SIAM J. APPL. MATH.
Vol. 34, No. 3, May 1978

STABILITY AND CONTROL OF STOCHASTIC SYSTEMS WITH WIDE-BAND NOISE DISTURBANCES. I*

G. BLANKENSHIP† AND G. C. PAPANICOLAOU‡

Abstract. For dynamical systems with external influences which are approximately white noise (wide-band noise), we show that stability and other properties of the white noise problem that depend on the infinite time interval, continue to hold away from white noise but not far from it.

1. Introduction.[1] The purpose of this work is to analyze the stability properties and the control of systems that are subjected to external noise disturbances. Specifically, we consider a family of systems labeled by a parameter $\varepsilon > 0$ such that as $\varepsilon \to 0$ the external disturbances become white noise. We assume that the limiting white noise problem has certain properties; for example, stability, recurrence, invariant distributions, etc. If these properties of the limiting problem hold in a sufficiently nonmarginal way, i.e., they hold in a sufficiently strong sense, then the corresponding problems for $\varepsilon > 0$ and sufficiently small also have these properties.

Thus, we confirm under specific conditions, what one expects to happen, namely, that computations based on the assumption that the disturbances are white noise are in fact robust relative to perturbations away from white noise but remaining in the wide-band regime.

The basic ideas underlying the problems at hand are due to Stratonovich [1]. The mathematical analysis of stochastic systems near the white noise limit is carried out in [2], [3] and elsewhere. However, in previous work the limit $\varepsilon \to 0$ was taken under the assumption that the time t remained in a bounded but arbitrary set, $0 \leqq t \leqq T < \infty$. Thus, questions of stability, etc., that depend on the infinite time interval could not be answered. In this paper we attempt to remove this deficiency.

In § 2 we formulate the problems and explain the nature of the family of systems, parameterized by $\varepsilon > 0$, which are near a white noise system. We also relate the parameter ε to the bandwidth of the local noise disturbances.

In § 3 we give in detail the perturbation analysis as $\varepsilon \to 0$ which is the basis for future constructions. The analysis is similar to the one in [4], [5] wherein additional references are cited to related work and methodology.

In § 4 we state and prove a weak convergence result similar to the one in [2] and [3]. The proof is different here and plays an important role in the stability

---

* Received by the editors June 16, 1976.

† Systems Research Center, Case Western Reserve University, Cleveland, Ohio 44106. This research was carried out while the author was a visiting member of the Courant Institute and supported by the Air Force Office of Scientific Research under Grant AFOSR-76-2884.

‡ Courant Institute of Mathematical Sciences, New York University, New York, New York 10012. The work of this author was supported by an Alfred P. Sloan Foundation Fellowship and by the Air Force Office of Scientific Research under Grant AFOSR-76-2884.

[1] This paper, part I, was written by G. Papanicolaou. A companion paper, part II, was written by G. Blankenship.

questions. The general scheme follows Kurtz's work [6]. We make effective use of martingales as in [7]; in the latter work it is no longer assumed that the coefficients (the external driving process) are ergodic and new phenomena arise.

Section 5 contains our results on stability. The theorem of § 5.1 is the nonwhite noise analogue of a well known white noise result [8, p. 325]. Section 5.2 is analogous to Khasminskii's result [9] for the white noise problem. We also employ Pinsky's device of introducing the function h [10, p. 320]. Section 5.3 contains an application of the theorem of § 5.2 to a harmonic oscillator. The analogous white noise problem at low frequency is treated in [11]. We treat here the high frequency case because it is easier and because it illustrates nicely the effect of averaging superimposed on the white noise or wide-band limit (cf. [3] and references therein for additional information on this point).

Section 6 contains our results on invariant measures. The theorem in § 6.1 is analogous to the one of § 5.1 and the one in § 6.2 is analogous to the one of § 5.2. In both results we employ a theorem of Beneš [12] concerning the existence of invariant measures. The theorem of § 6.2 is analogous to the white noise result of Zakai [13].

Section 7 contains an upper estimate for the probability of deviating far from the equilibrium point given that it is stable in a sufficiently strong sense. The white noise result and its proof is due to Pinsky [14] and it is presented as Theorem 1. Theorem 2 is the corresponding result for the wide-band noise systems.

In a companion paper, part II, we examine some related questions in control theory.

2. Stochastic systems with wide-band noise disturbances. Let $x(t) \in R^n$ be the state of a system at time $t \geqq 0$ and let $y(t) \in R^m$, say, be the state of some external process that influences the evolution of the system. Suppose that

$$\begin{aligned} \frac{dx(t)}{dt} &= F(x(t), y(t)), \qquad t>0, \\ x(0) &= x, \end{aligned} \tag{2.1}$$

where $F(x, y)$ is a smooth n-vector function on $R^n \times R^m$ so that $y(t)$ represents the random coefficients in (2.1). We are interested, typically, in the following questions:

(i) What is the probability law of $x(\cdot)$ given that of $y(\cdot)$?

(ii) Under what conditions is $x(t)$ stable as $t \to \infty$, with "stable" being appropriately defined?

(iii) How can systems such as (2.1) be controlled, i.e.,

$$\begin{aligned} \frac{dx(t)}{dt} &= F(x(t), y(t), u(t)), \qquad t>0, \\ x(0) &= x, \end{aligned} \tag{2.2}$$

where $u(\cdot)$ is the control (in some space) chosen to optimize some cost criterion?

These questions are too general to admit informative answers. When, however, the external process $y(t)$ is white noise, these questions have, as is well known, reasonably satisfactory answers collectively referred to as the theory of Itô stochastic differential equations (cf. [8], [15], [16], [17] and many other references cited in these).

We want to study questions (i)–(iii) when the external influences are not white noise but only approximately white noise. A very convenient way of defining what is meant by this is by introducing a small parameter $\varepsilon > 0$ which measures deviation from the white noise case. The situation is as follows.

Let $y(t)$, $t \geqq 0$, be a stationary m-vector valued process and let $F(x, y)$ and $G(x)$ be smooth n-vector functions on $R^n \times R^m$ and R^n respectively. Consider the problem

$$\frac{dx(t)}{dt} = F(x(t), y(t)) + G(x(t)), \qquad x(0) = x, \tag{2.3}$$

and assume that for each x fixed[2]

$$E\{F(x, y(t))\} = 0. \tag{2.4}$$

With (2.3) we associate the deterministic system

$$\frac{d\bar{x}(t)}{dt} = G(\bar{x}(t)), \qquad \bar{x}(0) = x. \tag{2.5}$$

Because of (2.4), the term $F(x(t), y(t))$ in (2.3) plays the role of fluctuations to the deterministic problem (2.5). Note that the fluctuations are locally dependent upon the solution, i.e., F depends on x.

Let $F^T(x, y)$ denote the transpose of $F(x, y)$ and let $R(x, s)$ be the $n \times n$ covariance matrix of the local fluctuations:

$$R(x, s) = E\{F(x, y(t))F^T(x, y(t+s))\}. \tag{2.6}$$

Let $S(x, \omega)$ be the power spectral density[3] of the fluctuations

$$S(x, \omega) = \int_{-\infty}^{\infty} e^{i\omega s} R(x, s)\, ds. \tag{2.7}$$

The basic premise of the wide-band noise approximation is that $S(x, \omega)$ is (effectively) band limited:

$$S(x, \omega) \equiv 0 \quad \text{for } |\omega| > \omega_0 > 0, \quad x \in R^n, \tag{2.8}$$

and that all relevant frequencies associated with the deterministic problem (2.5) are contained in $[-\omega_0, \omega_0]$, i.e., the support of S is wide enough.

Let $\varepsilon > 0$ be a parameter and define

$$F^\varepsilon(x, y^\varepsilon(t)) = \varepsilon^{-1} F(x, y(t/\varepsilon^2)). \tag{2.9}$$

---

[2] $E\{\cdot\}$ denotes expectation value.

[3] We assume it exists.

Then (2.6) yields

$$
\begin{aligned}
R^\varepsilon(x,s) &= E\{F^\varepsilon(x,y^\varepsilon(t))(F^\varepsilon(x,y^\varepsilon(t+s)))^T\}\\
&= \varepsilon^{-2}E\{F(x,y(t/\varepsilon^2))F^T(x,y((t+s)/\varepsilon^2))\}\\
&= \varepsilon^{-2}R(x,s/\varepsilon^2)\\
&\xrightarrow{\varepsilon\downarrow 0} \int_{-\infty}^{\infty} R(x,\sigma)\,d\sigma\,\delta(s) \qquad \text{(white noise)}
\end{aligned} \tag{2.10}
$$

and hence,

$$S^\varepsilon(x,\omega)=S(x,\varepsilon^2\omega). \tag{2.11}$$

Thus

$$S^\varepsilon(x,\omega)\equiv 0 \quad \text{for } |\omega|>\varepsilon^{-2}\omega_0 \tag{2.12}$$

and as $\varepsilon\to 0$ the bandwidth of S^ε tends to infinity.

We shall take as our wide-band noise system[4]

$$
\begin{aligned}
&\frac{dx^\varepsilon(t)}{dt}=\frac{1}{\varepsilon}F(x^\varepsilon(t),y^\varepsilon(t))+G(x^\varepsilon(t),y^\varepsilon(t)), \qquad t>0,\\
&x^\varepsilon(0)=x,\\
&y^\varepsilon(t)=y(t/\varepsilon^2),
\end{aligned} \tag{2.13}
$$

with $y(t)$ a given stationary process (other hypotheses introduced later), F and G smooth n-vector functions on $R^n\times R^m$ and F satisfying (2.4). The parameter $\varepsilon>0$ measures departure from the white noise approximation. Another interpretation for ε is that it differentiates between the time scale of fluctuations of the coefficients and the solution; the latter is much slower than the former.

In many applications one encounters other small parameters, in addition to ε, in (2.13). If as $\varepsilon\to 0$ they remain of order one, then it is reasonable to first take the limit $\varepsilon\to 0$ and consider other approximations afterwards. In many interesting cases the other small parameters are coupled to ε and one has, for example, rapid oscillations (averaging) or rapid decay superimposed on the wide band noise (white noise) limit. An example is given in § 5.3. For more details we refer to [3] and [18].

The behavior of the process $x^\varepsilon(t)$ of (2.13) when ε is small, and specifically stability or related questions, is the object of the following sections. If we rescale (2.13) as follows

$$t=\varepsilon^2\tau, \qquad \tilde{x}^\varepsilon(\tau)=x^\varepsilon(\varepsilon^2\tau), \tag{2.14}$$

then $\tilde{x}^\varepsilon(\tau)$, $\tau\geqq 0$, satisfies the equation

$$
\begin{aligned}
&\frac{d\tilde{x}^\varepsilon(\tau)}{d\tau}=\varepsilon F(\tilde{x}^\varepsilon(\tau),y(\tau))+\varepsilon^2G(\tilde{x}^\varepsilon(\tau),y(\tau)), \qquad \tau>0,\\
&\tilde{x}^\varepsilon(0)=x.
\end{aligned} \tag{2.15}
$$

[4] Letting $G=G(x,y)$ with $E\{G(x,y(t))\}\neq 0$ introduces no additional features into the problem.

In §§ 3 and 4 we analyze the limit $\varepsilon \downarrow 0$ in (2.13) or, $\varepsilon \downarrow 0$, $\tau \uparrow \infty$, $\varepsilon^2\tau = t$ in (2.15), *with t fixed in both cases and* $0 \leqq t \leqq T < \infty$ (T is arbitrary). This is the usual diffusion or white noise limit [2], [3]. In §§ 5, 6 and 7 we examine the limit $t \to \infty$ in (2.13) or, $\tau \to \infty$ in (2.15), *with* ε *fixed and* $0 < \varepsilon \leqq \varepsilon_0$, *where* ε_0 *is sufficiently small.*

Both scalings (2.13) and (2.15) are useful and *all* results can be stated in either scaling. In §§ 3 and 4 we employ (2.13). Later we also employ (2.15). In the context of specific applications one or the other scaling, (2.13) or (2.15), may be more appropriate or more natural. This should not obscure the fact that we are dealing with the same problem.

3. The perturbation expansion. In this section we describe the class of external influence processes we shall admit and then, under some simplifying assumptions that are removed in § 4, we describe the perturbation procedure that is used repeatedly in later sections.

Let $y(t)$ be a time homogeneous Markov process with values in a compact metric space S. We shall assume that $y(t)$ is ergodic and this is necessary for the results we want (see [7] for results without this assumption). This process will be our external influence process. To be specific, and without substantially restricting generality, we shall assume that $y(t)$ is a jump process as follows (for more general situations see [3] and [7]).

Let $P(t, y, A)$, $t \geqq 0$, $y \in S$, A a Borel subset of S, be the transition probability function of $y(t)$, $t \geqq 0$. We shall assume that for $f(y)$ a continuous real-valued function on S

$$Qf(y) \equiv \lim_{t \downarrow 0} \frac{1}{t} \int_S P(t, y, dz)(f(z) - f(y)) \tag{3.1}$$

has the form[5]

$$Qf(y) = q(y) \int \pi(y, dz) f(z) - q(y) f(y) \tag{3.2}$$

where

(3.3) $q(y)$ is a continuous function on S and there exist constants q_l, q_u such that $0 < q_l \leqq q(y) \leqq q_u < \infty$

and

(3.4) $\pi(y, A)$, $y \in S$, A a Borel subset of S, is a probability measure on S for each $y \in S$ and a continuous function of y for each Borel set A.

The operator Q of (3.1) and (3.2) is the infinitesimal generator of the process $y(t)$, $t \geqq 0$. It is easily seen that $P(t, y, A)$ is a Feller transition function and that $y(\cdot)$ can be taken as right continuous with left hand limits. Hence $y(\cdot)$ is a strong Markov process.

Let us assume, in addition to (3.4), that for some nontrivial reference probability measure μ on S, $\pi(y, \cdot)$ has a density

$$\pi(y, dz) = \tilde{\pi}(y, z)\mu(dz), \tag{3.5}$$

[5] The region of integration will be usually omitted in the following.

jointly measurable in $y, z \in S$ and such that

$$0<\tilde{\pi}_l \leqq \tilde{\pi}(y,z) \leqq \tilde{\pi}_u < \infty, \tag{3.6}$$

where $\tilde{\pi}_l$ and $\tilde{\pi}_u$ are constants. Then it follows easily by standard arguments that the process $y(t)$, $t \geqq 0$ is ergodic and that it has a unique invariant measure $\bar{P}(A)$, $A \subset S$

$$\bar{P}(S)=1, \qquad \bar{P}(A)=\int \bar{P}(dy)P(t,y,A). \tag{3.7}$$

Moreover the equation

$$Qf(y)=-g(y), \qquad y \in S, \tag{3.8}$$

with

$$\int \bar{P}(dy)g(y)=0 \tag{3.9}$$

has a bounded solution, unique up to an additive constant. The solution is given by

$$f(y)=-(Q^{-1}g)(y)\equiv \int_S \chi(y,dz)g(z), \qquad y \in S, \tag{3.10}$$

where

$$\chi(y,A)=\int_0^\infty dt(P(t,y,A)-\bar{P}(A)) \tag{3.11}$$

is the recurrent potential kernel; the integral is absolutely convergent since $P(t,y,A)$ approaches $\bar{P}(A)$ exponentially fast uniformly in y and A as a consequence of (3.3) and (3.6).

We shall refer to (3.8)–(3.11) by the statement that *the Fredholm alternative holds for the process* $y(t)$, $t \geqq 0$. The above description of the influence process $y(\cdot)$ and its properties will be used throughout in the sequel. We emphasize that we have introduced these details for the sake of being specific. Most results that follow hold under much more general conditions. However, as already mentioned, a minimum of ergodicity assumptions is necessary. Some form of the Fredholm alternative as above must hold or else different phenomena can arise.

Let $x^\varepsilon(t)$, $t \geqq 0$, be defined by (2.13) where $F(x,y)$, $G(x,y)$: $R^n \times S \to R^n$ are continuous in y and bounded and smooth as functions of x. Let $y^\varepsilon(t)$ be defined by (2.13) with $y(t)$ as above. The assumptions that F and G are bounded is removed in the next section so that linear systems can be included. We impose it in this section for simplicity.

