Journal of Statistical Physics, Vol. 1, No. 2, 1969

Mathematical Problems of Modeling Stochastic Nonlinear Dynamic Systems

Richard E. Mortensen[1]

Received March 20, 1969

The purpose of this report is to introduce the engineer to the area of stochastic differential equations, and to point out the mathematical techniques and pitfalls in this area. Topics discussed include continuous-time Markov processes, the Fokker–Planck–Kolmogorov equations, the Ito and Stratonovich stochastic calculi, and the problem of modeling physical systems.

KEY WORDS: stochastic differential equations; continuous-time Markov processes; Fokker–Planck–Kolmogorov equations; Ito and Stratonovich stochastic calculi; modeling physical systems.

1. INTRODUCTION

This paper discusses certain mathematical problems which arise in attempting to model a stochastic dynamic system by means of a set nonlinear ordinary differential equations with white-noise excitation. This approach has been advocated in engineering literature at various times over the past ten years. The appeal of this approach is that it is the natural extension to stochastic systems of the state-space approach to deterministic systems which has met so much success in optimal-control theory. Furthermore, the state vector in such a model turns out to be a vector Markov process, for which a substantial mathematical theory exists; in particular, there is the theory of the Kolmogorov or Fokker–Planck partial differential equation. In addition, as one would expect of a state-space approach, this method is especially suited to the study of the transient behavior of the stochastic system, with steady-state, or, more precisely, stationary behavior obtained as a limiting case.

[1] School of Engineering and Applied Science, University of California, Los Angeles, California.

The engineering literature tends to give one the impression that the major difficulties associated with this approach are computational. Although it is not denied that the computational difficulties are large, it is the main point of this paper to show that a fundamental difficulty may arise at an earlier phase of the analysis, namely, when the mathematical model itself is chosen. In a sense, this difficulty is not computational, but conceptual, i.e., there may be a basic divergence between the implications of the mathematical model and the facts of physical reality.

This difficulty arises from the properties of the heuristic mathematical idealization known as white noise, or its rigorous counterpart, Brownian motion, which is heuristically the time integral of white noise. The peculiar implications of the Brownian-motion stochastic process puzzled physicists of an earlier era, leading them to adopt a stochastic process with more "physical" properties, the Ornstein–Uhlenbeck stochastic process.

Mathematically, the trouble arises when one attempts to apply the usual rules of differential and integral calculus to functions of time which are actually sample functions of a stochastic process. The result has been that something of a controversy has appeared in recent literature concerning two possible ways of extending ordinary calculus to stochastic functions: the so-called Stratonovich calculus, in which the usual rules continue to apply, and the so-called Ito calculus, in which the rules are changed. Although this subject has been discussed in several papers in the last two or three years, reading some of these papers can leave one more bewildered than before one started.

The aim of this report to show, by means of examples which have been chosen to be as lucid as possible, the reasons for this divergence. Further, we will suggest an approach to the problems of mathematical modeling, analysis, and computation which seems to have the qualities of being both mathematically rigorous and consistent with physical reality.

2. THE ENGINEERING MODEL

Typically, the dynamic equations of motions that arise in the analysis of engineering systems are a statement of Newton's law of motion, $F = ma$, possibly augmented by the inclusion of known frictional or dissipative forces. Although the direct application of $F = ma$ yields second-order differential equations, it is well known that it is always possible, by adding more variables, to convert these to a set of coupled, first-order, and often nonlinear, differential equations of the form

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), t) \tag{1}$$

Here $\mathbf{x}$ and $\mathbf{f}$ are $\mathbf{n}$-vectors. The vector $\mathbf{x}(t)$ is called the state of the system at time t.

If now an engineer wishes to modify Eq. (1) to try to take account of random forces in the environment, a natural way to proceed is to write

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), t) + \mathbf{G}(\mathbf{x}(t), t)\,\mathbf{v}(t) \tag{2}$$

Here $\mathbf{v}(t)$ is an m-vector, representing the random force at time t, and $\mathbf{G}(\mathbf{x}(t), t)$ is an $n \times m$ matrix. It is allowed to be a function of $\mathbf{x}$ and t to take into account the possibility that the influence of the noise may depend on the state of the system.

The function $\mathbf{v}(t)$ is a random process, i.e., for each fixed t, the value of the function $\mathbf{v}(t)$ is a random variable. In the absence of any special knowledge about the nature of the random force, a commonly made assumption is that $\mathbf{v}(t)$ is a so-called Gaussian white-noise random process. This means that for each fixed t, the random variable has a Gaussian distribution with zero mean and infinite variance. Furthermore, for any two times t_1 and t_2, with $t_1 \neq t_2$, the two random variables $\mathbf{v}(t_1)$ and $\mathbf{v}(t_2)$ are completely independent of each other.

Let E denote expectation, i.e., averaging across the statistical ensemble. Let a prime denote the transpose of a vector or a matrix. Since $\mathbf{v}(t)$ is a column vector, $\mathbf{v}'(t)$ is a row vector. Mathematically, white noise is characterized by the conditions

$$E\{\mathbf{v}(t)\} \approx 0; \qquad E\{\mathbf{v}(t_1)\,\mathbf{v}'(t_2)\} = \mathbf{C}(t_1)\,\delta(t_1 - t_2) \tag{3}$$

Here $\mathbf{C}(t_1)$ is an $m \times m$ matrix, called the white-noise covariance matrix, which expresses how the components of the vector $\mathbf{v}(t_1)$ are correlated among themselves. It is meaningful to speak of such correlation even though each component has infinite variance.

In the case of stationary white noise, the matrix $\mathbf{C}$ is constant, independent of time. Strictly, it is only in this case that the name "white" can be justified, because only in this case can one define a power spectral density function. In this case, the power spectral density function is constant, independent of frequency, analogous to the spectrum of the white light.

White noise is much the same kind of mathematical pathology in the theory of random processes that the Dirac delta-function is in the theory of deterministic functions.

As is by now well appreciated, so long as one does only linear operations on a delta-function, it is usually possible to interpret the result in a meaningful way However, one runs into trouble in trying to do nonlinear things to a delta-function. The square or the logarithm of a delta-function is meaningless, for example.

A similar situation exists in the case of white noise. If the differential equation (2) is linear, i.e., of the form

$$\dot{\mathbf{x}}(t) = A(t)\,\mathbf{x}(t) + B(t)\,\mathbf{v}(t) \tag{4}$$

then it turns out that there is no difficulty in interpreting what is meant by a solution of this differential equation. As a function of t, $\mathbf{x}(t)$ turns out to be a Gaussian random process, and there is no controversy about how to compute the mean and the covariance of this process. The process $\mathbf{x}(t)$ is much better behaved than $\mathbf{v}(t)$, e.g., none of the components of $\mathbf{x}(t)$ has infinite variance.

However, when the differential equation (2) is nonlinear, a problem of interpretation arises. One might at first think that the nonlinear equation (2) is simply meaningless, as in the case of the square of a delta-function. However, this is not the case. It turns out that there are two distinct, meaningful ways of interpreting Eq. (2)

which appear in contemporary literature, and which are called, repsectively, the Ito and the Stratonovich interpretations.

As stated in the introduction, this report will explore this Ito–Stratonovich divergence. Each interpretation will be explained, as well as the reason for the divergence. The two interpretations will be shown to be equivalent, in the sense that it is possible to pass from the results obtained under one interpretation to the results for the other interpretation via a transformation formula. Finally, the problems of real-world modeling and computation will be discussed.

3. THE FOKKER–PLANCK EQUATION

Before discussing this divergence and the subtleties of the stochastic calculus, perhaps it will be well to review the area of the theory in which there is no controversy. For ease of exposition, henceforth we will consider onyl scalar-valued random processes, although the theory holds in the vector-valued case also. An introduction to the theory of the Fokker–Planck equation is given by Wang and Uhlenbeck[1] and Barrett,[2] who also give further references. This theory will not be developed here, but the major results will be stated.

