

# Introduction to the Asymptotic Analysis of Stochastic Equations

George C. Papanicolaou<sup>1</sup>

**1. Introduction.** The purpose of these notes is to bring to the attention of people interested in stochastic problems some tools for their analysis. It is easy to pose stochastic problems in applied science and engineering and it is also clear that they are more realistic than their deterministic counterparts. It is not easy, however, to analyze these problems. We shall attempt to show here that a class of interesting stochastic problems admit satisfactory solutions by using perturbation theory, in fact, perturbation theory in much the same format as for deterministic problems, multiple scales [1], the Fredholm alternative [2], etc.

We shall also attempt to show what it is about the stochastic problems under consideration that permits this analysis. Loosely speaking, it is the difference in the time scales of the observables as compared to the time scales of the noise fluctuations. Basic references for the material of §3, our main subject, are [3], [4], [5]. References to works on specific topics, mathematical or physical, are given in this section. The material of §2 is a general review on stochastic processes. The class of Markov processes plays the role of "canonical problems" for the problems we seek to analyze, i.e., the principal approximation will be Markovian and frequently a diffusion process. The assumption is, of course, that these processes are relatively easy to understand and their analysis reduces to reasonably standard mathematical problems, such as solving (deterministic) differential equations.

§3 deals with the asymptotics for ordinary differential equations with random coefficients following the basic work of Stratonovich [3]. Every result stated in this section has some mathematical theory behind it; there are few problems that are not understood here. Naturally, implementing the suggested methods to specific problems is quite challenging and more effort is needed in this direction.

Asymptotics for stochastic partial differential equations are, naturally, less well developed. A good deal of work has been done in connection with waves in random media and the emergence of transport theory as a first approximation [7], [8]. The analysis of the stochastic parts of the problem is formally similar to the one of §3 but there are many other issues that must be considered. A mathematical theory is lacking at present.

The formalism of §3 seems to apply also to a variety of problems in modern physics; see for example [4] and other work by M. Lax, [5], [6].

---

*AMS (MOS) subject classifications* (1970). Primary 60-02, 60H10.

<sup>1</sup>Partially supported by an Alfred P. Sloan Foundation Fellowship and by the Air Force Office of Scientific Research under Grant AFOSR-71-2013.

## REFERENCES

1. D. S. Cohen, *Lecture notes in perturbation theory*, Lectures in Appl. Math., vol. 16, Amer. Math. Soc., Providence, R.I., 1976, pp.
2. J. B. Keller, *Perturbation theory*, Lecture Notes, Dept. of Mathematics, Michigan State University, East Lansing, Michigan, 1968.
3. R. S. Stratonivič, *Topics in the theory of random noise*. Vols. 1, 2, "Sovetskoe Radio", Moscow, 1961; English transl., Gordon and Breach, New York, 1963. MR 28 # 1660.
4. M. Lax, *Classical noise*. IV. *Langevin methods*, Rev. Modern Phys. 38 (1966), 561-566.
5. F. Haake, *Statistical treatment of open systems by generalized master equations*, Springer-Tracts in Modern Physics, no. 66, Springer-Verlag, Berlin and New York, 1973.
6. H. Haken, *Cooperative phenomena in systems far from thermal equilibrium and in nonlinear systems*, Rev. Modern Phys. 47 (1975), 67-121.
7. A. S. Gurvič and V. I. Tatarskii, *Coherence and intensity fluctuations of light in the turbulent atmosphere*, Radio Science 10 (1975), 3-14.
8. Yu. N. Barabanenkov, A. G. Vinogradov, Yu. A. Kravtsov and V. I. Tatarskii, *Application of theory of multiple scattering to the radiation transfer equation for a statistically inhomogeneous medium*, Radiofizika 15 (1972), 1852-1860.

**2. Review of stochastic processes, diffusions and stochastic calculus.** References to the subsections that follow are [1]-[6] at the end of this section.

(i) *Stochastic processes, conditional expectations, Markov property.* A real-valued stochastic process is, at first, a consistent family of probability distribution functions, that is, for any  $n \geq 1$  and  $0 \leq t_0 < t_1 < \dots < t_n < \infty$ , functions

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n)$$

with  $x_i \in R$ ,  $i = 1, 2, \dots, n$ , are given which represent the probability that the process at  $t_1$  is less than  $x_1$ , at  $t_2$  less than  $x_2$ ,  $\dots$ , at  $t_n$  less than  $x_n$ . With this information at hand a probability space  $(\Omega, \mathfrak{F}, P)$  is constructed where the set of elementary events  $\Omega$  is the collection of real-valued functions on  $[0, \infty)$ , the events  $\mathfrak{F}$  are the  $\sigma$ -algebra generated by cylinder sets and  $P$  is the probability measure on  $\mathfrak{F}$  induced by the given distribution functions. Cylinder sets are sets of functions defined by  $\{x(\cdot) \in \Omega; x(t_1) \in A_1, x(t_2) \in A_2, \dots, x(t_n) \in A_n\}$  where  $A_1, A_2, \dots, A_n$  are Borel subsets of  $R$  and  $0 \leq t_1 < t_2 < \dots < t_n < \infty$ .

The  $\sigma$ -algebra generated by the cylinder sets with the restriction that  $0 \leq t_0 < t_1 < \dots < t_n \leq t$  is denoted by  $\mathfrak{F}_t$ . We say that  $\mathfrak{F}_t$  is the collection of events that occurs up to time  $t$ . Thus, along with  $(\Omega, \mathfrak{F}, P)$  we have also an increasing family  $\mathfrak{F}_t$ ,  $t \geq 0$ , of  $\sigma$ -algebras and the process, denoted by  $\{x(t), t \geq 0\}$ , is said to be adapted to  $\mathfrak{F}_t$ , by construction in this case.

Processes that are vector valued or take values in more general spaces can be considered without additional complication.

Under mild regularity conditions on the two-dimensional distributions  $F_{t_1, t_2}(x_1, x_2)$ , probabilities of events that depend on uncountably many  $t$  points can be computed using only a denumerable (dense) set of  $t$  points. The process is said to be separable in this case. If  $\omega \in \Omega$  let  $x(t, \omega)$  denote the value of  $\omega$  (a function on  $[0, \infty)$ ) at time  $t$ . Again, under mild restrictions on  $F_{t_1, t_2}(x_1, x_2)$  (stochastic continuity) it follows that  $x(t, \omega)$  can be considered a

random variable on  $[0, \infty) \times \Omega$ . In particular, if integrals exist, we may for  $P$ -almost all  $\omega$  integrate  $x(t, \omega)$  as a function of  $t$ . [3, Chapter III] gives a good account of these facts.

If  $Y$  is a random variable on  $(\Omega, \mathfrak{F}, P)$  taking real values, its expectation is denoted by

$$E\{Y\} = \int_{\Omega} Y(\omega) P(d\omega).$$

Since  $\Omega$  is the space of functions on  $[0, \infty)$ ,  $Y$  is a functional. We wish to define the conditional expectation of  $Y$  given  $\mathfrak{F}_t$ , i.e., given information about the process  $\{x(t), t \geq 0\}$  up to time  $t$ . This conditional expectation is denoted by  $E\{Y|\mathfrak{F}_t\}$  and it is a random variable that depends on the past up to time  $t$ , i.e., it is an  $\mathfrak{F}_t$  measurable random variable on  $(\Omega, \mathfrak{F}, P)$ . Its defining property is that if  $B$  is any event in  $\mathfrak{F}_t$ , then

$$\int_B E\{Y|\mathfrak{F}_t\} P(d\omega) = \int_B Y P(d\omega)$$

or

$$E\{\chi(B)E\{Y|\mathfrak{F}_t\}\} = E\{\chi(B)Y\}$$

where  $\chi(B)$  is the characteristic function of the set  $B$ . The random variable  $E\{Y|\mathfrak{F}_t\}$  is easily shown to exist but it is defined only for almost all  $\omega \in \Omega$ . Two basic properties of conditional expectations are these. If  $Z$  is an  $\mathfrak{F}_t$  measurable random variable then  $E\{ZY|\mathfrak{F}_t\} = ZE\{Y|\mathfrak{F}_t\}$  and if  $s < t$  (so that  $\mathfrak{F}_s \subset \mathfrak{F}_t$ ),  $E\{Y|\mathfrak{F}_s\} = E\{E\{Y|\mathfrak{F}_t\}|\mathfrak{F}_s\}$ . The latter is called the iterated expectation formula. A succinct presentation of the above facts is in [3, Chapter I].

It is clear that a general stochastic process is a complicated object. The measure  $P$  requires for its construction an enormous amount of information contained in the finite-dimensional distributions. To obtain useful information, special classes of processes are introduced such as Gaussian processes, stationary processes, martingales, Markov processes, etc. Their study is the main objective of the theory of stochastic processes.

Gaussian processes are characterized by their mean function  $E\{X(t)\} = m(t)$  and their covariance

$$\rho(t, s) = E\{(X(t) - m(t))(X(s) - m(s))\},$$

since if  $0 \leq t_1 < t_2 < \dots < t_n < \infty$ , the joint characteristic function of  $X(t_1), X(t_2), \dots, X(t_n)$  is given by

$$\exp\left[-\frac{1}{2} \sum_{p,q=1}^n \rho(t_p, t_q) z_p z_q + i \sum_{p=1}^n m(t_p) z_p\right],$$

where  $(z_1, z_2, \dots, z_n) \in R^n$ . Stationary processes are characterized by the fact that the joint distribution of  $X(t_1 + h), X(t_2 + h), \dots, X(t_n + h)$  is independent of  $h$ , i.e., the statistics are translation invariant. Martingales (see (vii) ahead) are defined by  $\sup_{t \geq 0} E\{|X(t)|\} < \infty$ , and  $E\{X(t+s)|\mathfrak{F}_t\}$

$= X(s)$  for almost all realizations. The significance of such processes will become apparent later. Finally, Markov processes, which are of principal interest to us, are characterized by the property that if  $A$  is any event in  $\mathfrak{F}$  but not in  $\mathfrak{F}_t$  then

$$P \{A | \mathfrak{F}_t\} = E \{ \chi(A) | \mathfrak{F}_t \}$$

is a function of  $X(t)$  only. In words, the conditional probability of an event in the future given the past up to the present  $t$  depends only on the present.

(ii) *Brownian motion.* This process  $\{w(t), t \geq 0\}$  is defined as follows.

(a)  $w(0) = 0$ .

(b) For any  $0 \leq s_1 < t_1 \leq s_2 < t_2 \leq \dots < s_n \leq t_n < \infty$ ,  $w(t_1) - w(s_1)$ ,  $w(t_2) - w(s_2), \dots, w(t_n) - w(s_n)$  are independent random variables.

(c)  $w(t) - w(s)$  has Gaussian distribution with mean zero and variance  $t - s$ .

From the properties of the Gaussian distribution we have

$$E \{ (w(t) - w(s))^{2n} \} = \int (y - x)^{2n} \frac{\exp \left[ -(y - x)^2 / 2(t - s) \right]}{(2\pi(t - s))^{1/2}} dx \tag{2.1}$$

$$= c_n (t - s)^n$$

where  $c_n$  is a constant. Now a sufficient condition that a process have with probability one continuous trajectories is [3, p. 186] that there exist  $\alpha, \beta > 0$  and  $C < \infty$  such that

$$E \{ |X(t) - X(s)|^\alpha \} \leq C |t - s|^{1+\beta},$$

with  $t, s \in [0, \infty)$ . It is clear therefore that Brownian motion has continuous trajectories with probability one. We can now take as basic probability space  $(\Omega, \mathfrak{F}, P)$  the set  $\Omega$  of bounded continuous functions on  $[0, \infty)$  since  $P$  is concentrated on this set. The  $\sigma$ -algebra of cylinder sets is the same as the  $\sigma$ -algebra of Borel sets of  $\Omega$  viewed as a metric space with  $\sup_t |x(t) - y(t)|$  as metric.

It turns out that Brownian motion is nowhere differentiable with probability one [5]. This is a phenomenon inherited from the independent increments property; it is the root of the difficulties associated with the stochastic calculus. One can see formally from (2.1) that derivatives will not exist since

$$E \left| \frac{w(t) - w(s)}{t - s} \right| = \frac{\tilde{c}_1}{|t - s|^{1/2}}$$

where  $\tilde{c}_1$  is a constant.

Since  $w(t)$  is Gaussian, for any  $f(x)$  bounded and continuous  $u(t, x) = E \{ f(w(t) + x) \}$  exists and

$$u(t, x) = \int_{-\infty}^{\infty} f(z + x) \frac{\exp \left[ -z^2 / 2t \right]}{(2\pi t)^{1/2}} dz$$

so that

$$\frac{\partial u}{\partial t} = \frac{1}{2} \frac{\partial^2 u}{\partial x^2}, \quad t > 0, u(0, x) = f(x). \quad (2.2)$$

Note that for  $t > 0$ ,  $u(t, x)$  is  $C^\infty$  in  $x$ .

Brownian motion is a process of independent increments and hence Markovian. Let  $P(t, x, y)$  denote the transition probability density that  $w(s + t)$  will be at  $y$  given that  $w(s)$  equals  $x$ . Clearly

$$P(t, x, y) = \frac{\exp[-(x - y)^2/2t]}{(2\pi t)^{1/2}} \quad (2.3)$$

and also,  $u(t, x) = \int_{-\infty}^{\infty} P(t, x, y) f(y) dy$ . Equation (2.2) is called the backward equation associated with Brownian motion.  $P(t, x, y)$  satisfies (2.2) as a function of both  $x$  and  $y$  so the forward and backward equations are the same for Brownian motion.

An important characterization of Brownian motion is the following [2], [4]. Let  $\Omega$  be the space of continuous trajectories on  $[0, \infty)$  and let  $\mathcal{F}_t$  be the  $\sigma$ -algebra generated by the trajectories up to time  $t$ . Suppose  $\{w(t), t \geq 0\}$  is a process with the following properties.

(a)  $w(\cdot) \in \Omega$ , i.e.,  $w(\cdot)$  is continuous with probability one.

(b)  $E\{w(t) - w(s) | \mathcal{F}_s\} = 0, t \geq s$ .

(c)  $E\{(w(t) - w(s))^2 | \mathcal{F}_s\} = t - s, t \geq s$ . Then  $\{w(t), t \geq 0\}$  is the Brownian motion process.

The point of this characterization is that the Gaussian distribution for the increments is deduced from (a), (b) and (c) above.

(iii) *Orstein-Uhlenbeck process*. Brownian motion plays a basic role in diffusion theory since a large class of diffusions can be obtained as functionals of it (see (vi) below). On the other hand as a physical process meant to describe a particle undergoing many collisions, it has many drawbacks. One is that it is not differentiable and the other is that its transition density (2.3) does not stabilize for  $t$  large to a limit that can be connected with phenomenological thermodynamics.

To overcome these difficulties we suppose [7] that the velocity  $\{v(t), t \geq 0\}$  of the particle satisfies Langevin's equation

$$dv(t) + \gamma v(t) dt = \alpha dw(t), \quad v(0) = v_0, \gamma > 0, \quad (2.4)$$

where  $\{w(t), t \geq 0\}$  is Brownian motion. The stochastic differential equation (2.4) is easily solved, even though the right-hand side has no formal meaning, and

$$v(t) = e^{-\gamma t} v_0 + \alpha \int_0^t e^{-\gamma(t-s)} dw(s). \quad (2.5)$$

In (2.5) the integral is defined by integration by parts and this then is taken as the definition for  $v(t)$  with (2.4) being merely symbolic. The constant  $\gamma$  is called the friction constant and  $\alpha$  the noise intensity level; (2.4) has the form  $F = ma$  with  $\alpha dw$  the external (symbolic) force and  $-\gamma v$  the frictional force.

