 2, and 3, but restrict $B_1, B_2,$ and B_3 of assumption 4 to dependence on $|x|$ alone. Further, we require the solution of the deterministic equation, $y_\varepsilon(t)$, to have the following stability property: when perturbed, it returns asymptotically to its original trajectory. This is the ‘‘physical’’ meaning of assumption 6 below. (It does not imply $y_\varepsilon(t) \rightarrow y_0$, for some constant y_0 .) Since the random solution persistently wanders from the deterministic trajectory, some such stability for the latter is necessary. In bounded linear systems, 6 is equivalent to exponential stability, and in this sense, it is a generalization of the assumption made in § 3.

Finally, we assume that $x_\varepsilon(t, \omega)$ is bounded in $t, \omega,$ and $\varepsilon > 0$. A more delicate analysis, following along the lines of Theorem 2, would perhaps eliminate this restriction, provided that the right hand side be bounded (at large x) by a linear function of x .

A simple example of an equation to which Theorem 4 applies is

$$(5.1) \quad \dot{x}_\varepsilon(t, \omega) = -x_\varepsilon(t, \omega) + \frac{\phi(t/\varepsilon, \omega)}{1 + x_\varepsilon(t, \omega)^2}, \quad x_\varepsilon(0, \omega) = 0.$$

If ϕ is bounded, Type II mixing, and (say) $E[\phi] = 0$, then

$$\lim_{\varepsilon \rightarrow 0} \sup_{t \geq 0} E x_\varepsilon(t)^2 = 0.$$

THEOREM 4. *In (4.1), assume 1, 2, 3, and:*

4. *There exist continuous functions $B_1(r), B_2(r),$ and $B_3(r),$ such that for all $i, j, k,$ $t \geq 0, \tau \geq 0,$ and ω :*

- a. $|F_i(x, \omega, t, \tau)| \leq B_1(|x|);$
- b. $|(\partial/\partial x_j)F_i(x, \omega, t, \tau)| \leq B_2(|x|);$
- c. $|(\partial^2/\partial x_j \partial x_k)F_i(x, \omega, t, \tau)| \leq B_3(|x|).$

5. *There exists a bounded region D such that*

$$x_\varepsilon(t, \omega) \in D$$

for all $\varepsilon > 0, t \geq 0,$ and $\omega.$

Let $g_\varepsilon(t, s, x)$ be the solution to

$$\frac{\partial}{\partial t} g_\varepsilon(t, s, x) = G_\varepsilon(g_\varepsilon(t, s, x), t), \quad g_\varepsilon(s, s, x) = x.$$

6. *There exists a constant $K,$ and continuous functions $M_1(t, s)$ and $M_2(t, s),$ such that:*

- a. $\sup_{x \in D} |g_\varepsilon(t, s, x)| \leq K$ for all $t \geq s \geq 0, \varepsilon > 0;$
- b. $K \geq M_1(t, s) \geq \sup_{x \in D} |\partial/\partial x| g_\varepsilon(t, s, x)|$ for all $t \geq s \geq 0, \varepsilon > 0;$

- c. $M_2(t, s) \geq \sup_{x \in D} |(\partial^2 / \partial x_i \partial x_j) g_\varepsilon(t, s, x)|$ for all $i, j, t \geq s \geq 0, \varepsilon > 0$;
- d. $\int_0^t M_1(t, s) ds \leq K$ for all $t > 0$;
- e. $\int_0^t M_2(t, s) ds \leq K$ for all $t > 0$.

Then

$$\lim_{\varepsilon \rightarrow 0} \sup_{t \geq 0} E|x_\varepsilon(t) - y_\varepsilon(t)|^2 = 0.$$

Remark. 6 can be replaced by 6a, and, for some $\alpha > 0$,

$$(5.2) \quad \sup_{|y| \leq K} x^T \left(\frac{\partial}{\partial x} G_\varepsilon(y, t) \right) x < -\alpha |x|^2$$

for all $x \in R^n, \varepsilon > 0$, and $t \geq 0$ (K as defined in 6a). The remainder of 6 is then a consequence of (5.2). In the example, (5.1), (5.2) holds with any $\alpha < 1$.