Consider the scalar stochastic differential equation

$$\dot{x}(t) = f(x(t), t) + g(t)\, v(t) \tag{5}$$

Here $v(t)$ is Gaussian white noise, with

$$E\{v(t)\} = 0, \qquad E\{v(t)\, v(s)\} = \delta(t - s) \tag{6}$$

We assume that $|g(t)| > 0$ for all t. We will also assume that $f(x, t)$ and $g(t)$ are at least piecewise continuous functions of t, that f is at least once differentiable with respect to x, and that f obeys the following conditions: there exists K_1, $K_2 < \infty$ such that $|f(x, t)| \leqslant K_1 + K_2 |x|$ for all t and all x.

Aside from the change from vector-valued functions to scalar-valued functions, the major difference between Eq. (2) and Eq. (5) is that in Eq. (5), the function $g(t)$ must be a function of t only, and not a function of x. That is, the white noise enters additively; it is not multiplied by any functions of the solution of the differential equation.

Under this restriction, the Ito and the Stratonovich interpretations of the solution of the differential equation coincide. The divergence only arises when the white noise is multiplied by a function of the solution of the equation.

In the earlier literature,[1] the stochastic differential Eq. (5) is called a Langevin equation. In the more-mathematical modern literature, Eq. (5) is rewritten in a more-rigorous manner. In order to avoid the mathematical pathology associated with white noise, its integral, the so-called Wiener or Brownian motion process $w(t)$ is introduced:

$$w(t) = \int_0^t v(\tau)\, d\tau \tag{7}$$

The process $w(t)$ can be defined independently of $v(t)$, merely by stating that it is Gaussian and that

$$E\{w(t)\} = 0; \qquad E\{w(t)\, w(s)\} = \min\{t, s\} \tag{8}$$

By "multiplying through by dt," Eq. (5) is recast in the form

$$dx(t) = f(x(t), t)\, dt + g(t)\, dw(t) \tag{9}$$

or, by integrating once, in the form of a stochastic integral equation

$$x(t) = x(0) + \int_0^t f(x(\tau), \tau)\, d\tau + \int_0^t g(\tau)\, dw(\tau) \tag{10}$$

The integral on the right, $\int_0^t g(\tau)\, dw(\tau)$, being an integral with respect to a Wiener process, is a new kind of integral, a so-called stochastic integral. However, so long as the function $g(t)$ is restricted as mentioned above, i.e., that it is a nonrandom function of t, then the Ito and Stratonovich interpretations of this integral agree. It may be defined, e.g., as the limit in probability of a sequence of sums of the form

$$\sum_{i=0}^{n-1} g(t_i)[w(t_{i+1}) - w(t_i)]$$

where $0 = t_0 < t_1 < t_2 \cdots < t_n = t$. So far, there is no problem; the usual rules of calculus continue to apply to this integral.

When the engineer tells the mathematician that what he really means by a solution to the Langevin equation (5) is a solution to the integral equation (10), then the mathematician is happy, because he can prove existence and uniqueness of solutions to Eq. (10) with probability 1. Furthermore, the mathematician's solution to Eq. (10) turns out to have the sort of properties that one intuitively expects that solutions to Eq. (5) might have, so the situation is good.

Since for each t, $x(t)$ is a random variable, it has a probability distribution associated with it. Furthermore, this distribution will be smooth enough so that it can be described by a probability density function $p(\xi, t)$. Here, the meaning of this function is that

$$p(\xi, t)\, d\xi = \text{Prob}\{\xi \leqslant x(t) < \xi + d\xi\} \tag{11}$$

The variable ξ is merely a parameter in the density function. It is not the same as the value of the process $x(t)$. In the density function $p(\xi, t)$, the two variables ξ and t are independent variables.

It turns out to be of great interest to study conditional densities, where we condition on the known value of the process at an earlier time. Therefore, define $p(\xi, t \mid \eta, s)$ for $s < t$ by

$$p(\xi, t \mid \eta, s)\, d\xi = \text{Prob}\{\xi \leqslant x(t) < \xi + d\xi \mid x(s) = \eta\} \tag{12}$$

The function $p(\xi, t \mid \eta, s)$ will be a function of all four independent parameters ξ, η, t, and s. When t, η, and s are held fixed, it is probability density function of ξ, e.g., $p(\xi, t \mid \eta, s) \geqslant 0$ and

$$\int_{-\infty}^{\infty} p(\xi, t \mid \eta, s)\, d\xi = 1 \tag{13}$$

Suppose we tried conditioning on several past events. Let $t_1 < t_2 < \cdots < t_n$. Consider the probability

$$\text{Prob}\{\xi \leqslant x(t_n) < \xi + d\xi \mid x(t_i) = \eta_i\,, \quad i = 1, 2,..., n-1\}$$

It turns out that for a process $x(t)$ obtained as the solution to a stochastic integral equation of the form of Eq. (10), this conditional probability is merely equal to $p(\xi, t_n \mid \eta_{n-1}\,, t_{n-1})\, d\xi$.
Written mathematically, what we are saying is

$$\begin{aligned}
&\text{Prob}\{\xi \leqslant x(t_n) < \xi + d\xi \mid x(t_i) = \eta_i\,, \quad i = 1, 2,..., n-1\} \\
&\quad = \text{Prob}\{\xi \leqslant x(t_n) < \xi + d\xi \mid x(t_{n-1}) = \eta_{n-1}\} \\
&\quad = p(\xi, t_n \mid \eta_{n-1}\,, t_{n-1})\, d\xi
\end{aligned} \tag{14}$$

Any process $x(t)$ for which Eq. (14) holds for every integer n, for every choice of $t_1\,, t_2\,,..., t_n$, provided only that $t_1 < t_2 < \cdots < t_{n-1} < t_n$, is called a Markov process. Stated in words, the defining property of a Markov process is that the single most recently observed value of the process contains as much information about the future evolution of the process as does knowledge of the entire past history of the process up to and including the most recently observed value.

The conditional probability density function $p(\xi, t \mid \eta, s)$ plays a fundamental role in the study of continuous Markov processes. This function is customarily called the transition density for the process. The transition density $p(\xi, t \mid \eta, s)$ may be obtained by solving the forward Fokker–Planck equation (also called the forward Kolmogorov equation)

$$\frac{\partial p(\xi, t \mid \eta, s)}{\partial t} = -\frac{\partial}{\partial \xi}\,[f(\xi, t)\, p(\xi, t \mid \eta, s)] + \frac{1}{2}\, g^2(t)\, \frac{\partial^2 p(\xi, t \mid \eta, s)}{\partial \xi^2} \tag{15}$$

$$-\infty < \xi < +\infty, \qquad t > s$$

with the boundary conditions

$$\lim_{t \to s} p(\xi, t \mid \eta, s) = \delta(\xi - \eta); \qquad \lim_{|\xi| \to \infty} p(\xi, t \mid \eta, s) = 0 \tag{16}$$

The transition density may be obtained equally well by solving the backward backward Fokker–Planck or Kolmogorov equation

$$-\frac{\partial p(\xi, t \mid \eta, s)}{\partial s} = f(\eta, s)\frac{\partial p(\xi, t \mid \eta, s)}{\partial \eta} + \frac{1}{2} g^2(t) \frac{\partial^2 p(\xi, t \mid \eta, s)}{\partial \eta^2} \tag{17}$$
$$-\infty < \eta < +\infty, \qquad s < t$$

with the boundary conditions

$$\lim_{s \to t} p(\xi, t \mid \eta, s) = \delta(\xi - \eta); \qquad \lim_{|\eta| \to \infty} p(\xi, t \mid \eta, s) = 0 \tag{18}$$

Equation (15) is a partial differential equation for p considered as a function of the independent variables ξ and t. The variables η and s are merely parameters which enter through the boundary conditions (16). On the other hand, Eq. (17) is a partial differential equation for p as a function of the independent variables η and s. Here ξ and t are merely parameters which enter through the boundary conditions (18). The coefficient functions f and g are the functions defined in Eqs. (5) and (6).

From an engineering standpoint, the situation may be summarized by saying that a complete probabilistic analysis of the properties of a stochastic dynamic system described by Eq. (5) may be made by finding the transition density $p(\xi, t \mid \eta, s)$ as a solution to one of the Fokker–Planck equations (if it satisfies one, it necessarily satisfies the other). This statement is accurate, provided one is careful what he does in such an analysis. The next sections will show what it means to be careful.