From (2.5) we deduce that  $v(t)$  is Gaussian and Markov and that its transition probability density is given by

$$P(t, v_0, v) = \frac{\exp\left[-\frac{\gamma(v - v_0 e^{-\gamma t})^2}{\alpha^2(1 - e^{-2\gamma t})}\right]}{(\pi(1 - e^{-2\gamma t})\alpha^2/\gamma)^{1/2}}. \quad (2.6)$$

To see how (2.6) arises it is enough to compute the mean and variance of  $v(t)$  from (2.5):

$$E\{v(t)\} = e^{-\gamma t}v_0,$$

$$E\{(v(t) - e^{-\gamma t}v_0)^2\} = \alpha^2 \int_0^t \int_0^t e^{-\gamma(t-s)} e^{-\gamma(t-\sigma)} E\{dw(s)dw(\sigma)\}.$$

Now  $E\{w(t)w(s)\} = \min(t, s)$ ; hence, formally,  $E\{dw(t)dw(s)\} = \delta(t - s) dt ds$  so that

$$E\{(v(t) - e^{-\gamma t}v_0)^2\} = \alpha^2 \int_0^t e^{-2\gamma(t-s)} ds = \frac{\alpha^2}{2\gamma} (1 - e^{-2\gamma t}).$$

Note that for  $t \ll 1$

$$P(t, v_0, v) \sim \frac{\exp[-(v - v_0)^2/2\alpha^2 t]}{(2\pi\alpha^2 t)^{1/2}},$$

and for  $t \gg 1$

$$P(t, v_0, v) \sim (\gamma/\pi\alpha^2)^{1/2} \exp[-(\gamma/\alpha^2)v^2]. \quad (2.7)$$

The friction constant  $\gamma$  can be related to the fluid density through which the particle is moving, its radius, assuming it spherical, etc. Thus  $\gamma$  is determinable from first principles. On the other hand  $\alpha$  comes into the problem in a somewhat artificial way. To determine it we assume that for  $t \gg 1$ , i.e., at thermodynamic equilibrium, the velocity of the particle has the Maxwell-Gibbs distribution  $ce^{-v^2/kT}$  where  $k$  is Boltzmann's constant and  $T$  is the absolute temperature. Matching (2.7) with this distribution gives

$$\alpha = \sqrt{kT\gamma} \quad (2.8)$$

and so the noise level is determined in terms of measurable quantities. Relation (2.8) is an instance of a fluctuation-dissipation theorem ( $\alpha$  measures the fluctuation,  $\gamma$  the dissipation) [8].

From (2.6) we deduce by direct computation that

$$u(t, v_0) = \int_{-\infty}^{\infty} P(t, v_0, v) f(v) dv$$

satisfies

$$\frac{\partial u}{\partial t} = \frac{\alpha^2}{2} \frac{\partial^2 u}{\partial v_0^2} - \gamma v_0 \frac{\partial u}{\partial v_0}, \quad u(0, v_0) = f(v_0)$$

which is the backward equation.  $P$  as a function of  $v$  satisfies also the forward equation

$$\frac{\partial P}{\partial t} = \frac{\alpha^2}{2} \frac{\partial^2 P}{\partial v^2} + \gamma \frac{\partial}{\partial v} (vP), \quad P(0, v_0, v) = \delta(v_0 - v).$$

Multi-dimensional versions of the above can be obtained easily. The class of Gaussian-Markov processes, although very special, has many applications primarily because it is perhaps the only class of processes for which the transition density is explicitly known and is relatively simple to work with.

An important problem arising from (2.4) is to find a mechanical model for this equation so that in some relevant limit the "white noise" driving force on the right-hand side is deduced, rather than postulated. This problem is solved in [9].

(iv) *Transition functions, backward equation, diffusion.* Markov processes are in most cases completely characterized by their transition function

$$P(t, x, A) = P\{X(t+s) \in A \mid X(s) = x\}.$$

Here  $A$  is a subset of  $E$ , the state space of  $\{X(t), t \geq 0\}$  which may be the real line, a discrete set, etc. We assume that  $P(t, x, A)$  is a probability measure for each  $x$ ,  $0 \leq P(t, x, A) \leq 1$ ,  $P(t, x, E) = 1$ ,  $t \geq 1$ , and that  $P(t, x, A_x) \rightarrow 1$  as  $t \rightarrow 0$  where  $A_x$  is any neighborhood of  $x$  and the limit is in a suitably uniform way in  $x$ . We also assume the Chapman-Kolmogorov relation,

$$P(t+s, x, A) = \int_E P(t, x, dy)P(s, y, A). \quad (2.9)$$

Let  $\mu_0(A)$  be the initial measure for  $X$ , i.e.,  $\mu_0(A) = P\{X(0) \in A\}$ . Then,

$$P\{X(t) \in A\} \equiv \mu_t(A) = \int_E \mu_0(dx)P(t, x, A), \quad t > 0,$$

defines a transformation  $S_t \mu_0 = \mu_t$  on probability measures on  $E$  such that, from (2.9),  $S_{t+s} = S_t S_s$ , i.e.,  $S_t$  is a semigroup. Similarly, for  $f(x)$  a bounded measurable function on  $E$

$$T_t f(x) \equiv E\{f(X(t)) \mid X(0) = x\} = \int_E P(t, x, dy)f(y), \quad t > 0,$$

defines a transformation that takes bounded measurable functions into bounded measurable functions and, from (2.9) again,  $T_{t+s} = T_t T_s$ , i.e., we have a semigroup.

It is convenient to study the process via the semigroup  $T_t$ ,  $t \geq 0$ , both because a certain amount of abstraction helps clarify the issues and because it provides well-organized means for discussing the problems; see [1, Volume II].

Assume now that  $E = R^n$  and that if  $C_0(R^n)$  denotes the space of bounded continuous functions on  $R^n$  vanishing at  $\infty$  then,

$$T_t f(x) = \int_{R^n} P(t, x, dy)f(y), \quad t > 0, \quad (2.10)$$

is also an element of  $C_0(R^n)$  and that

$$\lim_{t \downarrow 0} \|T_t f - f\| = \lim_{t \downarrow 0} \sup_{x \in R^n} |T_t f(x) - f(x)| = 0. \tag{2.11}$$

With these assumptions we have enough restrictions on  $P(t, x, A)$  so that it defines a strongly continuous contraction semigroup on the Banach space  $C_0(R^n)$ . The contractiveness is immediate.

$$\|T_t f\| = \sup_x \left| \int P(t, x, dy) f(y) \right| \leq \sup_x |f(x)| = \|f\|.$$

The point of semigroup theory is that, under favorable circumstances, the semigroup  $T_t$ , hence the transition function  $P(t, x, A)$  and hence the process, can be reconstructed from its infinitesimal generator  $A$  defined by

$$\lim_{h \downarrow 0} \frac{1}{h} (T_h f - f) \equiv A f$$

for all  $f$  for which the (norm) limit exists and which form the domain of  $A$ ,  $D(A)$ . From the point of view of physical applications, one usually has  $A$  available, as a differential or integral operator suggested from first principles, and one wants to solve for the transition function, i.e., solve the backward equation which is formally

$$(d/dt)(T_t f) = A(T_t f), \quad T_0 f = f,$$

or more carefully, for  $f \in D(A)$ ,

$$T_t f - f - \int_0^t T_s A f ds = 0. \tag{2.12}$$

Since  $A$  is usually given, the question is: does it determine  $T_t$  or not? This may mean that giving a differential operator is not enough if boundary conditions are not given as well. A useful result answering this question (see [1, Volume II]) is the Hille-Yoshida theorem:

In order that  $A$  be the infinitesimal generator of a strongly continuous contraction semigroup on  $C_0(R^n)$  it is necessary and sufficient that

- (a)  $D(A)$  be dense in  $C_0(R^n)$ ;
- (b) for  $\lambda > 0$ ,  $(\lambda - A)$  maps  $D(A)$  onto  $C_0(R^n)$ ;
- (c) for every  $f \in D(A)$  and  $\lambda > 0$ ,  $\|(\lambda - A)f\| \geq \lambda \|f\|$ . The necessity of these conditions is easy. If for  $\lambda > 0$  we let

$$R_\lambda f = \int_0^\infty e^{-\lambda t} T_t f dt,$$

then  $\|\lambda R_\lambda f\| \leq \|f\|$  which is the same as (c);  $R_\lambda f$  is in  $D(A)$ ,  $\lambda > 0$  and  $\lambda R_\lambda f = -f + \lambda R_\lambda f$  so (b) holds; and  $\lim_{\lambda \uparrow \infty} \lambda R_\lambda f = f$  so (a) holds also. The sufficiency requires more work [1, Volume II]. In specific problems condition (b) is usually the most difficult to verify.

Diffusion processes are characterized by the fact that for any  $\epsilon > 0$

$$\frac{1}{t} \int_{|x-y|>\epsilon} P(t, x, dy) \rightarrow 0 \quad \text{as } t \downarrow 0. \tag{2.13}$$



This condition alone [3, Chapter III] implies that  $\{X(t), t \geq 0\}$  is continuous with probability one. For  $\varepsilon > 0$  define  $a(x)$  and  $b(x)$  by (we let  $E = R$  for simplicity)

$$\lim_{t \downarrow 0} \frac{1}{t} \int_{|x-y| < \varepsilon} (x-y) P(t, x, dy) = b(x), \quad (2.14)$$

$$\lim_{t \downarrow 0} \frac{1}{t} \int_{|x-y| < \varepsilon} (x-y)^2 P(t, x, dy) = a(x). \quad (2.15)$$

Then we have that  $C^2(R)$ , functions with two continuous derivatives, is in  $D(A)$  and on  $C^2(R)$

$$Af(x) = \frac{1}{2} a(x) \partial^2 f(x) / \partial x^2 + b(x) \partial f(x) / \partial x, \quad (2.16)$$

so that the backward equation for  $u(t, x) = \int P(t, x, dy) f(y)$  is

$$\frac{\partial u(t, x)}{\partial t} = \frac{1}{2} a(x) \frac{\partial^2 u(t, x)}{\partial x^2} + b(x) \frac{\partial u(t, x)}{\partial x}, \quad (2.17)$$

$$t > 0, u(0, x) = f(x).$$

To verify these facts we note that

$$\begin{aligned} Af &= \lim_{h \downarrow 0} \frac{1}{h} (T_h f - f) = \lim_{h \downarrow 0} \frac{1}{h} \int P(h, x, dy) (f(y) - f(x)) \\ &= \lim_{h \downarrow 0} \left[ \frac{1}{h} \int_{|x-y| > \varepsilon} P(h, x, dy) (f(y) - f(x)) \right. \\ &\quad \left. + \frac{1}{h} \int_{|x-y| < \varepsilon} P(h, x, dy) \left[ f'(x)(x-y) \right. \right. \\ &\quad \left. \left. + \frac{1}{2} f''(x)(x-y)^2 + \rho(\varepsilon) \right] dy \right], \end{aligned}$$

where  $\rho(\varepsilon)$  goes to zero with  $\varepsilon$ . The hypotheses (2.13)–(2.15) now imply (2.16).

Let us note again that because no special assumptions have been made about  $a(x)$  and  $b(x)$ , there is no a priori reason that (2.17) be a well-posed problem. So, for example, the question of existence and uniqueness of the diffusion process given its infinitesimal mean  $b(x)$  and its infinitesimal variance  $a(x)$  is a difficult problem with many ramifications.

(v) *Trotter-Kato theorem.* This is a general semigroup approximation theorem that gives conditions for a sequence  $T_t^n$  of semigroups to converge as  $n \rightarrow \infty$  to another semigroup  $T_t$ . It is useful in proving, for example, diffusion approximations to random walks, convergence of finite difference approximations to the solution of differential equations, etc. For the kind of singular perturbations we have in mind it becomes useful after a preliminary adjustment.

Let  $T_t^n$ ,  $n = 1, 2, \dots$ , be strongly continuous contraction semigroups on a Banach space, say  $B$ . Let  $R_\lambda^n$ ,  $\lambda > 0$ , be the corresponding resolvents

$$R_\lambda^n = \int_0^\infty e^{-\lambda t} T_t^n dt.$$

Let  $T_t$  be another continuous semigroup and  $R_\lambda$  its resolvent. If

$$\sup_{0 < t < \tau} \|T_t^n f - T_t f\| \rightarrow 0, \quad \tau < \infty, f \in B, \quad (2.18)$$

as  $n \rightarrow \infty$  then clearly  $\|R_\lambda^n f - R_\lambda f\| \rightarrow 0$ ,  $\lambda > 0$ , as  $n \uparrow \infty$ . The converse of this (simplified somewhat) is one aspect of the Trotter-Kato theorem: If  $\|R_\lambda^n f - R_\lambda f\| \rightarrow 0$ ,  $\lambda > 0$ ,  $f \in B$ ,  $n \uparrow \infty$ , then the semigroups converge.

Naturally, showing that  $\|R_\lambda^n f - R_\lambda f\| \rightarrow 0$ ,  $n \uparrow \infty$ , is not easy. Suppose that  $A_n$  and  $A$  are the infinitesimal generators of  $T_t^n$  and  $T_t$  respectively. Suppose there exists a set  $D \subset B$  such that  $D$  is dense in  $B$ ,

$$A_n f \rightarrow A f, \quad f \in D, \quad (2.19)$$

and  $(\lambda - A)D$ ,  $\lambda > 0$ , is dense in  $B$ . Then the resolvents converge also. To see this we note that, for  $f \in D$ ,

$$T_t^n f - f - \int_0^t T_s^n A f ds = \int_0^t T_s^n (A_n - A) f ds,$$

and hence

$$\sup_{0 < t < \tau} \left\| T_t^n f - f - \int_0^t T_s^n A f ds \right\| \rightarrow 0.$$

From this it follows that, for  $f \in D$ ,

$$\|f - R_\lambda^n (\lambda - A) f\| \rightarrow 0, \quad \lambda > 0. \quad (2.20)$$

Let  $g \in (\lambda - A)D$ ,  $\lambda > 0$ , and put  $f = R_\lambda g$ . Then (2.20) implies that  $\|R_\lambda^n g - R_\lambda g\| \rightarrow 0$ ,  $\lambda > 0$ , for  $g \in (\lambda - A)D$ . But since  $R_\lambda$ ,  $R_\lambda^n$  are bounded and  $(\lambda - A)D$  is dense the result follows.

Condition (2.19) is impossible to satisfy in many problems of interest. This is because nonuniformities in the problem preclude the simple generator convergence (2.19). Kurtz [10] observed that it is enough that, exactly as above, for  $f \in D$  there exists a sequence  $f_n \in D$  such that

$$A_n f_n \rightarrow A f, \quad f_n \rightarrow f, \quad (2.21)$$

the extra freedom allowing for the removal of nonuniformities. To see how (2.21) implies (2.18) we note that

$$T_t^n f - f - \int_0^t T_s^n A f ds = T_t^n (f - f_n) - (f - f_n) + \int_0^t T_s^n (A^n f_n - A f) ds,$$

so

$$\sup_{0 < t < \tau} \left\| T_t^n f - f - \int_0^t T_s^n A f ds \right\| \rightarrow 0, \quad f \in D, \tau < \infty.$$

But this again implies (2.20) and hence, as above, the desired result (2.18).

The extension (2.21) of (2.19) may appear somewhat superficial at first but it is extremely useful as we shall see in §3 below.

(vi) *Itô's stochastic differential approach to diffusions.* Diffusion processes look locally like Brownian motion. This is so because in a small time interval  $\Delta t$  we have, from the above considerations, that  $x(t + \Delta t) - x(t)$  is approximately Gaussian with mean  $b(x(t))\Delta t$  and variance  $a(x(t))\Delta t$ . If  $w(t)$  is the Brownian motion process we may summarize this by writing

$$x(t + \Delta t) - x(t) \sim b(x(t))\Delta t + \sigma(x(t))(w(t + \Delta t) - w(t)) \quad (2.22)$$

where  $\sigma(x) = (a(x))^{1/2}$ . In the limit  $\Delta t \downarrow 0$  we may write formally

$$dx(t) = b(x(t))dt + \sigma(x(t))dw(t), \quad x(0) = x. \quad (2.23)$$

So far this is merely a restatement of the fact that  $A = \frac{1}{2}a(x)\partial^2/\partial x^2 + b(x)\partial/\partial x$  is the infinitesimal generator of  $\{x(t), t \geq 0\}$ . Can one make literal sense of this equation? If this is successful then we have a nice way to define diffusions pathwise as (nonlinear) functionals of Brownian motion via (2.23). Clearly the major difficulty is that  $w(t)$  is not differentiable and this is an essential aspect of its behavior as well as the pathology that emerges in studying (2.23).