Clearly $(x^\varepsilon(t), y^\varepsilon(t))$ are jointly a Markov process on $R^n \times S$. Let $P^\varepsilon(t,x,y,A)$, $x \in R^n$, $y \in S$, A a Borel subset of $R^n \times S$, be the transition probability function and $\mathcal{L}^\varepsilon$ the infinitesimal generator of this process. On functions $f(x,y)$ smooth in x and continuous in y, we find easily from (2.13) and (3.2) that

$$\mathcal{L}^\varepsilon f(x,y)=\frac{1}{\varepsilon^2}Qf(x,y)+\frac{1}{\varepsilon}F(x,y)\cdot\frac{\partial f(x,y)}{\partial x}+G(x,y)\cdot\frac{\partial f(x,y)}{\partial x}. \tag{3.12}$$

Here $\partial f/\partial x$ stands for the x-gradient of f and the dot stands for inner product of vectors in R^n. In the scaling (2.15) the process $(\tilde{x}^\varepsilon(\tau), y(\tau))$ has the infinitesimal generator

$$\tilde{\mathscr{L}}^\varepsilon f(x,y) = Qf(x,y) + \varepsilon F(x,y)\cdot\frac{\partial f(x,y)}{\partial x} + \varepsilon^2 G(x,y)\cdot\frac{\partial f(x,y)}{\partial x}, \tag{3.13}$$

which differs from $\mathscr{L}^\varepsilon$ of (3.12) merely by a factor ε^2. The backward Kolmogorov equation for

$$u^\varepsilon(t,x,y) = E_{x,y}\{f(x^\varepsilon(t), y^\varepsilon(t))\} = \int_{R^n}\int_S P^\varepsilon(t,x,y,d\xi\, du) f(\xi, u)$$

is

$$\begin{aligned} &\frac{\partial u^\varepsilon(t,x,y)}{\partial t} = \mathscr{L}^\varepsilon u^\varepsilon(t,x,y), \qquad t>0,\\ &u^\varepsilon(0,x,y) = f(x,y). \end{aligned} \tag{3.14}$$

For $\varepsilon > 0$ fixed the process $(x^\varepsilon(t), y^\varepsilon(t))$ has the following properties which are elementary consequences of the above definitions and assumptions:

(i) The transition function $P^\varepsilon(t,x,y,A)$ defines a Feller semigroup on $C(R^n \times S)$, the continuous functions on $R^n \times S$.

(ii) The semigroup maps $\hat{C}(R^n \times S)$ (continuous functions that vanish as $|x| \to \infty$) into itself. In fact for any compact set $K \subset R^n \times S$ and all

(3.15) $y \in S$, $t > 0$,

$$\lim_{|x|\to\infty} P^\varepsilon(t,x,y,K) = 0.$$

(iii) The process $(x^\varepsilon(t), y^\varepsilon(t))$ has a version which is right continuous with left hand limits. In fact, trivially, $x^\varepsilon(t)$ is continuous.

From the above it follows that $(x^\varepsilon(t), y^\varepsilon(t))$ is a strong Markov process.

We shall now proceed with the perturbation analysis of the backward equation (3.14) for ε small. Note first that $\mathscr{L}^\varepsilon$ of (3.12) has the form

$$\mathscr{L}^\varepsilon = \frac{1}{\varepsilon^2}\mathscr{L}_1 + \frac{1}{\varepsilon}\mathscr{L}_2 + \mathscr{L}_3 \tag{3.16}$$

where

$$\mathscr{L}_1 \equiv Q, \quad \mathscr{L}_2 = F\cdot\frac{\partial}{\partial x}, \quad \mathscr{L}_3 = G\cdot\frac{\partial}{\partial x}. \tag{3.17}$$

Let us now assume, as in (2.8), that

$$\int F(x,y)\bar{P}(dy) = 0. \tag{3.18}$$

Problem (3.14) can be analyzed as a problem entirely within the context of differential equations and semigroups following the general results of Kurtz [19].

The formal analysis behind such results follows the general rules of perturbation expansions and one can use the same general rules for a surprisingly diverse class of problems [20] in addition to the present ones. We continue now with the analysis.

Recall at first that we are interested mainly in the behavior of $x^\varepsilon(t)$ for ε small. Therefore we take f in (3.14) as a smooth function of x only,[6] i.e. $f=f(x)$. Because of (3.18) and the Fredholm alternative (3.8)–(3.11) the function

$$f_1(x,y)=\int_S \chi(y,dz)F(x,z)\cdot\frac{\partial f(x)}{\partial x} \tag{3.19}$$

satisfies the equation

$$Qf_1(x,y)+F(x,y)\cdot\frac{\partial f(x)}{\partial x}=0. \tag{3.20}$$

Now we define on smooth functions a linear operator $\mathscr{L}$ by setting

$$\mathscr{L}f(x)=\int_S \bar{P}(dy)F(x,y)\cdot\frac{\partial f_1(x,y)}{\partial x}+\int_S \bar{P}(dy)G(x,y)\cdot\frac{\partial f(x)}{\partial z}. \tag{3.21}$$

More explicitly, using (3.19), we have

$$\begin{aligned}\mathscr{L}f(x)=&\iint\limits_{S\times S}\chi(y,dz)\bar{P}(dy)F(x,y)\cdot\frac{\partial}{\partial x}\left(F(x,z)\cdot\frac{\partial f(x)}{\partial x}\right)\\&+\int\bar{P}(dy)G(x,y)\cdot\frac{\partial f(x)}{\partial x}.\end{aligned} \tag{3.22}$$

We assume now that the diffusion equation

$$\begin{aligned}&\frac{\partial u(t,x)}{\partial t}=\mathscr{L}u(t,x), \qquad t>0,\\&u(0,x)=f(x)\end{aligned} \tag{3.23}$$

has a unique smooth and bounded solution in $0\leqq t\leqq T<\infty$, whenever $f(x)$ is smooth and bounded. This is in fact true on account of the smoothness of the coefficients of the operator $\mathscr{L}$ which follows from the smoothness of F and G.

Under the above hypotheses we have the following.

THEOREM. *For $f(x)$ smooth and bounded let $u^\varepsilon(t,x,y)$ be the solution of* (3.14) *with $u^\varepsilon(0,x,y)=f(x)$, i.e.,*

$$u^\varepsilon(t,x,y)=E_{x,y}\{f(x^\varepsilon(t))\}. \tag{3.24}$$

Let $u(t,x)$ be the solution of (3.23) *with the same data f. Then*

$$|u^\varepsilon(t,x,y)-u(t,x)|=O(\varepsilon), \qquad 0\leqq t\leqq T<\infty, \tag{3.25}$$

uniformly in $x\in R^n$, $y\in S$.

Proof. The idea is to construct an expansion

$$u^\varepsilon=u_0+\varepsilon u_1+\varepsilon^2u_2+\cdots+\varepsilon^N u_N+R_N, \tag{3.26}$$

[6] Otherwise we have an initial layer which is handled easily.

where R_N is the remainder, and show that

$$|u^\varepsilon - u_0 - \varepsilon u_1 - \cdots - \varepsilon^N u_N| = O(\varepsilon^{N-1}), \qquad N = 0, 1, 2, \cdots. \tag{3.27}$$

Therefore, for example,

$$|u^\varepsilon - u_0| \leqq |u^\varepsilon - u_0 - \varepsilon u_1 - \varepsilon^2 u_2| + |\varepsilon u_1 + \varepsilon^2 u_2| = O(\varepsilon) + O(\varepsilon), \tag{3.28}$$

provided u_1 and u_2 are bounded. Naturally u_0 is identified with $u(t, x)$ of (3.23).

Now let u^ε satisfy (3.14), i.e., in view of (3.16),

$$\left(\frac{1}{\varepsilon^2}\mathcal{L}_1 + \frac{1}{\varepsilon}\mathcal{L}_2 + \mathcal{L}_3 - \frac{\partial}{\partial t}\right)u^\varepsilon = 0. \tag{3.29}$$

If we insert (3.26) into (3.29) and equate coefficients of equal powers of ε we obtain the following sequence of problems:

$$\mathcal{L}_1 u_0 = 0, \tag{3.30}$$

$$\mathcal{L}_1 u_1 = -\mathcal{L}_2 u_0, \tag{3.31}$$

$$\mathcal{L}_1 u_2 = -\mathcal{L}_2 u_1 - \left(\mathcal{L}_3 - \frac{\partial}{\partial t}\right)u_0, \cdots. \tag{3.32}$$

Recalling that $\mathcal{L}_1 = Q$ by (3.17), we conclude from (3.30) and the ergodic properties of Q that $u_0 = u_0(t, x)$ and does not depend on y. It will be identified with $u(t, x)$ of (3.23) later. Because of (3.18) and the Fredholm alternative, (3.31) has a bounded solution (up to a constant)

$$u_1 = -\mathcal{L}_1^{-1}\mathcal{L}_2 u_0. \tag{3.33}$$

Using this in (3.32) and applying the solvability condition (3.9) to the right side, we find that $u_0(t, x)$ must satisfy (3.23). Since $u^\varepsilon(0, x, y) = f(x)$, it follows that the initial data for u_0 is f and hence $u_0 = u$ of (3.23). Solving (3.32) now yields

$$u_2 = -\mathcal{L}_1^{-1}(\mathcal{L}_2 u_1 - (\mathcal{L}_3 - \mathcal{L})u), \tag{3.34}$$

and both u_2 and u_1 above are well defined smooth and bounded.

A direct computation and the above definitions (3.33) and (3.34) produce the following:

$$\begin{aligned} &\left(\mathcal{L}^\varepsilon - \frac{\partial}{\partial t}\right)(u^\varepsilon - u - \varepsilon u_1 - \varepsilon^2 u_2) \\ &\quad = \left(\frac{1}{\varepsilon^2}\mathcal{L}_1 + \frac{1}{\varepsilon}\mathcal{L}_2 + \mathcal{L}_3 - \frac{\partial}{\partial t}\right)(u^\varepsilon - u - \varepsilon u_1 - \varepsilon^2 u_2) \\ &\quad = -\left(\frac{1}{\varepsilon^2}\mathcal{L}_1 + \frac{1}{\varepsilon}\mathcal{L}_2 + \mathcal{L}_3 - \frac{\partial}{\partial t}\right)(u + \varepsilon u_1 + \varepsilon^2 u_2) \\ &\quad = -\varepsilon\left[\left(\mathcal{L}_3 - \frac{\partial}{\partial t}\right)u_1 + \mathcal{L}_2 u_2\right] - \varepsilon^2\left(\mathcal{L}_3 - \frac{\partial}{\partial t}\right)u_2. \end{aligned} \tag{3.35}$$

Therefore, by the maximum principle for $\mathcal{L}^3 - \partial/\partial t$ and the regularity, we obtain (3.27) with $N = 2$ and so (3.28). The proof is complete.

Let us note that the choice of u_1 and u_2 in (3.33) and (3.34) implies that (3.31) and (3.32) hold and hence the $O(1/\varepsilon)$ and $O(1)$ terms in (3.35) cancel. This is the essential point of the perturbation expansion and will be used repeatedly in the following sections.

4. Weak convergence. The result of § 3 shows that under suitable hypotheses, mostly smoothness and the ergodicity of $y(t)$,

$$E_{x,y}\{f(x^\varepsilon(t))\}\to E_x\{f(x(t))\}, \qquad 0\leqq t\leqq T<\infty, \tag{4.1}$$

as $\varepsilon\to 0$, with T arbitrary, where $x(t)$ is the diffusion Markov process associated with the operator $\mathcal{L}$ of (3.21). In order to conclude that the process $x^\varepsilon(t)$ converges weakly to $x(t)$, as a measure on $C([0, T]; R^n)$, $T<\infty$ but arbitrary, it is necessary to do a bit more. We also need information, useful in later sections, concerning moments of $x^\varepsilon(t)$. This is what is done in this section.

First we replace the boundedness assumptions on $F(x, y)$ and $G(x, y)$ by the following:

(i) $F(x, y)$ and $G(X, y)$ are functions on $R^n\times S\to R^n$, smooth in x and continuous in y;

(ii) There is a constant K such that for all $x\in R^n$ and $y\in S$

$$|F_i(x, y)|\leqq K(1+|x|), \qquad |G_i(x, y)|\leqq K(1+|x|) \tag{4.2}$$

$$\left|\frac{\partial F_i}{\partial x_j}(x, y)\right|\leqq K, \qquad \left|\frac{\partial G_i(x, y)}{\partial x_j}\right|\leqq K, \qquad i, j=1, 2, \cdots, n.$$

(iii) Higher order x-derivatives of F and G do not grow faster than powers of $|x|$ as $|x|\to\infty$ uniformly in $y\in S$.

We note that the assumptions (3.3) and (3.6) imply the exponential estimate

$$\sup_{y\in S}\sup_{A\subset S}|P(t, y, A)-\bar{P}(A)|\leqq e^{-\alpha t}, \qquad t\geqq 0, \tag{4.3}$$

for some $\alpha>0$.

LEMMA. *Under the above hypotheses, the solution $x^\varepsilon(t)$ of* (2.13) *satisfies*[7]

$$E_{x,y}\{|x^\varepsilon(t)|^p\}\leqq C_p(1+|x|^p), \qquad 0\leqq t\leqq T<\infty, \tag{4.4}$$

where $p\geqq 1$ is an integer and C_p is a constant depending on p and $T<\infty$ but not on $\varepsilon>0$.

Proof. It is enough to prove (4.4) with $G\equiv 0$ in (2.13) and with p replaced by $2p$.

We have[8]

$$|x^\varepsilon(t)|^{2p}=|x|^{2p}+\frac{2p}{\varepsilon}\int_0^t F_i(x^\varepsilon(s), y^\varepsilon(s))x_i(s)|x^\varepsilon(s)|^{2p-2}\,ds. \tag{4.5}$$

[7] $|x|^2=\sum_{i=1}^n x_i^2$.

[8] We use the summation convention.

Iterating this identity once, we obtain the identity

$$
\begin{aligned}
|x^\varepsilon(t)|^{2p} &= |x|^{2p} + \frac{2p}{\varepsilon}\int_0^t F_i(x, y^\varepsilon(s))x_i|x|^{2p-2}\,ds \\
&\quad + \frac{2p}{\varepsilon^2}\int_0^t \int_0^s F_i(x^\varepsilon(\gamma), y^\varepsilon(\gamma))\Big[x_j^\varepsilon(\gamma)\frac{\partial F_j(x^\varepsilon(\gamma), y^\varepsilon(s))}{\partial x_i}|x^\varepsilon(\gamma)|^{2p-2} \\
&\qquad + F_i(x^\varepsilon(\gamma), y^\varepsilon(s))|x^\varepsilon(\gamma)|^{2p-2} \\
&\qquad + (2p-2)x_i^\varepsilon(\gamma)F_j(x^\varepsilon(\gamma), y^\varepsilon(s))x_j^\varepsilon(\gamma)|x^\varepsilon(\gamma)|^{2p-4}\Big]\,d\gamma\,ds.
\end{aligned}
\tag{4.6}
$$

We now take expectation and, letting $\mathscr{F}_t$, $t \geqq 0$, stand for the σ-algebras generated by $(x^\varepsilon(s), y^\varepsilon(s))$ for $s \leqq t$, we obtain

$$
\begin{aligned}
E_{x,y}\{|x^\varepsilon(t)|^{2p}\} &= |x|^{2p} + \frac{2p}{\varepsilon}\int_0^t E_{x,y}\{F_i(x, y^\varepsilon(s))x_i|x|^{2p-2}\}\,ds \\
&\quad + \frac{2p}{\varepsilon^2}\int_0^t \int_0^s E_{x,y}\Big\{E\Big\{F_i(x^\varepsilon(\gamma), y^\varepsilon(\gamma))\Big[x_j^\varepsilon(\gamma)\frac{F_j(x^\varepsilon(\gamma), y^\varepsilon(s))}{\partial x_i} \\
&\qquad \cdot |x^\varepsilon(\gamma)|^{2p-2} + F_i(x^\varepsilon(\gamma), y^\varepsilon(s))|x^\varepsilon(\gamma)|^{2p-2} \\
&\qquad + (2p-2)x_i^\varepsilon(\gamma)F_j(x^\varepsilon(\gamma), y^\varepsilon(s))x_j^\varepsilon(\gamma)|x^\varepsilon(\gamma)|^{2p-4}\Big]\Big|\mathscr{F}_\gamma\Big\}\Big\}\,d\gamma\,ds.
\end{aligned}
\tag{4.7}
$$

Now we use (4.3) (recall that $y^\varepsilon(t) \equiv y(t/\varepsilon^2)$), the centering hypothesis (3.18) and (4.2) in (4.7) which yields

$$
\begin{aligned}
E_{x,y}\{|x^\varepsilon(t)|^{2p}\} &\leqq |x|^{2p} + \frac{4\varepsilon pK}{\alpha}(1+|x|^{2p}) \\
&\quad + \tilde{c}_p\frac{2pK^2}{\varepsilon^2}\int_0^t \int_0^s e^{-\alpha(s-\gamma)/\varepsilon^2}E_{x,y}\{1+|x^\varepsilon(\gamma)|^{2p}\}\,d\gamma\,ds,
\end{aligned}
\tag{4.8}
$$

where $\tilde{c}_p$ is a constant. Thus,

$$
\begin{aligned}
E_{x,y}\{|x^\varepsilon(t)|^{2p}\} &\leqq |x|^{2p} + \varepsilon\frac{4pK}{\alpha}(1+|x|^{2p}) \\
&\quad + \tilde{c}_p\frac{2pK^2}{\alpha}\int_0^t E_{x,y}\{1+|x^\varepsilon(\gamma)|^{2p}\}\,d\gamma.
\end{aligned}
\tag{4.9}
$$

Inequality (4.9) and Gronwall's lemma yield (4.4). The proof of the lemma is complete.