Proof. By now the approach is obvious, so only an outline is included. Using Lemma 1, bound

$$E|x_\varepsilon(t) - y_\varepsilon(t)|^2$$

by the usual 3 terms. Proceed as in Theorem 3, but use $M_1(t, s)$ or $M_2(t, s)$ whenever $((\partial / \partial x) g_\varepsilon(t, s, x))$ or $((\partial^2 / \partial x_i \partial x_j) g_\varepsilon(t, s, x))$ appears. Everything else is bounded by constants. The integrability of M_1 and M_2 allows the conclusion

$$\sup_{t \geq 0} E|x_\varepsilon(t) - y_\varepsilon(t)|^2 = O(\delta) + O(\rho(\delta/\varepsilon)),$$

and the theorem follows.

6. Examples. A. Consider an unforced damped spring-mass system:

$$(6.1) \quad \ddot{x}_\varepsilon(t, \omega) + \{c + f(t/\varepsilon, \omega)\} \dot{x}_\varepsilon(t, \omega) + \{k + g(t/\varepsilon, \omega)\} x_\varepsilon(t, \omega) = 0$$

where

1. c and k are positive constants,
2. $f(t, \omega)$ and $g(t, \omega)$ are zero mean, piecewise continuous, jointly type I mixing, and uniformly bounded in t and ω (for example generate f and g from independent Markov chains with finite state spaces and irreducible aperiodic transition matrices).

From Theorem 1 we conclude that for all ε sufficiently small

$$x_\varepsilon(t) \rightarrow 0$$

in mean square and almost surely. See Infante [8] and Blankenship [2] for discussions of stability in (6.1) from a slightly different point of view.

B. Suppose that we observe an R^n -valued stochastic process $\phi(t, \omega)$, together with an R^1 -valued stochastic process $z(t, \omega)$. We wish to choose a coefficient vector $x \in R^n$ such that

$$x \cdot \phi(t)$$

approximates, in the sense of minimum mean square error, $z(t)$. If the marginal distribution of $(\phi(t), z(t))$ is independent of t , then the problem has a well defined solution:

$$\hat{x} = E[\phi \phi^T]^{-1} E[\phi z]$$

(assuming $E[\phi \phi^T]$ is nonsingular). An obvious adaptation of the stochastic approximation algorithm for computing \hat{x} (cf. Duda and Hart [6] or Wasan [18]) is

$$\dot{x}_\varepsilon(t, \omega) = \varepsilon \phi(t, \omega) \{z(t, \omega) - \phi(t, \omega) \cdot x_\varepsilon(t, \omega)\}.$$

Assume (in addition to the above assumptions) that $(\phi(t), z(t))$ is piecewise continuous, Type I mixing, and uniformly bounded in t and ω . Theorem 2 implies

$$(6.2) \quad \lim_{\varepsilon \rightarrow 0} \overline{\lim}_{t \rightarrow \infty} E|x_\varepsilon(t) - \hat{x}|^2 = 0.$$

(Note: As is the case in discrete time, replacing ε by $1/t$ allows for the replacement of (6.2) by $x(t) \rightarrow \hat{x}$ a.s. This and other versions of continuous time stochastic approximation will be discussed in a future article. See Kushner [11] and Ljung [12], and references therein, for some recent publications in this area.)

C. As an illustration of the behavior discussed in remark 2 following Theorem 3, consider the Bernoulli equation with random coefficients:

$$(6.3) \quad \begin{aligned} \dot{x}_\varepsilon(t, \omega) + p(t/\varepsilon, \omega)x_\varepsilon(t, \omega) + q(t/\varepsilon, \omega)x_\varepsilon(t, \omega)^r &= 0, \\ x_\varepsilon(0, \omega) &= 1 \end{aligned}$$

where r is any constant except 0 or 1. Assume that $(p(t, \omega), q(t, \omega))$ is continuous, Type I mixing, and uniformly bounded in t and ω . Since the solution to (6.3) is

$$x_\varepsilon(t, \omega) = \left\{ e^{(r-1)\int_0^t p(s/\varepsilon, \omega) ds} + (r-1) \int_0^t e^{(r-1)\int_s^t p(u/\varepsilon, \omega) du} q(s/\varepsilon, \omega) ds \right\}^{-1/(r-1)},$$

$x_\varepsilon(t, \omega)$ can be singular at finite t . For each ω , define $x(t, \omega)$ to be zero after a singularity.

