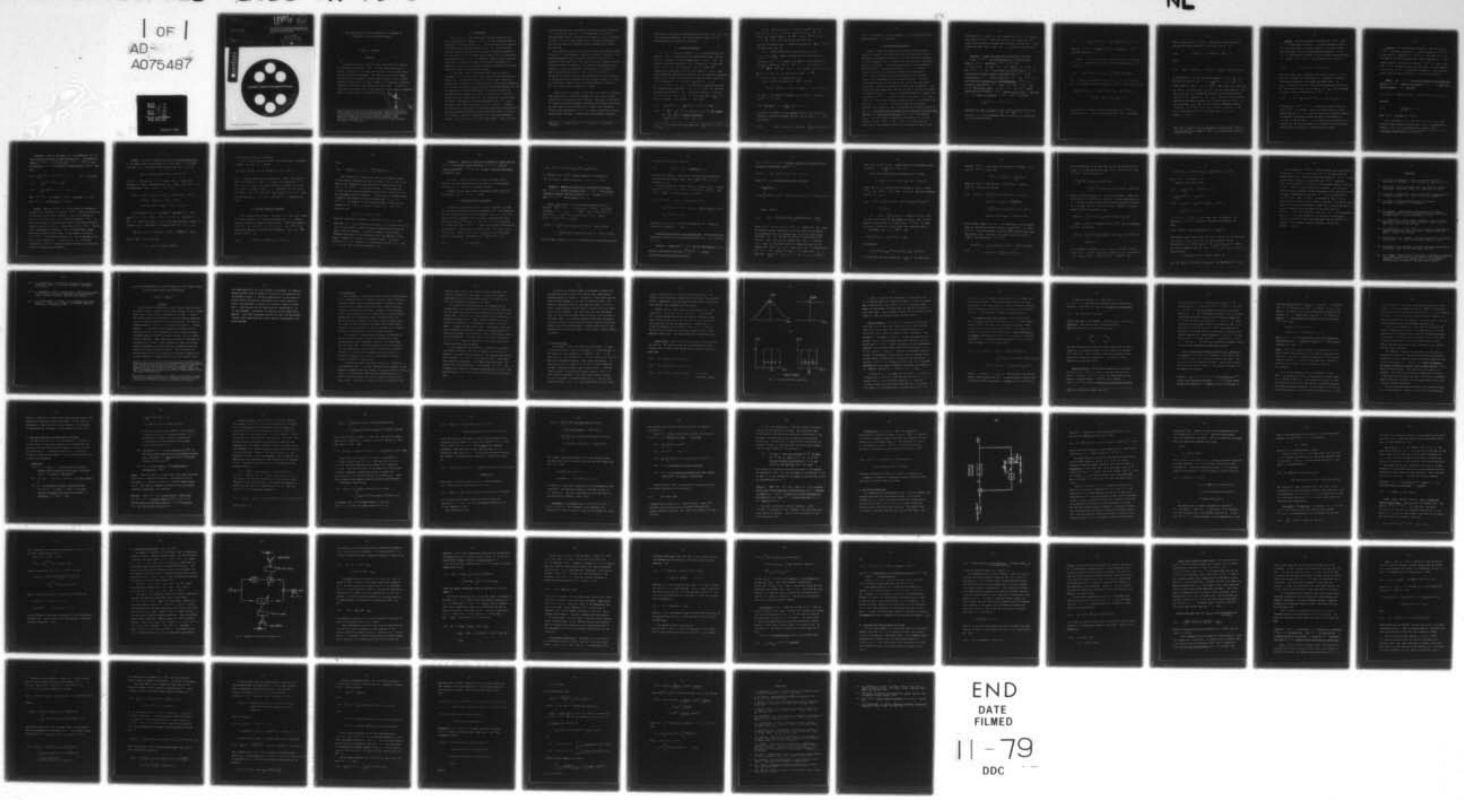
AD-A075 487 BROWN UNIV PROVIDENCE R I LEFSCHETZ CENTER FOR DYNAM--ETC F/G 12/1
TWO PAPERS ON JUMP DIFFUSION APPROXIMATIONS TO OUTPUT PROCESSES--ETC(U)
JUL 79 H J KUSHNER AFOSR-76-3063
UNCLASSIFIED LCDS-TR-79-3 NL

1 OF 1
AD-
A075487



END
DATE
FILMED
11-79
DDC

LCDS-TR-79-3

TWO PAPERS ON JUMP DIFFUSION APPROXIMATIONS TO OUTPUT PROCESSES OF NONLINEAR SYSTEMS WITH WIDE BAND INPUTS AND APPLICATIONS

HAROLD J. KUSHNER

JULY 1979

AFOSR-76-3063,
N00014-76-C-0279

Lefschetz Center for Dynamical Systems

Division of Applied Mathematics
Brown University Providence RI 02912

A MARTINGALE METHOD FOR THE CONVERGENCE OF A SEQUENCE OF PROCESSES TO A JUMP-DIFFUSION PROCESS

HAROLD J. KUSHNER[+]

Abstract

A convenient method for proving weak convergence of a sequence of non-Markovian processes $x^\epsilon(\cdot)$ to a jump-diffusion process is proved. Basically, it is shown that the limit solves the martingale problem of Strook and Varadhan. The proofs are relatively simple, and the conditions apparently weaker than required by other current methods (in particular, for limit theorems for a sequence of ordinary differential equations with random right hand sides). In order to illustrate the relative ease of applicability in many cases, a simpler proof of a known result on averaging is given.

[+]Brown University, Divisions of Applied Mathematics and Engineering and Lefschetz Center for Dynamical Systems. Research supported in part by the Air Force Office of Scientific Research under AFOSR AF-76-3063, by the National Science Foundation under NSF Eng. 73-03846A03 and in part by the Office of Naval Research ONR N00014-76-C-0279 P003.

1. Introduction

In [1], [2], [3], Kurtz developed a general theory for the weak convergence of a sequence $\{x^{\varepsilon}(\cdot)\}$ of (not necessarily Markov) processes to a Markov process $x(\cdot)$. The usefulness of his approach is attested to by the applications in [4], [5]. Kurtz's method involved a general approach to the problem of semigroup approximation and using it in [5] we were able to improve upon the approximations results of (e.g.) [6], [7] with both simpler proofs and better conditions. Reference [5] dealt with the weak convergence of a sequence of solutions to ordinary differential equations with random right hand sides to jump-diffusion process.

In this paper, some of the remarks and ideas on proving tightness in [3] are exploited to get a result of the same type in a simpler way. The general theory of [3] is not used. Instead of showing weak convergence to a Markov process, we show weak convergence to a solution of the martingale problem of Strook and Varadhan [8]-[9]. Our approach has several advantages: the proofs are short and straightforward, less care is required in the choice of the test functions (Kurtz's dense class $\mathscr{D}$), the conditions are often easier to verify - especially for the non-homogeneous case (e.g., we do not need to deal with the density of $(A-\lambda)\mathscr{D}$, see [3]). Finally, the extension to convergence of continuous parameter interpolations of discrete parameter processes is easier and more natural in that we do not need to treat the interpolation as a continuous parameter process - but only look at it at the "jump points". Also, for the discrete parameter case, our method often allows the perturbations+ $\{f^{\varepsilon}\}$ of the test function f to be obtained by relatively straightforward means. Some additional remarks on this case appear in Section 4, but a fuller exposition of the discrete parameter case will be given in a subsequent paper, together with several applications.

The problem is set up in Section 2 and the main theorem proved in Section 3. Some remarks on tightness are made in Section 3, the discrete parameter case is treated in Section 4 and Section 5 illustrates the simplicity of the idea by giving a simpler proof of a classical problem of averaging. One possible disadvantage of our approach is that tightness must be proved first. This does not seem to be a handicap however, since in many applications, one can do this quite easily by the approximation method developed in the sequel and the tightness results in Kurtz's paper [3]. This is illustrated in Theorem 5 or by the way tightness is proved in [5], [10], where Kurtz's scheme [3] is used.

Since writing this paper, it has come to the author's attention that Papanicolaou, Strook and Varadhan [13] have used a related method to get a variety of limit theorems. That work is restricted to a purely continuous parameter and Markovian framework and, except for some special results which do not overlap ours, the Markov driving noise (analogous to $y(\cdot)$ in (3.6a)) is required to have a strong ergodic property. Within their ergodic-Markov framework,

+We use $\varepsilon \to 0$ rather than $n \to \infty$ (as in [3]) to index the sequence.

they treat a wider class of problems than we do here. Our $x^{\varepsilon}(\cdot)$ are not necessarily Markovian, and various ideas of Kurtz [3] are adopted to prove the requisite tightness.

2. Problem Formulation

Let $\{x^{\varepsilon}(\cdot)\}$ denote a sequence of R^r valued processes with paths in $D^r[0,\infty)$ (endowed with the Skorokhod topology), and all (w.l.o.g.) defined on the same sample space $(\Omega, P, \mathcal{F})$. Let $\{\mathcal{F}_t\}$ denote an increasing sequence of sub σ-algebras of $\mathcal{F}$ such that $\{x^{\varepsilon}(s), s \leq t\}$ is $\mathcal{F}_t$ measurable. Let $\mathcal{L}$ denote the progressively measurable functions with respect to $\{\mathcal{F}_t\}$. There are progressively measurable versions of all the functions introduced below, and we assume that those are the versions used. Define $\overline{\mathcal{L}}$ to be the subset of $\mathcal{L}$ for which $\sup_t E|f(t)| < \infty$. Let $S_N = \{x: |x| \leq N\}$. Let $\mathcal{F}_t^{\varepsilon} \subset \mathcal{F}_t$ denote an increasing sequence of σ-algebras which measures $x^{\varepsilon}(s)$, $s \leq t$, and let E_t^{ε} denote expectation conditioned on $\mathcal{F}_t^{\varepsilon}$. Define the operators p-lim and $\hat{A}^{\varepsilon}$ as in [3] by: For f and $f^{\delta} \in \overline{\mathcal{L}}$, we say that $\text{p-lim}_{\delta \to 0} f^{\delta} = f$ iff (2.1) holds. (2.2) defines $\hat{A}^{\varepsilon}$ and $\mathcal{D}(\hat{A}^{\varepsilon})$.

(2.1) $\sup_{t,\delta} E|f^{\delta}(t)| < \infty$, $\lim_{\delta \to 0} E|f^{\delta}(t) - f(t)| = 0$ <u>each</u> t;

(2.2) $f \in \mathcal{D}(\hat{A}^{\varepsilon})$ <u>and</u> $\hat{A}^{\varepsilon} f = g$ <u>iff for</u> $f, g \in \overline{\mathcal{L}}$ <u>and adapted to</u> $\{\mathcal{F}_t^{\varepsilon}\}$ <u>and</u> g <u>p-right continuous</u>⁺,

$$\text{p-lim}_{\delta \to 0} \left[\frac{1}{\delta} (E_t^{\varepsilon} f(t+\delta) - f(t)) - g(t)\right] = 0.$$

⁺For $g \in \overline{\mathcal{L}}$, p-right continuity means that $\lim_{\delta \to 0} E|g(t+\delta) - g(t)| = 0$ for each $t \geq 0$.

Let $\hat{\mathcal{C}}$ denote the space of continuous bounded real-valued functions on $R^r \times R^+$ which are zero at infinity, $\hat{\mathcal{C}}_0$ the subset with compact support, and $\hat{\mathcal{C}}_0^{\alpha,\beta}$ (resp., $\hat{\mathcal{C}}^{\alpha,\beta}$) the subset of $\hat{\mathcal{C}}_0$ (resp., $\hat{\mathcal{C}}$) with bounded and continuous α t-partial derivatives and β x-partial derivatives. $\mathcal{B}(R^q)$ denotes the Borel field over R^q.

Let $\tilde{A}$ denote the diffusion operator $\sum_i b_i(x,t) \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i,j} a_{ij}(x,t) \frac{\partial^2}{\partial x_i \partial x_j}$ where we assume that the coefficients are continuous, let $\lambda(\cdot,\cdot)$ denote a non-negative real-valued bounded continuous function on $R^q \times R^+$. Let $\mu(\cdot,\cdot,\cdot)$ be a function on $R^r \times R^+ \times \mathcal{B}(R^q)$ such that $\mu(x,t,\cdot)$ is a measure of total mass unity on $\mathcal{B}(R^q)$ for each x,t, and such that for each $f \in \hat{\mathcal{C}}$, $\int f(x+y,t)\mu(x,t,dy)$ is continuous in x,t.

Let A denote the operator on $\hat{\mathcal{C}}_0^{1,2}$ given by

$$Af(x,t) = \tilde{A}f(x,t) + \lambda(x,t) \int [f(x+y,t) - f(x,t)]\mu(x,t,dy).$$

For each x let there be a measure P_x on $D^r[0,\infty)$ such that $P_x\{x(0) = x\} = 1$, and

(2.3) <u>for each</u> $T < \infty$, $P_x\{\sup_{t \le T} |x(t)| < \infty\} = 1$,

and which is assumed to be the <u>unique</u> solution to the martingale problem of Strook and Varadhan [8]-[9]: vis, for each $f(\cdot) \in \hat{\mathcal{C}}_0^{1,2}$, the function

(2.4) $$M_f(t) = f(x(t),t) - f(x(0),0) - \int_0^t (\frac{\partial}{\partial s} + A) f(x(s),s)ds$$

is a P_x martingale. By the uniqueness, x can be replaced by any random variable X in the above.

3. The Convergence Theorem

Theorem 1 requires tightness; since we do not work within a semigroup framework, we cannot first prove convergence of finite dimensional distributions as done in [3], and then prove tightness. For many problems (as will be seen) this is no barrier to effective use of our approach. It is often much easier to prove tightness when the $\{x^\varepsilon(t)\}$ are bounded uniformly in ε, t. In order to exploit this fact, we introduce a sequence of truncations of $\{x^\varepsilon(\cdot)\}$, prove convergence for each truncation, and then take limits to get convergence for the original sequence. Section 5 illustrates how such an idea can be used for at least one class of problems. See also the remark after the proof of Theorem 1. The truncation method seems quite natural and easy to use for the classes of problems in [5]-[7]. In fact, the general method for creating the truncations $x^{\varepsilon,N}(\cdot)$ is simply to stop or "slow down" $x^\varepsilon(\cdot)$ when $|x^\varepsilon(t)| \geq N$. The limits of these truncations (as $\varepsilon \to 0$) determine the limit as $\varepsilon \to 0$ of $x^\varepsilon(\cdot)$, under our uniqueness hypothesis. Because of the truncation idea, we introduce the following definition. Let $X(\cdot)$ denote the process defined by the solution to the martingale problem (2.4) with initial condition $X(0)$ (i.e., with measure $P_{X(0)}$). Let A^N denote an operator of the form of A with coefficients $a^N(\cdot,\cdot), b^N(\cdot,\cdot), \lambda^N(\cdot,\cdot), \mu^N(\cdot,\cdot,\cdot)$ satisfying the conditions in the paragraph above (2.3) and which are equal to the coefficients of A when $|x| \leq N$. Suppose that $X^N(\cdot)$ is a process with paths in $D^r[0,\infty)$ such that $X^N(0) \to X(0)$ weakly and $X^N(\cdot)$ solves (not necessarily uniquely) the martingale problem (2.4) - with initial condition $X^N(0)$. Then $X^N(\cdot)$ is called an N-truncation of $X(\cdot)$.

Theorem 1. Assume the conditions of Section 2. For each $\varepsilon > 0$, let $\{x^{\varepsilon,N}(\cdot)\}$ be a bounded[+] sequence of processes with paths in $D^r[0,\infty)$ satisfying $x^{\varepsilon,N}(t) = x^\varepsilon(t)$ for $t \leq \inf\{s: |x^\varepsilon(s)| > N\}$. Let $x^\varepsilon(0) \to X(0)$ weakly. If $|x^\varepsilon(0)| > N$, set $x^{\varepsilon,N}(t) \equiv 0$. Let $\mathcal{D}$ be a dense (in sup norm) set in $\hat{\mathcal{C}}$ of functions with compact support, and suppose that $X^N(\cdot)$ is an N-truncation of $X(\cdot)$ (initial condition $X(0)$). Let $\mathcal{F}_t^{\varepsilon,N} \subset \mathcal{F}_t$ be an increasing sequence of σ-algebras measuring $x^{\varepsilon,N}(s)$, $s \leq t$, and define the operator $\hat{A}^{\varepsilon,N}$ (using $\mathcal{F}_t^{\varepsilon,N}$ and $x^{\varepsilon,N}(\cdot)$) just as $\hat{A}^\varepsilon$ was defined (using $\mathcal{F}_t^\varepsilon$ and $x^\varepsilon(\cdot)$). Let $E_t^{\varepsilon,N}$ denote expectation conditioned on $\mathcal{F}_t^{\varepsilon,N}$. *Suppose that for each* N *and each* $f(\cdot) \in \mathcal{D}$, there is a sequence $f^{\varepsilon,N}(\cdot) \in \mathcal{D}(\hat{A}^{\varepsilon,N})$ such that

(3.1) $$\operatorname*{p-lim}_{\varepsilon\to 0}[f^{\varepsilon,N}(\cdot) - f(x^{\varepsilon,N}(\cdot),\cdot)] = 0$$

[+]Actually, all that is needed is $\lim_{K\to\infty} \sup_\varepsilon P\{\sup_{t\leq T}|x^{\varepsilon,N}(t)| \geq K\} = 0$ for each $T > 0$, but boundedness does not seem to be a restriction in our truncation method.

$$(3.2)\qquad \underset{\varepsilon\to 0}{\text{p-lim}}\, [\hat{A}^{\varepsilon,N} f^{\varepsilon,N}(\cdot) - (\frac{\partial}{\partial t} + A^N) f(x^{\varepsilon,N}(\cdot),\cdot)] = 0.$$

Then, if $\{x^{\varepsilon,N}(\cdot)\}$ _is tight in_ $D^r[0,\infty)$ _for each_ N, $\{x^{\varepsilon}(\cdot)\}$ _converges weakly to_ $X(\cdot)$.