We consider next the diffusion operator $\mathscr{L}$ of (3.21) and introduce a set of hypotheses corresponding to (4.2). First we write $\mathscr{L}$ in the form ($f \in C^2$)

$$
\mathscr{L}f(x) = \frac{1}{2}\sum_{i,j=1}^n a_{ij}(x)\frac{\partial^2 f(x)}{\partial x_i \partial x_j} + \sum_{j=1}^n b_j(x)\frac{\partial f(x)}{\partial x_j},
\tag{4.10}
$$

where, from (3.19)–(3.21), we have

$$(4.11)\quad \begin{aligned} a_{ij}(x) &= \text{symmetric part of } \left\{2\int\int \chi(y, dz)\bar{P}(dy)F_i(x, y)F_j(x, z)\right\} \\ &= \int_{-\infty}^{\infty} E\{F_j(x, y(t))F_i(x, y, (0))\}\, dt \\ &= \lim_{T\uparrow\infty} \frac{1}{T}\int_0^T \int_0^T E\{F_j(x, y(t))F_i(x, y(s))\}\, dt\, ds, \end{aligned}$$

$$(4.12)\quad b_j(x) = \int\int \chi(y, dz)\bar{P}(dy) \sum_{k=1}^{n} F_k(x, y)\frac{\partial F_j(x, z)}{\partial x_k} + \int \bar{P}(dy)G_j(x, y), \qquad i, j = 1, 2, \cdots, n.$$

Here $E\{\cdot\}$, without subscripts, denotes expectation relative to the stationary process $y(t)$, i.e., where $y(0)$ is distributed according to $\bar{P}$. The kernel χ is defined by (3.11). Note that the last expression in (4.11) displays the symmetry and nonnegative definiteness of $(a_{ij}(x)) = a(x)$.

In the notation (2.6) and (2.7), the matrix $a(x)$ has the form

$$(4.13)\quad a(x) = \int_{-\infty}^{\infty} R(x, s)\, ds = S(x, 0),$$

i.e., it is the power spectrum of F at zero frequency. Frequencies other than zero enter into the definition of $a(x)$ (and also of $b(x)$) when (2.13) has rapidly oscillating terms, i.e.,

$$x^{\varepsilon}(t) = e^{-At/\varepsilon^2}\hat{x}^{\varepsilon}(t)$$

where

$$\frac{d\hat{x}^{\varepsilon}(t)}{dt} = \frac{1}{\varepsilon^2}A\hat{x}^{\varepsilon}(t) + \frac{1}{\varepsilon}F(x^{\varepsilon}(t), y^{\varepsilon}(t)) + G(x^{\varepsilon}(t), y^{\varepsilon}(t)), \qquad t > 0,$$

$$\hat{x}^{\varepsilon}(0) = x,$$

and A has only imaginary eigenvalues. The corresponding problem (2.15) has an $O(1)$ linear term on the right hand side. We shall deal only with the case $A \equiv 0$ here and refer to § 5.3 for an example and to [3], [7], [18] for additional information on this point.

We assume that $(a_{ij}(x))$ and $(b_j(x))$ satisfy the following conditions

(i) They are smooth functions of x.

(ii) There is a constant K such that

$$(4.14)\quad \begin{aligned} &|a_{ij}(x)| \leqq K^2(1 + |x|^2), \qquad |b_j(x)| \leqq K(1 + |x|), \\ &\left|\frac{a_{ij}(x)}{\partial x_k}\right| \leqq K(1 + |x|), \qquad \left|\frac{\partial b_j(x)}{\partial x_k}\right| \leqq K, \\ &\left|\frac{a_{ij}(x)}{\partial x_k \partial x_l}\right| \leqq K, \qquad i, j, k, l = 1, 2, \cdots, n. \end{aligned}$$

(iii) Higher order derivatives of a_{ij} and b_j are dominated by powers of $|x|$ for $|x|$ large.

Under these conditions the diffusion equation

$$(4.15)\qquad \frac{\partial u(t,x)}{\partial t}=\mathcal{L}u(t,x), \quad t>0, \qquad u(0,x)=f(x),$$

has a unique smooth solution if $f(x)$ is smooth and $|f(x)|\leqq C(1+|x|^p)$, for some integer p. In fact we also have $|u(t,x)|\leqq C'(1+|x|^{\bar{p}})$, $0\leqq t\leqq T<\infty$, for T arbitrary and for some constant C' and integer $\bar{p}$. The process corresponding to $\mathcal{L}$ is denoted by $x(t)$ and it is a sample continuous strong Markov process (the corresponding semigroup is $\hat{C}$ (cf. (3.15)) and Feller).

THEOREM. *Under the above hypotheses the process $x^\varepsilon(t)$, $t\geqq 0$ defined by* (2.13) *converges weakly as $\varepsilon\to 0$ to the diffusion Markov process $x(t)$ generated by $\mathcal{L}$ of* (4.10)–(4.12). *Moreover, moments of $x^\varepsilon(t)$ converge to moments of $x(t)$ on any finite t interval.*

Proof. With the moment estimate (4.4) on hand, the theorem of § 3 extends easily to the present situation and we have that for f smooth such that $|f(x)|\leqq C(1+|x|^p)$, there is an integer $\bar{p}$ such that

$$(4.16)\qquad \lim_{\varepsilon\downarrow 0}\sup_{0\leqq t\leqq T}\sup_{x,y}\frac{|E_{x,y}\{f(x^\varepsilon(t))\}-u(t,x)|}{1+|x|^{\bar{p}}}=0,$$

where $T<\infty$ is arbitrary and $u(t,x)$ is the solution of (4.15).

It remains therefore to show that the processes $x^\varepsilon(t)$, with $0<\varepsilon\leqq 1$, say, are relatively weakly compact. Since $x^\varepsilon(t)$ and $x(t)$ are processes on $C([0,T];R^n)$, $T<\infty$, it suffices to show that they are relatively weakly compact in $D([0,T];R^n)$ (cf. [21, p. 150], [22]), i.e. to show that for some constants $\gamma>0$ and $\beta>0$

$$(4.17)\qquad E_{x,y}\{|x^\varepsilon(t)-x^\varepsilon(\sigma)|^\gamma|x^\varepsilon(\sigma)-x^\varepsilon(s)|^\gamma\}\leqq C(t-s)^{1+\beta}, \qquad 0\leqq s\leqq\sigma\leqq t\leqq T<\infty,$$

where C is a constant independent of s, σ, t, ε, x and y provided x, the starting point of $x^\varepsilon(t)$, ranges over a compact set in R^n. In addition to (4.17) we must show

$$(4.17')\qquad \lim_{N\uparrow\infty}\overline{\lim_{\varepsilon\downarrow 0}}P_{x,y}\Big\{\sup_{0\leqq t\leqq T}|x^\varepsilon(t)|>N\Big\}=0, \qquad T<\infty.$$

To prove (4.17) and (4.17′) we proceed roughly as in § 3. The preliminary considerations that follow are given in detail because they are relevant to later sections. For the proof of the present theorem it suffices to have f_1 in (4.18); f_2 is superfluous here, but is used in later sections.

Let $f(x)$ be a smooth function such that $|f(x)|\leqq C(1+|x|^p)$ and define $f_1(x,y)$, $f_2(x,y)$ by

$$(4.18)\qquad f_1(x,y)=\int_S \chi(y,dz)F(x,z)\cdot\frac{\partial f(x)}{\partial x},$$

$$(4.19)\qquad f_2(x,y)=\int_S \chi(y,dz)\Big[F(x,z)\cdot\frac{\partial f_1(x,z)}{\partial x}+G(x,z)\cdot\frac{\partial f(x)}{\partial x}-\mathcal{L}f(x)\Big],$$

so that

$$Qf_1 + F \cdot \frac{\partial f}{\partial x} = 0, \tag{4.18'}$$

$$Qf_2 + F \cdot \frac{\partial f_1}{\partial x} + G \cdot \frac{\partial f}{\partial x} - \mathcal{L}f = 0. \tag{4.19'}$$

Here Q is defined by (3.2) and χ by (3.11). Define

$$f^\varepsilon(x, y) = f(x) + \varepsilon f_1(x, y) + \varepsilon^2 f_2(x, y). \tag{4.20}$$

With $\mathcal{L}^\varepsilon$ defined by (3.12), it follows from (4.18′) and (4.19′) that

$$\mathcal{L}^\varepsilon f^\varepsilon(x, y) = \mathcal{L}f(x) + \varepsilon\left(F(x, y) \cdot \frac{\partial f_2(x, y)}{\partial x} + G(x, y) \cdot \frac{\partial f_1(x, y)}{\partial x}\right) + \varepsilon^2 G(x, y) \cdot \frac{\partial f_2(x, y)}{\partial x}. \tag{4.21}$$

Define M_{f^ε} by

$$M_{f^\varepsilon}(t) = f^\varepsilon(x^\varepsilon(t), y^\varepsilon(t)) - f^\varepsilon(x, y) - \int_0^t \mathcal{L}^\varepsilon f^\varepsilon(x^\varepsilon(s), y^\varepsilon(s))\, ds. \tag{4.22}$$

M_{f^ε} is a zero-mean, integrable (because of (4.4)), right-continuous martingale relative to $\mathcal{F}_t$, the σ-algebras associated with the paths $(x^\varepsilon(\cdot), y^\varepsilon(\cdot))$ up to time t. The increasing process $\langle M_{f^\varepsilon}, M_{f^\varepsilon}\rangle$ corresponding to M_{f^ε} is given by[9]

$$\begin{aligned}\langle M_{f^\varepsilon}(t)\rangle &= \int_0^t [\mathcal{L}^\varepsilon(f^\varepsilon)^2 - 2f^\varepsilon \mathcal{L}^\varepsilon f^\varepsilon](x^\varepsilon(s), y^\varepsilon(s))\, ds \\ &= \int_0^t [Q(f_1 + \varepsilon f_2)^2 - 2(f_1 + \varepsilon f_2)Q(f_1 + \varepsilon f_2)](x^\varepsilon(s), y^\varepsilon(s))\, ds \\ &\equiv \int_0^t H_{f^\varepsilon}(x^\varepsilon(s), y^\varepsilon(s))\, ds.\end{aligned} \tag{4.23}$$

By definition, $\langle M_{f^\varepsilon}\rangle$ satisfies

$$E\{(M_{f^\varepsilon}(t) - M_{f^\varepsilon}(s))^2 | \mathcal{F}_s\} = E\{\langle M_{f^\varepsilon}(t)\rangle - \langle M_{f^\varepsilon}(s)\rangle | \mathcal{F}_s\}, \tag{4.24}$$

for $0 \leqq s \leqq t \leqq T < \infty$. It can be verified by direct computation that (4.22) and (4.23) obey (4.24).

If we let g_1 and g_2 be the integrands multiplying χ in (4.18) and (4.19) respectively, we obtain the following expression for H_{f^ε} in (4.23):

$$H_{f^\varepsilon}(x, y) = q(y) \int_S \pi(y, dz) \left[\int_S (\chi(z, d\zeta) - \chi(y, d\zeta))(g_1 + \varepsilon g_2)(x, y)\right]^2. \tag{4.25}$$

Thus, $H_{f^\varepsilon}(x, y)$ is a smooth function of x and grows like a power of $|x|$ for x large, uniformly in y.

[9] For simplicity we write $\langle M_{f^\varepsilon}\rangle$ instead of $\langle M_{f^\varepsilon}, M_{f^\varepsilon}\rangle$.

From (4.20) and (4.21) it follows that the identity (4.22) can be written in the form

$$
\begin{aligned}
&f(x^\varepsilon(t))-f(x)-\int_0^t \mathscr{L}f(x^\varepsilon(s))\,ds \\
&\quad = M_{f^\varepsilon}(t)+\varepsilon f_1(x,y)+\varepsilon^2 f_2(x,y)-\varepsilon f_1(x^\varepsilon(t),y^\varepsilon(t))-\varepsilon^2 f_2(x^\varepsilon(t),y^\varepsilon(t)) \\
&\qquad +\int_0^t \left[\varepsilon\left(F\cdot\frac{\partial f_2}{\partial x}+G\cdot\frac{\partial f_1}{\partial x}\right)+\varepsilon^2 G\frac{\partial f_2}{\partial x}\right](x^\varepsilon(s),y^\varepsilon(s))\,ds.
\end{aligned}
\tag{4.26}
$$

This form displays clearly why indeed $x^\varepsilon(t)$ is well approximated by $x(t)$ generated by $\mathscr{L}$.

Let us return to estimates (4.17) and (4.17′). We choose $f(x)=x_i$, the coordinate functions, for $i=1,2,\cdots,n$ successively. Let the corresponding f_1 and f_2 of (4.18) and (4.19) be denoted by f_{1i} and f_{2i} respectively. Let the corresponding martingale M_{f^ε} be denoted by $M_i^\varepsilon(t)$. With the use of (4.12) the identity (4.26) becomes

$$
\begin{aligned}
x_i^\varepsilon(t)&=x_i+\int_0^t b_i(x^\varepsilon(s))\,ds+M_i^\varepsilon(t)+\varepsilon f_{1i}(x,y)+\varepsilon^2 f_{2i}(x,y) \\
&\quad -\varepsilon f_{1i}(x^\varepsilon(t),y^\varepsilon(t))-\varepsilon^2 f_{2i}(x^\varepsilon(t),y^\varepsilon(t)) \\
&\quad +\int_0^t\left[\varepsilon F\cdot\frac{\partial f_{2i}}{\partial x}+\varepsilon G\cdot\frac{\partial f_{1i}}{\partial x}+\varepsilon^2 G\cdot\frac{\partial f_{2i}}{\partial x}\right](x^\varepsilon(s),y^\varepsilon(s))\,ds, \\
&\qquad\qquad i=1,2,\cdots,n.
\end{aligned}
\tag{4.27}
$$

From (4.14), (4.4) and Kolmogorov's inequality for the martingale $M_i^\varepsilon(t)$, estimate (4.17′) follows. Now define

$$
\tilde{x}_i^\varepsilon(t)=x_i+\int_0^t b_i(x^\varepsilon(s))\,ds+M_i^\varepsilon(t),\qquad i=1,2,\cdots,n. \tag{4.28}
$$

From (4.4), (4.27) and Kolmogorov's inequality, it follows that for x in a compact set and $\varepsilon\leqq 1$, say,

$$
\overline{\lim_{\varepsilon\downarrow 0}}\, P_{x,y}\left\{\sup_{0\leqq t\leqq T}|x^\varepsilon(t)-\tilde{x}^\varepsilon(t)|>\delta\right\}=0,\qquad \forall\delta>0. \tag{4.29}
$$

Thus, it suffices to show that (4.17) holds for $(\tilde{x}_i^\varepsilon(t))$.

Let us first estimate conditional expectations. We have, in vector notation,[10]

$$
\begin{aligned}
E\{|\tilde{x}^\varepsilon(t)-\tilde{x}^\varepsilon(s)|^2|\mathscr{F}_s\}&=E\left\{\left|\int_s^t b(x^\varepsilon(\sigma))\,d\sigma+M^\varepsilon(t)-M^\varepsilon(s)\right|^2\Big|\mathscr{F}_s\right\} \\
&\leqq 2E\left\{\left|\int_s^t b(x^\varepsilon(\sigma))\,d\sigma\right|^2\Big|\mathscr{F}_s\right\}+2E\{|M^\varepsilon(t)-M^\varepsilon(s)|^2|\mathscr{F}_s\} \\
&\leqq C(t-s)(1+|x^\varepsilon(s)|^2),\qquad C\ \text{a constant}.
\end{aligned}
\tag{4.30}
$$

[10] $|x|^2=\sum_{i=1}^n x_i^2$.

The last step in (4.30) is elementary for the deterministic integral. For the martingale we have

$$E\{|M^\varepsilon(t)-M^\varepsilon(s)|^2|\mathcal{F}_s\}=\sum_{i=1}^n E\{(M_i^\varepsilon(t)-M_i^\varepsilon(s))^2|\mathcal{F}_s\}$$

$$=\sum_{i=1}^n E\{\langle M_i^\varepsilon(t)\rangle-\langle M_i^\varepsilon(s)\rangle|\mathcal{F}_s\}$$

$$=\sum_{i=1}^n E\left\{\int_s^t H_i^\varepsilon(x^\varepsilon(\sigma),y^\varepsilon(\sigma))\,d\sigma|\mathcal{F}_s\right\}$$

$$\leqq \tilde{c}(t-s)(1+|x^\varepsilon(s)|^2).$$

The last step here follows from (4.4) and the bound for H_i^ε that follows from (4.25).