We remark that the issue of making sense of (2.23) should not be confused with what "physical" stochastic differential equations should look like. In physical problems the random coefficients are always correlated processes and hence the resulting solutions are never diffusion processes. That they may be approximated well by diffusion processes calls for the analysis of some relevant asymptotic limits. It is not a matter subject to definitions.

To continue, we rewrite (2.23) as an integral equation

$$x(t) = x + \int_0^t b(x(s)) ds + \int_0^t \sigma(x(s)) dw(s). \quad (2.23')$$

We see that the first step in the analysis is the definition of the second integral on the right side of (2.23'), called an Itô stochastic integral. Integration by parts will not work since the integrand is also random. Furthermore, its definition must somehow include the intuitive idea expressed by (2.22), i.e., the Brownian increment must be in the future ( $\Delta t > 0$ ).

Let  $\Omega$  be the space of continuous trajectories on  $[0, T]$ ,  $T < \infty$ , and let  $\mathfrak{F}_t$ ,  $0 \leq t \leq T$ , be the  $\sigma$ -algebra generated by the cylinder sets up to time  $t$ . A real random variable on  $\Omega$  which is  $\mathfrak{F}_t$ -measurable, denoted by  $G(t) = G(t, \omega)$ ,  $\omega \in \Omega$ , is called a nonanticipating functional. Let  $\mathfrak{F}$  be the  $\sigma$ -algebra generated by cylinder sets and let  $P$  be the measure on  $\mathfrak{F}$  of the Brownian motion. We wish to define

$$\int_0^t G(s) dw(s), \quad 0 \leq t \leq T. \quad (2.24)$$

To begin with we define (2.24) for nonanticipating step functionals

$$\int_0^T G(s) dw(s) = \sum_{k=1}^n G(s_{k-1})[w(s_k) - w(s_{k-1})] \quad (2.25)$$

where  $s_0, s_1, \dots, s_n$  is a partition of  $[0, T]$  corresponding to the nonantic-

ipating step functional  $G$ . The point of the definition (2.25) is that the Brownian increments are in the future, which is what we wanted in (2.22). A different definition leads to different results contrary to what is the case with ordinary integrals.

We now state the following properties of Itô's integral, assuming that  $\int_0^T E \{G^2\} ds < \infty$ .

$$(a) \quad E \left\{ \int_0^t G dw \right\} = 0, \tag{2.26}$$

$$(b) \quad E \left\{ \left\{ \int_0^t G dw \right\}^2 \right\} = \int_0^t E \{G^2(s)\} ds.$$

(c)  $\int_0^t G dw$  is a.s. continuous as a function of  $t$  and it is a martingale (see (vii) below).

These properties are preserved under the limiting process necessary to define (2.24) for general nonanticipating functionals subject to  $\int_0^T E \{G^2\} ds < \infty$ ; see [4], [5].

With the definition and properties of stochastic integrals under control we return to (2.23). Assuming that  $b(x)$  and  $\sigma(x)$  are Lipschitz continuous, uniformly in  $x$ , then, an iteration argument resembling the usual one from ODE (but more complicated) leads to the existence and uniqueness of the diffusion process  $x(t)$  as a nonanticipating functional of Brownian motion.

What use can be made of this pathwise definition of diffusions? Most of what can be done depends on Itô's lemma, which is as follows.

Suppose  $f(x, t)$  has the bounded continuous derivatives. Then we have the identity

$$\begin{aligned} f(x(t), t) &= f(x(s), s) + \int_s^t \left( A + \frac{\partial}{\partial \gamma} \right) f(x(\gamma), \gamma) d\gamma \\ &+ \int_s^t \sigma(x(\gamma)) \frac{\partial f(x(\gamma), \gamma)}{\partial x} dw(\gamma), \quad 0 \leq s \leq t < \infty, \end{aligned} \tag{2.27}$$

where

$$A = \frac{1}{2} a(x) \partial^2 / \partial x^2 + b(x) \partial / \partial x, \quad a(x) = \sigma^2(x).$$

To see what this formula says we recall that if  $x(t)$  satisfies the deterministic ODE

$$dx(t)/dt = F(x(t), t), \quad x(0) = x,$$

then

$$f(x(t), t) = f(x(s), s) + \int_s^t \left( F(x(\gamma), \gamma) \frac{\partial}{\partial x} + \frac{\partial}{\partial \gamma} \right) f(x(\gamma), \gamma) d\gamma$$

is an obvious identity. So Itô's formula is a generalization of this fact for the stochastic equation (2.23).

A simple application of Itô's formula is this. Suppose  $u(t, x)$  is a solution of

$$\partial u(t, x) / \partial t = Au(t, x), \quad t > 0, u(0, x) = f(x). \tag{2.28}$$

Apply Itô's formula to  $u(t - s, x(s))$ , with  $t$  a parameter. This leads to

$$f(x(t)) = u(t, x) + \int_0^t \left( A + \frac{\partial}{\partial \gamma} \right) u(t - \gamma, x(\gamma)) d\gamma + \int_0^t \sigma(x(\gamma)) \frac{\partial u(t - \gamma, x(\gamma))}{\partial x} dw(\gamma).$$

Assuming that  $\sigma \partial u / \partial x$  is bounded then  $u(t, x) = E \{ f(x(t)) \}$ , since by (2.26)(a) the last term in Itô's formula drops out.

We have deduced by pathwise means that

$$u(t, x) = E \{ f(x(t)) \} = \int P(t, x, dy) f(y) \tag{2.29}$$

satisfies the backward equation (2.28) provided the latter has smooth solutions. Actually, one can deduce under appropriate conditions and using stochastic calculus that  $u(t, x)$  defined by (2.29) is smooth; therefore it must satisfy (2.28) [4], [5].

More interesting applications of Itô's formula are given in (viii).

(vii) *Martingales, stopping times, optional stopping theorem.* Let  $(\Omega, \mathfrak{F}, P)$  be a probability space and let  $\mathfrak{F}_t, t \geq 0$ , be an increasing family of  $\sigma$ -algebras contained in  $\mathfrak{F}$ . A process  $\{X(t), t \geq 0\}$  is a martingale if  $\sup_{t \geq 0} E \{|X(t)|\} < \infty$  and  $E \{X(t) | \mathfrak{F}_s\} = X(s)$  a.s., for all  $t \geq s$ . Note that Brownian motion is a martingale since  $E \{w(t) - w(s) | \mathfrak{F}_s\} = 0$ . Note also that stochastic integrals with, say,  $|G| \leq C$  are martingales

$$E \left\{ \int_0^t G(\sigma) dw(\sigma) | \mathfrak{F}_s \right\} = E \left\{ \int_s^t G(\sigma) dw(\sigma) | \mathfrak{F}_s \right\} + E \left\{ \int_0^s G(\sigma) dw(\sigma) | \mathfrak{F}_s \right\} = \int_0^s G(\sigma) dw(\sigma).$$

It is clear therefore that Itô's formula generates martingales in profusion.

In (ii) we saw that Brownian motion can be characterized by continuity of paths and

$$E \{ w(t) - w(s) | \mathfrak{F}_s \} = 0, \quad E \{ (w(t) - w(s))^2 | \mathfrak{F}_s \} = t - s,$$

which says that  $w(t)$  is a continuous square-integrable martingale with increasing process equal to  $t$ .

General diffusion processes can also be characterized in this martingale way, as is Brownian motion [11]. A number of interesting applications of this idea can be given in the area of asymptotics or limit theorems.

Aside from this connection with diffusions, we need for the next section the following two facts about martingales [1], [2], [5]:

- (a) the optional stopping theorem,
- (b) Kolmogorov's inequality.

Kolmogorov's inequality says that, for  $\epsilon > 0$ ,

$$P \left\{ \sup_{0 \leq t \leq T} X(t) > \epsilon \right\} \leq \frac{E \{ X(T)^+ \}}{\epsilon}$$

where  $X^+(t) = \max(0, X(t))$ . Without the sup inside  $P\{\}$  this is just Tchebyshev's inequality. The importance of having the sup inside is clear from the example of the next section.

A random variable  $t^*$  is called a stopping time if for all  $t \geq 0$  the event  $\{t^* > t\}$  belongs to  $\mathcal{F}_t$ , i.e., to decide if  $t^* > t$  or not, it is enough to look at information available up to time  $t$ . We denote by  $\mathcal{F}_{t^*}$  the  $\sigma$ -algebra of events generated by events of the form  $A \cap \{t^* > t\}$  with  $A \in \mathcal{F}$  and  $t \geq 0$ . The optional stopping theorem is now as follows.

Let  $s^* \leq t^*$  be two finite stopping times. Then

$$E\{X(t^*) | \mathcal{F}_{s^*}\} = X(s^*),$$

i.e., the definition of the martingale is still valid if  $t$  and  $s$  are finite stopping times.

The optional stopping theorem is most useful in carrying over to stochastic problems the method of characteristics, as we do in the next section.

General Markov processes, not necessarily diffusions, can be treated, under favorable circumstances, in much the same way as diffusions; but, because the backward equation will be an integrodifferential equation now, the results are not as detailed as for diffusions. The martingale characterization of these processes is again helpful, and Itô's formula holds again in a suitably generalized form.

(viii) *Applications of Itô's formula.* As mentioned above we shall give two applications. The first one concerns Lyapounov's criterion of stability. We recall the deterministic problem.

Let  $x(t)$  be the solution of

$$dx(t)/dt = F(x(t)), \quad x(0) = x,$$

where  $F(x)$  is a vector-valued function of  $x$  with  $F(0) = 0$ , i.e.,  $0$  is an equilibrium point for the system. Suppose that there is a Lyapounov function  $V(x)$ , i.e., a positive definite function  $V(x) \geq 0$ ,  $V(x) = 0 \Rightarrow x = 0$ , which is smooth, and

$$F(x) \cdot \partial V(x) / \partial x \leq 0, \quad x \text{ near zero.}$$

Then

$$V(x(t)) = V(x) + \int_0^t F(x(s)) \cdot \frac{\partial V(x(s))}{\partial x} ds \leq V(x),$$

and hence  $|x(t)|$  will remain near zero for all  $t \geq 0$  if the initial state  $x$  is sufficiently close to zero.

For stochastic (Itô) equations the problem is [4] to study the behavior of the diffusion  $x(t)$  when

$$dx(t) = b(x(t))dt + \sigma(x(t))dw(t), \quad x(0) = x,$$

and  $b(0) = 0$ ,  $\sigma(0) = 0$ , so that  $x = 0$  is again an equilibrium point (or a trap). Let  $A$  denote again the infinitesimal generator of  $\{X(t), t \geq 0\}$ . Suppose there exists a positive definite function  $V(x)$ , smooth near zero, such

that  $AV(x) \leq 0$ . Then for any  $\varepsilon_1, \varepsilon_2 > 0$  we can find a  $\delta > 0$  such that for  $|x| < \delta$ ,

$$P \left\{ \sup_{t>0} |X(t)| < x + \varepsilon_1 \right\} \geq 1 - \varepsilon_2, \quad (2.30)$$

i.e., the identically zero solution is stable in the above sense.

From Itô's formula we have

$$V(x(t)) = V(x) + \int_0^t AV(x(s)) ds + \int_0^t \sigma(x(s)) \frac{\partial V(x(s))}{\partial x} dw(s),$$

and we assume here that  $\sigma \partial V / \partial x$  is bounded. By hypothesis

$$0 \leq V(x(t)) \leq V(x) + \int_0^t \sigma(x(s)) \frac{\partial V(x(s))}{\partial x} dw(s) \equiv V(x) + f(t).$$

Now  $V(x) + f(t)$  is a nonnegative martingale, and hence by Kolmogorov's inequality we have

$$P \left\{ \sup_{t>0} (V(x) + f(t)) > V(x) + \varepsilon \right\} \leq \frac{V(x)}{\varepsilon + V(x)} \leq \frac{V(x)}{\varepsilon}.$$

Thus,

$$P \left\{ \sup_{t>0} V(x(t)) \leq V(x) + \varepsilon \right\} \geq 1 - \frac{V(x)}{\varepsilon}.$$

The positive definiteness of  $V(x)$  implies now the assertion (2.30).

The second application is the extension of the method of characteristics. We recall the deterministic situation.

Consider the first order equation

$$\frac{\partial u(t, x)}{\partial t} = b(x) \frac{\partial u(t, x)}{\partial x} + c(x)u(t, x) + d(x), \quad (2.31)$$

$$t > 0, u(0, x) = f(x).$$

Let  $x(t)$  be the solution of

$$dx(t)/dt = b(x(t)), \quad x(0) = x. \quad (2.32)$$

Then we can express the solution of (2.31) as

$$u(t, x) = \exp \left\{ \int_0^t c(x(s)) ds \right\} f(x(t)) + \int_0^t \exp \left\{ \int_0^s c(x(\sigma)) d\sigma \right\} d(x(s)) ds.$$

This is simply the method of characteristics. If (2.31) is to be solved in a region  $D \subset R^n$ , say, with boundary condition  $u(t, x) = g(x)$  for all  $x \in \partial D$  that can be reached by solutions of (3.32) with  $x(0) \in D$ , then we have

$$u(t, x) = \chi(t < t^*) \exp \left\{ \int_0^t c(x(s)) ds \right\} f(x(t))$$

$$+ \chi(t \geq t^*) \exp \left\{ \int_0^{t^*} c(x(s)) ds \right\} g(x(t^*))$$

$$+ \int_0^{t \wedge t^*} \exp \left\{ \int_0^s c(x(\sigma)) d\sigma \right\} d(x(s)) ds, \quad t \wedge t^* = \min(t, t^*).$$

Here  $t^*$  is the first time the solutions of (2.32) reach  $\partial D$  starting from  $x \in D$  and  $\chi(t < t^*) = 0$  if  $t \geq t^*$  and one otherwise. Similarly for  $\chi(t \geq t^*)$ .

The generalization of the above to the problem

$$\begin{aligned} \frac{\partial u}{\partial t} &= \frac{1}{2} a(x) \frac{\partial^2 u}{\partial x^2} + b(x) \frac{\partial u}{\partial x} + c(x)u + d(x), \quad t > 0, \\ u(0, x) &= f(x), \quad x \in D, \quad u(t, x) = g(x), \quad x \in \partial D, \end{aligned} \quad (2.33)$$

is immediate. Let  $\{x(t), t \geq 0\}$  be the diffusion process with stochastic differential equation

$$dx(t) = b(x(t))dt + \sigma(x(t))dw(t), \quad x(0) = x,$$

and let  $t^*$  = first time  $x(t)$  hits  $\partial D$  starting from the interior, which is a stopping time. In the vector case we assume that  $\sigma(x) = \sigma(x) \cdot (\sigma(x))^T$  where  $\sigma(x)$  is any Lipschitz continuous matrix. From Itô's formula and the optional stopping theorem with  $t \wedge t^*$  as the upper stopping time and zero as the lower one we readily obtain the representation

$$\begin{aligned} u(t, x) &= E \left\{ \chi(t < t^*) \exp \left\{ \int_0^t c(x(s)) ds \right\} f(x(t)) \right\} \\ &+ E \left\{ \chi(t \geq t^*) \exp \left\{ \int_0^{t^*} c(x(s)) ds \right\} g(x(t^*)) \right\} \\ &+ E \left\{ \int_0^{t \wedge t^*} \exp \left\{ \int_0^s c(x(s)) ds \right\} d(x(s)) ds \right\}. \end{aligned} \quad (2.34)$$

Representations such as (2.34) are useful in many ways, and sometimes they yield very simply results that are otherwise difficult to obtain, for example, in singular perturbation results when  $a(x)$  in (2.33) is replaced by  $\epsilon a(x)$  (small diffusion) and the limit  $\epsilon \downarrow 0$  is sought.