Proof. Fix N and $f(\cdot) \in \mathcal{D}$. Choose $\{f^{\varepsilon,N}(\cdot)\}$ satisfying (3.1)-(3.2). Equation (3.3) follows from the definition of $\hat{A}^{\varepsilon,N}$ (see Kurtz [3])

$$(3.3)\qquad E_t^{\varepsilon,N} f^{\varepsilon,N}(t+s) - f^{\varepsilon,N}(t) = \int_t^{t+s} E_t^{\varepsilon,N} \hat{A}^{\varepsilon,N} f^{\varepsilon,N}(u)du.$$

Let q denote an arbitrary integer, let $t_i \le t$, $i \le q$, and let $g(\cdot)$ be a bounded continuous function on R^{rq}. Then by (3.1)-(3.3),

$$\lim_{\varepsilon\to 0} Eg(x^{\varepsilon,N}(t_i),\ i \le q)[f(x^{\varepsilon,N}(t+s),t+s) - f(x^{\varepsilon,N}(t),t)$$

$$- \int_t^{t+s} (\frac{\partial}{\partial t} + A^N) f(x^{\varepsilon,N}(u),u)du] = 0.$$

Choose and fix a weakly convergent subsequence of $\{x^{\varepsilon,N}(\cdot)\}$, indexed also by ε and with limit denoted by $x^N(\cdot)$ and where

$x^N(0)$ = weak limit of $\{x^{N,\varepsilon}(0)\}$. Then, by the weak convergence+ and the continuity properties of the coefficients of A^N,

$$(3.4)\qquad Eg(X^N(t_i),\ i \le q)[M_f^N(t+s) - M_f^N(t)] = 0$$

where

$$(3.5)\qquad M_f^N(t) = f(X^N(t),t) - f(X^N(0),0) - \int_0^t (\frac{\partial}{\partial t} + A^N)f(X^N(u),u)du.$$

By the density of $\mathcal{D}$ and the arbitrariness+ of q, $\{t_i\}$ and $g(\cdot)$ and the fact that for each $T < \infty$, $P\{\sup_{t\le T}|X^N(t)| < \infty\} = 1$, $M_f^N(\cdot)$ is a martingale for each N, $f(\cdot,\cdot) \in \hat{C}_0^{1,2}$. Thus, there is a measure P_N on $D^r[0,\infty)$ such that under P_N, $x(0)$ has the distribution of $X^N(0)$ and P_N solves (not necessarily uniquely) the martingale problem (2.4) with A^N replacing A. Let $P_{X(0)}$ denote the unique measure solving the martingale problem (2.4) (initial condition $X(0)$). By the uniqueness, $P_{X(0)}$ and P_M must agree on measurable subsets of the set $\{x(\cdot): |x(t)| \le N, t \le T\} \in D^r[0,\infty)$ for each T and $M \ge N$. From this and (2.3), it follows that $x^\varepsilon(\cdot) \to X(\cdot)$ weakly. Q.E.D.

+For (3.4) to hold, it may be necessary to require that $t_i \notin T_0$, a null set which might depend on $X^N(\cdot)$, but this deletion does not affect the proof.

Remarks. One useful method for getting the $f^{\varepsilon,N}(\cdot)$ from $f(\cdot,\cdot)$ is developed in Section 5. A similar method was used in [5], [10], where Kurtz's original method [3] was used to prove weak convergence, and no N-truncations were required. In [5], $\{x^{\varepsilon}(\cdot)\}$ was defined as follows. For a stationary ϕ-mixing process $y(\cdot)$, define $y^{\varepsilon}(t) = y(t/\varepsilon^2)$ and for appropriate F,G, let

(3.6a) $$\dot{x}^{\varepsilon} = G(x^{\varepsilon},t,y^{\varepsilon}) + F(x^{\varepsilon},t,y^{\varepsilon})/\varepsilon.$$

The form (3.6a) uses a frequently used scaling (see [6], [7]) to get limit theorems for the solutions to a sequence of ordinary differential equations with random right hand sides. Let $q_N(\cdot)$ denote a real valued infinitely differentiable function on R^r which equals one on S_N, is zero on $R^r - S_{N+1}$, and is bounded by unity. Then, we might define $x^{\varepsilon,N}(\cdot)$ by

(3.6b) $$\dot{x}^{\varepsilon,N} = q_N(x^{\varepsilon,N})[G(x^{\varepsilon,N},t,y^{\varepsilon}) + F(x^{\varepsilon,N},t,y^{\varepsilon})/\varepsilon].$$

The sequence $\{x^{\varepsilon,N}(\cdot),\ \varepsilon > 0\}$ is obviously tight for each N, since $\dot{x}^{\varepsilon,N}(\cdot)$ is bounded. Such a truncation device is often useful when the $x^{\varepsilon}(\cdot)$ is given via an explicit dynamical equation. The method of this paper avoids the need for the density of $(\lambda-A)\mathcal{D}$ in $\hat{\mathcal{C}}$, for some $\lambda > 0$, as required by the method of [3]. Essentially, the density condition is replaced by the uniqueness condition for the solution to the martingale problem.

Tightness. We have assumed, in Theorem 1, that for each N, $x^{\varepsilon,N}(\cdot)$ is bounded uniformly in ε. Thus tightness of $\{x^{\varepsilon,N}(\cdot)\}$ is equivalent to tightness of $\{f(x^{\varepsilon,N}(\cdot))\}$ for each $f(\cdot) \in \hat{C}_0$, the space of continuous bounded real valued functions on R^r with compact support. We prefer to deal with $f(x^{\varepsilon,N}(\cdot))$ rather than with $x^{\varepsilon,N}(\cdot)$ directly because, as seen below, it often enables us to use the $f^{\varepsilon,N}(\cdot)$ constructed in Theorem 1. The following special case of Kurtz's theorem [3, (4.20)] will be useful.

Lemma 1. *Let* $\{y^{\varepsilon}(\cdot)\}$ *be a bounded sequence of processes with paths in* $D^r[0,\infty)$. *Suppose that for each* $T < \infty$, $\delta > 0$, *there is a random variable* $\gamma_{\varepsilon}(\delta)$ *such that*$^{+}$

$$(3.7)\quad E_t^{\varepsilon}\gamma_{\varepsilon}(\delta) \geq E_t^{\varepsilon}\min[1, |y^{\varepsilon}(t+u) - y^{\varepsilon}(t)|^2], \text{ all } u \leq \delta,\ t \leq T,$$

and that

$$(3.8)\qquad \lim_{\delta\to 0}\ \overline{\lim_{\varepsilon\to 0}}\ E\gamma_{\varepsilon}(\delta) = 0.$$

Then $\{y^{\varepsilon}(\cdot)\}$ *is tight in* $D^{\gamma}[0,\infty)$.

For our problem, $y^{\varepsilon}(\cdot) = f(x^{\varepsilon,N}(\cdot))$, for $f(\cdot) \in \hat{C}_0$. The following useful method of getting the $\gamma_{\varepsilon}(\delta)$ is suggested by a comment in [3], and is an extension of a result in [5].

$^{+}E_t^{\varepsilon}$ denotes conditioning on $y^{\varepsilon}(s)$, $s \leq t$.

Theorem 2. Fix N. For each $f(\cdot)$ in a dense set $\mathcal{D}_1 \subset \hat{C}_0$ which contains the square of each function in it, , let there be a sequence $\{f^{\varepsilon,N}(\cdot)\}$ in $\bar{\mathcal{F}}$ such that $f^{\varepsilon,N}(\cdot) \in \mathcal{D}(\hat{A}^{\varepsilon,N})$, and for each real $T > 0$ let there be a random variable $M_T^{\varepsilon,N}(f)$ such that

(3.9) $P\{\sup_{t \le T} |f^{\varepsilon,N}(t) - f(x^{\varepsilon,N}(t))| \ge \alpha\} \to 0$ as $\varepsilon \to 0$ for each $\alpha > 0$,

(3.10) $$\sup_{t \le T} |\hat{A}^{\varepsilon,N} f^{\varepsilon,N}(t)| \le M_T^{\varepsilon,N}$$

(3.11) $$\sup_{\varepsilon > 0} P\{M_T^{\varepsilon,N}(f) \ge K\} \to 0 \quad \text{as} \quad K \to \infty.$$

Then $\{f(x^{\varepsilon,N}(\cdot))\}$ is tight in $D^r[0,\infty)$ for each $f(\cdot) \in \mathcal{D}_1$ and $\{x^{\varepsilon,N}(\cdot)\}$ is also tight in $D^r[0,\infty)$.

Remark. Normally, we let $\mathcal{D}_1$ be the subset of functions in $\mathcal{D}$ which do not depend on t, and the $f^{\varepsilon,N}(\cdot)$ would be constructed as they would be for use in Theorem 1. This is illustrated in Section 5, and also in [5], [10], where the sequence $\{f^{\varepsilon}(\cdot)\}$ of perturbations must be constructed in order to use Kurtz's original method. Equations (3.9)-(3.11) require more of the $f^{\varepsilon,N}(\cdot)$ than the p-lim requirements of Theorem 1. But, as implied above, these additional properties often exist. In the example of Section 5 $\hat{A}^{\varepsilon,N} f^{\varepsilon,N}(t)$ is bounded uniformly in t and the sup term in (3.9) goes to zero uniformly in ω (w.p.1) as $\varepsilon \to 0$. Of course, tightness of $\{x^{\varepsilon,N}(\cdot)\}$ might be a-priori obvious, as in the case noted in the remark after Theorem 1.

Proof. In order to simplify the notation, drop the superscript N. By Lemma 1 and (3.9), for each $f(\cdot) \in \mathcal{D}_1$ and $T < \infty$, we need only find $\gamma_\varepsilon(\delta)$ satisfying (3.8) and such that for $u \leq \delta$, $t \leq T$,

$$E_t^\varepsilon \gamma_\varepsilon(\delta) \geq \min[1, E_t^\varepsilon |f(x^\varepsilon(t+u)) - f(x^\varepsilon(t))|^2].$$

Let $f(\cdot) \in \mathcal{D}_1$ and let $\{f^\varepsilon(\cdot)\}$ (resp., $\{f_2(\cdot)\}$) denote the sequences in $\mathcal{D}(\hat{A}^\varepsilon)$ associated with $f(\cdot)$ (resp., with $f^2(\cdot)$) by the hypotheses. Write $f(x^\varepsilon(t)) = f(t)$. Then

$$|f(t+u) - f(t)|^2 = (f_2^\varepsilon(t+u) - f_2^\varepsilon(t)) - 2f(t)(f^\varepsilon(t+u) - f^\varepsilon(t))$$

$$+ (f^2(t+u) - f_2^\varepsilon(t+u)) + (f_2^\varepsilon(t) - f^2(t))$$

$$+ 2f(t)[(f^\varepsilon(t+u) - f(t+u)) - (f^\varepsilon(t) - f(t))].$$

First suppose (Case 1): $\sup_\varepsilon EM_T^\varepsilon(f) + \sup_\varepsilon EM_T^\varepsilon(f^2) < \infty$ and $\sup_{t \leq T} (|f^\varepsilon(t) - f(t)| + |f_2^\varepsilon(t) - f^2(t)|) \equiv \alpha^\varepsilon(T) \to 0$ uniformly in ω as $\varepsilon \to 0$. Then we can get a $\gamma_\varepsilon(\delta)$ of the desired type if there is a function $\bar{\gamma}_\varepsilon(\delta)$ satisfying (3.8) and such that for $u \leq \delta$, $t \leq T$,

$$E_t^\varepsilon \bar{\gamma}_\varepsilon(\delta) \geq 2||f(\cdot)|| \; |E_t^\varepsilon f^\varepsilon(t+u) - f^\varepsilon(t)| + |E_t^\varepsilon f_2^\varepsilon(t+u) - f_2^\varepsilon(t)|.$$

By (3.3) and (3.10), we can use

$$\bar{\gamma}_\varepsilon(\delta) = \delta[2||f(\cdot)||M_T^\varepsilon(f) + M_T^\varepsilon(f^2)].$$

and the proof of Case 1 is completed.

Now, let the condition of Case 1 not hold and use a truncation argument. For each $\delta > 0$ define

$$T_\delta^\varepsilon = \min\{t: |\hat{A}^\varepsilon f^\varepsilon(s)| \geq 1/\delta \text{ or } |\hat{A}^\varepsilon f_2^\varepsilon(s)| \geq 1/\delta \text{ or } \alpha^\varepsilon(t) \geq \delta\}.$$

Let τ_ε denote a stopping time (relative to $\{\mathcal{F}_t^\varepsilon\}$) and note [3] that (3.3) holds if t and $t + s$, resp., are replaced by $t \cap \tau_\varepsilon$ and $(t+s) \cap \tau_\varepsilon$, resp. Now repeat the Case 1 argument to get that $\{f(\cdot \cap T_\delta^\varepsilon)\}$ is tight for each $\delta > 0$. Equations (3.9) and (3.11) (for both f and f^2) imply that $\lim_{\delta\to 0} \liminf_{\varepsilon\to 0} P\{T_\delta^\varepsilon > T\} = 1$. This and the tightness of $\{f(\cdot \cap T_\delta^\varepsilon)\}$ for each $\delta > 0$ completes the proof.

4. Discrete Parameter Processes

Only a sketch will be given. For each $\varepsilon > 0$, let $\{x_i^\varepsilon\}$ denote a discrete parameter R^r valued process and let $\{\tau_\varepsilon\}$ denote a sequence of positive numbers tending to zero. Define the interpolation $x^\varepsilon(\cdot)$ by $x^\varepsilon(t) = x_i^\varepsilon$ on $[i\tau_\varepsilon, i\tau_\varepsilon+\tau_\varepsilon)$. Let $x^{\varepsilon,N}(\cdot)$ be defined the same way it was defined in Theorem 1. For $f(\cdot) \in \mathcal{L}$ define $\hat{A}^{\varepsilon,N}f(\cdot)$ by

$$\hat{A}^{\varepsilon,N}f(t) = [E_t^{\varepsilon,N}f(t+\tau_\varepsilon) - f(t)]/\tau_\varepsilon. \tag{4.1}$$

Then

$$(4.2)\qquad E^{\varepsilon,N}_{i\tau_\varepsilon} f((i+n)\tau_\varepsilon) - f(i\tau_\varepsilon) = \sum_{j=i}^{i+n-1} E^{\varepsilon,N}_{i\tau_\varepsilon} \hat{A}^{\varepsilon,N} f(j\tau_\varepsilon)\tau_\varepsilon .$$

The suggestion for adapting the continuous parameter results to the discrete parameter case which was made in [3, p. 625] can be hard to use, because with that approach the construction of the $\{f^{\varepsilon}(\cdot)\}$ or $\{f^{\varepsilon,N}(\cdot)\}$ can be quite messy. It often seems best to deal with the discrete parameter case directly, using (4.2) and the martingale method of Theorem 1. In many cases, our method seems to simplify the problem of obtaining appropriate $\{f^{\varepsilon}(\cdot)\}$ or $\{f^{\varepsilon,N}(\cdot)\}$. For example, for the discrete parameter analog of the problems in [5], [12] or in Section 5 below, $f^{\varepsilon}(\cdot)$ (or $f^{\varepsilon,N}(\cdot)$) takes the form

$$f^{\varepsilon}(\cdot) = f(\cdot) + f_1^{\varepsilon}(\cdot) + f_2^{\varepsilon}(\cdot),$$

where all functions are constant on intervals $[i\tau_\varepsilon, i\tau_\varepsilon + \tau_\varepsilon)$ and where the $f_i(\cdot)$ are constructed analogously to the method used in [5], [12], or as in Section 5 below, but where the appropriate series are used in lieu of integrals. A simpler treatment of the discrete parameter case was one of the motivations for the approach of this paper. Examples will appear in a subsequent paper. Using the methods of [3], [5], reference [10] shows how to get the $f^{\varepsilon}(\cdot)$ from $f(\cdot)$ for one type of discrete parameter case.

Theorem 3. Assume the conditions of Theorem 1, except that the $f^{\varepsilon,N}(\cdot)$ are constant on the intervals $[i\tau_\varepsilon, i\tau_\varepsilon+\tau_\varepsilon)$, and the "discrete parameter" $\hat{A}^{\varepsilon,N}$ of (4.1) is used. Then the conclusions of Theorem 1 hold.

The proof is almost identical to that of Theorem 1 and is omitted. We need only work with (3.3) with t,s being integral multiples of τ_ε. Then (3.3) reduces to (4.2).

Theorem 2 can be used to prove tightness and the "sups" are over points $i\tau_\varepsilon \leq T$, $i = 0,1,\ldots,$ only.

5. An Application to Averaging

In order to illustrate the ease of use of Theorem 1, we redo the classical averaging problem of Khazminskii [11]. The conditions are not the best and can easily be weakened. Let $\xi(\cdot)$ be a R^q valued stationary bounded right continuous ϕ-mixing process with rate satisfying $\int_0^\infty \phi^{1/2}(s)ds < \infty$. Let $F(\cdot,\cdot,\cdot)$ be a continuous R^r valued function on R^{r+1+q}, whose first partial x-derivatives are continuous. Define $\bar{F}(x,t) = EF(x,t,\xi(v))$. Let $q_N(\cdot)$ be as below (3.6a). Define $\xi^\varepsilon(t) = \xi(t/\varepsilon)$, $F_N(x,t,\xi) = F(x,t,\xi)q_N(\xi)$, $\bar{F}_N(x,t) = \bar{F}(x,t)q_N(x)$ and define $x^\varepsilon(\cdot)$ and $x^{\varepsilon,N}(\cdot)$ by

$$\dot{x}^\varepsilon = F(x^\varepsilon,t,\xi^\varepsilon) \tag{5.1a}$$

(5.1b) $\dot{x}^{\varepsilon,N} = F(x^{\varepsilon,N},t,\xi^{\varepsilon})q_N(x^{\varepsilon,N}) = F_N(x^{\varepsilon,N},t,\xi^{\varepsilon})$.

In Theorems 4 and 5, $E_t^{\varepsilon,N}$ denotes expectation conditioned on $\mathcal{F}_t^{\varepsilon}$, the minimal σ-algebra measuring $\{\xi^{\varepsilon}(s),\ s \le t,\ x^{\varepsilon}(0)\}$.

Theorem 4. Assume the conditions of the previous paragraph, let $x^{\varepsilon}(0) \to x_0$ weakly and suppose that $\dot{x} = \bar{F}(x,t)$ has a unique (bounded on bounded intervals) solution $\bar{x}(\cdot)$ for each $x(0) \in R^r$. Then $x^{\varepsilon}(\cdot) \to \bar{x}(\cdot)$ weakly, where $\bar{x}(0) = x_0$.