4. AN APPARENT PARADOX

Let us consider the Wiener process introduced in Eq. (7). The preceding theory applies to this process, since by setting $x(0) = 0, f = 0, g = 1$, Eq. (10) becomes

$$x(t) = \int_0^t dw(\tau) \tag{19}$$

i.e., $x(t) = w(t)$. In order to make our point, it will suffice to consider only the forward equation (15), and to consider its solution only for the special case of $s = 0$, $\eta = 0$ in Eq. (16).

Denote this solution by $q(\xi, t)$. Thus,

$$q(\xi, t)\, d\xi = \text{Prob}\{\xi \leqslant x(t) < \xi + d\xi \mid x(0) = 0\} \tag{20}$$

where now $x(t)$ is a Wiener process.

It is well known that $q(\xi, t)$ is given by

$$q(\xi, t) = \frac{1}{(2\pi t)^{1/2}} \exp\left[-\frac{\xi^2}{2t}\right] \tag{21}$$

It is easily verified that this function obeys the forward equation

$$\partial q(\xi, t)/\partial t = \tfrac{1}{2}\, \partial^2 q(\xi, t)/\partial \xi^2 \tag{22}$$

and satisfies the boundary condition

$$\lim_{t\to 0} q(\xi, t) = \delta(\xi) \tag{23}$$

Now, suppose the Wiener process is passed through a memoryless nonlinear device to produce a new process $z(t)$. Since the device is memoryless, the process $z(t)$ will still be Markov, and the probability density for it will obey a Fokker–Planck equation. Specifically, suppose that

$$z(t) = \sinh[x(t)] \tag{24}$$

Define

$$p(\zeta, t)\, d\zeta = \text{Prob}\{\zeta \leqslant z(t) < \zeta + d\zeta \mid z(0) = 0\} \tag{25}$$

By the rule for change of variables in probability densities,

$$p(\zeta, t) = q(\xi, t) \frac{d\xi}{d\zeta}\Big|_{\xi=\sinh^{-1}\zeta} \tag{26}$$

Now, $(d/d\zeta) \sinh^{-1}\zeta = (1 + \zeta^2)^{1/2}$, so

$$p(\zeta, t) = \left[\frac{1 + \zeta^2}{2\pi t}\right]^{1/2} \exp\left[-\frac{(\sinh^{-1}\zeta)^2}{2t}\right] \tag{27}$$

Either by making the change of independent variable $\xi = \sinh^{-1}\zeta$ in Eq. (22) and using Eq. (26), or by direct differentiation of Eq. (27), one finds that the Fokker–Planck equation satisfied by $p(\zeta, t)$ is

$$\frac{\partial p(\zeta, t)}{\partial t} = -\frac{\partial}{\partial \zeta}\left[\frac{\zeta}{2} p(\zeta, t)\right] + \frac{1}{2}\frac{\partial^2}{\partial \zeta^2}\left[(1 + \zeta^2)\, p(\zeta, t)\right] \tag{28}$$

According to the theory given by Doob,[3] this is the forward equation corresponding to the stochastic differential equation

$$dz(t) = \tfrac{1}{2} z(t)\, dt + [1 + z^2(t)]^{1/2}\, dw(t) \tag{29}$$

This resembles the stochastic differential equation (9) discussed in the previous section. However, in regard to Eq. (9), it was specifically stated that the coefficient $g(t)$ which multiplies the noise had to be a nonrandom function of t only. In Eq. (29), the coefficient of the noise, namely $[1 + z^2(t)]^{1/2}$, is a function of $z(t)$.

Now, if we simply compute $dz(t)$ from Eq. (24) using the chain rule of ordinary calculus, we find

$$\begin{aligned} dz(t) &= \frac{\partial}{\partial x} \sinh x\, \Big|_{x(t)=\sinh^{-1}[z(t)]} dx(t) = \cosh x\, \Big|_{x(t)=\sinh^{-1}[z(t)]} dx(t) \\ &= [1 + \sinh^2 x]^{1/2}\, \Big|_{x=\sinh^{-1}[z(t)]} dx(t) \\ &= [1 + z^2(t)]\, dx(t) \end{aligned} \tag{30}$$

Since, in the present case, $x(t) = w(t)$, this may be rewritten

$$dz(t) = [1 + z^2(t)]^{1/2}\, dw(t) \tag{31}$$

The stochastic differential equations (29) and (31) differ by the term $\frac{1}{2}z\, dt$. The question is, which is the correct stochastic differential equation for generating the process $z(t)$ from a Wiener process?

Ito and Doob would say Eq. (29) is the correct equation. Stratonovich would say that Eq. (31) is the correct equation. Let us pinpoint the exact issue of disagreement by first stating the facts on which there is agreement:

1. The Wiener process is a well-defined process. Its probability density, given that the process starts at zero at time zero, is correctly given by Eq. (21), and this function satisfies Eqs. (22) and (23).

2. The process $z(t)$ defined by Eq. (24) is a well-defined process. Its density function, defined in Eq. (25), is correctly given by Eq. (27), and this function satisfies Eq. (28). Thus, in particular, both Stratonovich and Ito would agree that Eq. (28) is the correct Fokker–Planck equation for the $z(t)$ process defined by Eq. (24).

3. It is agreed that if we integrate Eq. (31) according to the rules of ordinary calculus, we do get $z(t) = \sinh[w(t)]$ as the solution, while if we integrate Eq. (31) according to the Ito calculus, we do not get this as the solution.

4. Ito and Stratonovich would both agree that if we integrate Eq. (29) according to the rules of Ito calculus, we do get $z(t) = \sinh[w(t)]$ as the solution, while if we integrate Eq. (31) according to ordinary calculus, we do not get this as a solution.

Therefore, the situation is that the one unambiguous way to specify a Markov process mathematically is to specify its transition density, or, equivalently, the Fokker–Planck equation obeyed by the transition density. The divergence arises when one wishes to generate the specified process as a solution to a stochastic differential equation forced by the differential of a Wiener process. The divergence boils down to two different ways of associating the coefficients in the Fokker–Planck equation with the coefficients in the stochastic differential equation, and, respectively, two ways of integrating this stochastic equation.

Each way is consistent within itself, as we have seen. Starting from the process $z(t)$ defined by the Fokker–Planck equation (28), the use of Stratonovich rules associates the stochastic differential equation (31) with Eq. (28). Integrating Eq. (31) by by the Stratonovich rules yields $z(t) = \sinh[w(t)]$.

On the other hand, the use of Ito rules will associate the stochastic differential equation (29) with the Fokker–Planck equation (28). However, integrating Eq. (29) by the Ito rules again yields $z(t) = \sinh[w(t)]$. Further, Ito would say that the computation of the differential $dz(t)$ in Eq. (30) is incorrect; if this computation is done by Ito rules, then Eq. (29) results. However, Stratonovich would say that Eq. (30) is a perfectly valid computation.

At first glance, it might seem academic to worry about this divergence between Ito rules and Stratonovich rules. Each set of rules is consistent within itself. If the same set of rules is consistently applied throughout the whole computation, both methods yields the same result.

The mathematician discusses Markov process by *starting* with the *transition density* for the process. He is able to associate a Fokker–Planck equation in an unambiguous way with this transition density. When he finds that he has two possible ways of modeling the process as the solution to a stochastic differential equation, he will choose the way which has the most mathematical elegance in its internal structure, and which is capable of the greatest generalization. Considered from this standpoint, the Ito calculus is the "right" choice. Indeed, the procedure just described is precisely, the one followed by Doob in his book.[3]

However, the question is not so simple for the engineer. He cannot resolve the issue on the basis of mathematical elegance alone. The engineer does *not* start with the transition density. As discussed in the earlier sections, the engineer *starts* with a *differential equation* which he has obtained on the basis of known physical laws. He then adds a white-noise forcing term to get a stochastic model. If the coefficient of the noise is itself random, then there are two possible ways of interpreting the equation, leading to two different Fokker–Planck equations and *two different processes.* The question is, *which* process does one "really" get in the physical world? Which kind of calculus does nature use?