A very interesting analysis, using the Cameron-Martin formula [4], [5] not mentioned here, concerns equations of the form (2.33), with  $a(x)$  replaced by  $\epsilon a(x)$  when the orbits of the deterministic problem (2.32) have a stable singular point in  $D$  and never exit from  $D$ , is given in [12].

#### REFERENCES

1. W. Feller, *An introduction to probability theory*. Vols. I, II, Wiley, New York, 1966. MR 35 #1048.
2. J. L. Doob, *Stochastic processes*, Wiley, New York; Chapman & Hall, London, 1953. MR 15, 445.
3. I. I. Gihman and A. V. Skorohod, *The theory of stochastic processes*. I, "Nauka", Moscow, 1971; English transl., Die Grundlehren der math. Wissenschaften, Band 210, Springer-Verlag, New York and Berlin, 1974.
4. ———, *Stochastic differential equations*, "Naukova Dumka", Kiev, 1968; English transl., Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 72, Springer-Verlag, New York and Berlin, 1972. MR 41 #7777.
5. H. P. McKean, Jr., *Stochastic integrals*, Probability and Math. Statist., no. 5, Academic Press, New York and London, 1969. MR 40 #947.



6. S. Varadhan, *Stochastic processes*, Courant Institute of Mathematical Sciences, New York University, New York, 1968. MR 41 #4657.
7. G. E. Uhlenbeck and L. S. Ornstein, *On the theory of Brownian motion*, Phys. Rev. 36 (1930), 823–841; Reprinted in *Selected papers on noise and stochastic processes*, N. Wax, Editor, Dover, New York, 1953.
8. R. Kubo, *Statistical mechanics*, North-Holland, Amsterdam; Interscience, New York, 1965. MR 31 #3139.
9. G. W. Ford, M. Kac and P. Mazur, *Statistical mechanics of assemblies of coupled oscillators*, J. Mathematical Phys. 6 (1965), 504–515. MR 31 #4512.
10. T. G. Kurtz, *Extensions of Trotter's operator semigroup approximation theorems*, J. Functional Analysis 3 (1969), 354–375. MR 39 #3351.
11. D. Stroock and S. R. S. Varadhan, *Diffusion processes with continuous coefficients*. I, II, Comm. Pure Appl. Math. 22 (1969), 345–400, 479–530. MR 40 #6641; #8130.
12. A. D. Ventsel' and M. I. Freidlin, *On small perturbations of dynamical systems*, Uspehi Mat. Nauk 25 (1970), no. 1 (151), 3–55 = Russian Math. Surveys 25 (1970), no. 1, 1–56. MR 42 #2123.

### 3. Asymptotics for initial value problems.

(i) *Formulation of the problem.* Physical problems generate stochastic processes in many ways, one of which is as solutions of differential equations with random coefficients. The prototype problem is of the form

$$dx(t)/dt = F(x(t), y(t), t) \quad x(0) = x, \quad (3.1)$$

where  $x(t)$  is an  $R^n$ -valued function,  $y(t)$  is an  $R^m$ -valued given stochastic process, and  $F$  is a sufficiently regular vector-valued function. Equation (3.1) generates the process  $x(t)$  from the process  $y(t)$ . A basic question is: given the statistical behavior of  $y(t)$ , the coefficients, can we deduce in some reasonably effective way the statistical behavior of the solution  $x(t)$  of (3.1)?

To motivate the kind of answer we have in mind consider how equations (3.1) arise frequently in physical problems. In the differential equation (3.1) the coefficients  $y(t)$  represent external influences of a complicated nature. One can always enlarge the state space to include the dynamics of  $y(t)$ , but this may complicate the problem too much. So one assumes that the coefficients are random and looks for results concerning (3.1) that depend as little as possible on detailed description of  $y(t)$ , i.e., they are model-independent. To make this statement precise we must differentiate between the characteristic time scales of evolution of the system  $x(t)$  and the external noise effects  $y(t)$ : the latter change much faster than the former. It is in the asymptotic limit where this behavior is fully developed that interesting results about  $x(t)$  emerge. We shall therefore concentrate attention on this problem and we shall use perturbation theory to analyze it.

The quantitative differentiation of time scales is effected by the introduction of a small parameter  $\epsilon$ . Let us assume that  $y(t)$  is an  $R^m$ -valued stationary stochastic process where the correlation time, typically defined by

$$\sum_{i=1}^m \int_0^\infty E \{ y_i(0) y_i(t) \} dt$$

with  $E \{ y(t) \} = 0$ , is a finite number. For the process  $y^\epsilon(t) \equiv y(t/\epsilon^2)$  (we use  $\epsilon^2$

to avoid square roots later) the correlation time is of order  $\varepsilon^2$ . We seek now to analyze the process  $x^\varepsilon(t)$  when  $\varepsilon$  is small where

$$dx^\varepsilon(t)/dt = F^\varepsilon(x^\varepsilon(t), y^\varepsilon(t), t), \quad x^\varepsilon(0) = x. \quad (3.2)$$

Here  $F^\varepsilon(x, y, t)$  is a vector-valued function on  $R^n \times R^m \times [0, \infty)$  depending on  $\varepsilon$ . It will turn out later that  $F^\varepsilon$  can have the form

$$F = \frac{1}{\varepsilon} F_0 + F_1 + \varepsilon F_2 + \dots$$

where, usually,  $E\{F_0(x, y(t), t)\} = 0$ . In addition  $F^\varepsilon$  may depend explicitly on  $t/\varepsilon^2$  as well as  $t$ . All these possibilities will be discussed in the sequel and examples will be given.

The limit  $\varepsilon \rightarrow 0$  with  $0 \leq t \leq T$ ,  $T < \infty$ , in the above problems will lead to a diffusion approximation for  $x^\varepsilon(t)$  so we may call it the diffusion limit or the white noise limit (because the scaling of  $y^\varepsilon(s)$  makes it tend to white noise). However, both names are somewhat unsatisfactory because upon changing variables we find that we are considering the weak-coupling limit as follows. Suppose the primary problem is of the form

$$d\tilde{x}(\tau)/d\tau = F_0(\tilde{x}(\tau)) + \varepsilon \tilde{F}_1(\tilde{x}(\tau), y(\tau)), \quad \tilde{x}(0) = x, \quad (3.3)$$

where  $y(\tau)$  is a stationary process with correlation time equal to one, say. Here  $\varepsilon$  plays the role of coupling constant for the system (3.3); coupling the unperturbed problem ( $\varepsilon = 0$ ) to the noise. For finite  $\tau$ -intervals the limit  $\varepsilon \rightarrow 0$  will simply produce the unperturbed problem. The effects of the coupling will not be felt unless  $\tau$  is large, usually of order  $1/\varepsilon^2$ . Furthermore, for such cumulative effects to be meaningful, the unperturbed problem must have special behavior. For example, it may admit a change of dependent variables  $\tilde{x} \rightarrow x$  and independent variable  $t = \varepsilon^2\tau$ , so that in the new representation it has the form (3.2) with suitable  $F^\varepsilon$  on the right-hand side.

All these points are well understood in the context of deterministic problems [1], [2], [3]. We shall see that the situation is quite similar for stochastic problems. Basic references for the analysis of stochastic problems from the point of view adopted here are [3], [4], [5] of the references listed in the introduction.

(ii) *Equations with Markov coefficients.* We shall begin with relatively familiar objects, stochastic differential equations whose coefficients form a stationary Markov process.

The following simple example has been analyzed by Goldstein [4]; see also the random evolution work in [5]. Let  $\{y(t), t \geq 0\}$  be the random telegraph process that takes the values  $\pm\alpha$  and with mean time between switching equal to  $1/\beta$ ,  $\beta > 0$ . The transition function of the Markov process  $y(t)$  is easily seen to be

$$\begin{aligned} & (P\{y(t+s) = \pm\alpha | y(s) = \pm\alpha\}) \\ &= \frac{1}{2} \begin{pmatrix} 1 + e^{-2\beta t} & 1 - e^{-2\beta t} \\ 1 - e^{-2\beta t} & 1 + e^{-2\beta t} \end{pmatrix}, \end{aligned} \quad (3.4)$$

so that

$$\frac{1}{2} \frac{d}{dt} \begin{pmatrix} 1 + e^{-2\beta t} & 1 - e^{-2\beta t} \\ 1 - e^{-2\beta t} & 1 + e^{-2\beta t} \end{pmatrix} \Big|_{t=0} = \begin{pmatrix} -\beta & \beta \\ \beta & -\beta \end{pmatrix} \equiv A. \quad (3.5)$$

The right-hand side of (3.5) is the infinitesimal generator (matrix)  $A$  of  $\{y(t), t \geq 0\}$ . Note that as  $t \rightarrow \infty$  the transition function (matrix) tends to

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \quad (3.6)$$

so that if  $y(0)$  takes the values  $\pm \alpha$  with probability  $\frac{1}{2}$ , then  $\{y(t), t \geq 0\}$  is a stationary Markov process.

Consider now the stochastic differential equation

$$dx(\tau)/d\tau = y(\tau), \quad x(0) = x, \quad (3.7)$$

which trivially integrates to

$$x(\tau) = x + \int_0^\tau y(s) ds.$$

We think of  $y(\tau)$  as the velocity of a particle, taking values  $\pm \alpha$ , and  $x(\tau)$  its position with  $x$  the initial position. Clearly  $(x(\tau), y(\tau))$  jointly form a Markov process. Let us compute the infinitesimal generator of this process. Let  $f(x, y)$  be a function on  $R \times \{\alpha, -\alpha\}$ , i.e., a 2-vector function of  $x$ , and let

$$u(\tau, x, y) = E_{x,y} \{f(x(\tau), y(\tau))\} \quad (3.8)$$

where  $(x, y)$  denotes the starting position  $x$  and velocity  $y (= \pm \alpha)$  of the process  $(x(\tau), y(\tau))$ . Using (3.4) we deduce easily that

$$\partial u / \partial \tau = y \partial u / \partial x + Au, \quad u(0, x, y) = f(x, y), \quad (3.9)$$

where  $A$  of (3.5) acts on  $u(\tau, x, y)$  with the latter considered as a 2-vector  $(u(\tau, x, +\alpha), u(\tau, x, -\alpha))$ .

Let us now consider what happens if we take the weak coupling limit in (3.7), i.e., consider  $dx(\tau)/d\tau = \epsilon y(\tau)$  and let  $\tau \sim 1/\epsilon^2$ . Let us define new variables

$$t = \epsilon^2 \tau, \quad y^\epsilon(t) = y(t/\epsilon^2), \quad x^\epsilon(t) = x(t/\epsilon^2). \quad (3.10)$$

Then we find that the equation for  $x^\epsilon(t)$  is

$$dx^\epsilon(t)/dt = (1/\epsilon)y^\epsilon(t), \quad x^\epsilon(0) = x. \quad (3.11)$$

Thus, as mentioned in (i), the study of (3.11) in the limit  $\epsilon \rightarrow 0$  could be called the diffusion or white noise limit since the velocity  $(1/\epsilon)y^\epsilon(t)$  tends to become white noise. The scaled version of (3.9) becomes

$$u^\epsilon(t, x, y) = E_{x,y} \{f(x^\epsilon(t), y^\epsilon(t))\},$$

$$\frac{\partial u^\epsilon}{\partial t} = \frac{1}{\epsilon} \frac{\partial u^\epsilon}{\partial x} + \frac{1}{\epsilon^2} Au^\epsilon, \quad u^\epsilon(0, x, y) = f(x, y). \quad (3.12)$$

Let us write (3.12) more explicitly with  $u(t, x, +\alpha) = u_+(t, x)$  and  $u(t, x, -\alpha) = u_-(t, x)$  as follows:

$$\begin{aligned} \frac{\partial}{\partial t} \begin{pmatrix} u_+(t, x) \\ u_-(t, x) \end{pmatrix} &= \frac{1}{\varepsilon} \begin{pmatrix} +\alpha & 0 \\ 0 & -\alpha \end{pmatrix} \frac{\partial}{\partial x} \begin{pmatrix} u_+(t, x) \\ u_-(t, x) \end{pmatrix} \\ &+ \frac{1}{\varepsilon^2} \begin{pmatrix} -\beta & \beta \\ \beta & -\beta \end{pmatrix} \begin{pmatrix} u_+(t, x) \\ u_-(t, x) \end{pmatrix} \end{aligned}$$

$$u_{\pm}(0, x) = f_{\pm}(x).$$

In this form the problem becomes a singular perturbation of a second order linear hyperbolic system. If  $f_+(x) = f_-(x)$ , so that no initial layers arise, it is easy to verify that  $w = u_+ + u_-$  satisfies the telegrapher's equation

$$\varepsilon^2 \frac{\partial^2 w}{\partial t^2} = \alpha^2 \frac{\partial^2 w}{\partial x^2} - 2\beta \frac{\partial w}{\partial t}, \quad w(0, x) = 2f(x), \quad \frac{\partial w}{\partial t}(0, x) = 0. \quad (3.13)$$

The limit  $\varepsilon \rightarrow 0$  in (3.12) will be studied formally in the next section for a large class of similar problems. From (3.13) it is clear however that we will end up with a diffusion equation which is no more than a statement of the central limit theorem for  $\int_0^t y^\varepsilon(s) ds$  as  $\varepsilon \rightarrow 0$ .

The general case of stochastic equations with Markovian coefficients is similar. We shall present the problem in scaled form directly, the other case being only superficially different. Suppose  $y(t)$  is an  $R^m$ -valued stationary Markov process, not necessarily a diffusion, with finite correlation time. We define  $y^\varepsilon(t) \equiv y(t/\varepsilon^2)$  and let  $x^\varepsilon(t)$  be the solution of

$$\begin{aligned} \frac{dx^\varepsilon(t)}{dt} &= \frac{1}{\varepsilon} F(x^\varepsilon(t), y^\varepsilon(t)) + G(x^\varepsilon(t), y^\varepsilon(t)), \quad t > 0, \\ x^\varepsilon(0) &= x. \end{aligned} \quad (3.14)$$

We are interested in the asymptotic behavior of the statistics of  $x^\varepsilon(t)$ , as  $\varepsilon \rightarrow 0$  and  $0 \leq t \leq T$ ,  $T < \infty$ . The case where

$$E\{F(x, y(t))\} = 0 \quad (3.15)$$

is typically assumed in (3.14) because then both terms on the right side produce comparable effects. Without (3.15) the  $F/\varepsilon$  term in (3.14) dominates. Both cases are treated later on.

To study (3.14) we consider the process  $(x^\varepsilon(t), y^\varepsilon(t))$  as a Markov process on  $R^n \times R^m$  and determine its infinitesimal generator. It is not difficult to verify that if  $A$  is the infinitesimal generator of  $y(t)$ , i.e. if for suitable functions  $f(y)$  on  $R^m$

$$Af(y) = \lim_{h \downarrow 0} \frac{E\{f(y(h)) | y(0) = y\} - f(y)}{h},$$

then the infinitesimal generator of  $(x^\varepsilon(t), y^\varepsilon(t))$ , acting on suitable functions  $f(x, y)$  on  $R^{n+m}$ , is given by

$$\begin{aligned} \lim_{h \downarrow 0} \frac{E\{f(x^\varepsilon(h), y^\varepsilon(h))\} - f(x, y)}{h} \\ = \left( \frac{1}{\varepsilon} F(x, y) + G(x, y) \right) \frac{\partial f(x, y)}{\partial x} + \frac{1}{2} Af(x, y). \end{aligned} \quad (3.16)$$

Here  $G \partial f / \partial x$  denotes the dot product of the vector  $G$  with the  $x$ -gradient of  $x$  and  $A$  acts on  $f$  as a function of  $y$  only. Since we are specifically interested in the limit of  $x^\epsilon(t)$ ,  $0 \leq t \leq T$ , as  $\epsilon \downarrow 0$ , we must analyze the asymptotic behavior of  $u^\epsilon(t, x, y) = E\{f(x^\epsilon(t))\}$  satisfying

$$\begin{aligned} \frac{\partial u^\epsilon(t, x, y)}{\partial t} &= \left( \frac{1}{\epsilon} F(x, y) + G(x, y) \right) \frac{\partial u^\epsilon(t, x, y)}{\partial x} \\ &+ \frac{1}{\epsilon^2} A u^\epsilon(t, x, y), \quad t > 0, \end{aligned} \quad (3.17)$$

$$u^\epsilon(0, x, y) = f(x).$$

Here  $f(x)$  must belong to a sufficiently rich class of functions (compare with the results in §2(v)).