Proof. The proof is a very straightforward application of Theorem 1. Fix N. The $\{x^{\varepsilon,N}(\cdot)\}$ is tight since $|\dot{x}^{\varepsilon,N}(t)|$ is bounded uniformly in ε,t. Fix $f(\cdot,\cdot) \in \hat{C}_0^{1,2}$. Then $\hat{A}^{\varepsilon,N}f(x^{\varepsilon,N}(t),t) = f_t(x^{\varepsilon,N}(t),t) + f_x'(x^{\varepsilon,N}(t),t)\ F_N(x^{\varepsilon,N}(t),t,\xi^{\varepsilon}(t))$. Define $f_1^{\varepsilon,N}(t) = f_1^{\varepsilon,N}(x^{\varepsilon,N}(t),t)$, where

$$f_1^{\varepsilon,N}(x,t) = \int_0^{\infty} E_t^{\varepsilon,N} f_x'(x,t+s)\,[F_N(x,t+s,\xi^{\varepsilon}(t+s)) - \bar{F}_N(x,t+s)]ds$$

$$= \varepsilon \int_0^{\infty} E_t^{\varepsilon,N} f_x'(x,t+\varepsilon s)\,[F_N(x,t+\varepsilon s,\xi(\tfrac{t}{\varepsilon}+s)) - \bar{F}_N(x,t+\varepsilon s)]ds.$$

By the compact support of $f(\cdot,\cdot)$, the centering of the integrand about its mean and the ϕ-mixing condition,

$$(5.2) \qquad f_1^{\varepsilon,N}(t) = O(\varepsilon) \quad \underline{\text{uniformly in}} \quad t,\omega.$$

The gradient $f_{1,x}^{\varepsilon,N}(x,t)$ also exists, and the differentiation can be done under the integral sign. Also, $f_{1,x}^{\varepsilon,N}(x^{\varepsilon,N}(t),t) = O(\varepsilon)$ uniformly in t,ω.

Define the operator A^N by $A^N f(x,t) = \bar{F}_N'(x,t)f_x(x,t)$. Define the function $f^{\varepsilon,N}(\cdot)$ by $f^{\varepsilon,N}(t) = f(x^{\varepsilon,N}(t),t) + f_1^{\varepsilon,N}(t)$. Then $f^{\varepsilon,N}(\cdot) \in \mathcal{D}(\hat{A}^{\varepsilon,N})$ and (abbreviate $x = x^{\varepsilon,N}(t)$)

$$(5.3) \quad \hat{A}^{\varepsilon,N} f^{\varepsilon,N}(t) = f_t(x,t) + f_x'(x,t)F_N(x,t,\xi^\varepsilon(t))$$

$$- f_x'(x,t)[F_N(x,t,\xi^\varepsilon(t)) - \bar{F}_N(x,t)] + (f_{1,x}^{\varepsilon,N}(x,t))'F_N(x,t,\xi^\varepsilon(t))$$

$$= (\frac{\partial}{\partial t} + A^N)f(x,t) + O(\varepsilon).$$

Equations (5.2) and (5.3) and the uniqueness of $\bar{x}(\cdot)$ imply the conditions of Theorem 1. Q.E.D.

<u>Linearization and first order perturbations</u>. We now examine the asymptotic properties of the centered solution $x^\varepsilon(t) - \bar{x}(\cdot) = \delta x^\varepsilon(t)$, as $\varepsilon \to 0$.

<u>Theorem 5</u>. <u>Assume that</u> $F(\cdot,t,\xi)$ <u>has two continuous in</u> (x,t,ξ) <u>partial x-derivatives, and that</u> $\dfrac{x^\varepsilon(0) - x_0}{\sqrt{\varepsilon}} \to 0$ <u>weakly</u>. <u>Then under the conditions of Theorem 4</u>,

$u^\varepsilon(\cdot) = (x^\varepsilon(\cdot) - \bar{x}(\cdot))/\sqrt{\varepsilon}$ converges weakly to the solution of the stochastic differential equation (5.4)

$$(5.4a) \qquad du = F_x(\bar{x}(t),t)u\,dt + dB, \quad u(0) = 0,$$

where $B(\cdot)$ is a Brownian notion with covariance

$$\frac{d\cdot E\, B(t)B'(t)}{dt} =$$

$$(5.4b) \int_{-\infty}^{\infty} E(F(\bar{x}(t),t,\xi(0)) - \bar{F}(\bar{x}(t),t))(F(\bar{x}(t),t,\xi(s)) - \bar{F}(\bar{x}(t),t))'ds$$

Proof. We have

$$(5.5) \qquad \dot{u}^\varepsilon(t) = \frac{F(x^\varepsilon(t),t,\xi^\varepsilon(t)) - F(\bar{x}(t),t,\xi^\varepsilon(t))}{\sqrt{\varepsilon}} + \frac{\delta F^\varepsilon(t)}{\sqrt{\varepsilon}}$$

where $\delta F^\varepsilon(t) = F(x(t),t,\xi^\varepsilon(t) - \bar{F}(\bar{x}(t),t)$. Although $x(0) = \bar{x}_0$ might be random it plays no important role in the proof and we suppose for convenience that it is constant and N will be large enough such that $|x_0| \le N$. We need only work on an arbitrary finite interval $[0,T]$ and hence can suppose that $|\bar{x}(t)|$ is bounded in t. Also, by a suitable choice of probability space, we can suppose that $\bar{x}^\varepsilon(0) \to x_0$ w.p.1. Thus, since $\bar{x}(\cdot)$ is determinisitic and $x^\varepsilon(\cdot) \to \bar{x}(\cdot)$ weakly, for each $\delta > 0$ there is an $\varepsilon_\delta > 0$ such that $P\{\sup_{t\le T}|x^\varepsilon(t) - \bar{x}(t)| \ge \delta\} \le \delta$ for $\varepsilon \le \varepsilon_\delta$. In order to prove (5.4) we can w.l.o.g., suppose that for an arbitrarily small but fixed $\delta > 0$, $\sup_{t \le T} |x^\varepsilon(t) - \bar{x}(t)| \le \delta$.

Using these simplications and expanding (5.5) yields

$$\dot{u}^\varepsilon = F_x(\bar{x},t,\xi^\varepsilon)u^\varepsilon + \sqrt{\varepsilon}\,(u^\varepsilon)' F^\varepsilon_{xx}(t)u^\varepsilon + \delta F^\varepsilon(t)/\sqrt{\varepsilon},$$

where $F^\varepsilon_{xx}(t)$ is a matrix which is bounded on $[0,T]$. Define $u^{\varepsilon,N}(\cdot)$ by $u^{\varepsilon,N}(0) = u^\varepsilon(0)$ when $|u^\varepsilon(0)| \le N$ and zero otherwise, and for $t > 0$ by

$$(5.6) \qquad \dot{u}^{\varepsilon,N} = [F_x(\bar{x},t,\xi^\varepsilon)u^{\varepsilon,N} + \sqrt{\varepsilon}(u^{\varepsilon,N})' F^\varepsilon_{xx}(t)u^{\varepsilon,N}]q_N(u^{\varepsilon,N})$$

$$+ \delta F^\varepsilon(t)/\sqrt{\varepsilon}.$$

If $B^\varepsilon(t) = \int_0^t \delta F^\varepsilon(s)ds/\sqrt{\varepsilon}$ is tight on $D^r[0,\infty)$ with all weak limits being the same Wiener processes, then we note merely in passing, that (5.6) implies that $\{u^{\varepsilon,N}(\cdot)\}$ is tight for each N and that all weak limits have continuous paths.

Proceeding, let $f(\cdot,\cdot) \in \hat{\mathcal{C}}^{1,3}_0$. Then

$$\hat{A}^{\varepsilon,N} f(u^{\varepsilon,N}(t),t) = f_t(u^{\varepsilon,N}(t),t) + f'_u(u^\varepsilon$$

The component

$$\sqrt{\varepsilon}\, f'_u(u^{\varepsilon,N}(t),t)q_N(u^{\varepsilon,N}(t))[(u^{\varepsilon,N}(t))' F^\varepsilon_{xx}(t)u^{\varepsilon,N}(t)]$$

of the second term above always has $\operatorname{p\,lim}_{\varepsilon \to 0} = 0$. We must find a sequence $\{f^{\varepsilon,N}(\cdot)\}$ satisfying the conditions of Theorem 1. We look for $f^{\varepsilon,N}(\cdot)$ of the form

$$f^{\varepsilon,N}(t) = f(u^{\varepsilon,N}(t),t) + \sum_{i=1}^{2} f_i^{\varepsilon,N}(t)$$

where the $f_i^{\varepsilon,N}(t)$ have the form $f_i^{\varepsilon,N}(u^{\varepsilon,N}(t),t)$ and the $f_i^{\varepsilon,N}(u,t)$ are now to be defined.

$$\begin{aligned}(5.7)\qquad f_1^{\varepsilon,N}(u,t) &= \int_0^\infty E_t^{\varepsilon,N} f_u'(u,t+s)[(F_x(\bar{x}(t+s),t+s,\xi^\varepsilon(t+s)) \\ &\quad - \bar{F}_x(\bar{x}(t+s),t+s))q_N(u) + \frac{\delta F^\varepsilon(t+s)}{\sqrt{\varepsilon}}]ds \\ &= \varepsilon \int_0^\infty E_t^{\varepsilon,N} f_u'(u,t+\varepsilon s)[F_x(\bar{x}(t+\varepsilon s),t+\varepsilon s,\xi(\tfrac{t}{\varepsilon}+s)) \\ &\quad - \bar{F}_x(\bar{x}(t+\varepsilon s),t+\varepsilon s))q_N(u) + \delta F^\varepsilon(t+\varepsilon s)/\sqrt{\varepsilon}]ds.\end{aligned}$$

By the compact support of $f(\cdot,\cdot)$, the centering of the integrand about its expectation and the ϕ-mixing, $\text{p-}\lim_{\varepsilon\to 0} f_1^{\varepsilon,N}(t) = 0$. Also, $f_1^{\varepsilon,N}(\cdot,t)$ is differentiable, and we can take the derivative under the integral sign. Using this, we readily show that $f_1^{\varepsilon,N}(\cdot) \in \mathcal{D}(\hat{A}^{\varepsilon,N})$ and (abbreviate $u = u^{\varepsilon,N}(t)$)

$$\begin{aligned}\hat{A}^{\varepsilon,N} f_1^{\varepsilon,N}(t) &= -f_u'(u,t)[(F_x(\bar{x}(t),t,\xi^\varepsilon(t)) - \bar{F}_x(\bar{x}(t),t))q_N(u) \\ (5.8)\qquad &\quad + \delta F^\varepsilon(t)/\sqrt{\varepsilon}] + f_{1,u}^{\varepsilon,N}(u,t)\cdot\dot{u}^{\varepsilon,N}(t).\end{aligned}$$

The only component of the last term of (5.8) which does not have $\underset{\varepsilon\to 0}{\text{p-lim}} = 0$ is that arising from the $\delta F^{\varepsilon}/\sqrt{\varepsilon}$ part of $\dot{u}^{\varepsilon,N}$ and is ($u^{\varepsilon,N}$ still written as u)

$$\frac{1}{\varepsilon}\int_0^{\infty} E_t^{\varepsilon,N}\delta F^{\varepsilon\prime}(t+s) f_{uu}(u,t+s)\delta F^{\varepsilon}(t)ds$$

$$= \int_0^{\infty} E_t^{\varepsilon,N}\delta F^{\varepsilon\prime}(t+\varepsilon s) f_{uu}(u,t+\varepsilon s)\delta F^{\varepsilon}(t)ds \equiv Q^{\varepsilon,N}f(u,t)$$

which is bounded uniformly in ω,t by the ϕ-mixing. Define the operator A_0^{ε} by $EQ^{\varepsilon,N}f(u,t) = A_0^{\varepsilon}f(u,t)$. By the stationarity of $\xi(\cdot)$ and the ϕ-mixing we have the following convergence uniformly in u,t, where $\delta F(t,s) = F(\bar{x}(t),t,\xi(t+s)) - \bar{F}(\bar{x}(t),t)$ and the integral defines the operator A_0:

$$A_0^{\varepsilon}f(u,t) \to \int_0^{\infty} E\ \delta F'(t,s))\ f_{uu}(u,t)\delta F\ (t,0)ds = A_0 f(u,t)$$

Next, in order to "average out" the $Q^{\varepsilon,N}$ term, we introduce $f_2^{\varepsilon,N}(u,t)$ in the form

$$f_2^{\varepsilon,N}(u,t) = \int_0^{\infty} ds E_t^{\varepsilon,N}\{\frac{1}{\varepsilon}\int_0^{\infty} dv\ E_{t+s}^{\varepsilon,N}\delta F^{\varepsilon\prime}(t+s+v) f_{uu}(u,t+s+v)\delta F^{\varepsilon}(t+s)$$

$$- A_0^{\varepsilon}f(u,t+s)\}.$$

By the change of variables $s/\varepsilon \to s$, $v/\varepsilon \to v$, and the ϕ-mixing condition, it is readily seen that $f_2^{\varepsilon,N}(u,t) = O(\varepsilon)$ uniformly in ω,t,u. Also, $f_2^{\varepsilon,N}(t) = f_2^{\varepsilon,N}(u^{\varepsilon,N}(t),t) \in \mathcal{D}(\hat{A}^{\varepsilon,N})$ and

$$\hat{A}^{\varepsilon,N} f_2^{\varepsilon,N}(t) = -Q^{\varepsilon,N} f(u^{\varepsilon,N}(t),t) + A_0^{\varepsilon} f(u^{\varepsilon,N}(t),t) + O(\sqrt{\varepsilon}),$$

where $O(\sqrt{\varepsilon})$ is uniform in ω,t.

Adding all the above yields

$$\text{(5.9)} \qquad \text{p-}\lim_{\varepsilon\to 0} [f^{\varepsilon,N}(\cdot) - f(u^{\varepsilon,N}(\cdot),\cdot)] = 0$$

$$\text{p-}\lim_{\varepsilon\to 0} [\hat{A}^{\varepsilon,N} f^{\varepsilon,N}(\cdot) - \{f_t(u^{\varepsilon,N}(\cdot),\cdot)$$

$$\text{(5.10)} \quad + f_u'(u^{\varepsilon,N}(\cdot),\cdot) q_N(u^{\varepsilon,N}(\cdot)) \bar{F}_x(\bar{x}(\cdot),\cdot) u^{\varepsilon,N}(\cdot)$$

$$+ A_0 f(u^{\varepsilon,N}(\cdot),\cdot)\}] = 0.$$

If $\{u^{\varepsilon,N}(\cdot)\}$ or $\{B^{\varepsilon}(\cdot)\}$ were tight, then via Theorem 1 any weak limit of $\{u^{\varepsilon,N}(\cdot)\}$ would be a diffusion with operator A^N where

$$\text{(5.11)} \quad A^N f(u,t) = f_u'(u,t) q_N(u) \bar{F}_x(\bar{x}(t),t) u + A_0 f(u,t)$$

and Theorem 1 would give us the limit (5.4). Then, to get the covariance form given in (5.4b) from (5.11), we just symmatrize the coefficients in A_0: Write $A_0 f(u,t) = \frac{1}{2} \sum_{i,j} a_{ij}(t) f_{u_i u_j}(u,t)$, where we take the symmetric form for $\{a_{ij}(\cdot)\} \equiv \Sigma(\cdot)$ (i.e., $a_{ij}(\cdot) = a_{ji}(\cdot)$). Then note that

$$(E\delta F(t,s)\,\delta F'(t,0))' = E\delta F(t,-s)\,\delta F'(t,0).$$

and use $A_0 f(u,t) = \frac{1}{2}\,\text{trace}\, \Sigma(t) f_{uu}(u,t)$ and $\frac{d}{dt} EB(t)B'(t) = \Sigma(t)$.

The tightness of $\{u^{\varepsilon,N}(\cdot)\}$ could be proved either by proving it directly for $\{u^{\varepsilon,N}(\cdot)\}$, or for the $\{B^{\varepsilon}(\cdot)\}$ and then using the remark below (5.6). To prove the weak convergence of $\{B^{\varepsilon}(\cdot)\}$ to the desired Wiener process, we could work with $B^{\varepsilon,N}(\cdot)$ defined by $\dot{B}^{\varepsilon,N} = \delta F^{\varepsilon}(t)q_N(B^{\varepsilon,N})/\sqrt{\varepsilon}$, $B^{\varepsilon,N}(0) = 0$, and proceed exactly as for the $\{u^{\varepsilon,N}(\cdot)\}$ above to show the desired convergence. An alternative method for both tightness and weak convergence follow the lines of the proceeding development but with the δF^{ε} in (5.6) multiplied by $q_N(u^{\varepsilon,N})$ also. Then $\{u^{\varepsilon,N}(\cdot)\}$ is bounded for each N and tightness can be proved by the method of Theorem 2. The details are the same as for the given development, except that a term proportional to $q_{N,u}(u)$ appears in the expression for the weak limit of $\{u^{\varepsilon,N}(\cdot), \varepsilon > 0\}$. But this term is zero for $|u| < N$ and disappears when $N \to \infty$. For this method the f_i^{ε} satisfy the same bounds $(O(\sqrt{\varepsilon}))$ as they do in the given development and the $M_T^{\varepsilon,N}(f)$ are bounded. Thus, Theorem 2 implies tightness. We omit the details. Q.E.D.

References

[1] T.G. Kurtz, "Extensions of Trotter's operator semigroup approximation theorems", J. Fnl. Analysis, 3, 1969, pp. 111-132.

[2] T.G. Kurtz, "A general theorem on the convergence of operator semigroups", Trans. Amer. Math. Soc., 148, 1970, pp. 23-32.

[3] T.G. Kurtz, "Semigroups of conditioned shifts and approximation of Markov processes", Ann. Prob., 3, 1975, pp. 618-642.

[4] T.G. Kurtz, "Diffusion approximations for branching processes" to appear Adv. in Probability.

[5] H.J. Kushner, "Jump-diffusion approximations for ordinary differential equations with wide-band random right hand sides", to appear SIAM J. on Control and Optimiz., 1979.

[6] G.C. Papanicolaou, and W. Kohler, "Asymptotic theory of mixing stochastic ordinary differential equations", Comm. Pure and Appl. Math., 27, 1974, pp. 641-668.

[7] R.Z. Khazminskii, "A limit theorem for solutions of differential equations with random right hand sides", Theory of Prob. and Applic., 11(1966), pp. 390-406.

[8] D.W. Strook, S.R.S. Varadhan, "Diffusion processes with continuous coefficients, I, II, Comm. Pure and Appl. Math., 22, 1969.

[9] D.W. Strook, S.R.S. Varadhan, "Diffusion processes with boundary conditions", Comm. Pure and Appl. Math., 24, 1971.

[10] H.J. Kushner, "Approximation of solutions to differential equations with random inputs by diffusion processes", Bonn Conference on Stochastic Control, Bonn, January 1979. To be published in Lecture Notes in Mathematics, Springer-Verlag, Berlin.

[11] R.Z. Khazminskii, "On stochastic processes with a small parameter", Theory of Prob. and its Applic., 11(1966), pp. 211-228.