The answer to this question hinges on whether white noise "really" exists, or whether the concept of white noise is only a convenient approximation which we use in place of a more-detailed knowledge of the properties of the noise process. The true situation is certainly the latter, since noise with a truly flat power density spectrum out to infinite frequency would carry infinite total power. However, this then implies that there is really no such thing as a Markov process either, and the whole theory of the Fokker–Planck equation goes down the drain.

Therefore, the whole theory of white noise, stochastic differential equation, Markov processes, and the Fokker–Planck equation must be approached from the standpoint of an approximate model rather than an exact model of physical reality. It is, of course, possible to use nonwhite noise in the model, but now one is faced with the problem of specifying the power density spectrum of the noise, which is usually completely unknown at high frequencies, even though it can be measured as flat at low frequencies. Furthermore, use of a nonflat high-frequency spectrum complicates the computations tremendously.

Once one realizes the kind of approximation that is being made, it turns out that it is possible to use either the Ito or the Stratonovich rules and obtain equally accurate results, provided that one is careful in setting up the mathematical model and that one is aware of the subtleties involved.

The paradox of obtaining two different stochastic processes as solutions to the same stochastic differential equation thus turns out to arise from the pathological nature of white noise. This paradox can be avoided by treating this pathology with propre respect. In the following sections, we will examine the situation in more detail.

5. THE ITO CALCULUS

In order to introduce the Ito calculus, let us begin by examining the Wiener process $w(t)$ more carefully. Let Δt be some very small, but not infinitesimal, increment of time. Define

$$\Delta w(t) = w(t + \Delta t) - w(t) \tag{32}$$

For fixed t and Δt, $w(t + \Delta t)$ and $w(t)$ are both Gaussian random variables, so $\Delta w(t)$ is also a Gaussian random variable.

Let $q(\xi, t \mid \eta, s)$ be the transition density for the Wiener process, i.e.,

$$q(\xi, t \mid \eta, s)\, d\xi = \text{Prob}\{\xi \leqslant w(t) < \xi + d\xi \mid w(s) = \eta\} \tag{33}$$

By the definition of the Wiener process, this density is given by

$$q(\xi, t \mid \eta, s) = \frac{1}{[2\pi(t - s)]^{1/2}} \exp\left[-\frac{(\xi - \eta)^2}{2(t - s)}\right] \tag{34}$$

With somewhat of an abuse of notation, define the conditional probability density

$$p_{\Delta w}(\Delta\xi \mid \xi)\, d(\Delta\xi) = \text{Prob}\{\Delta\xi \leqslant \Delta w(t) < \Delta\xi + d(\Delta\xi) \mid w(t) = \xi\} \tag{35}$$

Since we are conditioning on the fixed event $\{w(t) = \xi\}$, observe that

$$\begin{aligned}
&\text{Prob}\{\Delta\xi \leqslant \Delta w(t) < \Delta\xi + d(\Delta\xi) \mid w(t) = \xi\} \\
&\quad = \text{Prob}\{\xi + \Delta\xi \leqslant w(t) + \Delta w(t) < \xi + \Delta\xi + d(\Delta\xi) \mid w(t) = \xi\} \\
&\quad = \text{Prob}\{\xi + \Delta\xi \leqslant w(t + \Delta t) < \xi + \Delta\xi + d(\Delta\xi) \mid w(t) = \xi\}
\end{aligned} \tag{36}$$

It follows from Eqs. (33)–(36) that

$$\begin{aligned}
p_{\Delta w}(\Delta\xi \mid \xi) &= q(\xi + \Delta\xi, t + \Delta t \mid \xi, t) \\
&= \frac{1}{\{2\pi[(t + \Delta t) - t]\}^{1/2}} \exp\left\{\frac{[(\xi + \Delta\xi) - \xi]^2}{2[(t + \Delta t) - t]}\right\} \\
&= \frac{1}{(2\pi\, \Delta t)^{1/2}} \exp\left[-\frac{(\Delta\xi)^2}{2\Delta t}\right] = p_{\Delta w}(\Delta\xi)
\end{aligned} \tag{37}$$

In the last line of Eq. (37) we have written $p_{\Delta w}(\Delta\xi)$ to denote the unconditional probability density for the random variable $\Delta w(t)$, i.e.,

$$p_{\Delta w}(\Delta\xi)\, d(\Delta\xi) = \text{Prob}\{\Delta\xi \leqslant \Delta w(t) < \Delta\xi + d(\Delta\xi)\} \tag{38}$$

The important point is that

$$p_{\Delta w}(\Delta\xi \mid \xi) = p_{\Delta w}(\Delta\xi) \tag{39}$$

i.e., the distribution of the increment $\Delta w(t)$ is *independent* of $w(t)$, the state of the process at time t. This is not generally true of random processes, or even of Markov processes. The Wiener process $w(t)$ belongs to a special class of processes known as *processes with independent increments.*

From Eqs. (37) and (38), we see that the random variable $\Delta w(t)$ defined in Eq. (32) is Gaussian with mean zero and variance Δt. The fact that $E\{(\Delta w)^2\}$ is *first order* in Δt is what causes the peculiarities of the Ito stochastic calculus.

Let F be any smooth, real-valued, nonlinear function of a real variable. Consider $F(w(t+\Delta t))$ where $w(t)$ continues to denote the Wiener process. By Taylor series and Eq. (32),

$$\begin{aligned} F(w(t+\Delta t)) &= F(w(t)+\Delta w(t)) \\ &= F(w(t)) + F'(w(t))\,\Delta w(t) + \tfrac{1}{2}F''(w(t))[\Delta w(t)]^2 + \cdots \end{aligned} \tag{40}$$

Using the distribution of $\Delta w(t)$, we have

$$\begin{aligned} E\{(\Delta w)^k\} &= 0, && k \text{ odd} \\ &= 1.3.5 \ldots (k-1)(\Delta t)^{k/2}, && k \text{ even} \end{aligned} \tag{41}$$

Use the notation $O(\Delta t)$ to denote a remainder consisting only of terms of order $(\Delta t)^2$ and higher. Suppose now we tried to define

$$\frac{d}{dt} F(w(t)) = \lim_{\Delta t \to 0} \frac{F(w(t+\Delta t)) - F(w(t))}{\Delta t} \tag{42}$$

From Eqs. (40) and (41), it follows that

$$\begin{aligned} E\left\{\frac{d}{dt} F(w(t))\right\} &= \lim_{\Delta t \to 0} \frac{\tfrac{1}{2}E\{F''(w(t))\}\,\Delta t + O(\Delta t)}{\Delta t} \\ &= \frac{1}{2} E\{F''(w(t))\} \end{aligned} \tag{43}$$

On the other hand, if one computer the total differential $dF(w(t))$ using the chain rule of ordinary calculus, one has

$$dF(w(t)) = F'(w(t))\,dw(t) \tag{44}$$

By passing from an increment $\Delta w(t)$ to a differential $dw(t)$, it follows from Eq. (39) that $dw(t)$ is independent of $w(t)$. Therefore,

$$\begin{aligned} E\{dF(w(t))\} &= E\{F'(w(t))\,dw(t)\} \\ &= E\{F'(w(t))\}\,E\{dw(t)\} = 0 \end{aligned} \tag{45}$$

since $E\{dw(t)\} = 0$ by Eq. (41). Now, Eq. (45) would imply that

$$E\left\{\frac{d}{dt} F(w(t))\right\} = 0 \tag{46}$$

in contradiction to Eq. (43).

The point is that because of the fact that $E\{(\Delta w)^2\} = \Delta t$, the definition of the derivative (42) no longer leads to the usual rules of calculus. Ito was the first to show how the rules of calculus should be modified to handle this phenomenon. First of all, instead of computing the derivative as in (42), one should compute the differential $dF(t)$ because the differential $dw(t)$ can be rigorously interpreted, whereas the derivative $dw(t)/dt$ can not.