Example (3.14) leading to (3.17) is typical of the general situation that asymptotic problems for stochastic equations produce: singular perturbation problems of the form (3.17). As we shall see later (§3(vi)), non-Markovian problems give rise to problems formally similar to (3.17), and as mentioned in the introduction an even greater variety of problems also falls in the same category; see [5]–[8] of the introduction.

(iii) *Perturbation theory.* As we saw in the previous section, problems of the form

$$\frac{du^\epsilon}{dt} = \frac{1}{\epsilon^2} \mathcal{L}_1 u^\epsilon + \frac{1}{\epsilon} \mathcal{L}_2 u^\epsilon + \mathcal{L}_3 u^\epsilon, \quad u^\epsilon(0) = f, \quad (3.18)$$

arise where  $\mathcal{L}_1$ ,  $\mathcal{L}_2$  and  $\mathcal{L}_3$  are certain operators which are in general unbounded but for which (3.18) has solutions in a suitably generalized sense. The problem is to determine the asymptotic behavior of  $u^\epsilon$  as  $\epsilon \rightarrow 0$ , with  $0 \leq t \leq T$ ,  $T < \infty$ .

There are many formal procedures for treating (3.18), but they all amount, ultimately, to perturbation theory (usually second order) for  $\mathcal{L}_1 + \epsilon \mathcal{L}_2 + \epsilon^2 \mathcal{L}_3$  [6, Chapter IV, for example]. The method given in [2] of the introduction (for nonlinear problems there, in fact) will be discussed first.

We shall assume that  $\mathcal{L}_1$  has certain special properties corresponding to the fact that it is usually, but not always, the generator of a stationary (ergodic) Markov process. We assume specifically that the semigroup  $e^{\mathcal{L}_1 t}$  generated by  $\mathcal{L}_1$  converges to  $P$  as  $t \uparrow \infty$

$$e^{\mathcal{L}_1 t} \rightarrow P, \quad t \uparrow \infty, \quad (3.19)$$

where  $P$  is the projection operator into the null space of  $\mathcal{L}_1$ . The operators are all defined on a Banach space  $B$  and the range of  $P$ ,  $PB$ , is denoted by  $B_0$ . The details of the formalism that follows are in [7], [8] so we shall be rather brief here. Let us also suppose at first that

$$P \mathcal{L}_2 P = 0, \quad (3.20)$$

which corresponds to (3.15), since the expectation there is relative to the invariant distribution and  $P$  in (3.19) is, in case (3.14), expectation relative to the invariant distribution of the process  $y(t)$ .

Let  $u^\epsilon$  be formally represented by a power series

$$u^\epsilon = u_0 + \epsilon u_1 + \epsilon^2 u_2 + \dots$$

Inserting this into (3.18) and equating coefficients of equal powers of  $\epsilon$ , we obtain the following sequence of problems:

$$\mathcal{L}_1 u_0 = 0, \quad (3.21)$$

$$\mathcal{L}_1 u_1 = -\mathcal{L}_2 u_0, \quad (3.22)$$

$$\mathcal{L}_1 u_2 = -\mathcal{L}_2 u_1 - \mathcal{L}_3 u_0 + du_0/dt, \quad (3.23)$$

....

We shall assume that  $Pf = f$ , which for (3.17) means that  $f = f(x)$  and not a function of  $x$  and  $y$ . This typically is the interesting case in our problems. Let us first consider (3.21). It implies that

$$Pu_0 = u_0, \quad (3.24)$$

i.e.,  $u_0 = u_0(t)$  is in the range of  $P$ . Since  $u_0(0) = f$  from (3.18), the condition  $Pf = f$  avoids nonuniformities near  $t = 0$ , i.e., initial layers.

We insert (3.24) into (3.22) and consider the solvability of this equation for  $u_1$ . Clearly for the formal expansion to hold,  $u_1, u_2$ , etc., should be determinable in a recursive way; but so far  $u_0$  is not determined except that it must satisfy (3.24). The solvability condition for  $u_1, u_2$ , etc., will determine  $u_0$  and possibly other unknown functions, and this constitutes the essence of the method; see [1], [2] of the introduction. The solvability condition for (3.22) is formally  $P\mathcal{L}_2 Pu_0 = 0$ , and this is identically satisfied by (3.20). So we must continue with (3.23) to determine  $u_0$  (second order theory). Inserting  $u_1$  from (3.22) on the right-hand side of (3.23) and acting with  $P$  on the result yields the following equation for  $u_0$ .

$$P\mathcal{L}_2\mathcal{L}_1^{-1}\mathcal{L}_2 Pu_0 - P\mathcal{L}_3 Pu_0 + (d/dt)Pu_0 = 0, \quad u_0(0) = f.$$

Remembering that (3.24) holds and  $Pf = f$ , we rewrite this as an evolution equation in  $B_0 = PB$ :

$$du_0/dt = (P\mathcal{L}_3 P - P\mathcal{L}_2\mathcal{L}_1^{-1}\mathcal{L}_2 P)u_0 \equiv \bar{\mathcal{L}}u_0, \quad u_0(0) = f. \quad (3.25)$$

We usually stop with  $u_0$  in (3.20) because the solution of (3.25) is already quite complicated in most problems of interest. But the algorithm described above can continue without difficulty. In particular, up to terms in  $B_0$ ,  $u_1$  and  $u_2$  are given by

$$u_1 = -\mathcal{L}_1^{-1}\mathcal{L}_2 u_0, \quad u_2 = -\mathcal{L}_1^{-1}[\mathcal{L}_3 P - \mathcal{L}_2\mathcal{L}_1^{-1}\mathcal{L}_2 P - \bar{\mathcal{L}}]u_0, \quad (3.26)$$

assuming, of course, that the formal solvability conditions we have used (elimination of the component of the right-hand sides of (3.22), (3.23) in the null space of  $\mathcal{L}_1$ ) are truly correct. This actually turns out to be a delicate point with many problems which are formally of the form (3.18) (see §3(vi)).

Despite the fact that the above considerations are formal, we have all the elements of a rigorous proof at hand. The idea is to use (2.21) and then the Trotter-Kato theorem [7], [8]. We simplify the relevant steps here as follows:

Suppose that (3.25) has a nice solution  $u_0$ , and suppose  $u_1$  and  $u_2$  of (3.26) are well defined and bounded in the norm of our Banach space  $B$ . We wish to estimate  $u^\epsilon - u_0$ ; but, in view of what has just been assumed, it is enough to show that  $u^\epsilon - u_0 - \epsilon u_1 - \epsilon^2 u_2$  is small. This is a "bootstrap" approach familiar in singular perturbation. Now we have

$$\begin{aligned} & \left( \frac{d}{dt} - \frac{1}{\epsilon^2} \mathcal{L}_1 - \frac{1}{\epsilon} \mathcal{L}_2 - \mathcal{L}_3 \right) (u^\epsilon - u_0 - \epsilon u_1 - \epsilon^2 u_2) \\ &= \left( \frac{1}{\epsilon^2} \mathcal{L}_1 + \frac{1}{\epsilon} \mathcal{L}_2 + \mathcal{L}_3 - \frac{d}{dt} \right) (u_0 + \epsilon u_1 + \epsilon^2 u_2) \quad (3.27) \\ &= \epsilon \left[ \left( \mathcal{L}_3 - \frac{d}{dt} \right) u_1 + \mathcal{L}_2 u_2 \right] + \epsilon^2 \left( \mathcal{L}_3 - \frac{d}{dt} \right) u_2 = O(\epsilon), \end{aligned}$$

formally. Under favorable circumstances (3.27) implies that  $u - u_0 - \epsilon u_1 - \epsilon^2 u_2$ , and hence  $u^\epsilon - u_0$ , is  $O(\epsilon)$  and the desired approximation theorem follows. Note that the bootstrap idea fits exactly into the scheme (2.21). Moreover, (2.19) cannot possibly be valid for problems such as (3.18).

In the event that (3.20) is not valid, then the first approximation  $u_0$  will depend on  $\epsilon$  and we will denote it by  $u_0^\epsilon$ . The result is again that  $u^\epsilon - u_0^\epsilon \rightarrow 0$  as  $\epsilon \rightarrow 0$ ,  $0 \leq t \leq T$ , where  $u_0^\epsilon$  is in  $B_0$  and satisfies the evolution equation

$$\begin{aligned} \frac{du_0^\epsilon}{dt} &= \left( \frac{1}{\epsilon} P \mathcal{L}_2 P + P \mathcal{L}_3 P - P \mathcal{L}_2 \mathcal{L}_1^{-1} (\mathcal{L}_2 - P \mathcal{L}_2) \right) u_0^\epsilon = 0, \quad (3.28) \\ u_0^\epsilon(0) &= f. \end{aligned}$$

The details of this are given in [8].

The above analysis is quite satisfactory since it is simple and it gives a good clue for the actual estimation of the error. It is worthwhile, however, to give another formal procedure which is employed widely in the analysis of stochastic equations and other problems.

Let us assume for simplicity that  $\mathcal{L}_3 \equiv 0$  in (3.18) and that (3.20) holds. Define

$$v^\epsilon(t) = P u^\epsilon(t), \quad w^\epsilon(t) = (1 - P) u^\epsilon(t), \quad (3.29)$$

where  $P$  is the projection operator of (3.19). Equation (3.18) now takes the form

$$\begin{aligned} \frac{dv^\epsilon}{dt} &= \frac{1}{\epsilon} P \mathcal{L}_2 w^\epsilon, \quad v^\epsilon(0) = f \quad (Pf = f), \\ \frac{dw^\epsilon}{dt} &= \frac{1}{\epsilon^2} \mathcal{L}_1 w^\epsilon + \frac{1}{\epsilon} \mathcal{L}_2 v^\epsilon + \frac{1}{\epsilon} (1 - P) \mathcal{L}_2 w^\epsilon, \quad w^\epsilon(0) = 0. \end{aligned} \quad (3.30)$$

The second equation in (3.30) can be solved formally to yield

$$w^\epsilon(t) = \frac{1}{\epsilon} \int_0^t \exp\{(\mathcal{L}_1 + \epsilon(1 - P)\mathcal{L}_2)(t - s)/\epsilon^2\} \mathcal{L}_2 v^\epsilon(s) ds.$$

Inserting this into the first equation in (3.30) we obtain a closed equation for  $v^\epsilon(t)$ :

$$\frac{dv^\varepsilon(t)}{dt} = \frac{1}{\varepsilon^2} \int_0^t P \mathcal{L}_2 \exp\{(\mathcal{L}_1 + \varepsilon(1-P)\mathcal{L}_2)(t-s)/\varepsilon^2\} \mathcal{L}_2 v^\varepsilon(s) ds, \quad (3.31)$$

$$v^\varepsilon(0) = f.$$

This equation is exact, of course. The first order smoothing approximation, so-called, consists in dropping the  $\varepsilon(1-P)\mathcal{L}_2$  term in the exponent in (3.31) so that  $v_{\text{FOS}}^\varepsilon(t)$  satisfies

$$\frac{dv_{\text{FOS}}^\varepsilon(t)}{dt} = \frac{1}{\varepsilon^2} \int_0^t P \mathcal{L}_2 \exp\{\mathcal{L}_1(t-s)/\varepsilon^2\} \mathcal{L}_2 v_{\text{FOS}}^\varepsilon(s) ds, \quad (3.32)$$

$$v^\varepsilon(0) = f.$$

Usually this equation is too difficult to solve, so we consider the long-time Markovian approximation  $v_{\text{LTM}}(t)$  which satisfies the equation

$$\frac{dv_{\text{LTM}}(t)}{dt} = \int_0^\infty P \mathcal{L}_2 e^{\mathcal{L}_1 t'} \mathcal{L}_2 P dt' v_{\text{LTM}}(t) = -P \mathcal{L}_2 \mathcal{L}_1^{-1} \mathcal{L}_2 P v_{\text{LTM}}(t), \quad (3.33)$$

$$v_{\text{LTM}}(0) = f.$$

Thus, the LTM approximation is the same approximation as the  $u_0$  we obtained above. The name LTM comes from the fact that, in passing from (3.32) to (3.33), the unknown function  $v_{\text{FOS}}^\varepsilon(s)$  is pulled out of the integral and the integral is then stretched to  $\infty$  so that "memory" effects are lost. A good reference for these ideas is [5] of the introduction, where additional references are given; but the terminology is not used consistently in the literature; see also [9].

It is believed that the FOS approximation is better than the LTM approximation, since one is a special case of the other, insofar as the former gives a more accurate description at very small  $t$ . Experience with applications so far shows that the LTM approximation is hard enough to implement so that the extra accuracy of the FOS does not have significant impact.

(iv) *Applications.* Let us first apply the approximation (3.25) to (3.12). Here  $\mathcal{L}_1 = A$  and is given by (3.5) and  $e^{\mathcal{L}_1 t'}$  is given by (3.4) so that

$$P = \lim_{t' \uparrow \infty} e^{\mathcal{L}_1 t'} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \quad (3.34)$$

The operator  $\mathcal{L}_2$  is given by

$$\mathcal{L}_2 = \begin{pmatrix} \alpha & 0 \\ 0 & -\alpha \end{pmatrix} \frac{\partial}{\partial x},$$

and clearly  $P \mathcal{L}_2 P = 0$ . The operator  $\mathcal{L}_3 \equiv 0$ . An easy computation yields

$$-\mathcal{L}_1^{-1} = \int_0^\infty (e^{\mathcal{L}_1 t'} - P) dt' = \frac{1}{4\beta} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},$$

and hence

$$-P \mathcal{L}_2 \mathcal{L}_1^{-1} \mathcal{L}_2 P = \frac{\alpha^2}{4\beta} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \frac{\partial^2}{\partial x^2}.$$



Thus, if  $f_+(x) = f_-(x) = f(x)$ , then  $u_{\pm}^{\varepsilon}(t, x)$  converges as  $\varepsilon \rightarrow 0$  to  $u_0(t, x)$  where

$$\frac{\partial u_0(t, x)}{\partial t} = \frac{\alpha^2}{2\beta} \frac{\partial^2 u_0(t, x)}{\partial x^2}, \quad u_0(0, x) = f(x),$$

just as we expected from (3.13).

Naturally the example (3.12) is too simple to show the effectiveness of the approximation. Example (3.17), however, which corresponds to (3.14), is considerably more involved. The formal identification (3.17) with the notation of §3(iii) is as follows.  $\mathcal{L}_1 = A$ , the infinitesimal generator of  $\{y(t), t \geq 0\}$ ,  $P = E\{\cdot\}$  where  $E\{\cdot\}$  is expectation with respect to the invariant distribution of  $y(t)$ ,  $\mathcal{L}_2 = F(x, y)\partial/\partial x$  so that, by (3.5),  $P\mathcal{L}_2P = 0$  and  $\mathcal{L}_3 = G(x, y)\partial/\partial x$ . The operator  $\mathcal{L}_2^{-1}$  is the only thing that requires some discussion. Let  $P(t, y, dz)$  be the transition function of  $\{y(t), t \geq 0\}$  and assume that

$$\chi(y, dz) = \int_0^{\infty} [P(t, y, dz) - \bar{P}(dz)] dt \quad (3.35)$$

is well defined. Here  $\bar{P}(dz)$  is the invariant measure of  $\{y(t), t \geq 0\}$  and (3.35) is called the recurrent potential kernel of the process; it is the kernel of the integral operator  $-\mathcal{L}_1^{-1}$ , as can be easily verified. With these definitions we find that

$$\begin{aligned} \bar{\mathcal{L}}f(x) &= \int \int F(x, y) \frac{\partial}{\partial x} \left( F(x, z) \frac{\partial f(x)}{\partial x} \chi(y, dz) \bar{P}(dy) \right) \\ &\quad + \int G(x, y) \frac{\partial f(x)}{\partial x} \bar{P}(dy). \end{aligned} \quad (3.36)$$

Thus  $x^{\varepsilon}(t)$ , the solution of (3.14), converges to a diffusion process with generator  $\bar{\mathcal{L}}$  given by (3.36).