[12] G. Blankenship, and G.C. Papanicolaou, "Stability and control of stochastic systems with wide band noise disturbances, SIAM J. on Appl. and Math., 34(1978), pp. 437-476.

[13] G.C. Papanicolaou, D. Strook, S.R.S. Varadhan, Martingale approach to some limit theorems", Proceedings, Duke Univ. Conference on Turbulence, 1976.

DIFFUSION APPROXIMATIONS TO OUTPUT PROCESSES OF NON-LINEAR SYSTEMS WITH WIDE BAND INPUTS, AND APPLICATIONS

Harold J. Kushner*,**

Abstract

Many problems in communication theory involve approximations of a Markov type to outputs of non-linear (feedback or not) systems, often so that Fokker-Planck techniques can be used. A general and powerful method is presented for getting diffusion approximations to outputs of systems with wide band inputs. The input is parameterized by ε and as $\varepsilon \to 0$ the band width goes to ∞ (loosely speaking). It is proved, under reasonable conditions on the systems and noise, that the sequence of system output processes converges weakly to a Markov diffusion process, which is characterized completely. Many communication systems fit the model of the paper and, in order to make mathematical sense out of many common developments of system properties, assumptions such as those of this paper are often required. The usefulness and relative ease of use of the method is illustrated by application to three examples: (a) phase locked loop, where a Markov diffu-

---

*Brown University, Divisions of Applied Mathematics and Engineering and Lefschetz Center for Dynamical Systems. Research supported in part by the Air Force Office of Scientific Research under AFOSR AF-76-3063, by the National Science Foundation under NSF Eng. 73-03846A03 and in part by the Office of Naval Research under ONR N00014-76-C-0279 P003.

**The author's understanding of the examples was greatly increased thanks to discussions with Prof. Y. Bar-Ness of Tel Aviv University, whose assistance the author gratefully acknowledges.

sion approximation of the error process is developed, (b) adaptive antenna system, where an asymptotic analysis of the equations for the system is given, (c) diffusion approximation to the output of a hard limiter followed by a band pass filter; input-output S/N ratios are developed (a version of a classical problem of Davenport). Difficulties with the usual heuristic approaches to (a), (b) are discussed. The method is versatile and the models quite general. Since weak convergence methods are used, the approximate "limits" yield approximations to many types of functionals of the actual systems.

1. Introduction

Many problems in communication theory involve representations of (or approximations to) outputs of devices (linear, nonlinear, feedback) whose inputs are signals added to a relatively wide band noise; e.g., phase locked loops (PLL), adaptive antenna arrays or automatic gain controls. Normally, the development of the output representation (or approximation) requires special assumptions (e.g., sinusoidal inputs, Gaussian noise), and various heuristic arguments are usually needed to approximate the output by Markov diffusion processes whose Fokker-Planck equation is to be analyzed in order to get some sort of approximation to the statistics of the true output process.

In this paper, a rather powerful method is presented for getting either the usual or related approximations, under assumptions which are reasonable, explicit, and often weaker than the usual ones. The relative ease of use of the method is illustrated here by applications to three rather different problems: (a) the PLL, (b) an adaptive antenna array, (c) a version of Davenport's [1] result on the output of a band limiter followed by a zonal filter.

In particular, denote the input noise by $n^{\varepsilon}(\cdot)$ where as $\varepsilon \to 0$ the bandwidth (BW) $\to \infty$. Under conditions to be imposed, the sequence of outputs (with input signal $s(\cdot)$ plus noise $n^{\varepsilon}(\cdot)$) will converge to a process whose state variable representation is a Markov diffusion process, and we will readily be able to find that process. Implicitly or explicitly (as we indicate in the examples below) many of the current heuristic arguments use a similar assumption on the noise - at least, the "output approximation" may not make sense unless it is viewed as the limit of a sequence of outputs in our sense.

Our method has the advantage that the assumptions are clearly seen, it is applicable to a great variety of situations, and the terms which a more heuristic analysis would drop can be clearly seen. The limit is in the sense of weak convergence of probability measures [2]. Thus the distributions of a great variety of functionals of the sequence of outputs (with parameter ε) converge to that of the limit. Furthermore, under certain circumstances additional information on approximations to stationary measures can be obtained. Also, nonstationary inputs can be accommodated. The fact that $y(\cdot)$ can occur nonlinearly in (2.1) - (2.3) below is important in the applications which involve some nonlinear processing.

In the communications literature, the problem of obtaining the (Markov-diffusion) limit of the sequence of outputs of a system as the input noise BW tends to ∞ was perhaps initiated by Wong and Zakai [3], [4] in a very special case. Later cases were treated by Khazminskii [5], Papanicolaou and Kohler [6], Papanicolaou and Blankenship [7] and Kushner [8], [9]. The treatment here, based on a semigroup approximation of Kurtz [10], was developed in [8], [9] to get limit theorems of the desired type.

In Section 2, the basic model is discussed, together with the general scheme of Kurtz [10], and the main approximation theorems appear in Section 3. Sections 4, 5 and 6 deal with the three problem classes (a), (b) and (c) mentioned above. The theory is developed first for the canonical models (2.1)-(2.3). Often in applications, such as those in Sections 4-6, the models are a little different. But, as we will see, the development give for the canonical model tells us exactly how to proceed in the other cases. In a project currently under way, the method is used to study a class of PLL's with non-linear filters (which seems to have certain advantages), a problem which has not been treated and for which there seems to be no other "natural" method at present.

2. The Basic Model

Noise model. First, we derive the noise model. In order to conveniently get a class of processes $n^{\varepsilon}(\cdot)$ whose BW goes to ∞ and energy/unit BW converges to a constant $\neq 0$ as $\varepsilon \to 0$, we work with $n^{\varepsilon}(\cdot)$ of the form $n^{\varepsilon}(t) = y^{\varepsilon}(t)/\varepsilon$, where $y^{\varepsilon}(t) = y(t/\varepsilon^2)$, and y is a stationary process. Other forms are possible. In particular, see [9], where $n^{\varepsilon}(\cdot)$ is built up from a sequence of small correlated effects, each of whose "size" $\to 0$ and the number of which (in any finite interval) goes to ∞ as $\varepsilon \to 0$. Other forms are possible - and yield rather similar results. One way or another, an explicit model for $n^{\varepsilon}(\cdot)$ must be given which allows BW $\to \infty$ as $\varepsilon \to 0$. The selected model is one useful choice. As seen below, it has the desired properties. The method to be developed can handle many other useful noise models as well. Basically, the noise must be parametrized in such a way that Theorem 1 can be adapted to the problem.

Suppose that $y(\cdot)$ has a spectral distribution $S(w)$. Then that of $n^{\varepsilon}(\cdot)$ is $S(\varepsilon^2 w) \equiv S_{\varepsilon}(w)$. See Fig. 1. Note that the t/ε^2 scaling spreads the BW (and the "center" frequency in the band if there is one (as in Fig. 1b)) and the $1/\varepsilon$ scale keeps the energy per unit BW from degenerating. Without the $1/\varepsilon$ factor the energy per unit BW goes to zero as $\varepsilon \to 0$, and all the limits are "noiseless". We do not require that $y(\cdot)$ has a spectral distribution. The above remark is for motivation only.

<u>System models</u>. There are several canonical forms with which we can work. The system outputs can be representable (state variable form) by one of the related ODEs (ordinary differential equations)

(2.1) $\dot{x}^{\varepsilon} = G(x^{\varepsilon}, y^{\varepsilon}, t) + F(x^{\varepsilon}, y^{\varepsilon}, t)/\varepsilon,$

(2.2) $\dot{x}^{\varepsilon} = G_{\varepsilon}(x^{\varepsilon}, y^{\varepsilon}, t) + F_{\varepsilon}(x^{\varepsilon}, y^{\varepsilon}, t)/\varepsilon,$

(2.3) $\dot{x}^{\varepsilon} = G_{\varepsilon}(x^{\varepsilon}, t) + F_{\varepsilon}(x^{\varepsilon}, y^{\varepsilon}, t)/\varepsilon,$ $x^{\varepsilon}(0) \in R^r$, Euclidean r-space.

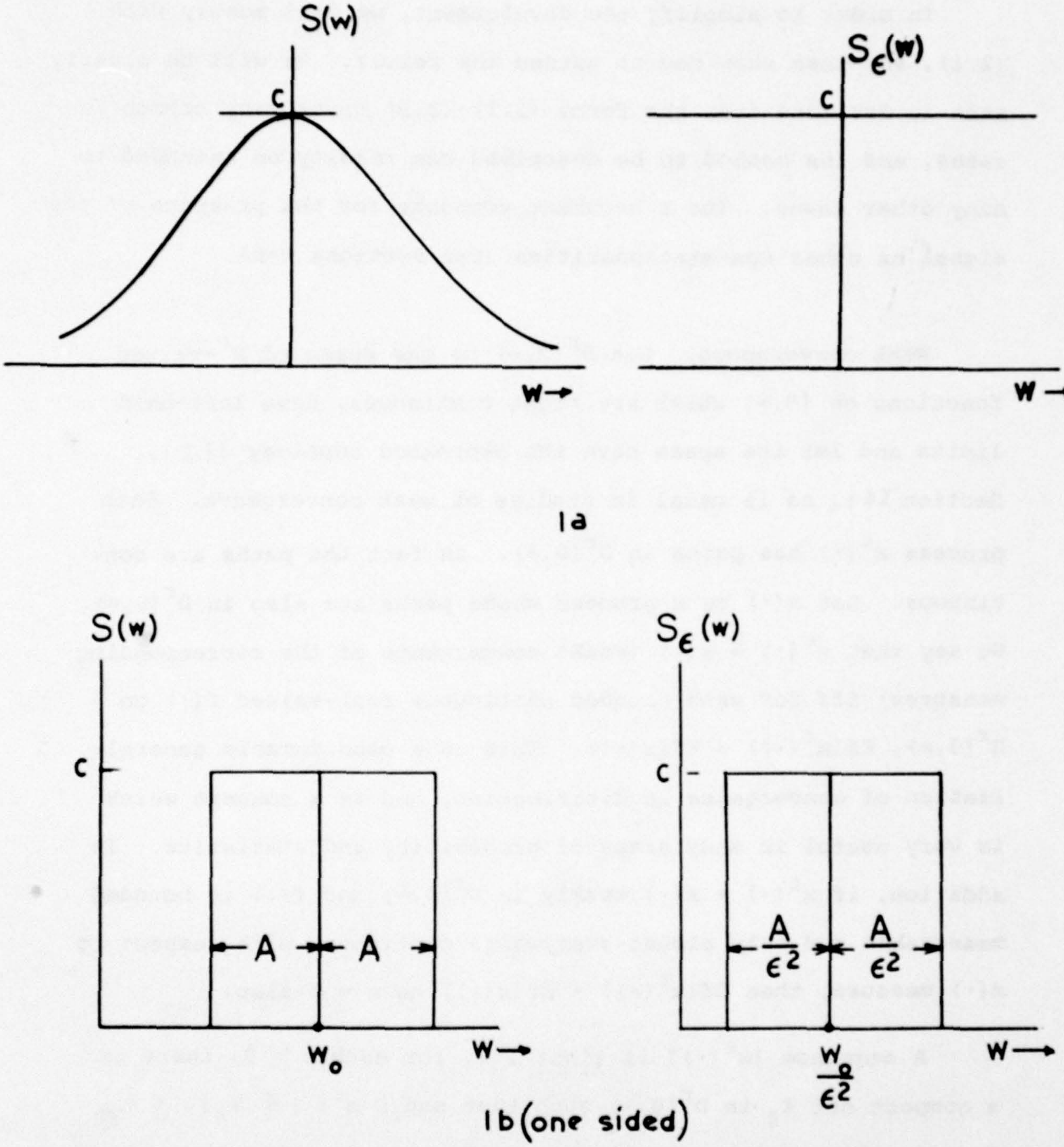


FIG. 1. ONE SIDED SPECTRAL DENSITIES

In order to simplify the development, we deal mostly with (2.1), and then show how to extend the result. As will be clearly seen in Sections 4-6, the forms (2.1)-(2.3) cover many common cases, and the method to be described can readily be extended to many other cases. The t argument accounts for the presence of the signal or other non-stationarities (see Sections 4-6).

Weak convergence. Let $D^r[0,\infty)$ be the space of R^r-valued functions on $[0,\infty)$ which are right continuous, have left-hand limits and let the space have the Skorokhod topology ([2], Section 14), as is usual in studies of weak convergence. Each process $x^\varepsilon(\cdot)$ has paths in $D^r[0,\infty)$. In fact the paths are continuous. Let $x(\cdot)$ be a process whose paths are also in $D^r[0,\infty)$. We say that $x^\varepsilon(\cdot) \overset{D}{\to} x(\cdot)$ (weak* convergence of the corresponding measures) iff for each bounded continuous real-valued $f(\cdot)$ on $D^r[0,\infty)$, $Ef(x^\varepsilon(\cdot)) \to Ef(x(\cdot))$. This is a considerable generalization of convergence in distribution, and is a concept which is very useful in many areas of probability and statistics. In addition, if $x^\varepsilon(\cdot) \to x(\cdot)$ weakly in $D^r[0,\infty)$ and $f(.)$ is bounded measurable and only almost everywhere continuous with respect to $x(\cdot)$ measure, then $Ef(x^\varepsilon(\cdot)) \to Ef(x(\cdot))$ as $\varepsilon \to 0$ also.

A sequence $\{x^\varepsilon(\cdot)\}$ is tight iff, for each $\delta > 0$, there is a compact set K_δ in $D^r[0,\infty)$ such that $\sup_\varepsilon P\{x^\varepsilon(\cdot) \notin K_\delta\} \leq \delta$. Suppose that $x(\cdot)$ has continuous paths w.p. 1. Then the two usual steps in proving weak convergence are: (i) showing convergence of finite-dimensional distribution of $\{x^\varepsilon(\cdot)\}$ to those of $x(\cdot)$; (ii) showing tightness of $\{x^\varepsilon(\cdot)\}$. These imply the weak convergence. The theorems given below do all this efficiently. Our limit process $x(\cdot)$ will be a Markov diffusion, and the theorems below allow us to calculate its infinitesimal operator. For more detail on weak convergence see [2].

<u>An example of the limit operator</u>. Let $\hat{\mathcal{C}}_0$ denote the set of bounded continuous functions on $[0,\infty) \times R^r$ with compact support, $\hat{\mathcal{C}}_0^{\alpha,\beta}$ the subset with continuous α-partial t-derivatives and β-partial x-derivatives, and let $\hat{\mathcal{C}}$ denote the closure of $\hat{\mathcal{C}}_0$ under uniform convergence. <u>Under conditions to be imposed</u> (including $EF(x,y(s),t) = 0$ for all x, t), and with the model (2.1), the infinitesimal operator $(\partial/\partial t + A)$ of the limit process $x(\cdot)$ is (acting on $\hat{\mathcal{C}}_0^{1,3}$):

$$(2.4)\quad (\partial/\partial t + A)f(x,t) = f_t(x,t) + Ef_x'(x,t)G(x,y(0),t) + \int_0^\infty EF'(x,y(0),t)(F'(x,y(s),t)f_x(x,t))_x ds$$

$$\equiv \sum_i b_i(x,t)f_{x_i}(x,t) + \frac{1}{2}\sum_{i,j} a_{ij}(x,t)f_{x_ix_j}(x,t),$$

where $b(\cdot,\cdot)$ and $a(\cdot,\cdot) = \{a_{ij}(\cdot,\cdot)\}$ are defined in the obvious manner, and we assume that $a(\cdot,\cdot)$ is symmetrized to conform with the usual form of the operator A. <u>Define</u> $\bar{G}(x,t) = EG(x,y(0),t)$.

If there is a matrix $\sigma(\cdot,\cdot)$ such that $a(\cdot,\cdot) = \sigma(\cdot,\cdot)\sigma(\cdot,\cdot)'/2$, then there is a standard vector-valued Wiener process $B(\cdot)$ such that $x(\cdot)$ has the Itô equation representation

$$dx = b(x,t)dt + \sigma(x,t)dB. \tag{2.5}$$

This is the case in the examples. Note that $b(x,t)$ contains two components - the first is $\bar{G}(x,t)$ and the second is $\int_0^\infty EF_x'(x,y(s),t)F(x,y(0),t)ds$, where we define

$$F_x' = \begin{Bmatrix} F_{1x_1} & \cdots & F_{1x_r} \\ F_{rx_1} & \cdots & F_{rx_r} \end{Bmatrix}.$$

The last term arises for the same reasons that cause the Wong-Zakai [3], [4] correction term; i.e. the interaction between $x^\varepsilon(t)$ and $n^\varepsilon(t)$. As it turns out, the typical heuristic arguments used to deal with problems (a), (b) obtain "limits" without the "correction term".

Some definitions. Let E_t^ε denote expectation conditioned on $n^\varepsilon(s)$, $s \le t$. If $k^\varepsilon(\cdot)$ is an ω,t function such that for each $T < \infty$ and for $(\omega,t) \in \Omega \times [0,T]$, it is measurable on the product σ-algebra* $\mathcal{B}[0,T] \times \mathcal{B}(n^\varepsilon(s), s \le T)$, we say that $k^\varepsilon \in \mathcal{L}^\varepsilon$, the class of progressively measurable functions.

---

* $\mathcal{B}[0,T]$ is the Borel algebra over $[0,T]$.

Let k_n^ε and k^ε be in $\mathscr{L}^\varepsilon$. We say that $\text{p-lim}_n k_n^\varepsilon = k^\varepsilon$ iff $\sup_n \sup_t E|k_n^\varepsilon(t)| < \infty$ and $E|k_n^\varepsilon(t)-k^\varepsilon(t)| \to 0$ as $n \to \infty$ for each t. Let $\hat{\mathscr{L}}^\varepsilon \subset \mathscr{L}^\varepsilon$ be the subclass of functions k such that $\sup_t E|k(t)| < \infty$. Let $\hat{\mathscr{L}}_0^\varepsilon$ denote the subset of $\hat{\mathscr{L}}^\varepsilon$ of p-right continuous functions; k being p-right continuous means that $k \in \hat{\mathscr{L}}$ and for each $t, E|k(t+s)-k(t)| \to 0$ as $s \downarrow 0$. If (some version of) $\text{p-lim}_{s\to 0}[E_t^\varepsilon k^\varepsilon(t+s)-k^\varepsilon(t)]/s$ exists in $\hat{\mathscr{L}}_0^\varepsilon$, it is called $\hat{A}^\varepsilon k^\varepsilon$ and we say that $k^\varepsilon \in \mathscr{D}(\hat{A}^\varepsilon)$, the domain of the operator $\hat{A}^\varepsilon$, an operator which is analogous to the weak infinitesimal operator of a Markov semigroup. If $k^\varepsilon \in \mathscr{L}^\varepsilon$, we say that p-lim $k^\varepsilon = 0$ if $\sup_{\varepsilon,t} E|k^\varepsilon(t)| < \infty$ and $E|k^\varepsilon(t)| \to 0$ as $\varepsilon \to 0$ for each t. The functions introduced in Theorem 2 and its proof have progressively measurable versions.