As given by Skorokhod,[4] the Ito rule for the stochastic differential in the present case is

$$d_I F(w(t)) = F'(w(t))\, dw(t) + \tfrac{1}{2} F''(w(t))\, dt \tag{47}$$

where now d_I means Ito differential. Note that this rule is now consistent with Doob's treatment of the Fokker–Planck equation. Let us apply Eq. (47) in the special case when $F(w(t)) = \sinh[w(t)]$. Now, $F'(w) = \cosh w$, $F''(w) = \sinh w$, so Eq. (47) says

$$d_I \sinh[w(t)] = \cosh w(t)\, dw(t) + \tfrac{1}{2} \sinh w(t)\, dt \tag{48}$$

Let us write $z = \sinh w$. Then, $\cosh w = [1 + z^2]^{1/2}$, so Eq. (48) can be rewritten as

$$d_I z(t) = \tfrac{1}{2} z(t)\, dt + [1 + z^2(t)]^{1/2}\, dw(t) \tag{49}$$

which is the same as Eq. (29). Thus, the Ito rule Eq. (47) for the total differential is consistent with the Fokker–Planck equation (28).

Since the rule for computing total differentials has now been changed from Eq. (44) to Eq. (47), we must expect a corresponding change in the rule for integration. Let us write $(I)\int$ when an integral is to be understood in the Ito sense, and continue to write just $\int$ for ordinary integrals.

We wish to preserve the fundamental property of calculus, that the integral can be interpreted as an antiderivative. Therefore, we *require* that

$$(I) \int_{t_0}^{t_1} d_I F(w(t)) = F(w(t_1)) - F(w(t_0)) \tag{50}$$

Applying this to Eq. (47) yields

$$\begin{aligned} (I) \int_{t_0}^{t_1} d_I F(w(t)) &= F(w(t_1)) - F(w(t_0)) \\ &= (I) \int_{t_0}^{t_1} F'(w(t))\, dw(t) + \frac{1}{2} \int_{t_0}^{t_1} F''(w(t))\, dt \end{aligned} \tag{51}$$

This may be rewritten as

$$(I) \int_{t_0}^{t_1} F'(w(t))\, dw(t) = F(w(t_1)) - F(w(t_0)) - \frac{1}{2} \int_{t_0}^{t_1} F''(w(t))\, dt \tag{52}$$

Now, let $g(x)$ be any once-differentiable function. Define

$$G(x) = \int_0^x g(\xi)\, d\xi \tag{53}$$

Then, using Eq. (52) with F replaced by G,

$$(I)\int_{t_0}^{t_1} g(w(t))\,dw(t) = (I)\int_{t_0}^{t_1} G'(w(t))\,dw(t)$$

$$= G(w(t)) - G(w(t_0)) - \frac{1}{2}\int_{t_0}^{t_1} G''(w(t))\,dt$$

$$= \int_{w(t_0)}^{w(t)} g(\xi)\,d\xi - \frac{1}{2}\int_{t_0}^{t_1} g'(w(t))\,dt \tag{54}$$

In Eq. (54), the notation $\int_{w(t_0)}^{w(t_1)} g(\xi)\,d\xi$ means compute $\int g(\xi)\,d\xi$ as an ordinary integral, treating ξ as a deterministic dummy variable of integration, and then evaluate between the random limits $w(t_1)$ and $w(t_0)$. This, incidentally, is essentially what Stratonovich has in mind in his definition of the stochastic integral.

Let $(S)\int$ denote the Stratonovich integral. Then, in the present context,

$$(S)\int_{t_0}^{t_1} g(w(t))\,dw(t) = \int_{w(t_0)}^{w(t_1)} g(\xi)\,d\xi \tag{55}$$

Therefore, Eq. (54) can be rewritten

$$(I)\int_{t_0}^{t_1} g(w(t))\,dw(t) = (S)\int_{t_0}^{t_1} g(w(t))\,dw(t) - \frac{1}{2}\int_{t_0}^{t_1} g'(w(t))\,dt \tag{56}$$

which is a special case of the formula given by Stratonovich[5] for the connection between Ito integrals and Stratonovich integrals.

The Ito calculus has some surprising consequences. For example, let $g(w(t)) = w(t)$ in Eq. (56). By the notation $g'(w(t))$ we mean, of course,

$$g'(w(t)) = \left.\frac{dg(\xi)}{d\xi}\right|_{\xi=w(t)} \tag{57}$$

so that in the present case $g'(w(t)) = 1$. Now,

$$\int \xi\,d\xi = \tfrac{1}{2}\xi^2 \tag{58}$$

so, using Eqs. (55) and (56), we obtain

$$(I)\int_{t_0}^{t_1} w(t)\,dw(t) = \tfrac{1}{2}w^2(t_1) - \tfrac{1}{2}w^2(t_0) - \tfrac{1}{2}(t_1 - t_0) \tag{59}$$

an example which is also given by Doob.

The presence of the $\frac{1}{2}(t_1 - t_0)$ term in Eq. (59) can be made more plausible by the following considerations. Let us consider

$$E\left\{(I)\int_{t_0}^{t_1} w(t)\,dw(t)\right\} = \int_{t_0}^{t_1} E\{w(t)\,dw(t)\} \tag{60}$$

As in Eq. (45), we have

$$E\{w(t)\,dw(t)\} = E\{w(t)\}\,E\{dw(t)\} = 0 \tag{61}$$

since the increment $dw(t)$ is understood to be independent of $w(t)$. Therefore, we conclude

$$E\left\{(I)\int_{t_0}^{t_1} w(t)\,dw(t)\right\} = 0 \tag{62}$$

Now, recall that the Wiener process was defined such that $w(0) = 0$. Thus, by Eq. (21), we have that

$$E\{w^2(t_1)\} = t_1\,; \qquad E\{w^2(t_0)\} = t_0 \tag{63}$$

Taking the expected value of both sides of Eq. (59) now gives

$$\begin{aligned} E\left\{(I)\int_{t_0}^{t_1} w(t)\,dw(t)\right\} &= \tfrac{1}{2}E\{w^2(t_1)\} - \tfrac{1}{2}E\{w^2(t_0)\} - \tfrac{1}{2}(t_1 - t_0) \\ &= \tfrac{1}{2}t_1 - \tfrac{1}{2}t_0 - \tfrac{1}{2}(t_1 - t_0) = 0 \end{aligned} \tag{64}$$

in agreement with Eq. (62). Thus, the $\frac{1}{2}(t_1 - t_0)$ can be viewed as a correction term which ensures that Eq. (62) holds.

However, these considerations also imply that for the Stratonovich integral,

$$E\left\{(S)\int_{t_0}^{t_1} w(t)\,dw(t)\right\} = \tfrac{1}{2}(t_1 - t_0) \neq 0 \tag{65}$$

Thus, for the Stratonovich integral, it cannot be true that $dw(t)$ is independent of $w(t)$, for we have just seen that this independence is what makes the expected value of the Ito integral always zero.

In fact, this is precisely the case. Stratonovich interpretes the differential $dw(t)$ in such a way that it is not independent of $w(t)$. The Ito calculus is based on the fact that the increment $\Delta w(t)$ defined in Eq. (32) is independent of $w(t)$, and has mean zero and variance Δt.

In contrast, Stratonovich works with a "Stratonovich increment" defined as

$$\Delta_S w(t) = w(t + \tfrac{1}{2}\Delta t) - w(t - \tfrac{1}{2}\Delta t) \tag{66}$$

This increment still has mean zero and variance Δt, but it is *not* independent of $w(t)$. We will examine the Stratonovich calculus in more detail in the next section.

This report is written in such a way as to be (we hope) pedagogically palatable to engineers. Consequently, our treatment of stochastic differential equations and stochastic calculus differs drastically from the rigorous mathematical treatment given by Doob[3] and Skorokhod.[4] Rather than carefully stating and proving theorems, we are trying to convey the basic ideas involved by considering only special cases and examining illustrative examples.

So far, we have discussed the Ito calculus by following the approach historically used in presenting ordinary calculus to students for the first time. Namely, we we introduced the derivative first, as the sort of limit given in Eq. (42). The integral was then introduced as an antiderivative.

In the modern, rigorous approach to calculus, which is usually presented to students only after their intuition has been sharpened, the integral is defined directly from first principles. The Riemann integral is defined as a limit of Riemann sums, and the Lebesgue integral is defined by use of measure theory.

Similarly, in a rigorous approach to stochastic calculus, the Ito integral is defined first, as a stochastic limit of Riemann-type sums. The Ito differential formula Eq. (47) is then derived as a consequence of this integral.