To prove (4.17) we use (4.30) and the moment estimate (4.4) as follows:

$$\begin{aligned}
&E_{x,y}\{|\tilde{x}^\varepsilon(t)-\tilde{x}^\varepsilon(\sigma)|^\gamma|\tilde{x}^\varepsilon(\sigma)-\tilde{x}^\varepsilon(s)|^\gamma\}\\
&\quad=E_{x,y}\{E\{|\tilde{x}^\varepsilon(t)-\tilde{x}^\varepsilon(\sigma)|^\gamma|\mathcal{F}_\sigma\}|\tilde{x}^\varepsilon(\sigma)-\tilde{x}^\varepsilon(s)|^\gamma\}\\
&\quad\leqq E_{x,y}\{E^{\gamma/2}\{|\tilde{x}^\varepsilon(t)-\tilde{x}^\varepsilon(\sigma)|^2|\mathcal{F}_\sigma\}|\tilde{x}^\varepsilon(\sigma)-\tilde{x}^\varepsilon(s)|^\gamma\}\\
&\quad\leqq C(t-\sigma)^{\gamma/2}E_{x,y}\{(1+|\tilde{x}^\varepsilon(\sigma)|^2)^{\gamma/2}|\tilde{x}^\varepsilon(\sigma)-\tilde{x}^\varepsilon(s)|^\gamma\}\\
&\quad=C(t-\sigma)^{\gamma/2}E_{x,y}\{E\{(1+|\tilde{x}^\varepsilon(\sigma)|^2)^{\gamma/2}|\tilde{x}^\varepsilon(\sigma)-\tilde{x}^\varepsilon(s)|^\gamma|\mathcal{F}_s\}\}\\
&\quad\leqq C(t-\sigma)^{\gamma/2}E_{x,y}\{E^{(2-\gamma)/2}\{(1+|\tilde{x}^\varepsilon(\sigma)|^2)^{\gamma/(2-\gamma)}\}\\
&\qquad\qquad\cdot E^{\gamma/2}\{|\tilde{x}^\varepsilon(\sigma)-\tilde{x}^\varepsilon(s)|^2|\mathcal{F}_s\}\}\\
&\quad\leqq C^2(t-\sigma)^{\gamma/2}(\sigma-s)^{\gamma/2}E_{x,y}\{E^{(2-\gamma)/2}\{(1+|\tilde{x}^\varepsilon(\sigma)|^2)^{2/(2-\gamma)}|\mathcal{F}_s\}\\
&\qquad\qquad\cdot(1+|\tilde{x}^\varepsilon(s)|^2)^{\gamma/2}\}.
\end{aligned}$$

Choosing $\gamma=\frac{3}{2}$ and applying (4.4) we obtain (4.17) with $\beta=\frac{1}{2}$. The proof of the theorem is complete.

We note that we shall have occasion in later sections to refer to (4.18)–(4.26). This is why the theorem is proved here in detail since it is not substantially different from the one in [3], for example.

5. Stochastic stability based on the wide-band noise approximation.

5.1. A Lyapunov theorem. Consider system (2.13) which defines $(x^\varepsilon(t), y^\varepsilon(t))$ (or (2.15) for $(\tilde{x}^\varepsilon(\tau), y(\tau))$) under the hypotheses of §4. As we have shown, $x^\varepsilon(t)$ converges weakly to the diffusion process $x(t)$ generated by $\mathcal{L}$ of (4.10)–(4.12). We shall examine properties of $x^\varepsilon(t)$ *with* $0<\varepsilon\leqq\varepsilon_0$, ε *fixed and* ε_0 *sufficiently small, as* $t\to\infty$. The objective is to establish results about $x^\varepsilon(t)$ which are based only on conditions upon the approximating diffusion $x(t)$. The first theorem is Lyapunov-like result for a class of Markov processes.

THEOREM. *Let $V(x)$ be a smooth function on R^n such that $V(x)\to+\infty$ as $|x|\to\infty$, it behaves like a polynomial in x for x large and it is positive definite*:

$$V(x)\geqq 0,\quad V(x)=0\Rightarrow x=0. \tag{5.1}$$

Suppose that the vector fields $F(x, y)$ and $G(x, y)$ satisfy, in addition to smoothness and (4.2),

$$F_i(0, y)=0, \quad G_i(0, y)=0, \quad i=1,2,\cdots,n, \quad y\in S, \tag{5.2}$$

so that, by (4.11), (4.12),

$$a_{ij}(x)=O(|x|^2), \qquad b_j(x)=O(|x|), \qquad |x|\downarrow 0. \tag{5.3}$$

Suppose that $V(x)$ satisfies

$$\mathscr{L}V(x)\leqq -\gamma V(x), \qquad \gamma>0, \quad x\in R^n. \tag{5.4}$$

Then for $0<\varepsilon\leqq\varepsilon_0$, ε fixed and ε_0 sufficiently small, $x^\varepsilon(t)$ is uniformly stochastically asymptotically stable as $t\to\infty$, i.e., for any $\eta_1>0$ and $\eta_2>0$, there is a $\delta>0$ such that if $|x(0)|=|x|<\delta$, then

$$P_{x,y}\{|x^\varepsilon(t)|\leqq \eta_2 e^{-\tilde{\gamma}t}, t\geqq 0\}\geqq 1-\eta_1, \tag{5.5}$$

for all $y\in S$, with $\tilde{\gamma}>0$ a constant. Furthermore,

$$P_{x,y}\left\{\lim_{t\uparrow\infty}|x^\varepsilon(t)|=0\right\}=1. \tag{5.6}$$

Remark 1. The global requirements on $V(x)$ are necessary only for (5.6) and not for the local result (5.5).

Remark 2. In the linear case (F and G linear in x), one can get very sharp results by deriving closed equations for moments (in the limit) and using well chosen quadratic functions for $V(x)$.

Proof. We refer repeatedly to the constructions (4.18)–(4.26). The function $f(x)$ in these constructions will now be the Lyapunov function $V(x)$.

Let $V_1(x, y)$ and $V_2(x, y)$ be the functions corresponding to (4.18) and (4.19) with $f=V$. Define V^ε as in (4.20) so that (4.21) holds.

We note that because of (4.2), (5.2) and (4.14), (5.3), V_1 and V_2 behave like $V(x)$ for small $|x|$ and large $|x|$ uniformly in $y\in S$. Thus, there is an ε_0 such that for $0<\varepsilon\leqq\varepsilon_0$

$$\tilde{c}_1 V(x)\leqq V^\varepsilon(x, y)\leqq \tilde{c}_2 V(x), \tag{5.7}$$

for some positive constants $\tilde{c}_1$ and $\tilde{c}_2$.

Let $\hat{\gamma}>0$ be a constant. From (4.21) we have

$$\begin{aligned}(\mathscr{L}^\varepsilon+\hat{\gamma})V^\varepsilon(x, y) &= \hat{\gamma}V^\varepsilon(x, y)+\mathscr{L}V(x)\\ &\quad+\varepsilon\left[(x, y)\cdot\frac{\partial V_2(x, y)}{\partial x}+G(x, y)\cdot\frac{\partial V_1(x, y)}{\partial x}\right]\\ &\quad+\varepsilon^2 G(x, y)\cdot\frac{\partial V_2(x, y)}{\partial x}\\ &\leqq(\mathscr{L}+\tilde{c}_2\hat{\gamma}+\varepsilon_0\tilde{c}_3)V(x),\end{aligned} \tag{5.8}$$

for some constant $\tilde{c}_3$. If we choose $\hat{\gamma}$ so that $\tilde{c}_2\hat{\gamma}+\varepsilon_0\tilde{c}_3\leqq\gamma$ we find that

$$(\mathscr{L}^\varepsilon+\hat{\gamma})V^\varepsilon(x, y)\leqq 0. \tag{5.9}$$

Now we write (4.22) with $f^\varepsilon = e^{\hat{\gamma}t}V^\varepsilon$:

$$e^{\hat{\gamma}t}V^\varepsilon(x^\varepsilon(t), y^\varepsilon(t)) = V^\varepsilon(x, y) + \int_0^t e^{s\hat{\gamma}}(\mathscr{L}^\varepsilon + \hat{\gamma})V^\varepsilon(x^\varepsilon(s), y^\varepsilon(s))\, ds + \tilde{M}_{V^\varepsilon}(t). \tag{5.10}$$

Here $\tilde{M}_{V^\varepsilon}(t)$ is a zero-mean, integrable, right-continuous martingale relative to $\mathscr{F}_t$. The increasing process associated with $\tilde{M}_{V^\varepsilon}$ has an extra factor $e^{2\hat{\gamma}t}$ multiplying H_{V^ε} in (4.23).

Using (5.7) and (5.9) in (5.10) we obtain

$$0 \leqq e^{\hat{\gamma}t}\tilde{c}_1 V(x^\varepsilon(t)) \leqq \tilde{c}_2 V(x) + \tilde{M}_{V^\varepsilon}(t), \tag{5.11}$$

and hence $\tilde{c}_2 V(x) + \tilde{M}_{V^\varepsilon}(t)$ is a nonnegative integrable martingale. By Kolmogorov's inequality we have that for each $\tilde{\eta}_2 > 0$

$$\begin{aligned} &P_{x,y}\Big\{\sup_{0\leqq t\leqq T} e^{\hat{\gamma}t}\tilde{c}_1 V(x^\varepsilon(t)) > \tilde{\eta}_2\Big\} \\ &\leqq P_{x,y}\Big\{\sup_{0\leqq t\leqq T} (\tilde{c}_2 V(x) + \tilde{M}_{V^\varepsilon}(t)) > \tilde{\eta}_2\Big\} \leqq \frac{\tilde{c}_2 V(x)}{\tilde{\eta}_2}. \end{aligned} \tag{5.12}$$

Letting $T \uparrow \infty$ in (5.12) yields

$$P_{x,y}\Big\{\sup_{t\geqq 0} e^{\hat{\gamma}t}\tilde{c}_1 V(x^\varepsilon(t)) > \tilde{\eta}_2\Big\} \leqq \frac{\tilde{c}_2 V(x)}{\tilde{\eta}_2}. \tag{5.13}$$

By the positive definiteness and smoothness of $V(x)$, there exist constants $c_1 > 0$ and $c_2 > 0$ and positive integers p_1 and p_2 such that

$$c_1|x|^{p_1} \leqq V(x) \leqq c_2|x|^{p_2}, \tag{5.14}$$

for $|x|$ small, say $|x| \leqq K$. Thus,

$$\Big\{c_1|x^\varepsilon(t)|^{p_1} \leqq e^{-\hat{\gamma}t}\frac{\tilde{\eta}_2}{\tilde{c}_1},\ t \geqq 0\Big\} \supset \Big\{V(x^\varepsilon(t)) \leqq e^{-t}\frac{\tilde{\eta}_2}{\tilde{c}_1},\ t \geqq 0\Big\},$$

and hence

$$P_{x,y}\Big\{|x^\varepsilon(t)| \leqq e^{-\tilde{\gamma}t}\Big(\frac{\tilde{\eta}_2}{\tilde{c}_1 c_1}\Big)^{1/p_1}, t \geqq 0\Big\} \geqq 1 - \frac{\tilde{c}_2 V(x)}{\tilde{\eta}_2}, \qquad \tilde{\gamma} \equiv \frac{\hat{\gamma}}{p_1}. \tag{5.15}$$

Let $\eta_1 > 0$ and $\eta_2 > 0$ be given. Choose $\tilde{\eta}_2$ so small that (5.15) yields

$$P_{x,y}\{|x^\varepsilon(t)| \leqq e^{-\tilde{\gamma}t}\tilde{\eta}_2, t \geqq 0\} \geqq 1 - \frac{\tilde{c}_2 V(x)}{\tilde{\eta}_2},$$

and then choose $\delta > 0$ so that for $|x| < \delta$, $V(x) < \tilde{c}_2^{-1}\eta_2\eta_1$. This proves (5.5).

To prove (5.6), we note that

$$\Big\{\lim_{t\uparrow\infty}|x^\varepsilon(t)| = 0\Big\} = \Big\{\lim_{t\uparrow\infty} V(x^\varepsilon(t)) = 0\Big\} \supset \Big\{\sup_{t\geqq 0} e^{\hat{\gamma}t}\tilde{c}_1 V(x^\varepsilon(t)) \leqq C\Big\}$$

where C is a constant. Thus, from (5.13),

$$P_{x,y}\left\{\lim_{t\uparrow\infty}|x^\varepsilon(t)|=0\right\}\geqq 1-\frac{\tilde{c}_2V(x)}{C}$$

and letting $C\uparrow\infty$ yields (5.6). The proof of the theorem is complete.

5.2. Linear systems. We shall consider the linear system

$$(5.16)\qquad \begin{aligned}&\frac{dx^\varepsilon(t)}{dt}=\frac{1}{\varepsilon}A(y^\varepsilon(t))x^\varepsilon(t)+B(y^\varepsilon(t))x^\varepsilon(t), \qquad t>0,\\ &x^\varepsilon(0)=x,\end{aligned}$$

or, in the scaling of (2.15),

$$(5.16')\qquad \begin{aligned}&\frac{d\tilde{x}^\varepsilon(\tau)}{d\tau}=\varepsilon A(y(\tau))\tilde{x}^\varepsilon(\tau)+\varepsilon^2B(y(\tau))\tilde{x}^\varepsilon(\tau), \qquad \tau>0,\\ &\tilde{x}^\varepsilon(0)=x,\end{aligned}$$

where $y(t)$ is as before (cf. §§ 3 and 4), $y^\varepsilon(t)=y(t/\varepsilon^2)$ and $A(y)$, $B(y)$ are $n\times n$ continuous matrix functions on S such that

$$(5.17)\qquad \int_S A(y)\bar{P}(dy)=0.$$

The theorems of §§ 4 and 5.1 specialize without changes to (5.16) and (5.16′). The limiting diffusion process has the infinitesimal generator $\mathscr{L}$ given by (4.10) where (4.11) and (4.12) are replaced by the following.

$$(5.18)\qquad \begin{aligned}a_{ij}(x)&=\text{symmetric part of}\left\{2\iint \chi(y,dz)\bar{P}(dy)\sum_{k,l=1}^n A_{ik}(y)A_{jl}(x)x_kx_l\right\}\\ &=\int_{-\infty}^{\infty}E\left\{\sum_{k,l=1}^n A_{ik}(y(t))A_{jl}(y(0))\right\}x_kx_l\,dt,\end{aligned}$$

$$(5.19)\qquad b_j(x)=\iint\chi(y,dz)\bar{P}(dy)\sum_{i,k=1}^n A_{ik}(y)A_{jk}(z)x_i+\int\bar{P}(dy)\sum_{k=1}^n B_{jk}(y)x_k, \qquad i,j=1,2,\cdots,n.$$

We prove next the following theorem which extends Khasminskii's result for linear Itô equations [9], [10].

THEOREM. *Let $x^\varepsilon(t)$ be the process defined by* (5.16) *and suppose that there exists a smooth function h on S^{n-1}, the unit sphere in R^n, such that if*

$$(5.20)\qquad f(x)=\log|x|+h(x/|x|),$$

then

$$(5.21)\qquad \mathscr{L}f(x)\leqq q<0, \qquad x\in R^n,$$

where q is a constant. Then, for $0<\varepsilon\leqq\varepsilon_0$, ε *fixed and* ε_0 *sufficiently small,*

$$P_{x,y}\left\{\lim_{t\uparrow\infty}|x^\varepsilon(t)|=0\right\}=1,\qquad x\in R^n,\quad y\in S. \tag{5.22}$$

If instead of (5.21) *we have*

$$\mathcal{L}f(x)\geqq q>0, \tag{5.23}$$

then (5.22) *is replaced by*

$$P_{x,y}\left\{\lim_{t\uparrow\infty}|x^\varepsilon(t)|=\infty\right\}=1,\qquad x\in R^n,\quad y\in S. \tag{5.24}$$

Remarks. 1. The conditions of this theorem are sharper than the ones of the theorem of § 5.1.

2. The device of introducing h is due to Pinsky [10]. Before proving the theorem we elaborate on (5.20) and (5.21); see also [10].

Let us introduce polar coordinates

$$x=\rho\xi,\quad \rho=|x|,\quad \xi=x/|x|, \tag{5.25}$$

and the following notation:

$$\begin{aligned} A^*(y,\xi)\cdot\frac{\partial f(\xi)}{\partial\xi}&=\sum_{i,j,k=1}^n A_{ik}(y)\xi_k(\delta_{kj}-\xi_k\xi_j)\frac{\partial f(\xi)}{\partial\xi_j},\\ B^*(y,\xi)\cdot\frac{\partial f(\xi)}{\partial\xi}&=\sum_{i,j,k=1}^n B_{ik}(y)\xi_k(\delta_{kj}-\xi_k\xi_j)\frac{\partial f(\xi)}{\partial\xi_j},\end{aligned} \tag{5.26}$$

$$\begin{aligned} a^*(y,\xi)&=\sum_{i,k=1}^n A_{ik}(y)\xi_k\xi_i,\\ b^*(y,\xi)&=\sum_{i,k=1}^n B_{ik}(y)\xi_k\xi_i,\end{aligned} \tag{5.27}$$

$$v(\xi)=\int\int\chi(y,dz)\bar{P}(dy)A^*(y,\xi)\cdot\frac{\partial a^*(z,\xi)}{\partial\xi}+\int\bar{P}(dy)b^*(y,\xi). \tag{5.28}$$

The infinitesimal generator $\mathcal{L}$ (cf. (4.10), (5.18) and (5.19)) has the following form in polar coordinates:

$$\begin{aligned}\mathcal{L}f(\rho,\xi)=\int\int\chi(y,dz)\bar{P}(dy)&\left[A^*(y,\xi)\cdot\frac{\partial}{\partial\xi}+a^*(y,\xi)\rho\frac{\partial}{\partial\rho}\right]\\ &\cdot\left[A^*(z,\xi)\cdot\frac{\partial f(\rho,\xi)}{\partial\xi}+a^*(z,\xi)\rho\frac{\partial f(\rho,\xi)}{\partial\rho}\right]\\ &+\int\bar{P}(dy)\left[B^*(y,\xi)\cdot\frac{\partial f(\rho,\xi)}{\partial\xi}+b^*(y,\xi)\rho\frac{\partial f(\rho,\xi)}{\partial\rho}\right].\end{aligned} \tag{5.29}$$

Therefore,

$$\mathscr{L}(\log \rho + h(\xi)) = v(\xi) + \mathscr{L}_\xi h(\xi), \tag{5.30}$$

where $\mathscr{L}_\xi$ is the infinitesimal generator of the angular process (on S^{n-1}) and

$$\begin{aligned}\mathscr{L}_\xi h(\xi) = \iint \chi(y, dz)\bar{P}(dy)A^*(y, \xi) \cdot \frac{\partial}{\partial \xi}\left(A^*(z, \xi) \cdot \frac{\partial h(\xi)}{\partial \xi}\right) \\ + \int \bar{P}(dy)B^*(y, \xi) \cdot \frac{\partial h(\xi)}{\partial \xi}.\end{aligned} \tag{5.31}$$

The operator $\mathscr{L}_\xi$ is an elliptic differential operator of second order defined on smooth functions on S^{n-1}. If it is uniformly elliptic, then the corresponding angular process is ergodic with $\bar{P}_\xi(A)$, $A \subset S^{n-1}$, its invariant measure. Moreover, the Fredholm alternative holds in the same way as in (3.8)–(3.11). Thus, in the ergodic case, if

$$q = \int_{S^{n-1}} v(\xi)\bar{P}_\xi(d\xi), \tag{5.32}$$

then $h(\xi)$ can be chosen so that it is smooth and

$$\mathscr{L}_\xi h(\xi) = -v(\xi) + q. \tag{5.33}$$

Combining this with (5.30) yields

$$\mathscr{L}(\log \rho + h(\xi)) = q. \tag{5.34}$$

Thus, (5.21) or (5.23) can be verified by computing the integral (5.32) in the ergodic case. In the nonergodic case, there are no general criteria assuring the existence of h with $q > 0$ or $q < 0$.