Kurtz's semigroup approximation theorem [10], adapted to our purposes. We treat t as a component of the state vector, in order to allow us to work with nonstationary cases. The conditions will be commented on below. They are more readily verifiable than may be apparent. The following theorem [3] is the basis of our method.

Theorem 1. Let $z^\varepsilon(\cdot) = (x^\varepsilon(\cdot), y^\varepsilon(\cdot))$ be a sequence of $R^{r+r'}$-valued right continuous processes, $x(\cdot)$ a (R^r-valued) Markov process with semigroup $T(\cdot)$ mapping $\hat{\mathscr{C}}$ into $\hat{\mathscr{C}}$ and which is strongly continuous on $\hat{\mathscr{L}}$ (sup norm). For some $\lambda > 0$ and dense set $\mathscr{D} \subset \hat{\mathscr{L}}$ (which will be $\hat{\mathscr{L}}_0^{1,3}$), let Range$(\lambda - A - \partial/\partial t|_{\mathscr{D}})$ be dense in $\hat{\mathscr{L}}$, where $A+\partial/\partial t$ is the infinitesimal operator of the process $(x(t),t)$. Suppose that, for each $f \in \mathscr{D}$, there is a sequence $\{f^\varepsilon\}$ of progressively measurable functions such that $f^\varepsilon \in \mathscr{D}(\hat{A}^\varepsilon)$ and

$$\text{p-lim}[f^\varepsilon(\cdot) - f(x^\varepsilon(\cdot))] = 0$$

$$\text{p-lim}[\hat{A}^\varepsilon f^\varepsilon(\cdot) - (A+\partial/\partial t) f(x^\varepsilon(\cdot),\cdot)] = 0.$$

Then if $x^\varepsilon(0) \to x(0)$ in distribution, the finite-dimensional distributions of $x^\varepsilon(\cdot)$ converge to those of $x(\cdot)$ (with initial condition $x(0)$) as $\varepsilon \to 0$.

Remark. There is a similar theorem for tightness of $\{x^\varepsilon(\cdot)\}$ which is particularly useful for the types of problems encountered here. In fact, if the finite-dimensional distributions converge (Theorem 1), a proof of tightness under reasonable conditions is not hard. See [8, Theorem 2, Part 4] for a method.

In (2.4) the operator $(A+\partial/\partial t)$ was defined on a set $\hat{\mathscr{L}}_0^{1,3}$. In Theorem 1, $(A+\partial/\partial t)$ is considered on a dense subset $\mathscr{D} \in \hat{\mathscr{L}}$ (which $\hat{\mathscr{L}}_0^{1,3}$ is). The question of concern is: does this restriction of $(A+\partial/\partial t)$ define the infinitesimal operator of a Markov semigroup uniquely? If the closure in $\hat{\mathscr{L}}$ of this restricted operator is the infinitesimal operator of $x(\cdot)$, then

$A+\partial/\partial t$ defines $T(\cdot)$ uniquely. Since we can only work (in the proofs) with nice classes such as $\mathscr{D} = \hat{\mathscr{C}}_0^{1,3}$ - and not with the domain of the infinitesimal operator of $x(\cdot)$, it is important to know if $\dot{\mathscr{D}}$ is big enough to yield the limit uniquely. In fact, the condition on the density of the range holds in all the cases of Sections 4-6, and is rather unrestrictive.

Due to lack of space, it is not possible to discuss the relative advantages of Theorem 1 (or Theorems 2 and 3) to the approximation problem over more classical semigroup approximation methods. It seems to be much easier to use in the usual problems encountered in control and communication theory and the relevant proofs (e.g. that of Theorem 2) are shorter and use better conditions.

The "density" condition together with the condition on strong continuity of $T(\cdot)$ can be eliminated by an alternative approach [11] which replaces them by the simple assumption that to the coefficients $a(\cdot,\cdot)$ and $b(\cdot)$ of A there corresponds a stochastic differential equation with a unique solution (in the sense of distributions). This condition also holds in our examples. The proof of the theorem corresponding to Theorem 2 in that case would be almost the same. We stick to an approach based on Theorem 1 because it is also applicable and the references are currently available.

We next give some specializations of the theorem suitable for our applications. Theorem 1 is given in the general form because it shows how to modify the following specializations when variants are required for particular cases. We omit explicit discussion of tightness, due to lack of space. Our conditions will guarantee the tightness, via the methods of proof of [8].

3. The Main Convergence and Approximation Theorems

We start with the form (2.1) and bounded $y(\cdot)$, because it is good enough for many applications and illustrates the technique with the least notational encumbrance. Then we discuss the case where $y(\cdot)$ is unbounded and $F(x,y,t)/\varepsilon = F(x,t)y/\varepsilon$. Finally, we remark on the cases (2.2)-(2.3), which actually occur in some of the examples.

Assumptions

(A1) $F(\cdot,\cdot,\cdot)$ and $G(\cdot,\cdot,\cdot)$ are continuous, the first (second, resp.) having two (one, resp.) continuous (in x, y, t) partial x-derivatives.

(A2) $|F(x,y,t)| + |G(x,y,t)| \le M(1+|x|)$, for some constant M.

(A3) $y(\cdot)$ is stationary, bounded, right-continuous, $EF(x,y(s),t) = 0$, each x, t, and $y(\cdot)$ is strong mixing in the sense that there is a function $\rho(\cdot)$ satisfying $\int_0^\infty \rho^{1/2}(s)ds < \infty$ and

$$\sup_{A,B,s} |P(B|A)-P(B)| \le \rho(t),$$

$$A \in \mathcal{B}(y(u),u\le s),\ B \in \mathcal{B}(y(u),u\ge s+t).$$

(Such a condition is quite common in the literature on applications of weak convergence theory. It is satisfied by truncated Gaussian processes with finite BW and continuous spectrum, by bounded ergodic Markov chains, etc.)

(A4) The operator $A+\partial/\partial t$ is the restriction to $\hat{\mathcal{C}}_0^{1,3}$ of the infinitesimal operator of a strong Markov process with semigroup $T(\cdot)$ mapping $\hat{\mathcal{C}}$ into $\hat{\mathcal{C}}$ and being strongly continuous on $\hat{\mathcal{C}}$.

(A5) $A+\partial/\partial t$ on its domain in $\hat{\mathcal{C}}$ is determined by its action on $\hat{\mathcal{C}}_0^{1,3}$.

Remark. (A4)-(A5) hold in our cases and in the usual situations which arise in communication theory. They pertain only to the limit $x(\cdot)$, and not to the $x^\varepsilon(\cdot)$. Further remarks appear in [8]. See also the comments at the end of the last section concerning simplifying the conditions.

Theorem 2. Let $x^\varepsilon(0) \to x(0)$ in distribution. Then, under (A1)-(A5), $\{x^\varepsilon(\cdot)\}$ converges weakly in $D^r[0,\infty)$ to $x(\cdot)$, a diffusion whose infinitesimal operator $(\partial/\partial t+A)$ is given by (2.4), and with initial condition $x(0)$.

Outline of proof. The proof for the non-time-dependent case appears in [8], and is a direct application of Theorem 1 (for convergence of finite-dimensional distributions), and another result of [10] for tightness. Given $f \in \hat{\mathcal{C}}_0^{1,3}$, the main object is to get the sequence $\{f^\varepsilon(\cdot)\}$ of Theorem 1, and to verify the p-lim requirement of that theorem. We outline this because various extensions of the method are needed for the examples, and it is useful to have an explicit outline for the time-dependent case, since the forms of some of the functions are a little different. Reference [8] dealt only with G, F not depending on time, but the time-dependent case is important for applications and requires only a few changes from the treatment in [8]. The method of getting the $f^\varepsilon(\cdot)$ is adapted from the averaging method in [7].

Let $f \in \hat{\mathcal{C}}_0^{1,3}$. We construct f^ε in the form $f^\varepsilon(t) = f(x^\varepsilon(t),t) + \sum_{i=0}^{2} f_i^\varepsilon(x^\varepsilon(t),t)$. Our f_1^ε is the $\varepsilon f_1^\varepsilon$ of [8], our f_0^ε is split off from the $\varepsilon^2 f_2^\varepsilon$ term in [8] and our f_2^ε is the remainder of the $\varepsilon^2 f_2^\varepsilon$ term there. The terms are modified to account for the explicit t-dependence of f, G, F. Note that (use $x = x^\varepsilon(t)$ for notational simplicity)

$$\hat{A}^\varepsilon f(x,t) = f_t(x,t) + f_x'(x,t)[G(x,y^\varepsilon(t),t) + F(x,y^\varepsilon(t),t)/\varepsilon]. \tag{3.1}$$

Define $f_0^\varepsilon(x,t)$ by

$$f_0^\varepsilon(x,t) = \int_0^\infty E_t^\varepsilon f_x'(x,t+s)[G(x,y^\varepsilon(t+s),t+s)-\overline{G}(x,t+s)]ds$$

$$= \varepsilon^2 \int_0^\infty E_t^\varepsilon f_x'(x,t+\varepsilon^2 s)[G(x,y(\frac{t}{\varepsilon^2}+s),t+\varepsilon^2 s)-\overline{G}(x,t+\varepsilon^2 s)]ds.$$

The integral exists for each $\varepsilon > 0$ by (A1), (A3) and the compact support of f. Also $f_0^\varepsilon(t) \equiv f_0^\varepsilon(x^\varepsilon(t),t) = O(\varepsilon^2)$ and $f_0^\varepsilon(\cdot) \in \mathcal{D}(\hat{A}^\varepsilon)$. Again use $x = x^\varepsilon(t)$. Then

(3.2) $\quad \hat{A}^\varepsilon f_0^\varepsilon(x,t) = -f_x'(x,t)G(x,y^\varepsilon(t),t) + f_x'(x,t)\overline{G}(x,t) + O(\varepsilon)$ terms

Note that $\hat{A}^\varepsilon[f(x,t)+f_0^\varepsilon(x,t)] = f_t(x,t) + f_x'(x,t)\overline{G}(x,t) +$ $O(\varepsilon)$ terms $+ f_x'(x,t)F(x,y^\varepsilon(t),t)/\varepsilon$. The term $G(x,y^\varepsilon(t),t)$ in (3.1) has thus been replaced by its average $\overline{G}(x,t)$ modulo an $O(\varepsilon)$ term. This was the reason for the addition of the f_0^ε term. A similar "averaging" scheme will be used to replace the $f_x'(x,t)F(x,y^\varepsilon(t),t)/\varepsilon$ term by the rest of Af modulo $O(\varepsilon)$. This will be done in two steps by using the f_1^ε and f_2^ε defined below.

Proceeding, define f_1^ε by

(3.3) $$f_1^\varepsilon(x,t) = \frac{1}{\varepsilon}\int_0^\infty E_t^\varepsilon f_x'(x,t+s)F(x,y^\varepsilon(t+s),t+s)ds$$

$$= \varepsilon \int_0^\infty E_t^\varepsilon f_x'(x,t+\varepsilon^2 s)F(x,y(\frac{t}{\varepsilon^2}+s),t+\varepsilon^2 s)ds = O(\varepsilon).$$

Furthermore, $f_1^\varepsilon(\cdot,t)$ is differentiable in x and $f_1^\varepsilon(\cdot) \equiv$ $f_1^\varepsilon(x^\varepsilon(\cdot),\cdot) \in \mathcal{D}(\hat{A}^\varepsilon)$ and (again, setting $x^\varepsilon(t) = x$)

$$(3.4)\quad \hat{A}^{\varepsilon} f_1^{\varepsilon}(x,t) = -f_x'(x,t)F(x,y^{\varepsilon}(t),t)/\varepsilon + O(\varepsilon)$$

$$+ (f_1^{\varepsilon}(x,t))_x'[G(x,y^{\varepsilon}(t),t) + F(x,y^{\varepsilon}(t),t)/\varepsilon].$$

It can be shown that the gradient $(f_1^{\varepsilon})_x$ can be obtained by differentiating with respect to x under the integral in (3.3). Also $(f_1^{\varepsilon}(x,t))_x'G(x,y^{\varepsilon}(t),t) = O(\varepsilon)$, and we ignore this component henceforth. The first term on the right-hand side of (3.4) cancels the last term of (3.1). To get the p-lim result required for Theorem 1, we need now only choose f_2^{ε} to "cancel the effect of"

$$(3.5)\quad (f_1^{\varepsilon}(x,t))_x'F(x,y^{\varepsilon}(t),t)/\varepsilon = \frac{1}{\varepsilon}\int_0^{\infty} E_t^{\varepsilon}(f_x'(x,t+s)F(x,y^{\varepsilon}(t+s),t+s))_x ds$$

$$\cdot\ [\frac{F(x,y^{\varepsilon}(t),t)}{\varepsilon}].$$

Define $A_0^{\varepsilon}f$ (the average value of (3.5) - change variables $s/\varepsilon^2 \to s$ and use the stationarity of $y(\cdot)$) by

$$(3.6)\quad A_0^{\varepsilon}f(x,t) = \int_0^{\infty} EF'(x,y(0),t)[f_x'(x,t+\varepsilon^2 s)F(x,y(s),t+\varepsilon^2 s)]_x ds.$$

(3.6) exists by the strong mixing (A3) and the fact that $EF(x,y(s),t) \equiv 0$. As $\varepsilon \to 0$, (3.6) converges uniformly in x, t to the integral in (2.4).

Now, define $f_2^{\varepsilon}(x,t)$ by

$$f_2^\varepsilon(x,t) = \int_0^\infty ds\{\int_0^\infty du\ E_t^\varepsilon F'(x,\frac{y^\varepsilon(t+s),t+s)}{\varepsilon}[f_x'(x,t+s+u)$$

$$\cdot F(x,\frac{y^\varepsilon(t+s+u),t+s+u)}{\varepsilon}]_x - A_0^\varepsilon f(x,t+s)\}$$

$$= \varepsilon^2\int_0^\infty ds\{\int_0^\infty du\ E_t^\varepsilon F'(x,y(\frac{t}{\varepsilon^2}+s),t+\varepsilon^2 s)[f_x'(x,t+\varepsilon^2 s+\varepsilon^2 u)$$

$$\cdot F(x,y(\frac{t}{\varepsilon^2}+s+u),t+\varepsilon^2 s+\varepsilon^2 u)]_x - A_0^\varepsilon f(x,t+\varepsilon^2 s)\}$$

$$= O(\varepsilon^2).$$

The integral exists and equals $O(\varepsilon^2)$ by the centering about the mean value $A_0^\varepsilon f$, the strong mixing (A3) and the compact support of f. Now,

$$\text{p-lim}[f_0^\varepsilon + f_1^\varepsilon + f_2^\varepsilon] = 0,$$

$$\text{p-lim}[\hat{A}^\varepsilon f^\varepsilon(\cdot) - (\partial/\partial t+A)f(x^\varepsilon(\cdot),\cdot)] = 0,$$

and Theorem 1 yields the convergence of finite-dimensional distributions. The tightness argument is the same as that in [8 , Theorem 2]. The proof concludes by noting that now all the conditions of Theorem 1 hold. Q.E.D.

<u>Unbounded</u> $y(\cdot)$ <u>and form</u> $F(x,y,t) = F(x,t)y$, $G(x,y,t) = \bar{G}(x,t) + G_0(x,t)y$. The treatment of the unbounded (e.g. Gaussian $y(\cdot)$) case is similar to that of the bounded $y(\cdot)$ case,

but somewhat more stringent conditions need to be imposed on the form of F.

Define $v(t) = \int_0^\infty E_t y(t+s)ds$, where E_t denotes conditioning on y(u), $u \leq t$. Let there be some $\rho > 0$ such that

(A1') $\sup_t E(\int_0^\infty |E_t y(t+s)|ds)^{2+\rho} < \infty$,

(A2') $E|y(t)|^{2+\rho} < \infty$, $Ey(t) \equiv 0$,

(A3') $\sup_t E(\int_0^\infty ds|E_t y(t+s)v'(t+s)-Ey(t+s)v'(t+s)|)^{2+\rho} < \infty$,

(A4') $y(\cdot)$ is stationary and right continuous,

(A5') F, $\overline{G}$, G_0 are continuous together with their second (first for $\overline{G}$, G_0) partial x-derivatives.

Conditions (A1')-(A4') are satisfied by any process which is a linear combination of the states of

(3.7) $du = Audt + Bdw$,

A asymptotically stable, $w(\cdot)$ = Wiener process. Since such processes constitute the class of Gaussian processes with rational spectral densities, (A1')-(A4') are certainly not restrictive.

In [8], the unbounded $y(\cdot)$ case was treated in Theorem 5, and $G_0 y$ was not explicitly included. The proof there goes through without any additional conditions or difficulty if $G_0 y$ is added - provided that G_0 has continuous x-first partial derivatives. In that proof, it was difficult to work with unbounded F, $\bar{G}$ when $y(\cdot)$ was unbounded, so the following assumption was added (adapted to our case here).

(A6') <u>For each</u> N, <u>there are functions</u> F^N, $\bar{G}^N$, G_0^N <u>equal to</u> F, $\bar{G}$ <u>and</u> G_0, <u>resp., in</u> $S_N = \{x: |x| \le N\}$, but <u>bounded and smooth (as smooth as</u> F, $\bar{G}$, G_0 <u>are) out of</u> S_N, <u>and such that</u> (A4), (A5) <u>hold</u>.

This condition is normally satisfied and holds in our examples. The reference to (A4), (A5) can be dropped if the approach in [11] is used (it is then replaced by uniqueness of the solution to the Itô representation of $x(\cdot)$).

<u>Theorem 3</u>. <u>Under</u> (A2), (A1')-(A6'), <u>and</u> $x^\varepsilon(0) \to x(0)$ <u>in distribution, the finite-dimensional distributions of</u> $\{x^\varepsilon(\cdot)\}$ <u>converge to those of</u> $x(\cdot)$. <u>If</u> $y(\cdot)$ <u>is given by a linear combination of the states of</u> (3.6), <u>then</u> $\{x^\varepsilon(\cdot)\}$ <u>is also tight and</u> $\{x^\varepsilon(\cdot)\} \to x(\cdot)$ <u>weakly in</u> $D^r[0,\infty)$.