Let us sketch briefly the definition of the Ito integral as a limit of sums.Let$w(t)$ be a Wiener process. Let $z(t)$ be any random process having the properties that for all t, $z(t)$ and $[w(\tau) - w(t)]$ are independent for all $\tau > t$, and that

$$\int_0^T z^2(t)\, dt < \infty \tag{67}$$

with probability one. Note that for $\tau \leqslant t$, $z(t)$ and $[w(\tau) - w(t)]$ may be dependent. Let $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$. Let

$$\Delta_n = \max_{1 \leqslant i \leqslant n} | t_i - t_{i-1} | \tag{68}$$

Choose any sequence of partitions t_0, t_1,..., t_n such that $\lim_{n\to\infty} \Delta_n = 0$. The object is to define the Ito integral

$$J = (I) \int_0^T z(t)\, dw(t) \tag{69}$$

Define

$$I_n = \sum_{k=1}^{n} z(t_{k-1})[w(t_k) - w(t_{k-1})] \tag{70}$$

Note that the integrand $z(t_{k-1})$ is always evaluated at the beginning of the interval $[t_{k-1}, t_k]$ over which the increment $[w(t_k) - w(t_{k-1})]$ is taken. Therefore, $z(t_{k-1})$ and $[w(t_k) - w(t_{k-1})]$ are always independent. Consequently,

$$E\{I_n\} = \sum_{k=1}^{n} E\{z(t_{k-1})\}\, E\{w(t_k) - w(t_{k-1})\} = 0 \tag{71}$$

It is now possible to prove that sequence of random variables I_n converges in probability to some limiting random variable J. This limit is called the Ito integral It has the property that $E\{J\} = 0$.

Note that the class of random processes $z(t)$ which may be used as integrands here is very broad. It is only required that $z(t)$ be square-integrable over the interval of integration and that the present value of $z(t)$ is always independent of all future increments of $w(t)$. In fact, there is not even any reason why the integrating process

$w(t)$ has to be a Wiener process. Doob[3] and Skorokhod[4] this in detail. The point is that the definition of the stochastic integral given by Ito is really quite general, much more so than our heuristic derivation of Eq. (56) would indicate.

6. THE STRATONOVICH CALCULUS

In the previous section, we asserted that a derivative defined as a limit of the form of Eq. (42) is consistent with an integral defined as a limit of sums of the form of Eq. (70), and we gave some examples to make this assertion plausible. The resulting stochastic calculus is called the Ito calculus. By examples such as Eqs. (48) and (59), it was illustrated that the rules of the Ito calculus differ from the usual rules of ordinary calculus.

Stratonovich[5] proposed a definition of the stochastic integral under rather restrictive conditions which leads to a stochastic calculus whose rules are the same as ordinary calculus. Basically, what Stratonovich did was to show that the formula (56) could be made rigorous. Thus, with the Ito integral on the left-hand side of Eq. (56) already well-defined, the Stratonovich integral on the right-hand side of Eq. (56) becomes well-defined.

Therefore, Stratonovich did not give a fundamental definition of a new stochastic integral, but only defined the new integral in terms of the already existing Ito integral. Furthermore, the new integral is not defined for forms as general as Eq. (69). It is only defined for the special case of Eq. (69) in which $z(t)$ is of the form

$$z(t) = g\big(w(t), t\big) \tag{72}$$

where $g(x, t)$ is a nonrandom function of the two arguments x, t. Consequently, the Ito integral remains both more fundamental and more general than the Stratonovich integral.

It is tempting to suppose that a fundamental definition of the Stratonovich integral could be given, in analogy with Eq. (70), by taking a sequence of sums of the form

$$I_n{}^* = \sum_{k=1}^{n} z\left(\frac{t_{k-1} + t_k}{2}\right) [w(t_k) - w(t_{k-1})] \tag{73}$$

Unfortunately, such a sequence of sums cannot be shown to converge, in general, even in such a weak sense as convergence in probability. The Stratonovich integral is not versatile enough to be suited for many application for which the Ito integral is suited.

The Stratonovich integral is just versatile enough to be suited to the integration of stochastic differential equations. Consider the following generalization of Eq. (9):

$$dx(t) = f\big(x(t), t\big)\, dt + g\big(x(t), t\big)\, dw(t) \tag{74}$$

The functions $f(x, t)$ and $g(x, t)$ are assumed to be jointly continuous in x and t, once-differentiable with respect to x, and to satisfy the following condition: there exist constants K_1, $K_2 < \infty$ such that

$$|f(x, t)| \leqslant K_1 + K_2|x| \qquad \text{and} \qquad |g(x, t)| \leqslant K_1 + K_2|x|$$

for all t and all x.

By rewriting Eq. (74) as an integral equation, one obtains

$$x(t) = x(0) + \int_0^t f(x(\tau), \tau)\, d\tau + \int_0^t g(x(\tau), \tau)\, dw(\tau) \tag{75}$$

The stochastic integral on the right has as its integrand $g(x(\tau), \tau)$, rather than $g(w(\tau), \tau)$ as is required by Eq. (72). However, by giving a multidimensional definition of his integral, Stratonovich was able to show how the integral in Eq. (75) could be recast in the desired form. Therefore, it is possible to say that the stochastic integral on the right-hand side of Eq. (75) can be interpreted as a Stratonovich integral. That is, Stratonovich integrals of the form $\int_0^T g(x(t), t)\, dw(t)$ can be defined, provided $dx(t)$ and $dw(t)$ are connected by a stochastic differential equation such as Eq. (74). This is apparently the most general situation for which the Stratonovich integral can be defined.

It is now possible to give an existence and uniqueness proof of solutions to the stochastic integral equation (75) when the stochastic integral is interpreted in the Stratonovich sense, in analogy to the type of proof using Picard iteration that Doob gives for the case of an Ito integral.

The Stratonovich and the Ito solutions of Eq. (75) will of course be different, because of the divergence between the two integrals indicated by Eq. (56). Call $x_I(t)$ the Ito solution and $x_S(t)$ the Stratonovich solution. Explicitly, we have

$$x_I(t) = x_I(0) + \int_0^t f(x_I(\tau), \tau)\, d\tau + (I) \int_0^t g(x_I(\tau), \tau)\, dw(\tau) \tag{76}$$

$$x_S(t) = x_S(0) + \int_0^t f(x_S(\tau), \tau)\, d\tau + (S) \int_0^t g(x_S(\tau), \tau)\, dw(\tau) \tag{77}$$

Although $x_I(t)$ and $x_S(t)$ are two different processes, they both still turn out to be Markov processes. Call $\rho_I(\xi, t \mid \eta, s)$ the transition density associated with $x_I(t)$, and $\rho_S(\xi, t \mid \eta, s)$ the transition density associated with $x_S(t)$.

As given by Doob, $\rho_I(\xi, t \mid \eta, s)$ obeys, respectively, the forward and backward Kolmogorov equations

$$\frac{\partial \rho_I(\xi, t \mid \eta, s)}{\partial t} = -\frac{\partial}{\partial \xi} [f(\xi, t)\, \rho_I(\xi, t \mid \eta, s)] + \frac{1}{2} \frac{\partial^2}{\partial \xi^2} [g(\xi, t)\, \rho_I(\xi, t \mid \eta, s)] \tag{78}$$

$$-\frac{\partial \rho_I(\xi, t \mid \eta, s)}{\partial s} = f(\eta, s) \frac{\partial \rho_I(\xi, t \mid \eta, s)}{\partial \eta} + \frac{1}{2} g^2(\eta, s) \frac{\partial^2 \rho_I(\xi, t \mid \eta, s)}{\partial \eta^2} \tag{79}$$

On the other hand, Stratonovich shows that $\rho_S(\xi, t \mid \eta, s)$ obeys, respectively, the forward and backward equations

$$\frac{\partial \rho_S(\xi, t \mid \eta, s)}{\partial t} = -\frac{\partial}{\partial \xi}[f(\xi, t)\,\rho_S(\xi, t \mid \eta, s)] + \frac{1}{2}\left\{\frac{\partial}{\partial \xi} g(\xi, t)\,\frac{\partial}{\partial \xi}[g(\xi, t)\,\rho_S(\xi, t \mid \eta, s)]\right\} \tag{80}$$

$$-\frac{\partial \rho_S(\xi, t \mid \eta, s)}{\partial s} = f(\eta, s)\,\frac{\partial \rho_S(\xi, t \mid \eta, s)}{\partial \eta} + \frac{1}{2} g(\eta, s)\,\frac{\partial}{\partial \eta}\left[g(\eta, s)\,\frac{\partial \rho_S(\xi, t \mid \eta, s)}{\partial \eta}\right] \tag{81}$$

We saw earlier that if Eq. (29) is interpreted in the Ito sense, then the appropriate forward Kolmogorov equation is (28). If one now uses the Stratonovich rule Eq. (80) for the forward equation, one finds that if Eq. (31) is interpreted in the Stratonovich sense, then the appropriate forward equation is again Eq. (28). This is as it should be, since the Ito solutionof Eq. (29) and the Stratonovich solution of Eq. (31) are the same process, namely,

$$z(t) = \sinh[w(t)] \tag{82}$$

as we saw earlier.