We rewrite  $\bar{\mathcal{L}}$  in another form that will be useful later.

$$\begin{aligned} \bar{\mathcal{L}}f(x) &= \int_0^{\infty} E \left\{ F(x, y(0)) \frac{\partial}{\partial x} \left( F(x, y(t)) \frac{\partial f(x)}{\partial x} \right) \right\} dt \\ &\quad + E \left\{ G(x, y(t)) \frac{\partial f(x)}{\partial x} \right\}. \end{aligned} \quad (3.37)$$

The identification of (3.36) and (3.37) is merely a matter of notation. We shall find in §3(vi) that the same result (3.37) holds for  $x^{\varepsilon}(t)$  of (3.14) even if  $y(t)$  is not Markovian but only stationary along with some additional assumptions discussed there.

As a final example consider the system

$$\begin{aligned} \frac{dx^{\varepsilon}(t)}{dt} &= \frac{1}{\varepsilon} y^{\varepsilon}(t), & x^{\varepsilon}(0) &= x, \\ \frac{dy^{\varepsilon}(t)}{dt} &= \frac{1}{\varepsilon} F(x^{\varepsilon}(t)) + \frac{1}{\varepsilon} \left( \frac{dw(t)}{dt} - y^{\varepsilon}(t) \right), & y^{\varepsilon}(0) &= y, \end{aligned} \quad (3.38)$$

where  $w(t)$  is the Brownian motion processes and  $dw/dt$  is formal white noise. The backward equation for the diffusion Markov process  $(x^\epsilon(t), y^\epsilon(t))$  is

$$\frac{\partial u^\epsilon}{\partial t} = \frac{1}{\epsilon} y \frac{\partial u^\epsilon}{\partial x} + \frac{1}{\epsilon} F(x) \frac{\partial u^\epsilon}{\partial y} + \frac{1}{\epsilon^2} \left( \frac{1}{2} \frac{\partial^2 u^\epsilon}{\partial y^2} - y \frac{\partial u^\epsilon}{\partial y} \right), \quad (3.39)$$

$$u^\epsilon(0, x, y) = f(x),$$

with  $u^\epsilon(t, x, y) = E\{f(x^\epsilon(t))\}$ . Thus,

$$\mathfrak{L}_1 = \frac{1}{2} \frac{\partial^2}{\partial y^2} - y \frac{\partial}{\partial y}, \quad \mathfrak{L}_2 = y \frac{\partial}{\partial x} + F(x) \frac{\partial}{\partial y}, \quad \mathfrak{L}_3 = 0.$$

The problem under consideration is called the Smoluchovski limit of the Ornstein-Uhlenbeck process in an external force field; see [10].

The perturbation analysis of §3(iii) yields the result that  $x^\epsilon(t)$  converges to a diffusion process, i.e.,  $u^\epsilon(t, x, y) \rightarrow u_0(t, x)$  and

$$\frac{\partial u_0}{\partial t} = \frac{1}{2} \frac{\partial^2 u_0}{\partial x^2} + F(x) \frac{\partial u_0}{\partial x}.$$

We leave the details to the reader.

Let us remark that the Banach spaces in the applications to stochastic equations with Markovian coefficients are usually spaces of bounded continuous functions with sup norm so that convergence of  $u^\epsilon$  to  $u_0$  is in the sup norm. When it is difficult to ascertain the properties of  $\mathfrak{L}_1^{-1}$ , the principal analytical and computational difficulty of our problem, other more appropriate spaces may be used. Some remarks on this are in [8].

(v) *Linear stochastic equations, averaging, master equations.* Linear equations are of particular interest so we discuss them separately.

Consider the system

$$\frac{d\tilde{x}_p(\tau)}{d\tau} = i\omega_p \tilde{x}_p(\tau) + i\epsilon \sum_{q=1}^n y_{pq}(\tau) \tilde{x}_q(\tau), \quad (3.40)$$

$$p = 1, 2, \dots, n, \quad \tilde{x}_p(0) = x_{p0}.$$

Here  $\omega_p$ ,  $p = 1, 2, \dots, n$ , are positive real numbers and  $(y_{pq}(t)) = (y_{qp}(t))$  is a real symmetric matrix-valued stochastic process which is Markovian with known infinitesimal generator. The Markovian assumption is removed in the next section.

We study (3.40) in the weak coupling limit, i.e., for  $\epsilon \rightarrow 0$  and  $\tau$  of order  $1/\epsilon^2$ . It is convenient therefore to introduce the scaled time and the adiabatic invariants

$$\tau = t/\epsilon^2,$$

$$x_p^\epsilon(t) = \exp[-i\omega_p t/\epsilon^2] \tilde{x}_p(t/\epsilon^2),$$

so that (3.40) yields

$$\frac{dx_p^\epsilon(t)}{dt} = \frac{i}{\epsilon} \sum_{q=1}^n y_{pq}(t/\epsilon^2) \exp[i(\omega_q - \omega_p)t/\epsilon^2] x_q^\epsilon(t), \tag{3.41}$$

$$t > 0, p = 1, 2, \dots, n, x_p^\epsilon(0) = x_{p0}.$$

The question of interest now is: what is the asymptotic distribution of  $(x_1^\epsilon(t), \dots, x_n^\epsilon(t))$  in the limit  $\epsilon \downarrow 0, 0 \leq t \leq T$ ?

We shall assume that the matrix-valued process  $(y_{pq}(t))$  is an ergodic Markov process with generator  $\mathcal{L}_1$  whose exact form is not important. Instead of writing  $P(t, y, A)$  for the transition function and  $\bar{P}$  for the invariant measure, we shall use  $E\{\cdot\}$  to express the answers. Implicit is the understanding that, in the final answers, we treat  $(y_{pq}(t))$  as a stationary process, i.e., we assume that the initial values have  $\bar{P}$  as their distribution. We also assume that  $E\{y(t)\} = 0$ .

To analyze our system, we proceed as follows. We consider  $(x^\epsilon(t), y^\epsilon(t))$ ,  $y^\epsilon(t) \equiv y(t/\epsilon^2)$ , jointly as a Markov process. Now, however, the right-hand side of (3.41) depends on  $t/\epsilon^2$  explicitly and  $x^\epsilon$  is complex valued. So we must change the format of §3(ii) somewhat to cope with this complication. The explicit  $t/\epsilon^2$  dependence in (3.41) is essential for the results we want.

Let  $\tau^\epsilon(t) = t/\epsilon^2$  and consider  $y(t) = (y_{pq}(t))$  jointly with  $\tau(t) = t$ . The generator of  $(y(t), \tau(t))$  is  $\mathcal{L}_1 + \partial/\partial\tau$ , and it acts on functions of real symmetric matrices and  $\tau \in (-\infty, \infty)$ . Note that  $\mathcal{L}_1$  and  $\partial/\partial\tau$  commute. Since  $x^\epsilon(t)$  is complex valued, we could treat its real and imaginary parts separately. It is more convenient, however, to deal with  $x^\epsilon(t)$  and  $x^{\epsilon*}(t)$ , its complex conjugate. If  $x = x^R + ix^I$ , we shall employ the notation

$$\frac{\partial}{\partial x} = \frac{1}{2} \left( \frac{\partial}{\partial x^R} - i \frac{\partial}{\partial x^I} \right), \quad \frac{\partial}{\partial x^*} = \frac{1}{2} \left( \frac{\partial}{\partial x^R} + i \frac{\partial}{\partial x^I} \right).$$

With this notation, the infinitesimal generator of the Markov process  $(x^\epsilon(t), y^\epsilon(t), \tau^\epsilon(t))$  is

$$\frac{1}{\epsilon^2} \left( \mathcal{L}_1 + \frac{\partial}{\partial \tau} \right) + \frac{1}{\epsilon} (\mathcal{L}_2 + \mathcal{L}_2^*), \tag{3.42}$$

where

$$\mathcal{L}_2 = i \sum_{p,q=1}^n y_{pq} \exp[i(\omega_q - \omega_p)\tau] x_q \frac{\partial}{\partial x_p},$$

$$\mathcal{L}_2^* = -i \sum_{p,q=1}^n y_{pq} \exp[-i(\omega_q - \omega_p)\tau] x_q^* \frac{\partial}{\partial x_p^*},$$

and (3.42) acts on functions of  $(x, y, \tau)$ ,  $x \in C^n$ ,  $y$  a real symmetric matrix,  $\tau \in (-\infty, \infty)$ .

Let  $f(x)$  be smooth bounded real-valued function of  $x \in C^n$  and let

$$u^\epsilon(t, x, y, \tau) = E\{f(x^\epsilon(t))\} \tag{3.43}$$

with  $x^\epsilon(0) = x, y^\epsilon(0) = y, \tau^\epsilon(0) = \tau$ . We wish to find the limit of  $u^\epsilon$ , in the appropriate sense, as  $\epsilon \rightarrow 0$ . The equation satisfied by  $u^\epsilon$ , the backward

equation, is

$$\frac{\partial u^\varepsilon}{\partial t} = \frac{1}{\varepsilon^2} \left( \mathfrak{L}_1 + \frac{\partial}{\partial \tau} \right) u^\varepsilon + \frac{1}{\varepsilon} (\mathfrak{L}_2 + \mathfrak{L}_2^*) u^\varepsilon, \quad u^\varepsilon(0, x, y, \tau) = f(x). \quad (3.44)$$

This problem is of the form (3.18) and so the perturbation theory applies. The operator  $\mathfrak{L}_1$  of (3.18) is  $\mathfrak{L}_1 + \partial/\partial\tau$  in (3.44) and the net effect of this is that we will perform an averaging [1] simultaneously with the other approximations.

If we assume that  $e^{\mathfrak{L}_1 t} \rightarrow E\{\cdot\}$  (expectation with respect to the invariant measure  $\bar{P}$ ) strongly and, say, exponentially fast, then, with  $\tilde{\mathfrak{L}}_1 = \partial/\partial\tau$ ,

$$e^{(\mathfrak{L}_1 + \tilde{\mathfrak{L}}_1)t} \rightarrow P = \lim_{T \uparrow \infty} \frac{1}{T} \int_0^T E\{\cdot\} ds,$$

which corresponds to (3.19). Thus the projection operator is averaging with respect to the invariant measure and time averaging; see [7], [13]. Since  $E\{y(t)\} = 0$ , it follows that  $P\mathfrak{L}_2 P = P\mathfrak{L}_2^* P = 0$ ; and hence, from (3.25),  $u^\varepsilon(t, x, y, \tau)$  of (3.43) converges as  $\varepsilon \rightarrow 0$ ,  $0 \leq t \leq T$ , to  $u_0(t, x)$  where

$$\begin{aligned} du_0/dt &= -P(\mathfrak{L}_2 + \mathfrak{L}_2^*)(\mathfrak{L}_1 + \tilde{\mathfrak{L}}_1)^{-1}(\mathfrak{L}_2 + \mathfrak{L}_2^*)u_0, \\ t > 0, u_0(0, x) &= f(x), \tilde{\mathfrak{L}}_1 = \partial/\partial\tau. \end{aligned} \quad (3.45)$$

We must now find the explicit form of the operator on the right-hand side of (3.45).

Using again the fact that  $e^{\mathfrak{L}_1 t} \rightarrow E\{\cdot\}$  rapidly, it is easy to see that

$$\begin{aligned} -(\mathfrak{L}_1 + \tilde{\mathfrak{L}}_1)^{-1}(\mathfrak{L}_2 + \mathfrak{L}_2^*)f &= \int_0^\infty e^{\mathfrak{L}_1 s} [e^{\mathfrak{L}_1 s} - P] ds (\mathfrak{L}_2 + \mathfrak{L}_2^*)f \\ &= i \sum_{p,q=1}^n \int_0^\infty E\{y_{pq}(s)|y(0)\} \exp[i(\omega_q - \omega_p)(\tau + s)] ds x_q \frac{\partial f}{\partial x_p} \\ &\quad - i \sum_{p,q=1}^n \int_0^\infty E\{y_{pq}(s)|y(0)\} \exp[-i(\omega_q - \omega_p)(\tau + s)] ds x_q^* \frac{\partial f}{\partial x_p^*}. \end{aligned}$$

Assuming now that  $\omega_1, \omega_2, \dots, \omega_n$  are distinct along with their sums and differences, equation (3.45) takes the following form.

$$\begin{aligned} \frac{\partial u_0(t, x)}{\partial t} &= \left[ (\delta_{pq} \delta_{ap'} + \delta_{q'p'} \delta_{pq}) \right. \\ &\quad \cdot \int_0^\infty E\{y_{pq}(s)y_{p'q'}(0)\} \exp[i(\omega_q - \omega_p)s] ds x_q \frac{\partial}{\partial x_{p'}} \left( x_q \frac{\partial}{\partial x_p} \right) \\ &\quad - (\delta_{q'q} \delta_{p'p} + \delta_{q'p'} \delta_{qp}) \\ &\quad \cdot \int_0^\infty E\{y_{pq}(s)y_{p'q'}(0)\} \exp[i(\omega_q - \omega_p)s] ds \\ &\quad \left. \cdot x_{q'} \frac{\partial}{\partial x_{p'}} \left( x_q^* \frac{\partial}{\partial x_p^*} \right) + \text{c.c.} \right] u_0(t, x), \quad u_0(0, x) = f(x). \end{aligned} \quad (3.46)$$

Here we have employed the summation convention and we have used the abbreviation c.c. for complex conjugate, i.e., the formal complex conjugate of the operator preceding c.c. in (3.46). Note also that  $u_0$  and  $f$  are thought of as functions of  $x$  and  $x^*$ ; they are real and the diffusion operator on the right side of (3.46) takes real functions into real functions.

One can deduce many interesting facts from the result that  $u^\epsilon \rightarrow u_0$ , the solution of (3.46). One result is as follows. Suppose  $f = f(x, x^*) = x_p x_p^*$  successively for  $p = 1, 2, \dots, n$ . Let

$$w_p(t) = \lim_{\epsilon \downarrow 0} E \{ |x_p(t)|^2 \}.$$

Then (3.46) yields the following equation for  $w_p(t)$ .

$$\frac{dw_p(t)}{dt} = \sum_{q=1}^n [A_{pq}w_q(t) - A_{qp}w_p(t)], \tag{3.47}$$

$$t > 0, w_p(0) = |x_{p0}|^2,$$

where

$$\begin{aligned} A_{pq} &= A_{qp} = \int_0^\infty E \{ y_{pq}(s)y_{pq}(0) \} \cos(\omega_q - \omega_p)s \, ds \\ &= \text{power spectrum of } \{ y(t), t \geq 0 \} \text{ at} \\ &\quad \text{the difference frequency } \omega_q - \omega_p. \end{aligned} \tag{3.48}$$

These equations have the form of master equations or transport equations. They control the way in which energy is exchanged among the oscillators (3.40) in the weak coupling limit. Further results of this kind, with Markovian or non-Markovian  $\{ y(t), t \geq 0 \}$  can be found in [11], [12], [13] and references cited therein. In [11] the role of the condition that  $\omega_1, \omega_2, \dots, \omega_n$  be distinct is examined closely, and in [12] a more complicated system of the form (3.40) is studied in connection with underwater sound propagation.

(vi) *Non-Markovian problems.* The case where the coefficients  $y^\epsilon(t) \equiv y(t/\epsilon^2)$  in (3.14) constitute a Markov process is typical of more general problems and it was treated first for illustrative purposes. We shall now consider more general situations as follows.