The proof is similar to that of Theorem 2. Given $f \in \hat{\mathcal{L}}_0^{1,3}$, we construct f^ε as in Theorem 2 and prove the p-lim requirements of Theorem 1. See [8] for the details in the non-time-varying case.

Extensions to (2.2), (2.3). When F and G depend on ε, the procedure is exactly the same. Given $f \in \hat{\mathcal{D}}_0^{1,3}$, we construct f^ε as done in Theorem 2, making sure that the integrals are well defined and of the proper order in ε, and replacing G, F by G_ε, F_ε. We need $EG_\varepsilon(x,y(0),t) \to \bar{G}(x,t)$, a continuous function, uniformly on bounded (x,t) sets, and that for each $f \in \hat{\mathcal{D}}_0^{1,3}$,

$$
(3.8)\quad f_t(x,t) + G'(x,t)f_x(x,t) + \int_0^\infty E\, F_\varepsilon'(x,y(0),t) \cdot [f_x(x,t+\varepsilon^2 s)F_\varepsilon(x,y(s),t+\varepsilon^2 s)]_x ds
$$

converges uniformly on bounded (x,t) sets to $(\partial/\partial t+A)f(x,t)$.

Example (c) (Section 6) requires a slightly different extension, but the general idea is the same.

4. The Phase Locked Loop

The standard PLL is represented in Fig. 2 and is, perhaps, the simplest application of the foregoing ideas. Via suitable choices of D, E, C, Q all the usual filters can be constructed. We first do the case D = E = C = 0, Q = 1, which yields the standard form of the first-order loop [12]. The general case is treated in precisely the same way, and is given below. The

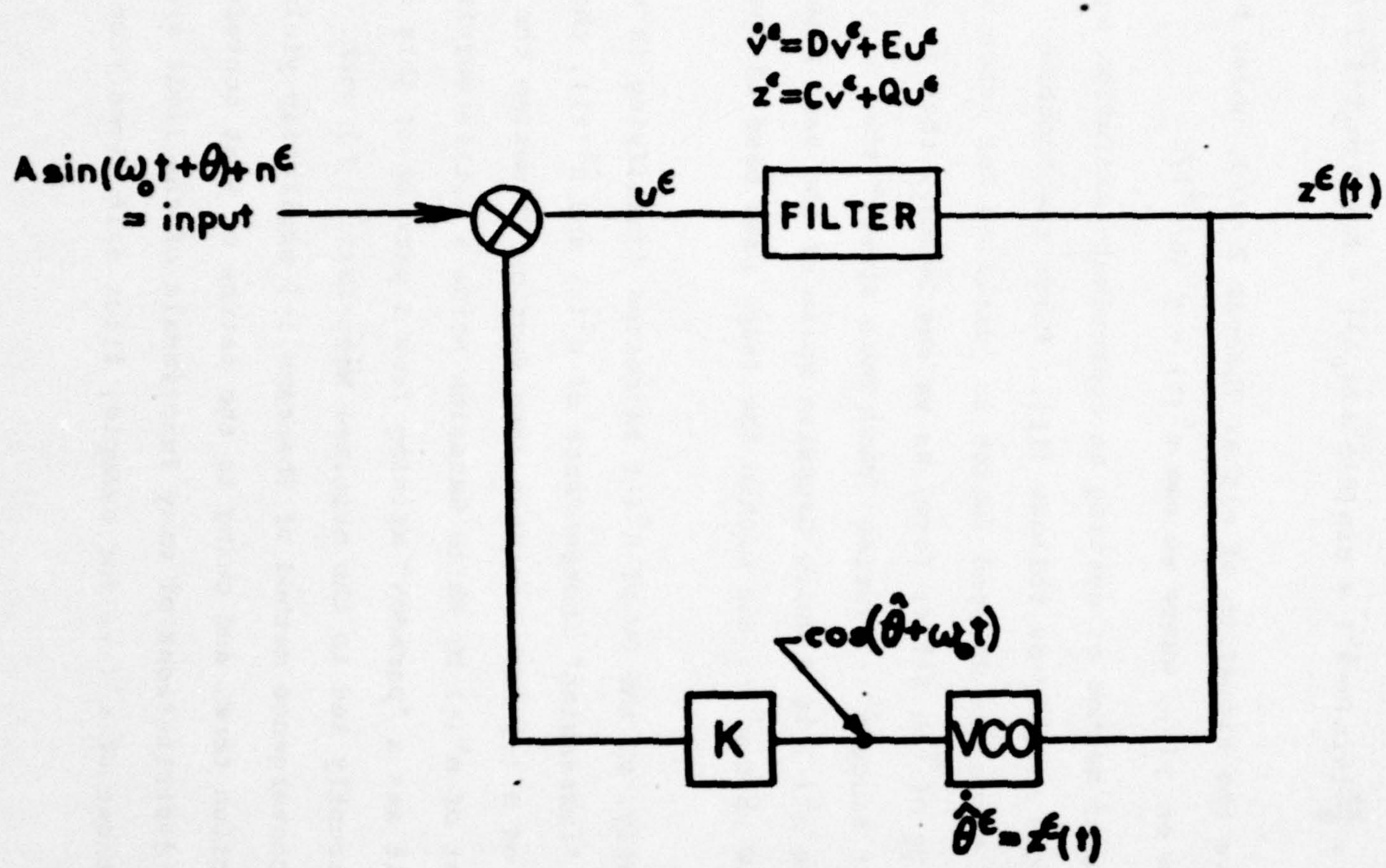


FIG. 2. PHASE LOCKED LOOP

function $\theta(\cdot)$ denotes the input phase process and $\hat{\theta}(\cdot)$ its estimate, as determined by the loop. Then

$$(4.1)\quad \dot{\hat{\theta}}^{\varepsilon} = \frac{AK}{2}[\sin(\theta-\hat{\theta}^{\varepsilon}) + \sin(\theta+\hat{\theta}^{\varepsilon}+2\omega_0 t)] + K\cos(\omega_0 t+\hat{\theta}^{\varepsilon})\cdot n^{\varepsilon}(t),$$

and we have the situation of either Theorem 2 or 3, under broad conditions on $y(\cdot)$, where we use $n^{\varepsilon}(t) = y^{\varepsilon}(t/\varepsilon^2)/\varepsilon$.

The usual method of getting an approximate diffusion equation proceeds roughly as follows [12]. First the "double frequency" terms are dropped (which is justified for large ω_0, irrespective of the filter form, as we see below), then it is argued that since $\hat{\theta}^{\varepsilon}(\cdot)$ varies "much more slowly" than $n^{\varepsilon}(\cdot)$, one can replace $n^{\varepsilon}(\cdot)$ by a white Gaussian noise of the same power per unit BW that $n^{\varepsilon}(\cdot)$ has within the (say) pass band of the filter.

Actually, as the BW of $n^{\varepsilon}(\cdot)$ increases (justifying in a sense the "increasing" independence of $n^{\varepsilon}(t)$ and $\hat{\theta}^{\varepsilon}(t)$), the magnitude of $n^{\varepsilon}(\cdot)$ must increase (see Section 2) making the replacement of $n^{\varepsilon}(\cdot)$ by white Gaussian noise a little worrisome. In fact, it was a "paradox" arising from a problem of this sort which apparently led to the original Wong-Zakai [3] work. The weak convergence method of Theorems 1-3 explicitly yields the correction term, and owing to the nature of weak convergence, the distributions of many functionals of the limit are close to those of $x^{\varepsilon}(\cdot)$; for example, first exit times from appropriate sets. Indeed, in order to study approximations to path properties of $x^{\varepsilon}(\cdot)$ via $x(\cdot)$, weak convergence seems to be the appropriate technique. This is an important advantage which the traditional methods do not have.

Define

$$R = \int_0^{\infty} Ey(0)y(s)ds,$$

and, in order to fix ideas, let $y(\cdot)$ satisfy the noise conditions of either Theorem 2 or 3. All the other conditions of these theorems hold. Then the theorems yield that $\{\hat{\theta}^{\varepsilon}(\cdot)\}$ converges weakly to the process $\hat{\theta}(\cdot)$ with the operator $(\partial/\partial t+A)$ given by

$$(4.2)\quad (\partial/\partial t+A)f(\hat{\theta},t) = f_t(\hat{\theta},t)$$

$$+ f_x(\hat{\theta},t)[\frac{AK}{2}\sin(\theta-\hat{\theta})+\sin(\theta+\hat{\theta}+2\omega_0 t)]$$

$$+ f_x(\hat{\theta},t)[-K^2R\cos(\omega_0 t+\hat{\theta})\sin(\omega_0 t+\hat{\theta})]$$

$$+ K^2R\cos^2(\omega_0 t+\hat{\theta})f_{xx}(\hat{\theta},t).$$

The quantity 2R is roughly the power per unit BW of $n^{\varepsilon}(\cdot)$ for small ε. Also, as $\varepsilon \to 0$, $\int_0^t n^{\varepsilon}(s)ds \equiv x^{\varepsilon}(t)$ converges to a Wiener process with infinitesimal covariance 2R. To see this, set $\dot{x}^{\varepsilon} = n^{\varepsilon}$, and use Theorem 2 or 3 as appropriate, to get that x^ε converges weakly to a process $x(\cdot)$ with infinitesimal operator $(\partial/\partial t + R\,\partial^2/\partial x^2)$, i.e., to a diffusion

$$dx = \sqrt{2R}\,dw$$

where $w(\cdot)$ is a standard Wiener process.

From the form of (4.2) it is easily seen that the limit $\hat{\theta}(\cdot)$ is a Markov diffusion with an Itô process representation. In particular, there is a standard Wiener process $B(\cdot)$ such that $\hat{\theta}(\cdot)$ is represented by

$$(4.3)\quad d\hat{\theta} = \frac{AK}{2}[\sin(\theta-\hat{\theta})+\sin(\theta+\hat{\theta}+2\omega_0 t)]dt$$

$$- K^2R\cos(\omega_0 t+\hat{\theta})\sin(\omega_0 t+\hat{\theta})dt + K\sqrt{2R}\cos(\omega_0 t+\hat{\theta})dB.$$

The "correction" term, the second one on the right, is not accounted for by the traditional analysis, and arises due to the non-independence of $\hat{\theta}^\varepsilon(t)$ and $n^\varepsilon(t)$. It is proportional to K^2R. For large power/unit BW of $n^\varepsilon(\cdot)$, or large system gain, this term might be of importance.

<u>The general</u> rth-<u>order loop</u>. For general D, E, C, Q in Fig. 2, $\hat{\theta}^\varepsilon = Cv^\varepsilon + Q\{\text{r.h.s. of (4.1)}\}$ and the limit process $(v(\cdot),\hat{\theta}(\cdot))$ is representable by an Itô equation of the form

$$(4.4)\quad d\binom{v}{\hat{\theta}} = \binom{D}{C}v\,dt + \binom{E}{Q}[\text{r.h.s. of (4.3)}].$$

The result for the general filter is just as easy to get as the result for (4.1), since the general filter only affects the G term (in the notation of (2.1).

The limit as $\omega_0 \to \infty$. We consider (4.3) as $\omega_0 \to \infty$. The same result holds for (4.4). It is not hard to see that the two middle terms of the right side of (4.3) should disappear as $\omega_0 \to \infty$, but it's a little harder to see what to do about the $\cos(\omega_0 t+\hat{\theta})$ coefficient of dB, since this coefficient depends on $\hat{\theta}$. The result will be the "traditional" one, but it is often dangerous to use heuristic methods to treat problems involving "products of white noise and state variables". Write the solution to (4.3) as $\hat{\theta}(\omega_0,\cdot)$. We have the following theorem.

Theorem 4. $\{\hat{\theta}(\omega_0,\cdot)\}$ is tight in $D[0,\infty)$ and as $\omega_0 \to \infty$, it converges weakly to the process $\hat{\theta}(\cdot)$ given by

$$(4.5) \qquad d\hat{\theta} = \frac{AK}{2}\sin(\theta-\hat{\theta})dt + K\sqrt{R}\,dB.$$

Proof. Apply Theorem 1 directly. The ω_0 indexes the sequence rather than ε. The proof will be outlined only. The state is $x^{\omega_0}(t) = \begin{pmatrix}\hat{\theta}(\omega_0,t)\\ t\end{pmatrix}$. All functions $(\cos(\omega_0 t+\hat{\theta})$, $\sin(\omega_0 t+\hat{\theta})$, etc.) in (4.3) are Lipschitz continuous in $\hat{\theta}$, uniformly in t, ω_0, and all are bounded. From this, we can easily show that there is a constant C such that

$$E|\hat{\theta}(\omega_0,t+s)-\hat{\theta}(\omega_0,t)|^4 \le Cs^2, \quad \text{all } t, s, \omega_0.$$

By [2, Theorem 12.3], this implies tightness of $\{\hat{\theta}(\omega_0,\cdot)\}$. Let E_t denote conditioning on $B(s)$, $s \leq t$.

Next, fix $f \in \hat{\mathscr{C}}_0^{1,3}$. Define

$$f^{\omega_0}(t) = \frac{\omega_0}{2\pi}\int_0^{2\pi/\omega_0} E_t f(\hat{\theta}(\omega_0,t+s)),t)ds.$$

Then it can be shown that $f^{\omega_0}(\cdot) \in \mathscr{D}(\hat{A}^{\omega_0})$ and that

$$\hat{A}^{\omega_0} f^{\omega_0}(t) = \frac{E_t f(\hat{\theta}(\omega_0,t+2\pi/\omega_0),t) - f(\hat{\theta}(\omega_0,t),t)}{2\pi/\omega_0}$$

$$+ \frac{2\pi}{\omega_0}\int_0^{2\pi/\omega_0} E_t f_t(\hat{\theta}(\omega_0,t+s),t)ds.$$

Owing to the uniform Lipschitz condition, we can show that

$$\text{p-lim}[f^{\omega_0}(\cdot) - f(\hat{\theta}(\omega_0,\cdot),\cdot)] = 0,$$

$$\text{p-lim}[\hat{A}^{\omega_0} f^{\omega_0}(\cdot) - (\partial/\partial t + A) f(\hat{\theta}(\cdot),\cdot)] = 0,$$

by which Theorem 1 guarantees convergence of finite-dimensional distributions. This, together with the tightness, guarantees the weak convergence. Q.E.D.

5. Adaptive Antenna Arrays [13], [14], [15]

For another illustration of the general idea, we consider a standard problem in adaptive antenna arrays. Let the array have r elements and input vector $z^{\varepsilon}(\cdot) = \{z_j^{\varepsilon}(\cdot), j=1,\ldots,r\}$, where $z_j^{\varepsilon}(\cdot) = s_j(\cdot) + n_j^{\varepsilon}(\cdot)$ and $s_j(t) = A\cos(\omega_0 t+\phi_j)$, where the $\{\phi_j\}$ are assumed known (known signal transmission direction). The input to the jth (each j) antenna is split into two parts, one part passing through an ideal $\pi/2$ phase lag device whose output we denote by $\tilde{z}_j^{\varepsilon}(\cdot) = \tilde{n}_j^{\varepsilon}(\cdot) + \tilde{s}_j(\cdot)$. The 2r outputs are weighted and added to yield the "array output" $X^{\varepsilon}(\cdot)$.

Define $Z^{\varepsilon} = (z_1^{\varepsilon},\ldots,z_r^{\varepsilon}, \tilde{z}_1^{\varepsilon},\ldots,\tilde{z}_r^{\varepsilon}) \equiv (z^{\varepsilon},\tilde{z}^{\varepsilon})$, denote the respective weights by $W = (w_1,\ldots,w_r, \tilde{w}_1,\ldots,\tilde{w}_r) = (w,\tilde{w})$ and set $S = (s_1,\ldots,s_r, \tilde{s}_1,\ldots,\tilde{s}_r) = (s,\tilde{s})$.

The object is to adaptively adjust W in order to adaptively maximize the signal-to-noise power ratio in the output $X^{\varepsilon} = w'z^{\varepsilon} + \tilde{w}'\tilde{z}^{\varepsilon} = W'Z^{\varepsilon}$. Again, for convenience, suppose that the noise takes the form $n^{\varepsilon}(t) = y(t/\varepsilon^2)/\varepsilon \equiv y^{\varepsilon}(t)/\varepsilon$. Let $\bar{M}_0$ and $\bar{M}^{\varepsilon}$ denote the covariance matrices of the vectors $(y(0),\tilde{y}(0))$ and $(n^{\varepsilon}(0), \tilde{n}^{\varepsilon}(0))$, resp. Then $\bar{M}^{\varepsilon} = \bar{M}_0/\varepsilon^2$. Assume that $\bar{M}_0 > 0$ (in the sense of positive definite matrices). Then the optimum weight vector equals $W_0 = k\bar{M}_0^{-1}S_0$, where $k \neq 0$ is any constant and $S_0 = \{\cos\phi_j, j\leq r, \sin\phi_j, j\leq r\}$. Define $Y^{\varepsilon}(\cdot) = (y^{\varepsilon}(\cdot),\tilde{y}^{\varepsilon}(\cdot))$, $Y(\cdot) = (y(\cdot),\tilde{y}(\cdot))$ and $M_s^{\varepsilon}(\cdot) = Z^{\varepsilon}(\cdot)(Z^{\varepsilon}(\cdot))'$. Then $\bar{M}_s^{\varepsilon}(\cdot) \equiv EM_s^{\varepsilon}(\cdot) = \bar{M}_0/\varepsilon^2 + (s(\cdot),\tilde{s}(\cdot))(s(\cdot),\tilde{s}(\cdot))'$. The scheme of Fig. 3 is a standard method [14], [15] of adaptively approximating