This would suggest that it ought to be possible to obtain the Ito solution $x_I(t)$ of Eq. (76) also as the solution of some Stratonovich equation, and *vice versa*. Indeed, this turns out to be the case. It was shown[5,8] that $x_I(t)$ also obeys

$$x_I(t) = x_I(0) + \int_0^t [f(x_I(\tau), \tau) - \tfrac{1}{2} g(x_I(\tau), \tau)\, g_1(x_I(\tau), \tau)]\, d\tau + (S)\int_0^t g(x_I(\tau), \tau)\, dw(\tau) \tag{83}$$

where

$$g_1(x, t) = \partial g(x, t)/\partial x \tag{84}$$

Similarly, the solution $x_S(t)$ of Eq. (77) also obeys

$$x_S(t) = x_S(0) + \int_0^t [f(x_S(\tau), \tau) + \tfrac{1}{2} g(x_S(\tau), \tau)\, g_1(x_S(\tau), \tau)]\, d\tau + (I)\int_0^t g(x_S(\tau), \tau)\, dw(\tau) \tag{85}$$

Therefore, although the Ito integral is more fundamental and more general than the Stratonovich integral, it turns out that when we restrict our attention to stochastic differential equations of the form of Eq. (74), the two definitions of the stochastic integral lead to two different, but interchangeable, theories.

7. MODELING THE REAL WORLD

We say in the last section that the stochastic differential equation (74) is ambiguous. The ambiguity may be removed by writing the equation in integral form with the type of integral definitely indicated, as in Eqs. (76) and (77).

We now return to the situation discussed at the beginning of this report. Suppose an engineer has a deterministic model of a dynamic system of the form of Eq. (1). Suppose that he now wants to include the effects of stochastic forces in the environment, and that physical reasoning suggests that a plausible stochastic model is Eq. (2). Which way should he interpret this equation, Ito or Stratonovich? Which kind of stochastic integration does Nature herfself perform?

In order to answer this question, it must be kept clearly in mind exactly what is the purpose of a mathematical model. Presumably, we have in front of us a physical dynamic system, i.e., a "black box," whose output is a random process. For simplicity, suppose this random process is scalar-valued, and call it $y(t)$.

In order to take advantage of the theory of Markov processes, one wishes to obtain $y(t)$ by means of a state-output relation of the form

$$y(t) = h(\mathbf{x}(t), t) \tag{86}$$

where $\mathbf{x}(t)$ is an n-dimensional vector-valued Markov process. The value of n, the statistics of the process $\mathbf{x}(t)$, and the deterministic function h are to be chosen in some suitable way so that the statistical properties of the process $y(t)$ obtained from Eq. (86) approximate to an acceptable degree of accuracy the sample statistics of the observed output of the black box.

It will further be convenient to obtain the Markov process $\mathbf{x}(t)$ by means of a stochastic differential equation of the form of Eq. (2). Once the statistics of the $\mathbf{x}(t)$ process have been specified, we have seen in the previous sections how the functions f and G may be chosen so that either the Ito or the Stratonovich interpretation may be used.

Since the form of the function h in Eq. (86) and the coordinatization of the state space are at our disposal, one may be able to make this choice in such a way that the matrix $\mathbf{G}$ in Eq. (2) is not a function of $\mathbf{x}(t)$, i.e., $\mathbf{G}$ would be a purely deterministic function of time. In this case, it is possible to avoid the Ito–Stratonovich divergence altogether, as we have seen.

The point of view being taken here is that the modeling problem consists of trying to make the statistics of the *output* of the *mathematical model* agree with the statistics of the *physically observable output* of a given black box. There is no claim that Eqs. (2) and (86) "really" portray what is "actually happening" *inside* the box, since the *inside* of the box is not observable to us.

This philosophical approach to the problem is generally known as the *phenomenological* approach, in contrast to what might be called an *axiomatic* approach.

If one adopts this phenomenological approach of working backward from the output with the only objective being to match the generated output with the observed data, then the choice between the Ito and Stratonovich calculi becomes merely a matter of personal preference. On this level, mathematicians will prefer the Ito calculus because of its elegance and generality, while engineers will prefer the Stratonovich calculus because of their familiarity with its rules.

It seems to the present author that this is perhaps the best resolution of the controversy, since it avoids having to answer the question of whether "Nature" prefers Ito integrals or Stratonovich integrals.

Another way of reaching essentially the same conclusion is to realize that true white noise cannot exist in the physical world. Any noise process, regardless of how flat its power density spectrum appears at low frequencies, must have a spectrum which eventually drops off to zero at sufficiently high frequencies, in order for the total power carried by the process to be finite. Physically, the dropping off the spectrum may occur because of quatum-mechanical effects, if for no other reason. White noise is reminiscent of the "ultraviolet catastrophe" which appeared when blackbody radiation was treated by classical physics.

Consequently, as pointed out previously, the concepts of the Wiener process and of a Markov process are mathematical idealizations which can only approximate physical reality.

Suppose we have a sequence of continuous-time stochastic processes, of finite total power, which become better and better approximations of white noise as one passes to the limit. The point has been made (1, 8) that the Ito and Stratonovich integrals behave differently under passage to the limit. Our point here is that this is no cause for concern, provided that one understands what is happening and views it appropriately, because Nature herself never passes to the limit.

For example, if one wishes to simulate Eq. (2) on a digital computer, since the digital computer operates necessarily in discrete time, the simulation output will be a discrete-time approximation to the desired continuous-time process. It is known how to program the computer so that its output will approximate either the Ito solution of Eq. (2) to any reasonable accuracy.

The same remarks apply to analog simulation. Now, the analog computer operates in continuous time, but since it must necessarily employ a physical noise generator, the spectrum of the noise cannot be truly white. This is in contrast to the digital computer, where it is possible to obtain true *discrete-time* white noise. Nevertheless, Kailath[9] mentioned a way of rigging the analog computer so that it will approximate either Ito integration or Stratonovich integration.

The above remarks still have not answered the question of what an engineer should do when he already has a deterministic model of a physical system and wants to convert it to a stochastic model. The safest answer is that he should throw away the deterministic model, and *remodel* the whole problem, with the objective being to get the statistics of the output of a Monte Carlo computer simulation to agree with the statistics of the observed data from the physical system. Any effort less than this is an attempt to find a short cut, and may yield an incorrect model.

As an example of the kind of situation that may occur in modeling, consider the planar motion of a particle of unit mass, subject to no deterministic forces.

In intertially fixed Cartesian coordinates, the dynamic equations of motion [analogous to Eq. (1)] are

$$\dot{v}_x(t) = 0, \qquad \dot{v}_y(t) = 0 \tag{87}$$

Suppose we introduce flight-path coordinates and write $v_x = V\cos\beta$, $v_y = V\sin\beta$. The flight-path equations of motion are

$$\dot{V} = 0, \qquad V\dot{\beta} = 0 \tag{88}$$

If one integrates both Eq. (87) and Eq. (88), starting from corresponding initial conditions, both Eq. (87) and Eq. (88) yield the same straight line for a trajectory.