Proof. We shall use polar coordinates and the notation (5.25)–(5.31). The infinitesimal generator of $(x^\varepsilon(t), y^\varepsilon(t))$ has the following form in polar coordinates:

$$\mathscr{L}^\varepsilon = \frac{1}{\varepsilon^2}Q + \frac{1}{\varepsilon}\left(A^* \cdot \frac{\partial}{\partial \xi} + a^* \rho \frac{\partial}{\partial \rho}\right) + \left(B^* \cdot \frac{\partial}{\partial \xi} + b^* \rho \frac{\partial}{\partial \rho}\right). \tag{5.35}$$

Let $f(x) = \log \rho + h(\xi)$ be as in (5.20) and define $f_1(\xi, y)$, $f_2(\xi, y)$ as in (4.18) and (4.19); they do not depend on ρ since (5.16) is linear:

$$f_1(\xi, y) = \int_S \chi(y, dz)\left[a^*(z, \xi) + A^*(z, \xi) \cdot \frac{\partial h(\xi)}{\partial \xi}\right], \tag{5.36}$$

$$\begin{aligned}f_2(\xi, y) = \int_S \chi(y, dz)\bigg[A^*(z, \xi) \cdot \frac{\partial f_1(\xi, z)}{\partial \xi} + B(z, \xi) \cdot \frac{\partial h(\xi)}{\partial \xi} \\ + b^*(z, \xi) - v(\xi) - \mathscr{L}_\xi h(\xi)\bigg].\end{aligned} \tag{5.37}$$

Define

$$f^\varepsilon(\rho, \xi, y) = \log \rho + h(\xi) + \varepsilon f_1(\xi, y) + \varepsilon^2 f_2(\xi, y). \tag{5.38}$$

Then, as in (4.21), we have[11]

$$\mathcal{L}^\varepsilon f^\varepsilon = \mathcal{L}f(x)+\varepsilon\left[A^*(y,\xi)\cdot\frac{\partial f_2(\xi,y)}{\partial\xi}+B^*(y,\xi)\cdot\frac{\partial f_1(\xi,y)}{\partial\xi}\right] + \varepsilon^2 B^*(y,\xi)\cdot\frac{\partial f_2(\xi,y)}{\partial\xi}. \tag{5.39}$$

The analogue of (4.22) is as follows:

$$\begin{aligned}
\log|x^\varepsilon(t)| &+ h\left(\frac{x^\varepsilon(t)}{|x^\varepsilon(t)|}\right)+\varepsilon f_1\left(\frac{x^\varepsilon(t)}{|x^\varepsilon(t)|},y^\varepsilon(t)\right)+\varepsilon^2 f_2\left(\frac{x^\varepsilon(t)}{|x^\varepsilon(t)|},y^\varepsilon(t)\right)\\
&=\log|x|+h\left(\frac{x}{|x|}\right)+\varepsilon f_1\left(\frac{x}{|x|},y\right)+\varepsilon^2 f_2\left(\frac{x}{|x|},y\right)\\
&\quad+\int_0^t\left[\mathcal{L}f(x^\varepsilon(s))+\varepsilon A^*\left(\frac{x^\varepsilon(s)}{|x^\varepsilon(s)|},y^\varepsilon(s)\right)\cdot\frac{\partial f_2\left(\frac{x^\varepsilon(s)}{|x^\varepsilon(s)|},y^\varepsilon(s)\right)}{\partial\xi}\right.\\
&\quad\left.+\varepsilon^2 B^*\left(\frac{x^\varepsilon(s)}{|x^\varepsilon(s)|},y^\varepsilon(s)\right)\cdot\frac{\partial f_2\left(\frac{x^\varepsilon(s)}{|x^\varepsilon(s)|},y^\varepsilon(s)\right)}{\partial\xi}\right]ds+M_{f^\varepsilon}(t),
\end{aligned} \tag{5.40}$$

where $M_{f^\varepsilon}(t)$ is a zero-mean, integrable, right-continuous martingale relative to $\mathcal{F}_t$. As in (4.23) and (4.25), its increasing process is given by

$$\langle M_{f^\varepsilon}(t)\rangle=\int_0^t H_{f^\varepsilon}\left(\frac{x^\varepsilon(s)}{|x^\varepsilon(s)|},y^\varepsilon(s)\right)ds \tag{5.41}$$

where

$$H_{f^\varepsilon}(\xi,y)=q(y)\int\pi(y,dz)\left[\int(\chi(z,d\zeta)-(y,d\zeta))(g_1+\varepsilon g_2)(\xi,\zeta)\right]^2, \tag{5.42}$$

and

$$\begin{aligned}
g_1(\xi,y)&=a^*(y,\xi)+A^*(y,\xi)\cdot\frac{\partial h(\xi)}{\partial\xi}\\
g_2(\xi,y)&=A^*(y,\xi)\cdot\frac{\partial f_1(\xi,y)}{\partial\xi}+B^*(y,\xi)\cdot\frac{\partial h(\xi)}{\partial\xi}\\
&\quad+b^*(y,\xi)-v(\xi)-\mathcal{L}_\xi h(\xi).
\end{aligned} \tag{5.43}$$

[11] We use interchangeably rectangular and polar coordinates to simplify notation.

Let us assume that (5.21) holds; the case (5.23) is similar. For $0<\varepsilon \leqq \varepsilon_0$ and with ε_0 sufficiently small, it follows that

$$q+\varepsilon \sup_{\xi,y}\left|A^*(y,\xi)\cdot\frac{\partial f_2(\xi,y)}{\partial\xi}+B^*(y,\xi)\cdot\frac{\partial f_1(\xi,y)}{\partial\xi}\right| + \varepsilon^2 \sup_{\xi,y}\left|B^*(y,\xi)\cdot\frac{\partial f_2(\xi,y)}{\partial\xi}\right| = \tilde{q}<0. \tag{5.44}$$

Thus, from (5.40) it follows that for ε fixed and $0<\varepsilon \leqq \varepsilon_0$,

$$\overline{\lim_{t\uparrow\infty}}\frac{1}{t}\log|x^\varepsilon(t)| \leqq \tilde{q}+\overline{\lim_{t\uparrow\infty}}\frac{1}{t}M_{f^\varepsilon}(t), \tag{5.45}$$

with probability one. The lemma that follows shows that $t^{-1}M_{f^\varepsilon}(t)\to 0$ as $t\to\infty$ with probability one. We have thus proved the following sharper result.

$$P_{x,y}\left\{\overline{\lim_{t\uparrow\infty}}\frac{1}{t}\log|x^\varepsilon(t)|\leqq\tilde{q}\right\}=1.$$

The proof of the theorem is complete.

LEMMA. *Let $M(t)$ be a zero-mean right-continuous martingale with $\langle M(t)\rangle$ its increasing process such that*

$$\langle M(t)\rangle=\int_0^t H(s)\,ds$$

with

$$\sup_{t\geqq 0} E\{H(t)\}\leqq C<\infty.$$

Then,

$$\frac{1}{t}M(t)\to 0 \quad as\ t\to\infty,$$

with probability one.

Proof. The proof is similar to the one in [23, p. 487]. Let $Y(t)=t^{-1}M(t)$, $t>0$. For $\alpha>1$, $m^\alpha\leqq t<(m+1)^\alpha$, $m=$ integer, we have

$$\frac{t}{m^\alpha}Y(t)=\frac{1}{m^\alpha}M(t)=\frac{1}{m^\alpha}M(m^\alpha)+\frac{1}{m^\alpha}[M(t)-M(m^\alpha)]$$
$$=Y(m^\alpha)+Z(m^\alpha,t).$$

Put

$$U(m^\alpha)=\sup_{m^\alpha\leqq t<(m+1)^\alpha}|Z(m^\alpha,t)|.$$

To show that $t^{-1}M(t)\to 0$ with probability one we must show:

(i) $Y(m^\alpha)\to 0$ as $m\to\infty$ with probability one,

(ii) $U(m^\alpha)\to 0$ as $m\to\infty$ with probability one.

We have

$$\sum_{m=1}^{\infty} E\{|Y(m^\alpha)|^2\} = \sum_{m=1}^{\infty} \frac{1}{m^{2\alpha}} E\{M^2(m^\alpha)\}$$

$$= \sum_{m=1}^{\infty} \frac{1}{m^{2\alpha}} \langle M(m^\alpha)\rangle \leqq C \sum_{m=1}^{\infty} \frac{1}{m^\alpha} < \infty, \quad \text{if } \alpha > 1.$$

Thus, $\sum_{m=1}^{\infty} |Y(m^\alpha)| < \infty$ and hence $Y(m^\alpha) \to 0$ with probability one. This proves (i).

With $\alpha > 1$ fixed we show that with probability one there exists an m_0 such that for some $\beta > 0$ fixed,

$$U(m^\alpha) \leqq \frac{1}{m^\beta}, \qquad m > m_0.$$

By the Borel–Cantelli lemma it suffices to show that the series below converges:

$$\sum_{m=1}^{\infty} P\left\{U(m^\alpha) > \frac{1}{m^\beta}\right\} = \sum_{m=1}^{\infty} P\left\{\sup_{m^\alpha \leqq t < (m+1)^\alpha} |M(t) - M(m^\alpha)| > m^{\alpha-\beta}\right\}$$

$$\leqq \sum_{m=1}^{\infty} \frac{E\{(M((m+1)^\alpha) - M(m^\alpha))^2\}}{m^{2\alpha-\beta}}$$

$$\leqq C \sum_{m=1}^{\infty} \frac{(m+1)^\alpha - m^\alpha}{m^{2(\alpha-\beta)}}.$$

The last series converges provided $\alpha > 2\beta + 1$. Clearly $\alpha = 3$ and $\beta = \frac{1}{2}$ satisfy all conditions and the lemma is proved.

5.3. The harmonic oscillator. In this section we shall consider in detail an example: the harmonic oscillator with random spring constant.

There are (at least) two different ways the wide-band noise limit ($\equiv$white noise limit) may be considered depending on whether the radian frequency of the oscillator is of order one or large (going to ∞) as $\varepsilon \to 0$. Both cases can be treated without difficulty. The former leads to the white noise problem studied in [11]. The latter leads to a much simpler white noise problem because an additional averaging is superimposed [3]. Therefore, we shall treat here the large frequency problem which illustrates at the same time how averaging can be handled in general. For multidimensional versions of this interaction between the wide-band noise limit and averaging see [3] and the references cited there.

As another variation of the general theme up to now, we shall consider the problem in the scaling (2.15). Thus, we have

$$\text{(5.46)} \quad \begin{aligned} &\frac{d^2\tilde{x}^\varepsilon(\tau)}{d\tau^2} + 2\varepsilon^2\gamma\frac{d\tilde{x}^\varepsilon(\tau)}{d\tau} + (\omega^2 + \varepsilon y(\tau))\tilde{x}^\varepsilon(\tau) = 0, \qquad \tau > 0, \\ &\tilde{x}^\varepsilon(0) = x_1, \qquad \frac{d\tilde{x}^\varepsilon(0)}{d\tau} = \omega x_2, \end{aligned}$$

where ω is the radian frequency[12] and γ is the damping constant of the oscillator. The process $y(\tau)$ is taken as the random telegraph process, i.e., the two state Markov process $y(\tau)=\pm\beta$, $\beta>0$, with α^{-1} the mean time between jumps. We are interested in the asymptotic behavior of the oscillator as $\tau\to\infty$, when ε is sufficiently small and fixed. We are also interested in the dependence of this behavior on the parameters, α, β, γ and ω and on the size of ε_0, the length of the ε interval.

Define

$$x_1^\varepsilon(\tau)=\tilde{x}^\varepsilon(\tau), \qquad x_2^\varepsilon(\tau)=\omega^{-1}\frac{\tilde{x}^\varepsilon(\tau)}{d\tau}$$

so that (5.46) becomes

$$\frac{d}{d\tau}\begin{pmatrix} x_1^\varepsilon(\tau) \\ x_2^\varepsilon(\tau)\end{pmatrix}=\begin{pmatrix} 0 & \omega \\ -\omega-\varepsilon\omega^{-1}y(\tau) & -2\varepsilon^2\gamma\end{pmatrix}\begin{pmatrix} x_1^\varepsilon(\tau) \\ x_2^\varepsilon(\tau)\end{pmatrix}, \qquad \tau>0,$$
$$x_1^\varepsilon(0)=x_1, \qquad x_2^\varepsilon(0)=x_2. \tag{5.47}$$

As usual, $(x_1^\varepsilon(\tau), x_2^\varepsilon(\tau), y(\tau))$ is a Markov process on $R\times R\times\{-\beta,\beta\}$. Its infinitesimal generator is defined on functions $f(x_1,x_2,y)$, $y=\pm\beta$, as follows:[13]

$$\begin{aligned}\mathcal{L}^\varepsilon f(x_1,x_2,y)&=\alpha(-f(x_1,x_2,y)+f(x_1,x_2,-y))\\ &\quad+\omega\left(x_2\frac{\partial f(x_1,x_2,y)}{\partial x_1}-x_1\frac{\partial f(x_1,x_2,y)}{\partial x_2}\right)\\ &\quad-\varepsilon\omega^{-1}yx_1\frac{\partial f(x_1,x_2,y)}{\partial x_2}-\varepsilon^2 2\gamma x_2\frac{\partial f(x_1,x_2,y)}{\partial x_2}.\end{aligned} \tag{5.48}$$

With the introduction of polar coordinates

$$x_1=e^r\cos\theta, \quad x_2=e^r\sin\theta, \qquad -\infty<r<\infty, \quad 0\leqq\theta<2\pi, \tag{5.49}$$

$\mathcal{L}^\varepsilon$ of (5.48) becomes

$$\begin{aligned}\mathcal{L}^\varepsilon f(r,\theta,y)&=\alpha[-f(r,\theta,y)+f(r,\theta,-y)]-\omega\frac{\partial f(r,\theta,y)}{\partial\theta}\\ &\quad-\varepsilon\omega^{-1}y\left[\frac{1}{2}\sin 2\theta\frac{\partial f(r,\theta,y)}{\partial r}+\cos^2\theta\frac{\partial f(r,\theta,y)}{\partial\theta}\right]\\ &\quad-\varepsilon^2 2\gamma\left[\sin^2\theta\frac{\partial f(r,\theta,y)}{\partial r}+\frac{1}{2}\sin 2\theta\frac{\partial f(r,\theta,y)}{\partial\theta}\right].\end{aligned} \tag{5.50}$$

Note that $\mathcal{L}^\varepsilon$ has the form (cf. (3.16), (3.17))

$$\mathcal{L}^\varepsilon=\mathcal{L}_1+\varepsilon\mathcal{L}_2+\varepsilon^2\mathcal{L}_3, \tag{5.51}$$

---

[12] Not to be confused with a point in a probability space.