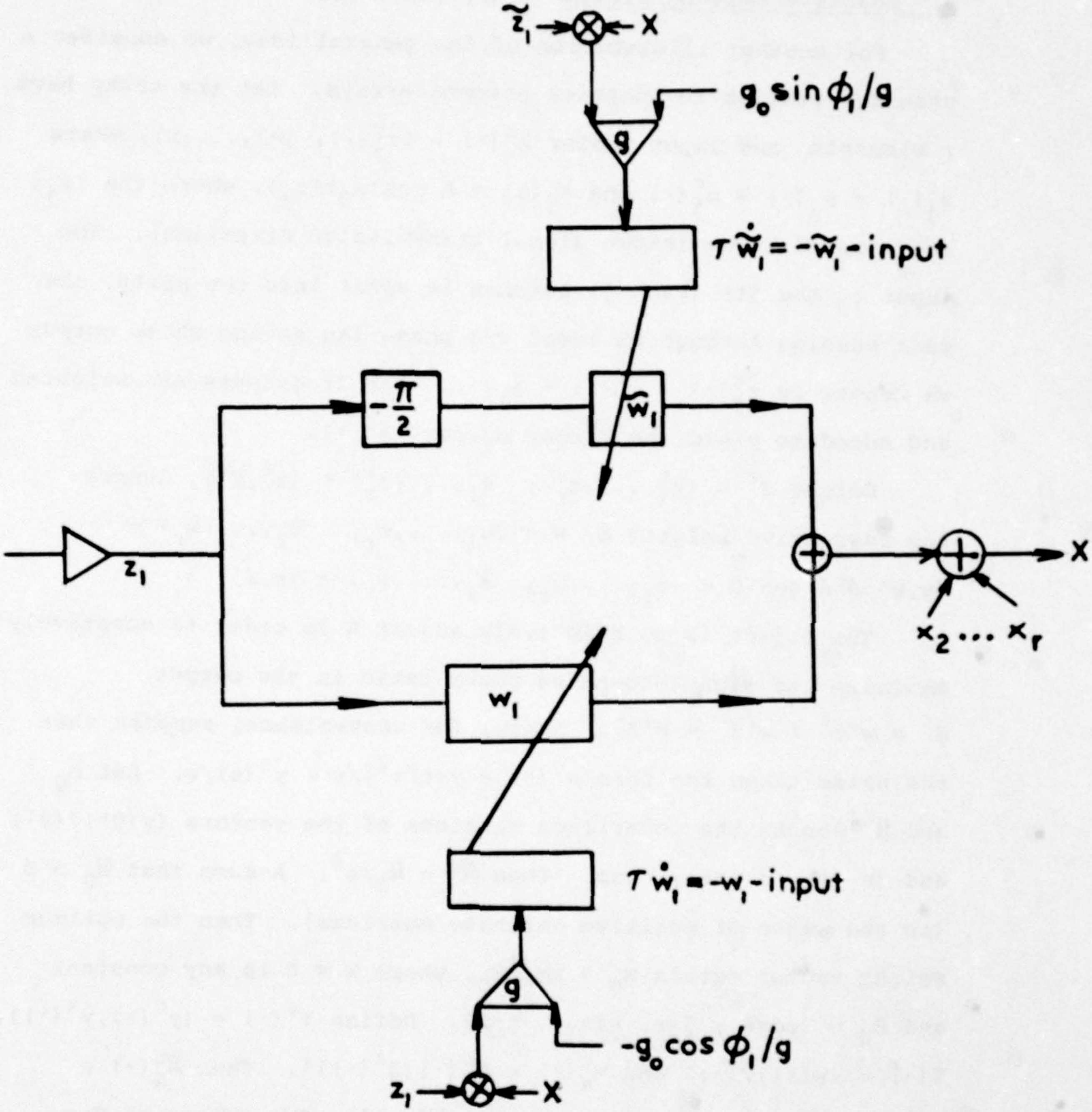


FIG. 3. SEGMENT OF AN ADAPTIVE ANTENNA ARRAY

the optimum W_0. The describing ordinary differential equation is (g, g_0 are positive constants, τ is a time constant) (5.1), where $W(\cdot)$ should, we hope, converge to something "close to" W_0.

$$(5.1) \qquad \tau\dot{W}^{\varepsilon} = -W^{\varepsilon} - gZ^{\varepsilon}X^{\varepsilon} + g_0S_0$$

$$= -[gZ^{\varepsilon}(Z^{\varepsilon})'+I]W^{\varepsilon} + g_0S_0$$

A standard method of treating (5.1) (see, e.g. [14] - or papers in [15]) involves first dropping the signal component of Z^{ε}, and then arguing as follows: since $M^{\varepsilon}(\cdot)$ is wide-band and $W^{\varepsilon}(\cdot)$ is much smoother than $M^{\varepsilon}(\cdot)$, the two are essentially independent, so assume this, take expectations in (5.1) and replace (5.1) by the resulting equation (5.2), which ought to be (approximately) an equation for the mean value $\overline{W}^{\varepsilon}(\cdot)$ of $W^{\varepsilon}(\cdot)$.

$$(5.2) \qquad \tau\dot{\overline{W}}^{\varepsilon} = -[g\overline{M}^{\varepsilon}+I]\overline{W}^{\varepsilon} + g_0S_0.$$

The asymptotic solution to (5.2) is $\overline{W}^{\varepsilon} = g_0[g\overline{M}^{\varepsilon}+I]^{-1}S_0$ which is close to the optimal value if g is large.

From a mathematical point of view, there are some difficulties with this line of reasoning - even allowing for the usually justifiable neglect of the $s(\cdot)$ terms in (5.1). As the BW of $n^{\varepsilon}(\cdot)$ increases, thereby "justifying" the "almost independence" assertion, the covariance $\overline{M}^{\varepsilon}$ must also increase (see

Section 2), so it's not immediately clear what one can say about the expectation of the product of $M^{\varepsilon}(t)$ and $W^{\varepsilon}(t)$. To see the problem more clearly, consider the scalar case $\tau\dot{W}^{\varepsilon} = -(g(n^{\varepsilon}(s))^2+1)W^{\varepsilon}+g_0S_0$ (where we set $s(\cdot) = 0$), solve it and take expectations to get

$$(5.3)\qquad EW^{\varepsilon}(t) = E(\exp - \int_0^t (g(n^{\varepsilon}(s))^2+1)ds)W^{\varepsilon}(0)$$

$$+ \int_0^t ds\ E\exp - \int_s^t [g(n^{\varepsilon}(u))^2+1]du\ g_0S_0,$$

which can differ considerably from the solution to (5.2) for small ε.

We now set the problem up in a way that admits an asymptotic analysis (as $\varepsilon \to 0$). Clearly g must be inversely proportional to ε for otherwise $\bar{M}^{\varepsilon} \to \infty$ as $\varepsilon \to 0$. Suppose that an automatic gain control mechanism of some sort is available and that we obtain an estimate of the power in $M^{\varepsilon}(\cdot)$, which is proportional to $1/\varepsilon^2$. Thus, let $g = \varepsilon^2K$ for some $K > 0$. Write $\delta M_s^{\varepsilon}(\cdot) = M_s^{\varepsilon}(\cdot) - EM_s^{\varepsilon}(\cdot)$, and rewrite (5.1) in the form

$$(5.4)\qquad \tau\dot{W}^{\varepsilon} = -[\varepsilon^2K\bar{M}_s^{\varepsilon} + \varepsilon^2K\delta M_s^{\varepsilon} + I]W^{\varepsilon} + g_0S_0$$

$$= -[K\bar{M}_0 + \varepsilon^2KSS' + \varepsilon K(Y^{\varepsilon}S'+SY^{\varepsilon\prime}) + K(Y^{\varepsilon}Y^{\varepsilon\prime}-\bar{M}_0)+I]W^{\varepsilon}$$

$$+ g_0S_0.$$

Define $\delta M_0(\cdot) = Y(\cdot)Y'(\cdot)-\bar{M}_0$ and $\delta M_0^\varepsilon(t) = \delta M_0(t/\varepsilon^2)$. Then it is clear that $(Y(\cdot),\delta M_0(\cdot))$ plays the role of the noise $y(\cdot)$ of Theorems 2 and 3. Theorem 3 (extended to (2.2)) is applicable here, and its conditions are better than those of Theorem 1 for this case. If $y(\cdot)$ satisfies the (reasonable) conditions of Theorem 3, then $\{W^\varepsilon(\cdot)\}$ is tight and converges weakly to the solution of (5.5) as $\varepsilon \to 0$. Here the limiting diffusion is degenerate because there are no $1/\varepsilon$ terms in (5.4):

$$(5.5) \qquad \tau\dot{\hat{W}} = -[K\bar{M}_0+I]\hat{W} + g_0S_0.$$

This type of argument, with the appropriate scaling of g, justifies the end result of the traditional treatment, namely going from (5.1) to (5.2). Note that, owing to the degeneracy (the limit is an ODE), the f_2^ε component of f^ε in the proofs of Theorems 2 or 3 is not needed, and owing to this (A3') can be dropped from Theorem 3. We note in passing that the scaling $g \to$ (K/average power) is often used in practice due to "dynamic range" considerations. So our scaling conforms with practice - even if this particular practice is not traditionally used in the development of (5.5), it is actually required for its justification.

<u>First-order noise effects</u>. The system (5.4) can readily be centered and scaled in order to get the first-order noise effect. Define $U^\varepsilon(\cdot) = [W^\varepsilon(\cdot)-\hat{W}(\cdot)]/\varepsilon$. A comparison of the

following development with that, say, in [14] reveals some of the mathematical shortcomings of the usual, more heuristic approach. Then

$$(5.6)\qquad \dot{U}^{\varepsilon} = -K[\bar{M}_0+I]U^{\varepsilon} - K(Y^{\varepsilon}S'+SY^{\varepsilon}{}')W^{\varepsilon} - K\varepsilon SS'W^{\varepsilon}$$

$$- K(\delta M_0^{\varepsilon}/\varepsilon)\hat{W} - K\delta M_0^{\varepsilon}U^{\varepsilon}, \qquad U^{\varepsilon}(0) = 0.$$

Theorems 2 or 3 can be applied and, again, $(Y(\cdot),\delta M_0(\cdot))$ plays the role of $y(\cdot)$ in those theorems. If $(Y(\cdot),\delta M_0(\cdot))$ satisfies the conditions on the $y(\cdot)$ of those theorems, then $\{U^{\varepsilon}(\cdot)\}$ is tight and converges weakly to a process $U(\cdot)$ with the Itö equation representation

$$(5.7)\qquad \tau dU = -K[\bar{M}_0+I]Udt + K\, dB,$$

where $B(\cdot)$ is a non-standard Wiener process whose covariance can be obtained from the $\{a_{ij}\}$ in the operator A in (2.4) in the following way.

The operator $(A+\partial/\partial t)$ is given by

$(A+\partial/\partial t)f(U,t) = f_t(U,t) + f_x'(U,t)[-K(\bar{M}_0+I)U] + (5.8)$, where (5.8) is the integral term in (2.4) (note $\hat{W}(\cdot)$ is not random)

$$(5.8)\quad K^2 \int_0^\infty E\hat{W}'(t)\delta M_0'(0) f_{uu}(U,t)\delta M_0(s)\hat{W}(t)ds$$

$$= K^2 \text{ trace } f_{uu}(U,t) \int_0^\infty E\delta M_0(0)\hat{W}(t)\hat{W}'(t)\ \delta M_0'(s)ds$$

$$\equiv \frac{K^2}{2} \sum_{i,j} \bar{a}_{ij}(t) f_{u_i u_j}(U,t),$$

(If the $\{\bar{a}_{ij}(t)\}$ in (5.8) is not symmetric, then symmetrize it so that $\bar{a}_{ij} = \bar{a}_{ji}$.) The "infinitesimal" covariance of $B(t+dt)-B(t)$ is $\{\bar{a}_{ij}(t)\}dt$. Then, to first-order terms and with wide-band input noises, $W(t) = \hat{W}(t) + \varepsilon U(t)$. Note that the limit equation (5.7) does not have a "correction" term since the $1/\varepsilon$ term in (5.6) does not involve U^ε. The lack of a "correction" term is not a priori obvious, however.

<u>Convergence on</u> $[0,\infty)$. Normally, the part of $U^\varepsilon(\cdot)$ that is of most interest is the "tail". We would like to know, for example, that the distributions of $U^\varepsilon(t)$, $t \geq T$, are close to those of the stationary solution to (5.7) for small ε and large enough T. Weak convergence does not quite give this type of result. However, in this case, a useful result is not hard to get. We only state it - the details of proof of a similar case are in [9].

If $y(\cdot)$ is a <u>bounded process</u>, then it can be shown that

$$(5.9)\quad \sup_{t>0,\ \varepsilon \text{ small}} E|U^\varepsilon(t)|^2 \text{ is bounded}$$

and

(5.10) $\{U^\varepsilon(T+\cdot),\ T>0,\ \varepsilon \text{ small}\}$ is tight in $D^{2r}[0,\infty)$

and $U^\varepsilon(T+\cdot)$ tends weakly to the stationary solution to (5.7), as $\varepsilon \to 0$ and $T \to \infty$ in any way at all.

A proof is in [9]. (5.8) is the key, and is obtained via a Liapunov function stability analysis of (5.6).

The Liapunov function $V^\varepsilon(\cdot)$, constructed according to a method in [9], has the form $V^\varepsilon(\cdot) = V(\cdot) + V_0^\varepsilon(\cdot) + V_1^\varepsilon(\cdot) + V_2^\varepsilon(\cdot)$, where $V(\cdot)$ is a Liapunov function for the deterministic system (5.5) and V^ε is a perturbation calculated from $V(\cdot)$ more or less the way f^ε was calculated from $f(\cdot)$ in Theorem 2. The ability to obtain results of the type stated in (5.8) and below it is a very useful byproduct of the method discussed in this paper. In fact, the traditional method of analysis of this problem assumes some sort of asymptotic stationarity and stability [14].

6. Filtered Hard Limited Signal Plus Noise

For the final example, we consider the case of a continuous signal plus noise $s(t) + n^\varepsilon(t)$ passing through a hard limiter (level L) followed by a band pass filter. In a classical paper, Davenport [16] treated a form of this problem where $s(t) = A \cos pt$ and $n^\varepsilon(\cdot)$ had total power N and a spectrum in a fixed band centered around p. He obtained specific values for the ratio

(6.1) $[(\text{signal power})_{out}/\text{Noise power out}$ - $/[(\text{signal power})_{in}/N]$
in a band around p

in the two limiting cases of $N \to 0$ and $N \to \infty$.

Our assumptions are a little different. Here, $s(\cdot)$ is an arbitrary continuous function. Again we set $n^\varepsilon(t) = y(t/\varepsilon^2)/\varepsilon$ and pass $s(t) + n^\varepsilon(t)$ through a hard limiter and then through any filter which has a linear differential (or even nonlinear, if we wish) equation representation. The stochastic differential equation which represents the output is derived, and from it we can readily obtain a limit value for an input-output ratio similar to (6.1). In order to keep the notation simple, we first suppose that the limiter is followed only by an integrator. As in Section 4, the general case is handled in exactly the same way; the form of the filter does not affect the method. The output $x^\varepsilon(\cdot)$ is given by

$$\dot{x}^\varepsilon = K_\varepsilon \, sign(s(t) + n^\varepsilon(t)),$$

where K_ε is a scale factor whose value will not affect the power ratios. In fact, it is convenient to use $K_\varepsilon = L/\varepsilon$, which we will do. Then

$$\dot{x}^\varepsilon = \frac{L}{\varepsilon} \, sign[s(t) + y(t/\varepsilon^2)/\varepsilon]. \tag{6.2}$$

Although (6.2) differs from the forms (2.1) used in Theorems 2 and 3, particularly because of the $1/\varepsilon$ factor appearing both inside and outside the sign function, Theorem 1 can still be used and a proof similar to that of Theorem 2 gets the correct limit and the construction of the $f^{\varepsilon}(\cdot)$. We go through some of the details below, in order to illustrate the versatility and robustness of the technique. Since $x^{\varepsilon}(\cdot)$ is not involved as either an argument or coefficient of the sign function, the scheme is not hard to use.

To facilitate computation, we let $y(\cdot)$ be Gaussian with correlation function $\sigma^2 \exp -a|t|$ $(a > 0)$. It will be shown that as $\varepsilon \to 0$, $x^{\varepsilon}(\cdot)$ converges weakly to a process $x(\cdot)$ which has the Itô representation

$$(6.3) \qquad dx = L\sqrt{2/\pi}\,\left(\frac{s(t)}{\sigma}\right)dt + L\sqrt{2 \ln 2/a}\; dB.$$

If a filter of the form used in Fig. 2 follows the limiter (where we set $Q = 0$ to avoid white noise in the output), then the limit equation is

$$(6.4) \qquad dv = Dv\,dt + E\,dx$$

$$z(t) = \text{output} = Cv(t).$$

Input-output signal-to-noise ratios. The integrated input noise $\int_0^t n^\varepsilon(s)ds$ converges weakly (as in Section 4) to a Wiener process whose covariance at time t is $2\sigma^2 t/a = 2Rt = 2t\int_0^\infty Ey(0)y(s)ds$. Thus, as $\varepsilon \to 0$, the input power per unit BW (in any finite frequency range) converges to $2\sigma^2/a$. In order to get a concrete power ratio comparison with Davenport's result, set $s(t) = A \cos pt$ here only. Consider the form (6.4) where the system dimension and D, E, C are chosen to get a good approximation to a zonal filter whose pass band includes p. The filter gain is unimportant since it does not affect the ratio, so we assume that it is unity in the pass band. The noise power per unit BW in dx/dt is $L^2\, 2 \ln 2/a$, the limit output noise power per unit BW as $\varepsilon \to 0$. The signal power in dx/dt is $L^2A^2/\pi\sigma^2$, the limit output signal power as $\varepsilon \to 0$. Thus

(Signal/Noise power per unit BW)$_{out}$/(Signal/Noise power per unit BW)$_{in}$

$$(6.5) \qquad = \left(\frac{L^2A^2}{\pi\sigma^2}\right)\Big/\left(\frac{L^2\ 2\ln 2}{a}\right)\Big/\left(\frac{A^2}{2}\Big/\frac{2\sigma^2}{a}\right) = \frac{2}{\pi \ln 2},$$

which is slightly greater than Davenport's [1] limit ratio (as his $N \to \infty$) of $\pi/4$.

This closeness of the two results is very pleasing. Since our assumptions are different, it suggests that our scheme might yield results that are meaningful under other circumstances where similar averaging phenomena occur. In our case, the input noise energy per unit BW is held constant and the BW increased. In [16], the BW is held fixed and the power per unit BW $\to \infty$ (to get the $\pi/4$ limit ratio). In [16], the ratio seems to decrease as the input noise power increases, which is consistent with our result $2/\pi \ln 2 > \pi/4$, since our power/unit BW is held fixed. The "averaging" phenomena in both cases are similar - in that the existence of the limit makes implicit use of the "wild" fluctuations and "large" magnitude of the noise.

In a promising study currently under way, a phase-locked loop with a saturator like non-linearity is being studied and compared (favorably) to the more standard systems. Asymptotic methods such as described here are used. They seem to be the only available tool.

Now an outline of the proof that the $x^\varepsilon(\cdot)$ of (6.2) converges weakly to the solution of (6.3) will be given. The proof for the general filter case with limit (6.4) is about the same.