If $\bar{p}(t) = E[p(t, \omega)]$ and $\bar{q}(t) = E[q(t, \omega)]$, then the averaged equation has solution

$$y_\varepsilon(t) = \left\{ e^{(r-1)\int_0^t \bar{p}(s/\varepsilon) ds} + (r-1) \int_0^t e^{(r-1)\int_s^t \bar{p}(u/\varepsilon) du} \bar{q}(s/\varepsilon) ds \right\}^{-1/(r-1)}.$$

To avoid singularities in $y_\varepsilon(t)$, assume that for all $t \geq 0$

$$\begin{aligned} \bar{q}(t) &\geq 0 \quad \text{if } r > 1 \quad \text{or} \\ \bar{q}(t) &\leq 0 \quad \text{if } r < 1. \end{aligned}$$

For any $T > 0$ and $\eta > 0$, Theorem 3 asserts that

$$\lim_{\varepsilon \rightarrow 0} P\left\{ \sup_{t \in [0, T]} |x_\varepsilon(t) - y_\varepsilon(t)| > \eta \right\} = 0.$$

Acknowledgment. My initial proof of Lemma 1 was much too long and much too obscure. I owe a special thanks to Professor Michael Woodroffe for convincing me that something simpler must be going on. Also, it is a pleasure to acknowledge the many valuable suggestions offered by Professor Frank Kozin.

REFERENCES

- [1] P. BILLINGSLEY, *Convergence of Probability Measures*, John Wiley, New York, 1968.
- [2] G. BLANKENSHIP, *Stability of linear differential equations with random coefficients*. IEEE Trans. Automatic Control, AC-22, 5 (1977), pp. 834-838.
- [3] G. BLANKENSHIP AND G. C. PAPANICOLAOU, *Stability and control of stochastic systems with wide-band noise disturbances*, this Journal, 34 (1978), pp. 437-476.
- [4] R. W. BROCKETT, *Finite Dimensional Linear Systems*, John Wiley, New York, 1970.
- [5] R. COGBURN AND R. HERSH, *Two limit theorems for random differential equations*, Indiana Univ. Math. J., 22 (1973), pp. 1067-1089.
- [6] R. O. DUDA AND P. E. HART, *Pattern Classification and Scene Analysis*, John Wiley, New York, 1973.
- [7] P. HARTMAN, *Ordinary Differential Equations*, John Wiley, New York, 1964.
- [8] E. F. INFANTE, *On the stability of some linear nonautonomous random systems*, J. Appl. Mech., 35 (1968), pp. 7-12.

- [9] R. Z. KHASHMINSKII, *On stochastic processes defined by differential equations with a small parameter*, Theor. Probability Appl., 11 (1966), pp. 211–228.
- [10] ———, *A limit theorem for the solutions of differential equations with random right-hand sides*, Ibid., 11 (1966), pp. 390–406.
- [11] H. J. KUSHNER, *Convergence of recursive adaptive and identification procedures via weak convergence theory*, IEEE Transactions Automatic Control, AC-22 (1977), pp. 921–930.
- [12] L. LJUNG, *Strong convergence of a stochastic approximation algorithm*, Ann. Statist., 6 (1978), pp. 680–696.
- [13] E. R. LYBRAND, *Method of averaging for stochastic systems*, Ph.D. thesis, Electrical Engineering Dept., Polytechnic Institute of Brooklyn, June, 1975.
- [14] G. C. PAPANICOLAOU AND W. KOHLER, *Asymptotic theory of mixing stochastic ordinary differential equations*, Comm. Pure Appl. Math., 27 (1974), pp. 641–668.
- [15] M. ROSENBLATT, *A central limit theorem and a strong mixing condition*, Proc. Nat. Acad. Sci., 42 (1956), pp. 43–47.
- [16] V. A. VOLKONSKII AND YU. A. ROZANOV, *Some limit theorems for random functions I*, Theor. Probability Appl., 4 (1959), pp. 178–197.
- [17] I. VRKOC, *Extension of the averaging method to stochastic equations*, Chekh. Mat. Zh., 16 (1966), pp. 518–540.
- [18] M. T. WASAN, *Stochastic Approximations*, Cambridge University Press, Cambridge, 1969.
- [19] B. S. WHITE, *Some limit theorems for stochastic delay-differential equations*, Comm. Pure Appl. Math., 29 (1976), pp. 113–141.