[13] Because of the scaling (2.15) $\mathcal{L}^\varepsilon$ of (3.12) is multiplied by ε^2.

where

$$\mathscr{L}_1 = Q - \omega\frac{\partial}{\partial\theta}, \qquad Qf(y) \equiv \alpha[-f(y) + f(-y)], \tag{5.52}$$

$$\mathscr{L}_2 = -\omega^{-1} y\left[\frac{1}{2}\sin 2\theta\frac{\partial}{\partial r} + \cos^2\theta\frac{\partial}{\partial\theta}\right], \tag{5.53}$$

$$\mathscr{L}_3 = -2\gamma\left[\sin^2\theta\frac{\partial}{\partial r} + \frac{1}{2}\sin 2\theta\frac{\partial}{\partial\theta}\right]. \tag{5.54}$$

From the form of $\mathscr{L}_1$ above, it is clear that the effect of large radian frequency is the extra term $-\omega\partial/\partial\theta$. In the problem corresponding to [11] this term would appear as part[14] of $\mathscr{L}_3$. We must now study the ergodic properties of $\mathscr{L}_1$ in (5.52). Since Q and $\omega\partial/\partial\theta$ commute this presents no difficulties. In fact if $g(y, \theta)$ is a function on $\{-\beta, \beta\}\times[0, 2\pi)$ such that

$$\int \bar{P}(dy)g(y, \theta) \equiv \tfrac{1}{2}g(\beta, \theta) + \tfrac{1}{2}g(-\beta, \theta) = 0, \qquad 0 \leqq \theta < 2\pi, \tag{5.55}$$

then

$$-\mathscr{L}_1^{-1}g(y, \theta) = \frac{1}{2}\int_0^\infty e^{-2\alpha t}[g(y, \theta - \omega t) - g(-y, \theta - \omega t)]\, dt. \tag{5.56}$$

Let $f(r)$ be a smooth function on $-\infty < r < \infty$. We define $f_1(r, \theta, y)$ and $f_2(r, \theta, y)$, as usual, by (4.18) and (4.19) which now yield

$$f_1(r, \theta, y) = \frac{-y}{4\omega(\omega^2 + \alpha^2)}(\alpha \sin 2\theta - \cos 2\theta)\frac{\partial f(r)}{\partial r}, \tag{5.57}$$

$$f_2(r, \theta, y) \equiv 0. \tag{5.58}$$

Thus, the generator of the limiting diffusion process to which $r^\varepsilon(\tau) = \log(x_1^\varepsilon(\tau) + x_2^\varepsilon(\tau))^{1/2}$ converges as $\varepsilon \to 0$ is given by

$$\mathscr{L}f(r) = \frac{\alpha\beta^2}{16\omega^2(\omega^2 + \alpha^2)}\frac{\partial^2 f(r)}{\partial r^2} + \left(\frac{\alpha\beta^2}{8\omega^2(\omega^2 + \alpha^2)} - \gamma\right)\frac{\partial f(r)}{\partial r}. \tag{5.59}$$

The process $r(\tau)$ generated by $\mathscr{L}$ in (5.59) is one dimensional Brownian motion with variance

$$\frac{\alpha\beta^2}{8\omega^2(\omega^2 + \alpha^2)}$$

and drift

$$\frac{\alpha\beta^2}{8\omega^2(\omega^2 + \alpha^2)} - \gamma.$$

[14] It can also appear as part of $\mathscr{L}_2$, as it does in some applications [7], [24]. We still have averaging, however.

Now choose $f(r)=r$. Then,

$$\mathcal{L}f(r)=\frac{\alpha\beta^2}{8\omega^2(\omega^2+\alpha^2)}-\gamma=q \tag{5.60}$$

and hence the *stability condition* for the limit problem ($q<0$) is now[15]

$$\frac{\alpha\beta^2}{8\omega^2(\omega^2+\alpha^2)}<\gamma. \tag{5.61}$$

From the expression corresponding to (5.44) we estimate the relevant ε_0 as follows: stability (with probability one) persists for (5.47), provided $0<\varepsilon<\varepsilon_0$ where

$$\varepsilon_0=\frac{4\gamma\omega[2\omega(\omega^2+\alpha^2)-(\alpha+1)\beta]-\alpha\beta^2}{8\omega^2(\omega^2+\alpha^2)}, \tag{5.62}$$

and where α, β, γ and ω satisfy (5.61). This result corresponds to (5.22). The instability result (5.24) follow analogously. When $\gamma=0$ the system is unstable, as is well known.

One may ask the following question regarding (5.47) and problems similar to it. If (5.61) holds, is it true that (5.47) is stable for any $\varepsilon>0$ not just for $0<\varepsilon<\varepsilon_0$? We do not have an answer to this question at present (cf. also [11]).

Stability of moments and explicit bounds on the range of variation of ε can be obtained easily in much the same way done above.

6. Asymptotic distributions for large time based on the wide-band noise approximation.

6.1. A Lyapunov theorem. We shall prove a theorem closely related to the one in § 5.1, the main difference being that (5.2) and (5.3) do not hold here. We shall work with the scaling (2.15) and with the assumptions of § 4.

THEOREM. *Let $(\tilde{x}^\varepsilon(\tau), y(\tau))$ be the process defined by* (2.15) *jointly with the external influence process and let $\mathcal{L}$ be the infinitesimal generator of the limiting diffusion process defined by* (4.10)–(4.12).

Assume that there exists a smooth function $V(x)$ such that

(i) *$V(x)\geqq 0$, $V(x)\to\infty$ as $|x|\to\infty$ and $V(x)$ and its derivatives are bounded by powers of $|x|$ for $|x|$ large,*

(ii) *for some $\gamma>0$ and $K>0$,*

$$\mathcal{L}V(x)\leqq-\gamma V(x), \quad \text{for } |x|>K. \tag{6.1}$$

Then, for $0<\varepsilon\leqq\varepsilon_0$, ε fixed and ε_0 sufficiently small, $(\tilde{x}^\varepsilon(\tau), y(\tau))$ has invariant probability measures on $R^n\times S$. If the diffusion process $x(t)$, with generator $\mathcal{L}$, has a unique invariant measure μ, if μ^ε is the marginal measure on R^n of any invariant measure of $(\tilde{x}^\varepsilon(\tau), y(\tau))$ and if $f(x)$ is a bounded continuous function, then

$$\lim_{\varepsilon\downarrow 0}\int f(x)\mu^\varepsilon(dx)=\int f(x)\mu(dx). \tag{6.2}$$

[15] This result was first obtained formally in [1].

Remarks. 1. Conditions (i) and (ii) imply that $x(t)$ has an invariant probability measure. This follows from a theorem of Beneš [12] since $x(t)$ is a $\hat{C}$ Feller process (cf. (3.15)).

2. If $V(x)$ behaves like $|x|^p$, p an integer, for $|x|$ large, then any limit measure (as t or $\tau\uparrow\infty$) μ and μ^ε of $x(t)$ and $\tilde{x}^\varepsilon(\tau)$ respectively, has moments of order $\bar{p}<p$.

Proof. We construct $V^\varepsilon(x, y)$ as in (4.20) with $f=V$ exactly as in § 5.1. The inequalities (5.11) hold again for $|x|>K$ and there is a $\hat{\gamma}$ and an $\varepsilon_0>0$ sufficiently small so that (5.9) holds for $|x|>K$ and $0<\varepsilon\leqq\varepsilon_0$. With the use of (4.4), the martingale term drops out in (4.10) on taking expectations and we obtain (allowing for the (2.15) scaling)

$$\begin{aligned}
&\tilde{c}_1E_{x,y}\{e^{\hat{\gamma}\varepsilon^2\tau}V(\tilde{x}^\varepsilon(\tau))\}\\
&\quad\leqq E_{x,y}\{e^{\hat{\gamma}\varepsilon^2\tau}V^\varepsilon(\tilde{x}^\varepsilon(\tau), y(\tau))\}\\
(6.3)\qquad&\quad= V^\varepsilon(x,y)+\varepsilon^2\int_0^\tau e^{\hat{\gamma}\varepsilon^2 s}E_{x,y}\{(\mathcal{L}^\varepsilon+\hat{\gamma})V^\varepsilon(\tilde{x}^\varepsilon(s), y(s))\}\,ds\\
&\quad\leqq\tilde{c}_2V(x)+\varepsilon^2\int_0^\tau e^{\hat{\gamma}\varepsilon^2 s}E_{x,y}\{(\mathcal{L}^\varepsilon+\hat{\gamma})V^\varepsilon(\tilde{x}^\varepsilon(s), y(s)), |\tilde{x}^\varepsilon(s)|>K\}\,ds\\
&\qquad+\varepsilon^2\int_0^\tau e^{\hat{\gamma}\varepsilon^2 s}E_{x,y}\{(\mathcal{L}^\varepsilon+\gamma)V^\varepsilon(\tilde{x}^\varepsilon(s), y(s)), |\tilde{x}^\varepsilon(s)|\leqq K\}ds.
\end{aligned}$$

Now when $|x|\leqq K$, $(\mathcal{L}^\varepsilon+\hat{\gamma})V^\varepsilon$ is bounded by a constant C_K. Thus, from (6.3), we obtain

$$(6.4)\qquad E_{x,y}\{V(\tilde{x}^\varepsilon(\tau))\}\leqq\frac{\tilde{c}_2}{\tilde{c}_1}e^{-\hat{\gamma}\varepsilon^2\tau}+\frac{C_K}{\tilde{c}_1\hat{\gamma}},$$

with $0<\varepsilon\leqq\varepsilon_0$, and ε_0 sufficiently small.

Inequality (6.4) implies that the probability distributions of $\tilde{x}^\varepsilon(\tau)$, $\tau\geqq0$, with x, y and ε fixed, are tight. Since $V(x)\to+\infty$ as $|x|\to\infty$, it follows that there is a function $R(K)\to\infty$ as $K\to\infty$ such that $|x|>K$ implies $V(x)>R(K)$. Thus,

$$\begin{aligned}
(6.5)\qquad P_{x,y}\{|\tilde{x}^\varepsilon(\tau)|>K\}&\leqq P_{x,y}\{V(\tilde{x}^\varepsilon(\tau))>R(K)\}\\
&\leqq\frac{1}{R(K)}E_{x,y}\{V(\tilde{x}^\varepsilon(\tau))\},
\end{aligned}$$

and this along with (6.4) gives tightness.

We can now apply a theorem of Beneš [12] to the Markov process $(\tilde{x}^\varepsilon(\tau), y(\tau))$ on $R^n\times S$ and deduce[16] that as $t\to\infty$, $0<\varepsilon\leqq\varepsilon_0$, it has invariant probability measures ν^ε on $R^n\times S$. Note that we make no statement about uniqueness; there may be many limits. For any limit measure, we have that $\bar{P}(A)=\nu^\varepsilon(R^n\times A)$, A is a Borel subset of S, since $\bar{P}(A)$ is the unique invariant measure of $y(\tau)$.

Let $\mu^\varepsilon(B)=\nu^\varepsilon(B\times S)$, B a Borel subset of R^n, let $f(x)$ be the bounded continuous function on R^n and let μ be the unique invariant measure of the diffusion $x(t)$ (we assume it is unique). By the ergodic theorem for $x(t)$ we have

[16] By (3.15), $(\tilde{x}^\varepsilon(\tau), y(\tau))$ is a $\hat{C}$ Feller process as required in [12]. Recall that S is compact.

that for almost all x

$$\lim_{t\uparrow\infty} E_x\{f(x(t))\} = \int f(\bar{x})\mu(d\bar{x}). \tag{6.6}$$

Thus, for $t>0$ fixed,

$$\begin{aligned}
&\lim_{\varepsilon\downarrow 0}\left|\int f(\bar{x})\mu^\varepsilon(d\bar{x}) - \int f(\bar{x})\mu(d\bar{x})\right| \\
&\quad = \lim_{\varepsilon\downarrow 0}\left|\iint \nu^\varepsilon(dx\,dy)E_{x,y}\{f(\tilde{x}^\varepsilon(t/\varepsilon^2))\} - \int f(\bar{x})\mu(d\bar{x})\right| \\
&\quad \leqq \lim_{\varepsilon\downarrow 0}\left|\iint \nu^\varepsilon(dx\,dy)[E_{x,y}\{f(\tilde{x}^\varepsilon(t/\varepsilon^2))\} - E_x\{f(x(t/\varepsilon^2))\}]\right| \\
&\qquad + \lim_{\varepsilon\downarrow 0}\left|\int \mu^\varepsilon(dx)\left[E_x\{f(x(t/\varepsilon^2))\} - \int f(\bar{x})\mu(d\bar{x})\right]\right| \\
&\quad = 0.
\end{aligned} \tag{6.7}$$

The second term on the right side of the last inequality in (6.7) is zero in view of (6.6) and the dominated convergence theorem. The first term is zero by the weak convergence theorem of § 4 and the dominated convergence theorem. This proves (6.2) and the proof of the theorem is complete.

6.2. Linear systems. As in § 5.2, one can be considerably more specific when dealing with linear problems. We shall consider the following analogue of (5.16):

$$\begin{aligned}
\frac{dx^\varepsilon(t)}{dt} &= \frac{1}{\varepsilon}[A(y^\varepsilon(t))x^\varepsilon(t) + \tilde{A}(y^\varepsilon(t))] \\
&\quad + B(y^\varepsilon(t))x^\varepsilon(t) + \tilde{B}(y^\varepsilon(t)), \qquad t>0, \quad x^\varepsilon(0)=x,
\end{aligned} \tag{6.8}$$

where $y^\varepsilon(t) \equiv y(t/\varepsilon^2)$ as in (2.13), $A(y)$ and $B(y)$ are $n\times n$ continuous matrix functions on S and $\tilde{A}(y)$ and $\tilde{B}(y)$ are continuous n-vector functions on S such that

$$\int_S A(y)\bar{P}(dy) = 0, \qquad \int_S \tilde{A}(y)\bar{P}(dy) = 0. \tag{6.9}$$

In the scaling (2.15), the process $\tilde{x}^\varepsilon(\tau)$ satisfies the system

$$\begin{aligned}
\frac{d\tilde{x}^\varepsilon(\tau)}{d\tau} &= \varepsilon[A(y(\tau))\tilde{x}^\varepsilon(\tau) + \tilde{A}(y(\tau))] \\
&\quad + \varepsilon^2[B(t(\tau))\tilde{x}^\varepsilon(\tau) + \tilde{B}(y(\tau))], \qquad \tau>0, \quad x^\varepsilon(0)=x,
\end{aligned} \tag{6.8'}$$

which corresponds to (5.16′).

Note that for both (6.8) and (6.8′) the origin $x=0$ is no longer an equilibrium point as it was in § 5.2.

For the processes $x^\varepsilon(t)$ of (6.8) (or $\tilde{x}^\varepsilon(\tau)$ of (6.8′)) the limiting diffusion process has, according to the theorem of § 4, generator $\mathscr{L}$ given by

$$\mathscr{L}f(x)=\mathscr{L}_1 f(x)+\mathscr{L}_2 f(x), \tag{6.10}$$

where

$$\begin{aligned}\mathscr{L}_1 f(x)=&\iint \chi(y,dz)\bar{P}(dy)\left[A(y)x\cdot\frac{\partial}{\partial x}\left(A(z)x\cdot\frac{\partial f(x)}{\partial x}\right)\right]\\&+\int \bar{P}(dy)B(y)x\cdot\frac{\partial f(x)}{\partial x}\end{aligned} \tag{6.11}$$

and

$$\begin{aligned}\mathscr{L}_2 f(x)=&\iint \chi(y,dz)\bar{P}(dy)\left[\tilde{A}(y)\cdot\frac{\partial}{\partial x}\left(A(z)x\cdot\frac{\partial f(x)}{\partial x}\right)\right.\\&\left.+A(y)x\cdot\frac{\partial}{\partial x}\left(\tilde{A}(z)\cdot\frac{\partial f(x)}{\partial x}\right)+\tilde{A}(y)\cdot\frac{\partial}{\partial x}\left(\tilde{A}(z)\cdot\frac{\partial f(x)}{\partial x}\right)\right]\\&+\int \bar{P}(dy)\tilde{B}(y)\cdot\frac{\partial f(x)}{\partial x}.\end{aligned} \tag{6.12}$$

Note that $\mathscr{L}_1$ is the same as the operator denoted by $\mathscr{L}$ in § 5.2, given by (5.29) in polar coordinates.

THEOREM. *Let $\tilde{x}^\varepsilon(\tau)$ be the process defined by* (6.8′) *and assume that there exists a smooth function h on S^{n-1}, the unit sphere in R^n, such that if*

$$\tilde{f}(x)=\log|x|+h(x/|x|), \tag{6.13}$$

then

$$\mathscr{L}_1\tilde{f}(x)\leqq q<0, \qquad |x|>K, \tag{6.14}$$

where q and K are constants.

Then for $0<\varepsilon\leqq\varepsilon_0$, ε fixed and ε_0 sufficiently small, $(\tilde{x}^\varepsilon(\tau), y(\tau))$ has invariant probability measures on $R^n\times S$. If the process $x(t)$ with generator $\mathscr{L}_1+\mathscr{L}_2$ in (6.10) *has a unique invariant measure μ, if μ^ε is the marginal measure on R^n of any invariant measure of $(\tilde{x}^\varepsilon(\tau), y(\tau))$ and if $f(x)$ is a bounded continuous function on R^n, then*

$$\lim_{\varepsilon\downarrow 0}\int f(x)\mu^\varepsilon(dx)=\int f(x)\mu(dx). \tag{6.15}$$

Remarks. 1. The conditions (6.13), (6.14) are essentially the same as (5.20) and (5.21) since $\mathscr{L}_1$ is the same as $\mathscr{L}$ of (5.29). Thus, we are demanding that (6.8′) be stable when $\tilde{A}$ and $\tilde{B}$ are removed.