Theorem 4. Let $s(\cdot)$ be continuous, $y(\cdot)$ Gaussian with covariance $\sigma^2 \exp - a|t|$ and mean zero. Then $\{x^\varepsilon(\cdot)\}$ is tight and as $\varepsilon \to 0$, converges weakly to the process $x(\cdot)$ given by (6.3) (integrator only used) or (6.4) (general filter following the limiter used).

Proof. We stick to the integrator case. The general case requires only carrying an extra "drift" term, and is done in exactly the same way. E_t^ε denotes conditioning on $y(u/\varepsilon^2)$, $u \le t$.

<u>Part 1</u>. Set $y(\cdot) = \sigma z^a(\cdot)$, where $z^a(\cdot)$ has correlation $\exp -a|t|$ and let z denote a random variable with the normal N(0,1) distribution. The factor L is unimportant, so set L = 1 here. We evaluate $G_\varepsilon(s) = E\,\mathrm{sign}(s+\sigma z/\varepsilon)$. Then

$$(6.6) \qquad G_\varepsilon(s) = [\mathrm{erf}(\frac{s\varepsilon}{\sigma}) - \mathrm{erf}(\frac{-s\varepsilon}{\sigma})] = \sqrt{2/\pi}\ \frac{s}{\sigma}\,\varepsilon + o(\varepsilon)$$

where $o(\varepsilon)/\varepsilon \to 0$ as $\varepsilon \to 0$, uniformly in s in any bounded set. Define $G(s) = \sqrt{2/\pi}\ (s/\sigma)$ and

$$F_\varepsilon(s,\sigma z^a(t)/\varepsilon) = [\mathrm{sign}(s+\sigma z^a(t)/\varepsilon) - E\,\mathrm{sign}(s+\sigma z^a(t)/\varepsilon)]$$

$$= [\mathrm{sign}(s+\sigma z^a(t)/\varepsilon)] - G_\varepsilon(s).$$

Then

$$(6.7) \qquad \dot{x}^\varepsilon = \sqrt{2/\pi}\ \frac{s}{\sigma} + O(\varepsilon) + F_\varepsilon(s,\sigma z^a(t/\varepsilon^2)/\varepsilon)/\varepsilon.$$

Some details will be omitted. We note that $F_\varepsilon(s,t)$ is p-right continuous since $y(\cdot)$ has a continuous density - similarly with other functions below for which this property is needed.

The aim now is to apply Theorem 1. Given $f \in \hat{\mathcal{L}}_0^{1,3}$, $\{f^\varepsilon\}$ must be found such that the "p-lim" requirements of Theorem 1 hold. The other conditions of Theorem 1 are satisfied, where A is the operator of the process (6.3). The method of proof of Theorem 2 will be used to get both $\{f^\varepsilon\}$ and A.

Similarly to the situation in Theorem 2, f^ε will have the form (no f_0^ε is needed, since G_ε is not random) $f^\varepsilon(t) = f(x^\varepsilon(t)) + f_1^\varepsilon(x^\varepsilon(t),t) + f_2^\varepsilon(x^\varepsilon(t),t)$. Setting $x = x^\varepsilon(t)$, $s = s(t)$ for notational simplicity, we have

$$(6.8) \quad \hat{A}^\varepsilon f(x,t) = f_t(x,t) + f_x(x,t)[G_\varepsilon(s) + F_\varepsilon(s, \sigma z^a(t/\varepsilon^2)/\varepsilon)/\varepsilon].$$

Define

$$f_1^\varepsilon(x,t) = \frac{1}{\varepsilon}\int_0^\infty f_x(x,t+u)E_t^\varepsilon F_\varepsilon(s(t+u),\sigma z^a(\frac{t+u}{\varepsilon^2})/\varepsilon)du$$

$$= \varepsilon \int_0^\infty f_x(x,t+\varepsilon^2 u)E_t^\varepsilon F_\varepsilon(s(t+\varepsilon^2 u),\sigma z^a(\frac{t}{\varepsilon^2}+u)/\varepsilon)du.$$

Owing to the fact that F_ε is "centered" about its expectation and to the exponential correlation of $y(\cdot)$, the integral exists and p-lim $f_1^\varepsilon = 0$. Also, $f_1^\varepsilon \in \mathcal{D}(\hat{A}^\varepsilon)$ and (use $x = x^\varepsilon(t)$)

$$(6.9) \quad \hat{A}^\varepsilon f_1^\varepsilon(t) = -f_x(x,t)F_\varepsilon(s(t), \sigma z^a(t/\varepsilon^2)/\varepsilon)/\varepsilon$$

$$+ \frac{1}{\varepsilon}\int_0^\infty f_{xx}(x,t+u)E_t^\varepsilon F_\varepsilon(s(t+u),\sigma z^a(\frac{t+u}{\varepsilon^2})/\varepsilon)du$$

$$\cdot [\frac{F_\varepsilon(s(t),\sigma z^a(t/\varepsilon^2)/\varepsilon)}{\varepsilon} + G_\varepsilon(s(t))].$$

Calculations such as showing $f_1^\varepsilon \in \mathscr{D}(\hat{A}^\varepsilon)$ are not hard here, since F_ε does not depend on the state x. Similarly for f_2^ε below.

The first term of (6.9) cancels the last term of (6.8) (which, of course, is the reason for introducing $f_1^\varepsilon(\cdot)$). The integral of (6.9) exists and equals (change variables $u/\varepsilon^2 \to u$)

$$(6.10)\quad \int_0^\infty f_{xx}(x,t+\varepsilon^2 u)E_t^\varepsilon F_\varepsilon(s(t+\varepsilon^2 u),\sigma z^a(\frac{t}{\varepsilon^2}+u)/\varepsilon)du$$

$$\cdot\, [F_\varepsilon(s(t),\sigma z^a(t/\varepsilon^2)/\varepsilon)+\varepsilon G_\varepsilon(s(t))].$$

The $\varepsilon G_\varepsilon$ term goes to zero in the p-lim sense as $\varepsilon \to 0$, and it is ignored henceforth. Still following the method of Theorem 2, let $A_0^\varepsilon f(x,t)$ denote the expectation of (6.10) (minus the $\varepsilon G_\varepsilon$ term). Then (the integral exists, again by the centering of F_ε and the exponential decrease in the correlation function of $y(\cdot)$)

$$A_0^\varepsilon f(x,t) = \int_0^\infty f_{xx}(x,t+\varepsilon^2 u)EF_\varepsilon(s(t+\varepsilon^2 u),\sigma z^a(u)/\varepsilon)F_\varepsilon(s(t),\sigma z^a(0)/\varepsilon)du.$$

Next, define $f_2^\varepsilon(x,t)$, the "centered and averaged" last term of (6.9) (minus the G_ε term), by

$$f_2^\varepsilon(x,t) = \int_0^\infty dv\{\int_0^\infty du\,[f_{xx}(x,t+u+v)\frac{1}{\varepsilon^2}E_t^\varepsilon F_\varepsilon(s(t+u+v),\frac{\sigma}{\varepsilon}z^a(\frac{t+u+v}{\varepsilon^2}))$$

$$F_\varepsilon(s(t+v),\frac{\sigma}{\varepsilon}z^a(\frac{t+v}{\varepsilon^2}))]-A_0^\varepsilon f(x,t+v)\}.$$

It can be shown that the integral exists, again by use of the centering about the mean $A_0^\varepsilon f(x,t+v)$ of the inner integral, and the exponential correlation function. Via the usual change of variables $u/\varepsilon^2 \to u$, $v/\varepsilon^2 \to v$, we get that $E|f_2^\varepsilon(x,t)| = O(\varepsilon^2)$. Also $f_2^\varepsilon \in \mathcal{D}(\hat{A}^\varepsilon)$ and (use $x = x^\varepsilon(t)$)

$$\hat{A}^\varepsilon f_2^\varepsilon(x,t) = \text{minus first term on r.h.s. of (6.10) plus terms whose (absolute) expectation is } O(\varepsilon) + A_0^\varepsilon f(x,t).$$

Then, concluding,

$$\text{(6.11)} \quad \text{p-lim}[f^\varepsilon(\cdot) - f(x^\varepsilon(\cdot),\cdot)] = 0,$$

$$\text{p-lim}[\hat{A}^\varepsilon f^\varepsilon(\cdot) - f_t(x^\varepsilon(\cdot),\cdot) - A_0^\varepsilon f(x^\varepsilon(\cdot),\cdot) - G_\varepsilon(s) f_x(x^\varepsilon(\cdot),\cdot)] = 0.$$

In Parts 2 and 3 below, it is shown that ((6.12) defines A_0)

$$\text{(6.12)} \quad A_0^\varepsilon f(x,t) \to \frac{f_{xx}(x,t)\ln 2}{a} \equiv A_0 f(x,t) \text{ uniformly in } x \text{ for each } t.$$

This, together with $G_\varepsilon(s) \to G(s)$ uniformly on bounded s-sets (hence $f_x G_\varepsilon \to f_x G$ uniformly in x for each t) yields the theorem, since the process of (6.3) is the unique process corresponding to the operator

$$(\partial/\partial t + A) = \left(\frac{\partial}{\partial t} + G(s(t))\frac{\partial}{\partial x} + \frac{1}{2}\left(\frac{2\ \ln\ 2}{a}\right)\frac{\partial^2}{\partial x^2}\right).$$

Part 2. Evaluation of $A_0^\varepsilon f$. Let $z(\cdot)$ denote a Gaussian process with correlation function $\exp -|t|$. Changing variables u(old) → u(new)/a yields

$$A_0^\varepsilon f(x,t) = \frac{1}{a} I_0^\varepsilon(x,t)$$

where

$$(6.13)\quad I_0^\varepsilon(x,t) = \int_0^\infty f_{xx}(x,t+\varepsilon^2 u/a)[Q^\varepsilon(t,u) - R^\varepsilon(t,u)]du,$$

where

$$Q^\varepsilon(t,u) = E\ \text{sign}[s(t+u\varepsilon^2/a) + \frac{\sigma}{\varepsilon} z(u)]\ \text{sign}[s(t)+\frac{\sigma}{\varepsilon}z(0)],$$

$$R^\varepsilon(t,u) = E\ \text{sign}[s(t+\frac{u\varepsilon^2}{a})+\frac{\sigma}{\varepsilon}z(u)]\ E\ \text{sign}[s(t)+\frac{\sigma}{\varepsilon}z(0)].$$

Owing to the properties of the joint distribution of $(z(0),z(u))$, $\int_0^T |Q^\varepsilon(t,u)-R^\varepsilon(t,u)|du$ is bounded uniformly in ε and T and converges uniformly in ε as $T \to \infty$. (In fact, the integrand goes to zero at an exponential rate as $u \to \infty$.) Using this and the smoothness and compact support of $f_{xx}(\cdot,\cdot)$, we can replace $f_{xx}(x,t+\varepsilon^2 u/a)$ by $f_{xx}(x,t)$ in $I_0^\varepsilon(x,t)$ without altering the limit as $\varepsilon \to 0$.

By the above arguments, if $Q^\varepsilon(t,u)-R^\varepsilon(t,u)$ has a limit for each t as $\varepsilon \to 0$, then

$$(6.14)\quad \lim_{\varepsilon\to 0} A_0^\varepsilon f(x,t) = \frac{1}{a}\int_0^\infty \lim_{\varepsilon\to 0}[Q^\varepsilon(t,u)-R^\varepsilon(t,u)]du$$

and also that in order to show (6.12) it is enough to show that the integral on the right equals ln 2. First the existence of the limit will be shown. Let $s^+ = s(t+\varepsilon^2 u/a)$ and $s = s(t)$. Then

$$Q^\varepsilon(t,u) = P\{z(u)>-s^+\varepsilon/\sigma, z(0)>-s\varepsilon/\sigma\}+P\{z(u)<-s^+\varepsilon/\sigma, z(0)<-s\varepsilon/\sigma\}$$

$$-P\{z(u)>-s^+\varepsilon/\sigma, z(0)<-s\varepsilon/\sigma\}-P\{z(u)<-s^+\varepsilon/\sigma, z(0)>-s\varepsilon/\sigma\},$$

$$R^\varepsilon(t,u) = [P\{z(u)>-s^+\varepsilon/\sigma\}-P\{z(u)<-s^+\varepsilon/\sigma\}]$$

$$\cdot [P\{z(0)>-s\varepsilon/\sigma\}-P\{z(0)<-s\varepsilon/\sigma\}].$$

Obviously as $\varepsilon \to 0$, $R^\varepsilon(t,u) \to 0$ (even uniformly on bounded s, s^+, t, u sets, although we don't need this). Also (even uniformly as above)

$$Q^\varepsilon(t,u) \to P\{z(u)>0,z(0)>0\} + P\{z(u)<0,z(0)<0\}$$

$$- P\{z(u)<0,z(0)>0\} - P\{z(u)>0,z(0)<0\}$$

$$= 2[P\{z(u)>0,z(0)>0\} - P\{z(u)<0,z(0)>0\}]$$

$$\equiv 2J(u).$$

Define

$$J_0 = \int_0^\infty J(u)\,du.$$

Then we have proved that

$$\hat{A}_0^\varepsilon f(x,t) \to \frac{2f_{xx}(x,t)}{a}\int_0^\infty J(u)\,du = A_0 f(x,t)$$

uniformly in x for each t. We need only evaluate J_0.

Part 3. Proof that $J_0 = (\ln 2)/2$. Use polar coordinates and write $\rho = e^{-u}$. Then the joint density of $(z(0), z(u))$ is

$$\frac{1}{2\pi(1-\rho^2)^{1/2}}(\exp -r^2 g(\theta)/2)r,$$

$$g(\theta) = \frac{1}{1-\rho^2}[\cos^2\theta - 2\rho\sin\theta\cos\theta + \sin^2\theta] = \frac{1}{1-\rho^2}[1-\rho\sin 2\theta].$$

Also

$$J^+(u) \equiv P\{z(0)>0, z(u)>0\} = \int_0^{\pi/2}\int_0^\infty \frac{r}{2\pi(1-\rho^2)^{1/2}}(\exp -\frac{r^2}{2}g(\theta))\,dr\,d\theta,$$

$$J^-(u) \equiv P\{z(0)>0, z(u)<0\} = \int_{-\pi/2}^{0}\int_0^\infty \frac{r}{2\pi(1-\rho^2)^{1/2}}(\exp -\frac{r^2}{2}g(\theta))\,dr\,d\theta.$$

Integrating with respect to r yields

$$J^+(u) = \int_0^{\pi/2}\frac{d\theta(1-\rho^2)}{2\pi(1-\rho^2)^{1/2}(1-\rho\sin 2\theta)} = \frac{1}{2}\int_0^\pi \frac{d\theta}{1-\rho\sin\theta}\,\frac{(1-\rho^2)^{1/2}}{2\pi}.$$

By [16, eqn. 298],

$$J^{\pm}(u) = \frac{1}{\pi}\tan^{-1}\frac{1\mp\rho}{(1-\rho^2)^{1/2}} - \frac{1}{\pi}\tan^{-1}\frac{\mp\rho}{(1-\rho^2)^{1/2}}$$

Using $\tan^{-1}x - \tan^{-1}y = \tan^{-1}(x-y)/(1+xy)$ [17, p. 48], we have

$$J^{+}(u) - J^{-}(u) = \frac{1}{\pi}\tan^{-1}\frac{-\rho}{(1-\rho^2)^{1/2}} - \frac{1}{\pi}\tan^{-1}\frac{-\rho}{(1-\rho^2)^{1/2}}$$

$$+ \frac{1}{\pi}\tan^{-1}\frac{\rho}{(1-\rho^2)^{1/2}}$$

$$= \frac{1}{\pi}\tan^{-1}\frac{\rho}{(1-\rho^2)^{1/2}} = \frac{1}{\pi}\sin^{-1}\rho.$$

Next, let $\rho = e^{-t}$ and change variables $v = e^{-t}$, $t = -\ln v$, to get

$$J_0 = \frac{1}{\pi}\int_0^{\infty}\sin^{-1}(e^{-t})dt = \frac{1}{\pi}\int_0^1 \frac{\sin^{-1} v}{v}\,dv$$

With $v = \sin w$ [17, p. 417],

$$J_0 = \frac{1}{\pi}\int_0^{\pi/2} w.\text{ctn}\, w\, dw = \frac{1}{2}\ln 2. \qquad \text{Q.E.D.}$$

References

1. W.B. Davenport, "Signal to noise ratios in band pass limiters", J. of Appl. Physics, 24, 720-727, June 1953.

2. P. Billingsley, Convergence of probability measures, John Wiley and Sons, New York, 1968.

3. E. Wong, M. Zakai, "On the convergence of ordinary integrals to stochastic integrals", Ann. Math. Stat., 36, pp. 1560-1564, 1965.

4. E. Wong, M. Zakai, "On the relationship between ordinary and stochastic differential equations", Int. J. Engin. Sci., 3, pp. 213-229, 1965.

5. R.Z. Khasminskii, "A limit theorem for solutions of differential equations with random right-hand sides", Theory of Prob. and Applic., 11, pp. 390-406, 1966.

6. G.C. Papanicolao, W. Kohler, "Asymptotic theory of mixing ordinary differential equations", Comm. Pure and Appl. Math., 27, pp. 641-668, 1974.

7. G. Blankenship, G.C. Papanicolaou, "Stability and control of stochastic systems with wide-band noise disturbances", SIAM J. on Appl. Math., 34, pp. 437-476, 1978.

8. H.J. Kushner, "Jump-diffusion approximations for ordinary differential equations with wide-band random right-hand sides", to appear SIAM J. on Control and Optimiz.

9. H.J. Kushner, "Approximation of solutions to differential equations with random inputs by diffusion processes", January 1979 Bonn (Germany) conference on stochastic control. Proceedings to be published in Lecture Notes in Mathematics by Springer.

10. T.G. Kurtz, "Semigroups of conditional shifts and approximation of Markov processes", Ann. Prob., 4, pp. 618-642, 1975.

11. H.J. Kushner, "Martingale methods for approximations of processes by jump-diffusion processes", in preparation.

12. A.J. Viterbi, Principles of coherent communication, McGraw-Hill, New York, 1966.

13. W.F. Gabriel, "Adaptive arrays, an introduction", Proc. of the IEEE, 64, pp. 239-272.

14. L.E. Brennan, E.L. Pugh, I.S. Reed, "Control loop noise in adaptive antenna arrays", IEEE Trans. Aerosp. Electron. Syst., AES-7, pp. 254-262, 1971.

15. IEEE Trans. on Antennas and Propagation, AP-24, special issue on adaptive antenna arrays, 1976.

16. B.O. Pierce, A short table of integrals, Ginn and Co., Boston, 1929.

17. I.S. Gradshteyn, I.M. Ryzhik, Tables of integrals, series and products, Academic Press, New York, 1965.