Consider the stochastic differential equation

$$\frac{dx^\epsilon(t)}{dt} = \frac{1}{\epsilon} F(x^\epsilon(t), \omega^\epsilon(t)) + G(x^\epsilon(t), \omega^\epsilon(t)), \tag{3.49}$$

$$t > 0, x^\epsilon(0) = x,$$

where, to distinguish it from the Markovian case (3.14), we denote the random "coefficients" by  $\omega^\epsilon(t) \equiv \omega(t/\epsilon^2)$ . The "coefficients"  $\omega(t)$  are defined on a probability space  $(\Omega, \mathfrak{F}, P)$  so that  $\omega \rightarrow \omega(t)$  ( $\omega(0) = \omega$ ),  $-\infty < t < \infty$ , is a one-parameter group of measure-preserving transformations on  $\Omega$ . This means that if  $A \in \mathfrak{F}$  and if  $A_t = \{ \omega \in \Omega | \omega(-t) \in A \}$ , then

$$P(A_t) = P(A), \quad -\infty < t < \infty.$$

The functions  $F(x, \omega)$  and  $G(x, \omega)$  on  $R^n \times \Omega \rightarrow R^m$  are appropriately restricted so that (3.49) has solutions for all  $t \geq 0$  when  $\varepsilon > 0$ .

We shall refer to the motion  $\omega \rightarrow \omega(t)$  as the bath and to  $x^\varepsilon(t)$  as the system. This terminology is used to emphasize the point that (3.49) represents a classical dynamical system driven by interaction with the motion  $\omega(t)$ . The state of the system  $x^\varepsilon(t) = x^\varepsilon(t, x, \omega)$ , i.e., it depends on the initial state of the system and the initial state of the bath. The initial state of the bath is distributed randomly with measure  $P$ . Thus, the only way randomness enters into (3.49) is via the initial distribution of the bath.

Let

$$F(x, t, \omega) = F(x, \omega(t)), \quad G(x, t, \omega) = G(x, \omega(t)). \quad (3.50)$$

For each  $x$ ,  $F(x, t, \omega)$  and  $G(x, t, \omega)$  are stationary random functions, often abbreviated  $F(x, t)$ ,  $G(x, t)$ , and we may write (3.49) in the form

$$\frac{dx^\varepsilon(t)}{dt} = \frac{1}{\varepsilon} F(x^\varepsilon(t), t/\varepsilon^2) + G(x^\varepsilon(t), t/\varepsilon^2), \quad (3.51)$$

$$t > 0, \quad x^\varepsilon(0) = x.$$

This is the familiar form in stochastic differential equations [14], [15]. We shall assume that

$$E \{ F(x, t, \cdot) \} = \int_{\Omega} F(x, \omega) P(d\omega) = 0 \quad (3.52)$$

which corresponds to the case  $P \mathbb{E}_2 P = 0$  in (3.20). The extension to the other case is not difficult.

How do we proceed with the asymptotic analysis of (3.51) as  $\varepsilon \downarrow 0$ ? We attempt to follow the pattern of the Markovian case and establish an equation of the form (3.17), i.e., a backward equation. However,  $(x^\varepsilon(t), \omega^\varepsilon(t))$  do not constitute a Markov process, so we must approach the problem in a more general way. This is done in [7] in the manner of [16]. Direct but more involved procedures are [14], [15]. We shall not give details here but we shall summarize the facts as follows.

First of all the final results are the same as in the Markovian case. That is, if  $f(x)$  is a smooth bounded function on  $R^n$ , then  $E \{ f(x^\varepsilon(t)) \}$ , with  $x^\varepsilon(0) = x$ , converges as  $\varepsilon \rightarrow 0$ ,  $0 \leq t \leq T < \infty$ , to the solution  $u(t, x)$  of the diffusion equation

$$\partial u(t, x) / \partial t = \bar{\mathbb{E}} u(t, x), \quad u(0, x) = f(x),$$

with  $\bar{\mathbb{E}}$  given by (3.37). In (3.37) we replace  $y(t)$  by  $\omega(t)$ , i.e., the expectations are taken relative to the stationary processes defined by (3.50). This identity of the results in the Markovian and non-Markovian cases does not, of course, tell us what the necessary assumptions on  $F$  and  $G$  of (3.50) are for its validity. In the Markovian case, ergodicity of  $\{y(t), t \geq 0\}$  in a sufficiently strong sense (a Fredholm alternative for  $\mathbb{E}_1$ ) was sufficient. In the non-Markovian case sufficient conditions are given in [7]. A typical one expressed

in words says: The best estimate of the state of the bath at time  $t + s$  as observed, i.e.,  $F(x, \omega(t + s))$ , given all observable information up to some  $t$ , tends to zero (because of (3.52)) sufficiently rapidly as  $s \uparrow \infty$ . This “unpredictability” of the bath is what makes the system go to a diffusion process in the weak coupling limit.

We conclude this section by stating again that, under appropriate conditions, the asymptotics for Markovian and non-Markovian coefficients are the same. Thus, all of our above examples, suitably reinterpreted, serve as non-Markovian examples as well. In particular, the results of §3(v) on the randomly coupled oscillators carry over immediately.

(vii) *Case  $P \mathbb{E}_2 P \neq 0$ . Small diffusion.* When condition (3.15) or (3.52) does not hold then we set

$$\bar{F}(x) = E \{ F(x, t, \cdot) \}, \quad \bar{G}(x) = E \{ G(x, t, \cdot) \} \quad (3.53)$$

and the result is that  $E \{ f(x^\epsilon(t)) \}$  with  $x^\epsilon(0) = x$  is approximated well when  $\epsilon \ll 1$  by the solution  $\bar{u}^\epsilon(t, x)$  of the diffusion equation

$$\partial \bar{u}^\epsilon(t, x) / \partial t = \bar{\mathbb{L}}^\epsilon \bar{u}^\epsilon(t, x), \quad \bar{u}^\epsilon(0, x) = f(x), \quad (3.54)$$

where, from (3.28),

$$\begin{aligned} \mathbb{L}^\epsilon f(x) = & \frac{1}{\epsilon} \bar{F}(x) \frac{\partial f(x)}{\partial x} + \bar{G}(x) \frac{\partial f(x)}{\partial x} \\ & + \int_0^\infty E \left\{ (F(x, 0, \cdot) - \bar{F}(x)) \frac{\partial}{\partial x} \left( (F(x, t, \cdot) - \bar{F}(x)) \frac{\partial f(x)}{\partial x} \right) \right\} dt. \end{aligned} \quad (3.55)$$

The approximation of  $E \{ f(x^\epsilon(t)) \}$  by  $\bar{u}^\epsilon(t, x)$  is valid in an interval  $0 \leq t \leq T$  with  $T < \infty$  but arbitrary.

The analysis of the diffusion equation (3.54), a small diffusion problem, is of special interest independently of the way it was obtained here. Specifically, on rescaling the time we may write it as

$$\frac{\partial \bar{u}^\epsilon(t, x)}{\partial t} = \epsilon \sum_{i,j=1}^n a_{ij}(x) \frac{\partial^2 \bar{u}^\epsilon(t, x)}{\partial x_i \partial x_j} + \sum_{i=1}^n (\bar{F}_i(x) + \epsilon b_i(x)) \frac{\partial \bar{u}^\epsilon(t, x)}{\partial x_i}, \quad (3.56)$$

$$\bar{u}^\epsilon(0, x) = f(x),$$

with the coefficients identified from (3.55). We may now ask several questions about (3.56) or boundary value problems associated with it. If the solutions of

$$d\bar{x}(t)/dt = \bar{F}(\bar{x}(t)), \quad \bar{x}(0) = x, \quad (3.57)$$

are periodic orbits for all  $x \in R^n$ , then one may employ the method of averaging on (3.56). First we introduce new coordinates in space, corresponding to “action-angle” variables, and then do the averaging. The analysis of this step is carried out in [18].

If the structure of the solutions of (3.57) is more complex, for example if

there are stable or unstable equilibria, limit cycles or other invariant manifolds, etc., then approximations to (3.56) lead to a very rich and potentially extremely useful theory. The simplest form of that theory is the Gauss-Markov theory discussed in the next section. There we study the solution of (3.56) near the orbits of (3.57). The behavior of the diffusion process associated with (3.56) away from the orbits of (3.57) is a problem in the theory of large deviations. It has been studied in detail in the fundamental work of Ventsel and Freidlin ([12] of §2). Interesting applications and many illuminating remarks can be found in [19].

(viii) *Gauss-Markov limit.* As we mentioned above this is a relatively crude theory which is used when (3.15) or (3.52) do not hold and when the orbits of (3.57) are complicated but results are needed only near these orbits. Because of the importance of the results we shall start the analysis afresh and we shall give a self-contained treatment following [20] and [21]. The analysis extends to stochastic delay differential equations [21] but the limiting process is not Markovian in this case.

Consider again the stochastic equation

$$dx^\epsilon(t)/dt = F(x^\epsilon(t), \omega^\epsilon(t)), \quad x^\epsilon(0) = x, \quad (3.58)$$

where  $\omega^\epsilon(t) \equiv \omega(t/\epsilon)$  and  $\omega(t)$  is the motion defined on  $(\Omega, \mathfrak{F}, P)$  in §3(vi). Note, however, that the scaling of (3.58) is different from that of (3.49) or (3.51). We may also write (3.58) in the form

$$dx^\epsilon(t)/dt = F(x^\epsilon(t), t/\epsilon), \quad x^\epsilon(0) = x, \quad (3.59)$$

with  $F(x, t, \omega) = F(x, \omega(t))$ . We define  $\bar{F}(x)$  by

$$\bar{F}(x) = E\{F(x, \cdot)\} = E\{F(x, t, \cdot)\} \quad (3.60)$$

and  $\bar{x}(t)$  as the solution

$$d\bar{x}(t)/dt = \bar{F}(\bar{x}(t)), \quad \bar{x}(0) = x. \quad (3.61)$$

By an elementary application of the perturbation procedure of §3(iii), or better by direct analysis, it is easy to conclude that

$$\lim_{\epsilon \downarrow 0} \sup_{0 < t \leq T} E\{|x^\epsilon(t) - \bar{x}(t)|\} = 0. \quad (3.62)$$

What happens on a longer time interval ( $t \sim 1/\epsilon$ ) corresponds to the analysis of §3(vii) (with  $G \equiv 0$  and  $t \rightarrow \epsilon t$  there). Let us ask, however, for more modest information. How does the fluctuation process, defined by

$$(x^\epsilon(t) - \bar{x}(t))/\sqrt{\epsilon} = z^\epsilon(t), \quad (3.63)$$

behave in  $0 \leq t \leq T$  as  $\epsilon \downarrow 0$ ? First we find a differential equation for  $z^\epsilon(t)$ . From (3.63), (3.59) and (3.61) it follows that

$$\frac{dz^\epsilon(t)}{dt} = \frac{1}{\sqrt{\epsilon}} \left[ F(\bar{x}(t) + \sqrt{\epsilon} z^\epsilon(t), t/\epsilon) - \bar{F}(\bar{x}(t)) \right], \quad z^\epsilon(0) = 0. \quad (3.64)$$

Next we expand the right-hand side of (3.64) in powers of  $\sqrt{\epsilon}$ .



$$dz^\epsilon(t)/dt = (1/\sqrt{\epsilon}) [F(\bar{x}(t), t/\epsilon) - \bar{F}(\bar{x}(t))] + \frac{\partial F(\bar{x}(t), t/\epsilon)}{\partial x} z^\epsilon(t) + O(\sqrt{\epsilon}), \quad z^\epsilon(0) = 0. \tag{3.65}$$

Now the result is clear. Under appropriate conditions [20], [21]  $z^\epsilon(t)$  will converge to a Gaussian process because

$$\frac{1}{\sqrt{\epsilon}} \int_0^t [F(\bar{x}(s), s/\epsilon) - \bar{F}(\bar{x}(s))] ds \tag{3.66}$$

will converge as  $\epsilon \rightarrow 0$  to a Gaussian process  $w(t)$ , say, and hence  $z^\epsilon(t) \rightarrow \bar{z}(t)$  where

$$\bar{z}(t) = w(t) + \int_0^t \frac{\partial \bar{F}(\bar{x}(s))}{\partial x} \bar{z}(s) ds. \tag{3.67}$$

Actually (3.66) converges to a Gaussian process with independent increments, i.e., a time-inhomogeneous Brownian motion with  $E\{w(t)\} = 0$ ; and if (\* = transpose)

$$A(x) = \lim_{T \uparrow \infty} \frac{1}{T} \int_0^T \int_0^T E \{ (F(x, s) - \bar{F}(x)) \cdot (F(x, \sigma) - \bar{F}(x))^* \} d\sigma ds, \tag{3.68}$$

then  $E\{w(t)w^*(t)\} = A(\bar{x}(t))$  is the covariance matrix of  $w(t)$ . Thus  $\bar{z}(t)$ , the limiting fluctuation process defined by (3.67), is a time-inhomogeneous Ornstein-Uhlenbeck process. We may call the approximation  $\bar{x}(t) + \sqrt{\epsilon} \bar{z}(t)$ ,  $0 \leq t \leq T$ , of  $x^\epsilon(t)$  the guided O.U. approximation because the mean follows the orbit  $\bar{x}(t)$  and the fluctuation is an O.U. process centered at the orbit  $\bar{x}(t)$  and with covariance matrix  $C(t)$  computed as follows.

Let  $U(t, s)$  be the fundamental matrix solution of the linear equations

$$\frac{dU(t, s)}{dt} = \frac{\partial \bar{F}(\bar{x}(t))}{\partial x} U(t, s), \quad t > s, U(s, s) = I.$$

Then (3.67) yields  $\bar{z}(t) = \int_0^t U(t, s) dw(s)$ , and hence

$$E\{\bar{z}(t)\bar{z}^*(t)\} = \int_0^t U(t, s) A(\bar{x}(s)) U^*(t, s) ds = C(t). \tag{3.69}$$

Thus,  $C(t)$  can be computed explicitly if we have access to the fundamental matrix  $U$ .

As a simple application we consider the symmetric competition two-species model of interacting populations [21] in a noisy environment.

Let  $x(t)$  and  $y(t)$  denote the population of species one and two respectively and assume that they satisfy the evolution equations

$$\begin{aligned} \frac{dx^\epsilon(t)}{dt} &= x^\epsilon(t)(k + \gamma_1(t/\epsilon) - x^\epsilon(t) - \alpha y^\epsilon(t)), & x^\epsilon(0) &= x, \\ \frac{dy^\epsilon(t)}{dt} &= y^\epsilon(t)(k + \gamma_2(t/\epsilon) - y^\epsilon(t) - \alpha x^\epsilon(t)), & y^\epsilon(0) &= y. \end{aligned} \tag{3.70}$$

Here  $\alpha \in [0, 1]$  is the competition coefficient and measures the degree of interaction between the species,  $k$  is the deterministic carrying capacity of the environment and  $\gamma_1(t/\epsilon)$ ,  $\gamma_2(t/\epsilon)$  are zero-mean stochastic processes which fluctuate rapidly since  $\epsilon \ll 1$ . We assume they are stationary and denote the covariance by  $R_{ij}(t) = E\{\gamma_i(t+s)\gamma_j(s)\}$ . If  $\mathbf{x} = (x, y)$  and

$$F(\mathbf{x}, t/\epsilon) = \begin{bmatrix} x(k + \gamma_1(t/\epsilon) - x - \alpha y) \\ y(k + \gamma_2(t/\epsilon) - y - \alpha x) \end{bmatrix},$$

then

$$\bar{F}(\mathbf{x}) = \begin{bmatrix} x(k - x - \alpha y) \\ y(k - y - \alpha x) \end{bmatrix}.$$

We note that  $\bar{F}(\mathbf{x})$  has two equilibrium points,  $(0, 0)$  and the point  $(k/(1 + \alpha), k/(1 + \alpha))$ , the latter being stable.