2. As in Remark 2 of § 5.2, we shall work with polar coordinates (5.25). In addition to the notation (5.26)–(5.31) (with $\mathcal{L}$ in these formulas being $\mathcal{L}_1$ of (6.11) in polar coordinates) we need the following definitions:

$$\tag{6.16} \begin{aligned} \tilde{A}^*(y,\xi)\cdot\frac{\partial f(\xi)}{\partial \xi} &= \sum_{i,j=1}^n \tilde{A}_i(y)(\delta_{ij}-\xi_i\xi_j)\frac{\partial f(\xi)}{\partial \xi_j},\\ \tilde{B}^*(y,\xi)\cdot\frac{\partial f(\xi)}{\partial \xi} &= \sum_{i,j=1}^n \tilde{B}_i(y)(\delta_{ij}-\xi_i\xi_j)\frac{\partial f(\xi)}{\partial \xi_j}, \end{aligned}$$

$$\tag{6.17} \begin{aligned} \tilde{a}^*(y,\xi) &= \sum_{i=1}^n \tilde{A}_i(y)\xi_i,\\ \tilde{b}^*(y,\xi) &= \sum_{i=1}^n \tilde{B}_i(y)\xi_i. \end{aligned}$$

The operator $\mathcal{L}_1$ is identical with $\mathcal{L}$ of (5.29) as mentioned already. The operator $\mathcal{L}_2$ of (6.12) has the following form in polar coordinates:

$$\tag{6.18} \begin{aligned} \mathcal{L}_2 f(\rho,\xi) = \iint \chi(y,dz)\bar{P}(dz)\Big[&\Big(\frac{1}{\rho}\tilde{A}^*(y,\xi)\cdot\frac{\partial}{\partial\xi}+\tilde{a}^*(y,\xi)\frac{\partial}{\partial\rho}\Big)\\ &\cdot\Big(A^*(z,\xi)\cdot\frac{\partial f(\rho,\xi)}{\partial\xi}+a^*(z,\xi)\rho\frac{\partial f(\rho,\xi)}{\partial\xi}\Big)\\ &+\Big(A^*(y,\xi)\cdot\frac{\partial}{\partial\xi}+a^*(y,\xi)\rho\frac{\partial}{\partial\rho}\Big)\\ &\cdot\Big(\frac{1}{\rho}\tilde{A}^*(z,\xi)\cdot\frac{\partial f(\rho,\xi)}{\partial\xi}+\tilde{a}^*(z,\xi)\frac{\partial f(\rho,\xi)}{\partial\rho}\Big)\\ &+\Big(\frac{1}{\rho}\tilde{A}^*(y,\xi)\cdot\frac{\partial}{\partial\xi}+\tilde{a}^*(y,\xi)\frac{\partial}{\partial\rho}\Big)\\ &\cdot\Big(\frac{1}{\rho}\tilde{A}^*(z,\xi)\cdot\frac{\partial f(\rho,\xi)}{\partial\xi}+\tilde{a}^*(z,\xi)\frac{\partial f(\rho,\xi)}{\partial\rho}\Big)\Big]\\ &+\int \bar{P}(dy)\Big[\frac{1}{\rho}\tilde{B}^*(y,\xi)\cdot\frac{\partial f(\rho,\xi)}{\partial\xi}\\ &\qquad+\tilde{b}^*(y,\xi)\frac{\partial f(\rho,\xi)}{\partial\rho}\Big]. \end{aligned}$$

Proof. Let $\phi(\rho)$ be a C^∞ function on $[0,\infty)$ such that

$$\phi(\rho)=\begin{cases}0 & \text{if } \rho\leqq 1,\\ 1 & \text{if } \rho\geqq 2,\end{cases}$$

and define $f(x)$ by

$$\tag{6.19} f(x)=\phi(|x|)\log|x|+h\Big(\frac{x}{|x|}\Big).$$

Since (6.14) is not altered by adding a constant to h we may assume that

$$f(x) > 0. \tag{6.20}$$

From the boundedness and smoothness of the various coefficients in $\mathcal{L}_2$ of (6.18) and from their specific dependence on ρ, it follows from (6.14) that

$$\mathcal{L}f(x) = \mathcal{L}_1 f(x) + \mathcal{L}_2 f(x) \leqq \tilde{q} < 0, \qquad |x| > \tilde{K}, \tag{6.21}$$

where $\tilde{K}$ is sufficiently large and $|\tilde{q}|$ is smaller than $|q|$.

We construct next the functions $f_1(\rho, \xi, y)$ and $f_2(\rho, \xi, y)$ as in (5.36) and (5.41) but now, because of the $\tilde{A}$ and $\tilde{B}$ terms, they do depend on ρ. It is easily seen that they are uniformly bounded, however, along with their ρ derivatives. For $0 < \varepsilon \leqq \varepsilon_0$ and ε_0 sufficiently small, we can arrange to have

$$f^\varepsilon(\rho, \xi, y) = f(\rho, \xi) + \varepsilon f_1(\rho, \xi, y) + \varepsilon^2 f_2(\rho, \xi, y) \geqq 0, \tag{6.22}$$

and

$$\begin{aligned} &\mathcal{L}^\varepsilon f^\varepsilon(\rho, \xi, y) \leqq M < \infty \quad \text{if } \rho \leqq \tilde{K}, \\ &\mathcal{L}^\varepsilon f^\varepsilon(\rho, \xi, y) \leqq \hat{q} < 0 \quad \text{if } \rho > \tilde{K}. \end{aligned} \tag{6.23}$$

Therefore, using (4.4) to let the martingale term drop, we have

$$\begin{aligned} 0 &\leqq E_{x,y}\left\{ f^\varepsilon\left(|\tilde{x}^\varepsilon(\tau)|, \frac{\tilde{x}^\varepsilon(\tau)}{|\tilde{x}^\varepsilon(\tau)|}, y(\tau) \right) \right\} \\ &= f^\varepsilon\left(|x|, \frac{x}{|x|}, y \right) + \varepsilon^2 E_{x,y}\left\{ \int_0^\tau \mathcal{L}^\varepsilon f^\varepsilon\left(|\tilde{x}^\varepsilon(s)|, \frac{\tilde{x}^\varepsilon(s)}{|\tilde{x}^\varepsilon(s)|}, y(s) \right) \right\} ds \\ &\leqq f^\varepsilon\left(|x|, \frac{x}{|x|}, y \right) + \varepsilon^2 M \int_0^\tau P_{x,y}\{ |\tilde{x}^\varepsilon(s)| \leqq \tilde{K} \}\, ds \\ &\qquad + \varepsilon^2 \hat{q} \int_0^\tau P_{x,y}\{ |\tilde{x}^\varepsilon(s)| > \tilde{K} \}\, ds, \end{aligned} \tag{6.24}$$

and hence[17]

$$-f^\varepsilon\left(|x|, \frac{x}{|x|}, y \right) - \varepsilon^2 \hat{q}\tau \leqq (\varepsilon^2 M - \varepsilon^2 \hat{q}) \int_0^\tau P_{x,y}\{ |\tilde{x}^\varepsilon(s)| \leqq \tilde{K} \}\, ds. \tag{6.25}$$

From (6.25) it follows that for x ranging over a compact set and any $y \in S$ $(0 < \varepsilon \leqq \varepsilon_0)$

$$\varlimsup_{\tau \uparrow \infty} \frac{1}{\tau} \int_0^\tau P_{x,y}\{ |\tilde{x}^\varepsilon(s)| \leqq \tilde{K} \}\, ds > 0. \tag{6.26}$$

Now we can use again the theorem of Beneš [12] to establish the existence of an invariant measure for $(\tilde{x}^\varepsilon(\tau), y(\tau))$. The rest of the argument is as in § 6.1. The proof is complete.

In [13], Zakai used condition (6.26) and the theorem of Beneš in the same manner as above for general (not necessarily linear) Itô stochastic differential

[17] Recall $\hat{q}$ is negative.

equations. It is easily seen that his results have direct analogs in the present context; the linear systems provide a more concrete situation.

7. An estimate for large deviations of stable linear systems. Consider the systems (5.16) or (5.16′) under the usual hypotheses and in particular[18] (5.20), (5.21). Thus $\tilde{x}^\varepsilon(\tau)$ (we shall use the (2.15) scaling, or (5.16′)) is stable as $\tau \to \infty$ in the sense of (5.22). In this section we shall estimate the quantity

$$P_{x,y}\left\{\sup_{s \geqq \tau} |x^\varepsilon(s)| \geqq R\right\}$$

for $0<\varepsilon \leqq \varepsilon_0$, ε_0 sufficiently small. We shall show (Theorem 2 below) that it is exponentially small in $t=\varepsilon^2\tau$, $\tau \geqq 0$.

The analysis we shall follow is motivated by the setup of §§ 4 and 5 and results of Pinsky [14] on the analogous large deviations problem for Itô stochastic equations. Since the white noise results are of independent interest, we shall reproduce them here in Theorem 1 below.

Let us consider the following system of linear Itô stochastic differential equations:

$$\begin{aligned} dx(t) &= Bx(t)\,dt + \sum_{k=1}^{m} A^k x(t)\,dw_k(t), \qquad t>0, \\ x(0) &= x. \end{aligned} \tag{7.1}$$

Here $x(t)$ takes values in R^n, $A^1, A^2, \cdots, A^m$ and B are $n \times n$ constant matrices and $(w_1(t), \cdots, w_m(t))$ is the standard m dimensional Brownian motion. The infinitesimal generator L of the diffusion process $x(t)$ has the form

$$Lf(x) = \frac{1}{2} \sum_{i,j,l,r=1}^{n} \sum_{k=1}^{m} A^k_{il} x_l A^k_{jr} x_r \frac{\partial^2 f(x)}{\partial x_i \partial x_j} + \sum_{i,j=1}^{n} B_{ij} x_j \frac{\partial f(x)}{\partial x_i}. \tag{7.2}$$

We assume that $x=0$ is a stable equilibrium point of $x(t)$ in the sense that there exists a smooth function $h(x/|x|)$ on S^{n-1}, the unit sphere in R^n, such that if

$$f(x) = \log|x| + h(x/|x|), \tag{7.3}$$

then

$$Lf(x) \leqq -q < 0, \tag{7.4}$$

where q is a positive constant.

Applying Itô's formula to $f(x(t))$ yields

$$\begin{aligned} \log|x(t)| + h\left(\frac{x(t)}{|x(t)|}\right) &= \log|x| + h\left(\frac{x}{|x|}\right) + \int_0^t Lf(x(s))\,ds \\ &\quad + \int_0^t \sum_{j,l=1}^{n} \sum_{k=1}^{m} \frac{\partial f(x(t))}{\partial x_j} A^k_{lj} x_l(t)\,dw_k(t). \end{aligned} \tag{7.5}$$

[18] The case (5.23) leads to entirely analogous results and requires no special treatment.

Direct computation leads to the following expression for the integrand of the stochastic integral in (7.5)

$$(7.6)\qquad H_k\left(\frac{x}{|x|}\right)=\sum_{j,l=1}^{n}A_{jl}^{k}\left[\frac{x_jx_l}{|x|^2}+\sum_{r=1}^{n}\frac{\partial h}{\partial x_r}\left(\frac{x}{|x|}\right)\left(\frac{x_l\delta_{rj}}{|x|}-\frac{x_rx_jx_l}{|x|^3}\right)\right],\qquad k=1,2,\cdots,m,$$

and these are smooth functions on S^{n-1}. Define also

$$(7.7)\qquad H\left(\frac{x}{|x|}\right)=\sum_{k=1}^{n}H_k^2\left(\frac{x}{|x|}\right).$$

We may now rewrite (7.5) in the form

$$(7.8)\qquad \begin{aligned}&\log|x(t)|+h\left(\frac{x(t)}{|x(t)|}\right)=\log|x|+h\left(\frac{x}{|x|}\right)+\int_0^t Lf(x(s))\,ds+M(t),\\ &M(t)=\int_0^t\sum_{k=1}^{m}H_k\left(\frac{x(s)}{|x(s)|}\right)dw_k(s),\\ &\langle M(t)\rangle=\int_0^t H\left(\frac{x(s)}{|x(s)|}\right)ds,\end{aligned}$$

where $M(t)$ is a zero-mean continuous martingale and $\langle M(t)\rangle$ ($\langle M(t)\rangle\equiv\langle M(t), M(t)\rangle$ for simplicity) is its increasing process.

From the strong law of large numbers for $M(t)$ (see the lemma at end of § 5.2 or use a time substitution) and (7.4), it follows that

$$(7.9)\qquad \overline{\lim_{t\uparrow\infty}}\frac{1}{t}\log|x(t)|\leqq -q<0,$$

with probability one.

The following result is due to Pinsky [14].

THEOREM 1. *Under the above hypotheses*

$$(7.10)\qquad P_x\left\{\max_{s\geqq t}|x(s)|\geqq R\right\}\leqq\left(\frac{|x|}{R}\right)^{q/\bar H}\frac{e^{(2q\bar h+q^2/2)/\bar H}}{1-e^{-q^2/(2\bar H)}}e^{-q^2t/(2\bar H)},$$

where q is as in (7.4) *and*

$$(7.11)\qquad \bar H\equiv\sup_{|x|=1}H(x),\qquad \bar h=\sup_{|x|=1}h(x).$$

Remarks. 1. Clearly (7.10) implies that

$$(7.12)\qquad P_x\{|x(t)|\geqq R\}\leqq Ce^{-q^2t/(2\bar H)},$$

with C a constant as above. For (7.12) the constant can be improved a bit by an elementary direct argument as in the proof below.

2. Let T_R be the time of *last* entrance of $x(t)$ into $|x|\leqq R$ from the outside. Then

$$P_x\{T_R>t\}=P_x\left\{\sup_{s\geqq t}|x(s)|\geqq R\right\},$$

and so (7.10) gives an upper estimate for the distribution of T_R.

Proof. For any $\beta > 0$, $\exp(\beta M(t) - \beta^2 \langle M(t) \rangle / 2)$ is an integrable, continuous, nonnegative martingale with mean equal to one.

We have the following:

$$P_x\Big\{\max_{s>t} |x(s)| \geqq R\Big\} \leqq \sum_{\nu=[t]}^{\infty} P\Big\{\max_{\nu \leqq t < \nu+1} |x(s)| \geqq R\Big\} \tag{7.13}$$

and for $\nu = [t], [t]+1, \cdots, \beta > 0$,

$$P_x\Big\{\max_{\nu < s < \nu+1} |x(s)| \geqq R\Big\} = P_x\Big\{\max_{\nu \leqq s < \nu+1} \beta \log |x(s)| \geqq \beta \log R\Big\}. \tag{7.14}$$

We also have

$$\begin{aligned} &\max_{\nu \leqq s < \nu+1} \beta \log |x(s)| \\ &\quad \leqq \max_{\nu \leqq s < \nu+1} \beta\Big[\log |x(s)| + h\Big(\frac{x(s)}{|x(s)|}\Big) - \log |x| - h\Big(\frac{x}{|x|}\Big) - \int_0^s Lf(x(\gamma))\, d\gamma\Big] \\ &\quad + \max_{\nu \leqq s < \nu+1} \beta\Big[-h\Big(\frac{x(s)}{|x(s)|}\Big) + \log |x| + h\Big(\frac{x}{|x|}\Big) + \int_0^s Lf(x(\gamma))\, d\gamma\Big] \\ &\quad \leqq \max_{\nu \leqq s < \nu+1} \Big[\beta M(s) - \frac{\beta^2}{2} \langle M(s) \rangle\Big] + \beta\Big[2\bar{h} + \log |x| - q\nu + \frac{\beta}{2} \bar{H}(\nu+1)\Big]. \end{aligned} \tag{7.15}$$

Combining (7.14) and (7.15) and using Kolmogorov's inequality we obtain ($\beta > 0$ is fixed)

$$\begin{aligned} P_x\Big\{\max_{\nu \leqq s < \nu+1} |x(s)| \geqq R\Big\} &\leqq P\Big\{\max_{\nu \leqq s < \nu+1} \Big(\beta M(s) - \frac{\beta^2}{2} \langle M(s) \rangle\Big) \\ &\qquad \geqq \beta\Big(\log \frac{R}{|x|} - 2\bar{h} + q\nu - \frac{\beta}{2} \bar{H}(\nu+1)\Big)\Big\} \\ &\leqq \Big(\frac{|x|}{R}\Big)^{\beta} e^{2\bar{h}\beta + \beta^2 \bar{H}/2} e^{-\nu(q\beta - \beta^2 \bar{H}/2)}. \end{aligned} \tag{7.16}$$

The choice $\beta = q/\bar{H}$ optimizes the inequality (7.16). Using the result in (7.13) we find that

$$\begin{aligned} P_x\Big\{\max_{s \geqq t} |x(s)| \geqq R\Big\} &\leqq \Big(\frac{|x|}{R}\Big)^{q/\bar{H}} e^{(2q\bar{h} + q^2/2)/\bar{H}} \sum_{\nu=[t]}^{\infty} e^{-\nu q^2/(2\bar{H})} \\ &\leqq \Big(\frac{|x|}{R}\Big)^{q/\bar{H}} \frac{e^{(2q\bar{h} + q^2)/\bar{H}}}{1 - e^{-q^2/(2\bar{H})}} e^{-q^2 t/(2\bar{H})}. \end{aligned} \tag{7.17}$$

By beginning with $\beta = q/\bar{H}$ in (7.15) we can improve a bit the constant as in (7.10). The proof of Theorem 1 is complete.

We turn now to the analogous problem for the linear system (5.16′). We employ systematically the notation of § 5.2; in particular the polar coordinates of the representation of $\mathscr{L}$ in (5.29) and of $\mathscr{L}^{\varepsilon}$ in (5.35).

At first we must establish the analogue of the exponential martingale which was employed in Theorem 1. Let $u(x, y) > 0$ be a bounded smooth function in x and continuous in y and suppose that for each $\tau \geqq 0$ the nonnegative random variable

$$M^\varepsilon(\tau) = u(\tilde{x}^\varepsilon(\tau), y(\tau)) \exp\left\{-\int_0^\tau \frac{\varepsilon^2 \mathcal{L}^\varepsilon u(\tilde{x}^\varepsilon(s), y(s))}{u(\tilde{x}^\varepsilon(s), y(s))} ds\right\} \tag{7.18}$$

is integrable. By direct computation we find that

$$E_{x,y}\{M^\varepsilon(\tau)\} = u(x, y), \tag{7.19}$$

which is nothing other than the Feynman–Kac formula with the potential equal to $-\mathcal{L}u/u$, u being the initial data. It follows from this that $M^\varepsilon(\tau)$ is a right-continuous, integrable martingale with mean equal to $u(x, y)$. In the case of diffusions in R^n, proper choice of u yields the usual martingales of Stroock and Varadhan [25].

In the spirit of the considerations of the previous sections, we construct a martingale $M^\varepsilon(\tau)$ by (7.18) with an appropriate choice of the function u. The u will depend on ε and it will be denoted by $u^\varepsilon = u^\varepsilon(|x|, x/|x|, y)$, in polar coordinates. First we state the result.

THEOREM 2. *Let $\tilde{x}^\varepsilon(\tau)$ be the process defined by* (5.16′). *Suppose, as in* § 5.2, *that there is a smooth function $h(\xi)$, $\xi \in S^{n-1}$, such that if*

$$f(x) = \log|x| + h(x/|x|), \tag{7.20}$$

then

$$\mathcal{L}f(x) \leqq -q < 0, \tag{7.21}$$

where q is a positive constant and $\mathcal{L}$ is defined, in polar coordinates, by (5.29).

Under these conditions there is an $\varepsilon_0 > 0$ such that for each $\varepsilon \in (0, \varepsilon_0]$ and all $\tau \geqq 0$,

$$\begin{aligned} &P_{x,y}\left\{\max_{s \geqq \tau} |\tilde{x}^\varepsilon(s)| \geqq R\right\} \\ &\leqq \left(\frac{|x|}{R}\right)^{q/\bar{H}} \frac{1+\alpha_1(\varepsilon)}{1-\alpha_1(\varepsilon)} \exp\left\{2q\bar{h} + \frac{q^2}{2} - \varepsilon^2 \alpha_2(\varepsilon)\right\} \\ &\quad \cdot [1 - \exp(q^2/(2\bar{H} - \alpha_2(\varepsilon)))]^{-1} \exp\{-\varepsilon^2 \tau q^2/(2\bar{H} - \alpha_2(\varepsilon))\}. \end{aligned} \tag{7.22}$$

Here $\alpha_1(\varepsilon)$ and $\alpha_2(\varepsilon)$ are positive functions which go to zero as $\varepsilon \to 0$ and are defined by (7.32) *and* (7.33) *below,*

$$\bar{h} = \sup_{|x|=1} h(x), \qquad q \textit{ is as in } (7.21), \tag{7.23}$$

and

$$\bar{H} = \sup_{|x|=1} H(x)$$

(7.24)
$$= \sup_{|x|=1} 2 \iint \chi(y, dz)\bar{P}(dy)\left[A^{\#}(x, y) \cdot \frac{\partial h(x)}{\partial \xi} + a^{\#}(x, y)\right] \cdot \left[A^{\#}(x, z) \cdot \frac{\partial h(x)}{\partial \xi} + a^{\#}(x, z)\right],$$

in the notation of § 5.2.

Remarks. 1. Note that (7.10) and (7.22) are very similar. In fact as $\varepsilon \to 0$ with $\varepsilon^2 \tau = t$ fixed we recover (7.10) from (7.22) with appropriate definition of the other constants, as should be.

2. Remarks 1 and 2 of Theorem 1 apply verbatim to this theorem as well.

Proof. Let $f(x) = \log \rho + h(\xi)$ be as in (7.20) and define $f_1(\xi, y)$ and $f_2(\xi, y)$ by (5.36) and (5.37) respectively. Define $f_3(\xi, y)$ by

(7.25)
$$f_3(\xi, y) = \int \chi(y, dz)\left[f_1(z, \xi)\left(A^{\#}(z, \xi) \cdot \frac{\partial h(\xi)}{\partial \xi} + a^{\#}(z, \xi)\right) - H(\xi)\right]$$

where $H(\xi)$ is given by the right hand side of (7.24) without the sup.

Define

(7.26)
$$u^{\varepsilon}(\rho, \xi, y) = \rho^{q/\bar{H}} e^{qh(\xi)/\bar{H}}\left[1 + \frac{\varepsilon q}{\bar{H}} f_1(\xi, y) + \frac{\varepsilon^2 q}{\bar{H}} f_2(\xi, y) + \frac{\varepsilon^2 q^2}{\bar{H}^2} f_3(\xi, y)\right].$$

A lengthy but straightforward calculation yields the following result (with $\mathscr{L}^{\varepsilon}$ given by (5.39))

(7.27)
$$\frac{\varepsilon^2 \mathscr{L}^{\varepsilon} u^{\varepsilon}(\rho, \xi, y)}{u^{\varepsilon}(\rho, \xi, y)} = \varepsilon^2\left[\frac{q}{\bar{H}} \mathscr{L} f(\xi) + \frac{q^2}{2\bar{H}^2} H(\xi) + \varepsilon U^{\varepsilon}(\xi, y)\right],$$

where, omitting the arguments ξ and y,

(7.28)
$$U^{\varepsilon} = \frac{G_1 + \varepsilon G_2 - G_3^{\varepsilon}}{1 + \varepsilon g_1 + \varepsilon^2 g_2},$$

$$G_1 = A^{\#} \cdot \frac{\partial f_2}{\partial \xi} + B^{\#} \cdot \frac{\partial f_1}{\partial \xi} + \frac{q}{\bar{H}}\left(A^{\#} \cdot \frac{\partial f_3}{\partial \xi} + f_2 A^{\#} \cdot \frac{\partial h}{\partial \xi} + f_1 B^{\#} \cdot \frac{\partial h}{\partial \xi} + f_2 a^{\#} + f_1 b^{\#}\right) + \left(\frac{q}{\bar{H}}\right)^2 f_3\left(A^{\#} \cdot \frac{\partial h}{\partial \xi} + a^{\#}\right),$$

$$G_2 = B^{\#} \cdot \frac{\partial f_2}{\partial \xi} + \frac{q}{\bar{H}}\left(B^{\#} \cdot \frac{\partial f_3}{\partial \xi} + f_2 B^{\#} \cdot \frac{\partial h}{\partial \xi} + f_2 b^{\#}\right)$$

(7.29)
$$+\left(\frac{q}{\bar{H}}\right)^2 f_3\left(B^\# \cdot \frac{\partial h}{\partial \xi}+b^\#\right),$$

$$G_3^\varepsilon=\left(\mathscr{L}f+\frac{q}{2\bar{H}}H\right)\left(\frac{q}{\bar{H}}f_1+\varepsilon\frac{q}{\bar{H}}f_2+\varepsilon\left(\frac{q}{\bar{H}}\right)^2 f_3\right),$$

$$g_1=\frac{q}{\bar{H}}f_1, \qquad g_2=\frac{q}{\bar{H}}f_2+\left(\frac{q}{\bar{H}}\right)^2 f_3.$$

The important thing to observe here is that $U^\varepsilon(\xi, y)$ is a bounded function and that u^ε is a bounded function of (ξ, y) multiplied by $\rho^{q/\bar{H}}$. Therefore, by an elementary moment estimate like (4.4), it follows that $M^\varepsilon(\tau)$ of (7.18) with $u=u^\varepsilon$ of (7.26) is an integrable, right-continuous martingale.

We now continue as in Theorem 1 assuming that $0<\varepsilon\leqq\varepsilon_0$ and ε_0 is sufficiently small

$$P_{x,y}\left\{\max_{s\geqq\tau}|\tilde{x}^\varepsilon(s)|\geqq R\right\}$$

(7.30)
$$\leqq \sum_{\nu=[\varepsilon^2\tau]}^{\infty} P_{x,y}\left\{\max_{\nu/\varepsilon^2\leqq s\leqq(\nu+1)/\varepsilon^2}|\tilde{x}^\varepsilon(s)|\geqq R\right\}$$

$$=\sum_{\nu=[\varepsilon^2\tau]}^{\infty} P_{x,y}\left\{\max_{\nu/\varepsilon^2\leqq s<(\nu+1)/\varepsilon^2}|\tilde{x}^\varepsilon(s)|^{q/\bar{H}}\geqq R^{q/\bar{H}}\right\}.$$

We also have that

$$\max_{\nu/\varepsilon^2\leqq s<(\nu+1)/\varepsilon^2}|\tilde{x}^\varepsilon(s)|^{q/\bar{H}}$$

$$=\max_{\nu/\varepsilon^2\leqq s<(\nu+1)/\varepsilon^2}\left\{|\tilde{x}^\varepsilon(s)|^{q/\bar{H}}\frac{e^{qh/\bar{H}}[1+\varepsilon g_1+\varepsilon^2 g_2]\exp\left[-\int_0^s\frac{\varepsilon^2\mathscr{L}^\varepsilon u^\varepsilon}{u^\varepsilon}\,d\gamma\right]}{e^{qh/\bar{H}}[1+\varepsilon g_1+\varepsilon^2 g_2]\exp\left[-\int_0^s\frac{\varepsilon^2\mathscr{L}^\varepsilon u^\varepsilon}{u^\varepsilon}\,d\gamma\right]}\right\}$$

(7.31)
$$\leqq\frac{\max_{\nu/\varepsilon^2\leqq s(\nu+1)/\varepsilon^2}\left\{u^\varepsilon\left(|\tilde{x}^\varepsilon(s)|,\frac{\tilde{x}^\varepsilon(s)}{|\tilde{x}^\varepsilon(s)|},y(s)\right)\exp\left[-\int_0^s\frac{\varepsilon^2\mathscr{L}^\varepsilon u^\varepsilon}{u^\varepsilon}\left(\frac{\tilde{x}^\varepsilon(\gamma)}{|\tilde{x}^\varepsilon(\gamma)|},y(\gamma)\right)d\gamma\right]\right\}}{\min_{\nu/\varepsilon^2\leqq s<(\nu+1)/\varepsilon^2}\left\{e^{qh/\bar{H}}[1+\varepsilon g_1+\varepsilon^2 g_2]\exp\left[-\int_0^s\frac{\varepsilon^2\mathscr{L}^\varepsilon u^\varepsilon}{u^\varepsilon}\left(\frac{\tilde{x}^\varepsilon(\gamma)}{|\tilde{x}^\varepsilon(\gamma)|},y(\gamma)\right)d\gamma\right]\right\}}$$

$$\leqq\frac{\max_{\nu/\varepsilon^2\leqq s<(\nu+1)/\varepsilon^2}M^\varepsilon(s)}{e^{-q\bar{h}/\bar{H}}[1-\alpha_1(\varepsilon)]\,e^{-\nu[-q^2/2\bar{H}+\alpha_2(\varepsilon)]}}.$$

Here $\alpha_1(\varepsilon)$ and $\alpha_2(\varepsilon)$ are defined by

(7.32)
$$\alpha_1(\varepsilon)=\varepsilon\sup_{y,\xi}|g_1(\xi,y)+\varepsilon g_2(\xi,y)|,$$

$$\alpha_2(\varepsilon) = \varepsilon \sup_{\xi, y} |U^\varepsilon(\xi, y)|. \tag{7.33}$$

Thus

$$\begin{aligned} &P_{x,y}\left\{\max_{\nu/\varepsilon^2 \leqq s < (\nu+1)/\varepsilon^2} |\tilde{x}^\varepsilon(s)|^{q/\bar{H}} \geqq R^{q/\bar{H}}\right\} \\ &\leqq P_{x,y}\left\{\max_{\nu/\varepsilon^2 \leqq s < (\nu+1)/\varepsilon^2} M^\varepsilon(s) \geqq R^{q/\bar{H}} e^{-q\bar{h}/\bar{H}}[1-\alpha_1(\varepsilon)]\, e^{-\nu[-q^2/(2\bar{H}+\alpha_2(\varepsilon))]}\right\} \\ &\leqq \left(\frac{|x|}{R}\right)^{q/\bar{H}} \frac{1+\alpha_1(\varepsilon)}{1-\alpha_2(\varepsilon)} e^{2q\bar{h}/\bar{H}} e^{-\nu(q^2/(2\bar{H}+\alpha_2(\varepsilon)))} \end{aligned} \tag{7.34}$$

and returning to (7.30), using (7.34) and summing the series yields (7.22). The proof of Theorem 2 is complete.

Acknowledgment. We thank Mark Pinsky for discussions on the results of § 7 and making available to us his manuscript of Theorem 1 of that section. We also thank S. R. S. Varadhan for numerous discussions on the problems considered here over the span of several months. His ideas and specific recommendations, especially the use of martingales, have proved very useful.

REFERENCES

[1] R. L. STRATONOVICH, *Topics in the Theory of Random Noise*, English transl., Gordon and Breach, New York, 1963.

[2] R. Z. KHASMINSKII, *A limit theorem for solutions of differential equations with a random right hand side*, Theor. Probability Appl., 11 (1966), pp. 390–406.

[3] G. C. PAPANICOLAOU AND W. KOHLER, *Asymptotic theory of mixing stochastic ordinary differential equations*, Comm. Pure Appl. Math., 27 (1974), pp. 641–668.

[4] G. C. PAPANICOLAOU, *Introduction to the asymptotic analysis of stochastic equations*, Proceedings of SIAM-AMS Symposium (R.P.I., July 1975); Lectures in Applied Mathematics, vol. 16, American Mathematical Society, Providence, RI, to appear.

[5] ———, *Some probabilistic problems and methods in singular perturbations*, Rocky Mountain Math. J., 6 (1976), pp. 653–674.

[6] T. G. KURTZ, *Semigroups of conditioned shifts and approximation of Markov processes*, Ann. Probability, 3 (1975), pp. 618–642.

[7] G. C. PAPANICOLAOU, D. STROOCK AND S. R. S. VARADHAN, *Martingale approach to some limit theorems*, Statistical Mechanics, Dynamical Systems and the Duke Turbulence Conference, M. Reed, ed., Duke University Mathematics Series, vol. 3, Durham, NC, 1977.

[8] I. I. GIKHMAN AND A. V. SKOROKHOD, *Stochastic Differential Equations*, Springer-Verlag, New York, 1972.

[9] R. Z. KHASMINSKII, *Necessary and sufficient conditions for the asymptotic stability of linear stochastic systems*, Theor. Probability Appl., 12 (1967), pp. 144–147.

[10] M. PINSKY, *Stochastic stability and the Dirichlet problem*, Comm. Pure Appl. Math., 27 (1974), pp. 311–350.

[11] R. R. MITCHELL AND F. KOZIN, *Sample stability of second order linear differential equations with wide band noise coefficients*, this Journal, 27 (1974), pp. 571–605.

[12] V. BENEŠ, *Finite regular invariant measures for Feller processes*, J. Appl. Probability, 5 (1968), pp. 203–209.

[13] M. ZAKAI, *A Lyapounov criterion for the existence of stationary probability distributions for systems perturbed by noise*, SIAM J. Control, 7 (1969), pp. 390–397.

[14] M. PINSKY, unpublished notes, private communication.
[15] R. Z. KHASMINSKII, *Stability of Systems of Differential Equations under Random Perturbations of their Parameters*, "Nauka", Moscow, 1969.
[16] W. M. WONHAM, *Random differential equations in control theory*, Probabilistic Methods in Applied Mathematics, vol. 2, A. T. Bharucha-Reid, ed., Academic Press, New York, 1970, pp. 131–212.
[17] W. FLEMING, *Optimal continuous-parameter stochastic control*, SIAM Rev., 11 (1969), pp. 470–509.
[18] G. C. PAPANICOLAOU AND W. KOHLER, *Asymptotic analysis of deterministic and stochastic equations with rapidly varying components*, Comm. Math. Phys., 45 (1975), pp. 217–232.
[19] T. G. KURTZ, *A limit theorem for perturbed operator semigroups with applications to random evolutions*, J. Functional Anal., 12 (1973), pp. 55–67.
[20] A. BENSOUSSAN, J. L. LIONS AND G. C. PAPANICOLAOU, *book in preparation.*
[21] P. BILLINGSLEY, *Convergence of Probability Measures*, John Wiley, New York, 1968.
[22] I. I. GIHMAN AND A. V. SKOROKHOD, *The Theory of Stochastic Processes I*, Springer-Verlag, New York, 1974.
[23] M. LOÈVE, *Probability Theory*, Van Nostrand, Princeton, NJ, 1963.
[24] W. KOHLER AND G. C. PAPANICOLAOU, *Power statistics for wave propagation in one dimension III*, to appear.
[25] D. STROOCK AND S. R. S. VARADHAN, *Diffusion processes with boundary conditions*, Comm. Pure Appl. Math., 24 (1971), pp. 147–225.