Let us assume that (3.70) is solved with initial conditions  $(x, y) =$

$$(k/(1 + \alpha), k/(1 + \alpha)) = \mathbf{x}^0.$$

Then,  $\bar{\mathbf{x}}(t) = \mathbf{x}^0$  for all  $t > 0$  and

$$\frac{\partial \bar{F}(\mathbf{x})}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\mathbf{x}^0} = \frac{-k}{1 + \alpha} \begin{bmatrix} 1 & \alpha \\ \alpha & 1 \end{bmatrix} \equiv B. \quad (3.71)$$

Furthermore, from (3.68) we find that (assuming finite integrals)

$$A(\mathbf{x}^0) = \left( \frac{k}{1 + \alpha} \right)^2 \int_0^\infty dt \begin{bmatrix} R_{11}(t) & R_{12}(t) \\ R_{12}(t) & R_{22}(t) \end{bmatrix}. \quad (3.72)$$

We are ready now to compute the covariance of the limit fluctuation process. We have chosen the initial conditions  $(x, y) = \mathbf{x}^0 =$  the stable equilibrium point in order to make the mean  $\bar{\mathbf{x}}(t)$  as simple as possible. The fluctuation process  $\bar{\mathbf{z}}(t)$  will be a Gaussian process with independent increments, mean zero and covariance given by (3.69). To simplify the calculations we assume further that the noise intensities for the two species are the same so we may write  $A(\mathbf{x}^0)$  in the form

$$A = \left( \frac{k}{1 + \alpha} \right)^2 \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Now however  $A$  and  $B$  of (3.71) commute so  $C(t)$  of (3.69) is given by

$$C(t) = \int_0^t e^{2Bs} ds A. \quad (3.73)$$

From this formula one finds easily that the ellipses of constant probability for the Gaussian process  $\bar{\mathbf{z}}(t)$  are rotated by  $45^\circ$  and that in this  $45^\circ$  frame of reference the covariance matrix is diagonal and is given by

$$\left[ \begin{array}{cc} \frac{\sigma^2 k}{2} \frac{1 + \rho}{(1 + \alpha)^2} (1 - e^{-2kt}) & 0 \\ 0 & \frac{\sigma^2 k}{2} \frac{1 - \rho}{1 - \alpha^2} (1 - e^{-2k((1-\alpha)/(1+\alpha))t}) \end{array} \right]. \quad (3.74)$$

When  $t = \infty$  this simplifies to

$$\left[ \begin{array}{cc} \frac{\sigma^2 k}{2} \frac{1 + \rho}{(1 + \alpha)^2} & 0 \\ 0 & \frac{\sigma^2 k}{2} \frac{1 - \rho}{1 - \alpha^2} \end{array} \right], \quad (3.75)$$

and a number of interesting conclusions can be drawn from this simple result. We note however, since the results hold for  $0 \leq t \leq T < \infty$ , it is not, strictly speaking, proper to set  $t = \infty$  in (3.74). The significance of (3.75) is, however, relative; it is simply an approximation to (3.74) which is appropriate and useful in certain ranges of the parameters.

(ix) *Higher order terms, larger intervals of validity, etc.* Throughout the preceding discussion, approximations were valid as  $\epsilon \downarrow 0$ ,  $0 \leq t \leq T < \infty$ . It is natural to ask if the results are valid in  $0 \leq t < \infty$ . This has considerable practical significance since solving diffusion equations to obtain information about the limiting process is frequently a hopeless task. On the other hand, finding time-independent solutions is simpler. The question is, then, what is the significance of such equilibrium solutions of the limiting diffusion equation relative to the original problem? (See [17].)

An interesting answer to this question is given by the following result of Norman [22], which we state in the context of diffusion approximations for equations with Markov coefficients.

Let  $u^\epsilon(t, x, y)$  be the solution of (3.17), and suppose that

$$\lim_{\epsilon \downarrow 0} \sup_{0 \leq t \leq T} \sup_{x, y} |u^\epsilon(t, x, y) - \bar{u}(t, x)| = 0, \quad (3.76)$$

where  $\bar{u}(t, x)$  satisfies the equation

$$\partial \bar{u}(t, x) / \partial t = \bar{\mathbb{E}} \bar{u}(t, x), \quad \bar{u}(0, x) = f(x), \quad (3.77)$$

with  $\bar{\mathbb{E}}$  defined by (3.36). Suppose that there is an  $\bar{f}(x)$  such that

$$\lim_{t \uparrow \infty} \sup_x |\bar{u}(t, x) - \bar{f}(x)| = 0. \quad (3.78)$$

Let  $\bar{u}^\epsilon(t, x, y)$  be the solution of (3.17) with  $\bar{u}^\epsilon(0, x, y) = \bar{f}(x)$  and suppose that

$$\lim_{\epsilon \downarrow 0} \sup_{t > 0} \sup_{x, y} |\bar{u}^\epsilon(t, x, y) - \bar{f}(x)| = 0. \quad (3.79)$$

Then (3.76) holds with  $T = \infty$ . The proof of this result is quite elementary [22].

Clearly (3.79) is the hypothesis that is hardest to verify since it looks about as difficult as the original problem. It may happen, however, that the limit of  $\bar{u}(t, x)$  as  $t \uparrow \infty$  is very simple, for example,  $\bar{f}(x) = \text{constant}$ . In this case (3.79) is trivially true and hence we have the result (3.76) with  $T = \infty$ .

As an example, consider (3.41), the linear randomly coupled oscillators, and take  $f(x) = x_p x_p^*$ ,  $p = 1, 2, \dots, n$ . Assuming nondegeneracy, (3.74) yields  $\bar{u}(t, x) \rightarrow 1/n$  as  $t \rightarrow \infty$  if  $\sum_{p=1}^n |x_{p0}|^2 = 1$ , say. Thus,

$$E \left\{ |x_p^\varepsilon(t)|^2 \right\} \xrightarrow{\varepsilon \downarrow 0} w_p(t),$$

the solution of (3.47), and the result holds for all  $t \geq 0$ . This example is typical of what one may expect in general. It should be contrasted with the situation described in §3(vii) where it is expected that the above argument will fail.

Concerning higher order terms in the expansions, valid in the same time interval, the procedure of §3(iii) is perfectly adequate. It goes without saying that calculations become prohibitively complex very rapidly.

(x) *Boundary value problems and shooting; one-dimensional waves.* Linear two-point (or multi-point) boundary value problems for stochastic equations can be treated by the simple device of obtaining the statistical properties of the fundamental solution matrix. From this one can obtain all relevant information about the statistics of the solution to the boundary value problem. This is the method of shooting.

We shall describe an example, one-dimensional wave propagation, that illustrates the general ideas. The example is also of independent interest, and many other remarks and references to the literature can be found in [23], [24].

Consider a one-dimensional random medium occupying the interval  $0 \leq x \leq l$ . Let  $u(x)$  and  $n(x)$  be the wave field (with the time factor  $e^{-i\omega t}$  omitted) and the index of refraction at location  $x$ , respectively. We assume that  $u(x)$ ,  $-\infty < x < \infty$ , satisfies the reduced wave equation

$$d^2u(x)/dx^2 + k^2n^2(x)u(x) = 0, \quad -\infty < x < \infty, \quad (3.80)$$

$$n^2(x) = \begin{cases} n_1^2, & x < 0, \\ 1 + \varepsilon\mu(x), & 0 \leq x \leq l, \\ n_2^2, & x > l. \end{cases}$$

$u(x)$  and  $du(x)/dx$  continuous,  $x \in (-\infty, \infty)$ .

Here  $\mu(x)$  denotes a real zero-mean stationary random process with  $x$  the "time" variable and  $\varepsilon \ll 1$  characterizing the size of the fluctuations of the refractive index in  $0 \leq x \leq l$ . If we denote by  $R$  and  $T$  the complex-valued reflection and transmission coefficients, then we have

$$\begin{aligned} u(x) &= e^{ikn_1x} + R e^{-ikn_1x}, & x < 0, \\ u(x) &= T e^{ikn_2x}, & x > l. \end{aligned} \quad (3.81)$$

From (3.80) and (3.81) it follows that  $u(x)$  satisfies the following stochastic two-point boundary value problem.

$$\frac{d^2u(x)}{dx^2} + k^2[1 + \epsilon\mu(x)]u(x) = 0, \quad 0 \leq x \leq l,$$

$$\frac{1}{2} \left[ u(0) + \frac{1}{ikn_1} \frac{du(0)}{dx} \right] = 1, \quad \frac{1}{2} \left[ u(l) - \frac{1}{ikn_2} \frac{du(l)}{dx} \right] = 0. \quad (3.82)$$

To reduce (3.82) to a first order system we define  $A(x)$  and  $B(x)$  by

$$u(x) = e^{ikx}A(x) + e^{-ikx}B(x),$$

$$du(x)/dx = ik[e^{ikx}A(x) - e^{-ikx}B(x)], \quad 0 \leq x \leq l.$$

It follows that  $A$  and  $B$  satisfy the stochastic two-point boundary value problem

$$\frac{dA(x)}{dx} = \epsilon \frac{ik\mu(x)}{2} [A(x) + B(x)e^{-2ikx}],$$

$$\frac{dB(x)}{dx} = -\epsilon \frac{ik\mu(x)}{2} [A(x)e^{2ikx} + B(x)], \quad 0 \leq x \leq l, \quad (3.83)$$

$$A(0) = E_g + \Gamma_g B(0), \quad B(l) = \Gamma_l A(l)$$

with

$$E_g = \frac{2n_1}{1+n_1}, \quad \Gamma_g = \frac{1-n_1}{1+n_1}, \quad \Gamma_l = e^{2ikl} \frac{1-n_2}{1+n_2}.$$

Since  $E\{\mu(x)\} = 0$  in (3.83) we are naturally interested in using the methods of previous sections to study the statistics of  $A$  and  $B$  when  $\epsilon \ll 1$  and  $l \sim 1/\epsilon^2$ , i.e., the weak coupling limit.

Let us express the solution of (3.83) in terms of fundamental matrix solutions. Let us assume for simplicity that  $n_1 = n_2 = 1$  so that  $\Gamma_g = \Gamma_l = 0$  in (3.83). Let  $Y(x)$  be the  $2 \times 2$  matrix solution of

$$\frac{dY(x)}{dx} = \epsilon \frac{ik\mu(x)}{2} \begin{bmatrix} 1 & e^{-2ikx} \\ -e^{2ikx} & -1 \end{bmatrix} Y(x), \quad Y(0) = I. \quad (3.84)$$

It is easily verified that  $Y(x)$  is of the form

$$Y = \begin{bmatrix} a & b \\ \bar{b} & \bar{a} \end{bmatrix}, \quad |a|^2 - |b|^2 = 1, \quad (3.85)$$

where  $a = a(x)$  and  $b = b(x)$  are complex random functions. Note that  $Y = Y(x, 0)$ , i.e., it is the fundamental matrix or propagator from location 0 to location  $x$ .

Let

$$Y_1 = \begin{pmatrix} a_1 & b_1 \\ \bar{b}_1 & \bar{a}_1 \end{pmatrix} = Y(x, 0) \quad \text{and} \quad Y_2 = \begin{pmatrix} a_2 & b_2 \\ \bar{b}_2 & \bar{a}_2 \end{pmatrix} = Y(l, x).$$

Then, it can be verified by direct computation that the solution  $A(l, x)$ ,  $B(l, x)$  of (3.83) (with  $\Gamma_g = \Gamma_l = 0$ ) is given by

$$A(l, x) = \frac{\bar{a}_2}{\bar{b}_2 b_1 + \bar{a}_2 \bar{a}_1}, \quad B(l, x) = \frac{-\bar{b}_2}{\bar{b}_2 b_1 + \bar{a}_2 \bar{a}_1}, \quad 0 \leq x \leq l. \quad (3.86)$$

Formulas (3.86) give the desired expression of  $A$  and  $B$  in terms of fundamental matrices.

The stochastic analysis can now begin with the matrix initial value problem (3.84). The procedures of the previous sections apply directly and we are led quickly to a diffusion equation for the transition probabilities of the limiting (matrix valued) Markov process.

The solution of this diffusion equation turns out to be obtainable in closed form in terms of known functions. Using this explicit solution we can then return to (3.86) and compute statistical averages of  $A$ ,  $B$  and higher powers. The calculations are straightforward but somewhat lengthy, so we refer to [23], [24] for details and discussion of the results.

#### REFERENCES

1. N. N. Bogoljubov and Ju. A. Mitropolskiĭ, *Asymptotic methods in the theory of nonlinear oscillators*, 2nd rev. ed., Fizmatgiz, Moscow, 1958; English transl., Internat. Monographs on Advanced Math. and Phys., Hindustan, Delhi, 1961; Gordon and Breach, New York, 1962. MR 20 #6812; 25 #5242.
2. V. M. Volosov, *Averaging in systems of ordinary differential equations*, Uspehi Mat. Nauk 17 (1962), no. 6 (108), 3–126 = Russian Math. Surveys 17 (1962), no. 6, 1–126. MR 26 #3976.
3. J. Cole, *Perturbation methods in applied mathematics*, Blaisdell, Waltham, Mass., 1968. MR 39 #7841.
4. S. Goldstein, *On diffusion by discontinuous movements, and on the telegraph equation*, Quart. J. Mech. Appl. Math. 4 (1951), 129–156. MR 13, 960.
5. R. Hersh, *Random evolutions: a survey of results and problems*, Rocky Mountain Math. J. 14 (1974), 443–496.
6. W. Heitler, *The quantum theory of radiation*, Oxford Univ. Press, Oxford, 1954.
7. G. C. Papanicolaou, *Some probabilistic problems and methods in singular perturbations*, Rocky Mountain J. Math. 6 (1976), 653–674.
8. T. G. Kurtz, *A limit theorem for perturbed operator semigroups with applications to random evolution*, J. Functional Analysis 12 (1973), 55–67.
9. U. Frisch, *Wave propagation in random media*, Probabilistic Methods in Appl. Math., vol. 1, Academic Press, New York, 1968, pp. 75–198. MR 42 #4088.
10. H. Kramers, *Brownian motion in a field of force and the diffusion model of chemical reactions*, Physica 7 (1940), 284–304. MR 2, 140.
11. G. C. Papanicolaou and R. Burridge, *Transport equations for the Stokes parameters starting from Maxwell's equations in a random medium*, J. Mathematical Phys. 16 (1975), 2074–2085.
12. W. Kohler and G. C. Papanicolaou, *Sound propagation in a randomly inhomogeneous ocean* (to appear).
13. G. C. Papanicolaou, *A kinetic theory for power transfer in stochastic systems*, J. Mathematical Phys. 13 (1972), 1912–1918. MR 47 #4577.
14. R. Z. Has'minskiĭ, *A limit theorem for solutions of differential equations with a random right hand part*, Teor. Veroyatnost. i Primenen. 11 (1966), 444–462 = Teor. Probability Appl. 11 (1966), 390–406. MR 34 #3637.
15. G. C. Papanicolaou and W. Kohler, *Asymptotic theory of mixing stochastic ordinary differential equations*, Comm. Pure Appl. Math. 27 (1974), 641–668.

16. T. G. Kurtz, *Semigroups of conditioned shifts and approximations of Markov processes*, Ann. of Prob. **3** (1975), 618–642.
17. G. Blankenship and G. C. Papanicolaou, *Stability and control of stochastic systems with wide band noise disturbances*, SIAM J. Appl. Math. (to appear).
18. R. Z. Has'minskii, *The averaging principle for parabolic and elliptic differential equations and Markov processes with small diffusion*, Teor. Veroyatnost. i Primenen. **8** (1963), 3–25 = Theor. Probability Appl. **8** (1963), 1–21. MR **28** #4253.
19. D. Ludwig, *Persistence in dynamical systems under random perturbations*, SIAM Rev. **17** (1975), 605–640.
20. R. Z. Has'minskii, *On stochastic processes defined by differential equations with a small parameter*, Teor. Veroyatnost. i Primenen. **11** (1966), 240–259 = Theor. Probability Appl. **11** (1966), 211–228. MR **34** #3636.
21. B. White, Dissertation, N.Y.U., 1974.
22. M. F. Norman, *Ergodicity of diffusion and temporal uniformity of diffusion approximations*, Ann. Probability (to appear).
23. W. Kohler and G. C. Papanicolaou, *Power statistics for wave propagation in one dimension and comparison with radiative transport theory*, J. Mathematical Phys. **14** (1973), 1733–1745. MR **49** #4442.
24. ———, *Power statistics for wave propagation in one dimension and comparison with radiative transport theory. II*, J. Mathematical Phys. **15** (1974), 2186–2197.

COURANT INSTITUTE OF MATHEMATICAL SCIENCES, NEW YORK UNIVERSITY, NEW YORK, NEW YORK 10012