Now, consider making the jump from Eq. (1) to Eq. (2). Let $n_x(t)$ and $n_y(t)$be independent Gaussian white noises, each of unity power density. Equation (87) becomes

$$\dot{v}_x(t) = n_x(t), \qquad \dot{v}_y(t) = n_y(t) \tag{89}$$

In order to write these equations in Ito form, introduce the two Wiener processes

$$w_x(t) = \int_0^t n_x(\tau)\, d\tau, \qquad w_y(t) = \int_0^t n_y(\tau)\, d\tau \tag{90}$$

The Ito form of Eq. (89) is

$$dv_x(t) = dw_x(t), \qquad dv_y(t) = dw_y(t) \tag{91}$$

and, of course, the velocity vector of the particle is a two-dimensional Wiener process.

The defining relations for the flight-path coordinates may be written

$$V^2(t) = v_x^2(t) + v_y^2(t), \qquad \beta(t) = \tan^{-1}[v_y(t)/v_x(t)] \tag{92}$$

If one computes total time differentials according to the rules of ordinary calculus, one obtains

$$\begin{aligned} dV &= \frac{v_x}{V}\, dv_x + \frac{v_y}{V}\, dv_y = \cos\beta\, dv_x + \sin\beta\, dv_y \\ d\beta &= -\frac{v_y}{V^2}\, dv_x + \frac{v_x}{V^2}\, dv_y = -\frac{\sin\beta}{V}\, dv_x + \frac{\cos\beta}{V}\, dv_y \end{aligned} \tag{93}$$

By Eqs. (91) and (93), therefore, the Stratonovich form of the stochastic equation of motion in flight path coordinates is

$$\begin{bmatrix} dV(t) \\ d\beta(t) \end{bmatrix} = \begin{bmatrix} \cos\beta(t) & \sin\beta(t) \\ -\dfrac{\sin\beta(t)}{V(t)} & \dfrac{\cos\beta(t)}{V(t)} \end{bmatrix} \begin{bmatrix} dw_x(t) \\ dw_y(t) \end{bmatrix} \tag{94}$$

Now, suppose that one computes the total time differential of Eq. (92) according to the rules of the Ito stochastic calculus, or, alternatively, one computes the Ito correction term for Eq. (94) according to the rule given by Wong and Zakai.[8] Either way, the Ito differentail equation corresponding to Eq. (94) is

$$\begin{bmatrix} dV(t) \\ d\beta(t) \end{bmatrix} = \begin{bmatrix} \dfrac{1}{V(t)} \\ 0 \end{bmatrix} dt + \begin{bmatrix} \cos\beta(t) & \sin\beta(t) \\ -\dfrac{\sin\beta(t)}{V(t)} & \dfrac{\cos\beta(t)}{V(t)} \end{bmatrix} \begin{bmatrix} dw_x(t) \\ dw_y(t) \end{bmatrix} \tag{95}$$

Thus, the $d\beta$ equation is the same in both Ito and Stratonovich forms, but the dV equation differs by a term $(1/V)\,dt$.

Let p be the transition density for the (V, β) process. The forward partial differential equation obeyed by this density can be written down from Eq. (95) using the rule given by Doob, or it can be written down from Eq. (94) using the rule given by Stratonovich. Either way, one finds that the equation is

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial V}\left(\frac{1}{V}\,p\right) + \frac{1}{2}\frac{\partial^2 p}{\partial V^2} + \frac{1}{2V^2}\frac{\partial^2 p}{\partial \beta^2} \tag{96}$$

Equation (96) is the equation obeyed by the transition density of a two-dimensional Wiener process expressed in polar coordinates, as can be verified by starting with the diffusion equation in rectangular coordinates and applying the rules for change of variables in probability densities.

Summarizing what we have so far, the stochastic differential equation of motion of a particle of unit mass whose velocity vector is a planar Wiener process is given in Cartesian coordinates by Eq. (91), in Stratonovich form in flight-path coordinates by Eq. (94), and in Ito form in flight-path coordinates by Eq. (95). In Cartesian coordinates, the Ito and Stratonovich forms of the equations coincide; in flight-path coordinates, they do not coincide. The choice of which one to use is entirely a matter of personal preference, because Eqs. (91), (94), and (95) are merely three different, but equivalent, ways of describing exactly the same process.

In Eq. (89), it was implicity assumed that $n_x(t)$ and $n_y(t)$ are independent of $v_x(t)$ and $v_y(t)$, or, stated more rigorously, in Eq. (91), the increments $dv_x(t)$ and $dv_y(t)$ are independent of $v_x(t)$ and $v_y(t)$. Physically, we have a white-noise force field which is fixed in inertial coordinates, through which the particle moves. When the situation is viewed in flight-path coordinates, the force on the particle appears to be correlated with the flight-path angle $\beta(t)$.

Since the Stratonovich equation (94) can be manipulated according to the rules of ordinary calculus, let us reintroduce the white-noise forces n_x and n_y and rewrite Eq. (94) in engineering fashion as

$$\begin{bmatrix} \dot{V}(t) \\ V\dot{\beta}(t) \end{bmatrix} = \begin{bmatrix} \cos\beta(t) & \sin\beta(t) \\ -\sin\beta(t) & \cos\beta(t) \end{bmatrix} \begin{bmatrix} n_x(t) \\ n_y(t) \end{bmatrix} \tag{97}$$

Both components of this vector now have the physical dimensions of force. Let $n_\parallel$ and $n_\perp$, respectively, be the forces parallel and perpendicular to the flight path. Thus,

$$\begin{bmatrix} n_\parallel(t) \\ n_\perp(t) \end{bmatrix} = \begin{bmatrix} \cos\beta(t) & \sin\beta(t) \\ -\sin\beta(t) & \cos\beta(t) \end{bmatrix} \begin{bmatrix} n_x(t) \\ n_y(t) \end{bmatrix} \tag{98}$$

By definition of n_x and n_y,

$$E\left\{\begin{bmatrix} n_x(t) \\ n_y(t) \end{bmatrix}\right\} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{99}$$

$$E\left\{\begin{bmatrix} n_x(t) \\ n_y(t) \end{bmatrix}\begin{bmatrix} n_x(\tau) & n_y(\tau) \end{bmatrix}\right\} = \delta(t-\tau)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{100}$$

Now, consider

$$E\left\{\begin{bmatrix} n_\parallel(t) \\ n_\perp(t) \end{bmatrix}\right\} = E\left\{E\left\{\begin{bmatrix} n_\parallel(t) \\ n_\perp(t) \end{bmatrix}\middle|\, \beta(t)\right\}\right\}$$

$$= E\left\{\begin{bmatrix} \cos\beta(t) & \sin\beta(t) \\ -\sin\beta(t) & \cos\beta(t) \end{bmatrix} E\left\{\begin{bmatrix} n_x(t) \\ n_y(t) \end{bmatrix}\middle|\, \beta(t)\right\}\right\} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{101}$$

$$E\left\{\begin{bmatrix} n_\parallel(t) \\ n_\perp(t) \end{bmatrix}[n_\parallel(\tau) \quad n_\perp(\tau)]\middle|\, \beta(t), \beta(\tau)\right\}$$

$$= \begin{bmatrix} \cos\beta(t) & \sin\beta(t) \\ -\sin\beta(t) & \cos\beta(t) \end{bmatrix} E\left\{\begin{bmatrix} n_x(t) \\ n_y(t) \end{bmatrix}[n_x(\tau) \quad n_y(\tau)]\middle|\, \beta(t)\, \beta(\tau)\right\}\begin{bmatrix} \cos\beta(\tau) & -\sin\beta(\tau) \\ \sin\beta(\tau) & \cos\beta(\tau) \end{bmatrix}$$

$$= \delta(t-\tau)\begin{bmatrix} \cos\beta(t) & \sin\beta(t) \\ -\sin\beta(t) & \cos\beta(t) \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \cos\beta(\tau) & -\sin\beta(\tau) \\ \sin\beta(\tau) & \cos\beta(\tau) \end{bmatrix}$$

$$= \delta(t-\tau)\begin{bmatrix} \cos[\beta(t)-\beta(\tau)] & \sin[\beta(t)-\beta(\tau)] \\ -\sin[\beta(t)-\beta(\tau)] & \cos[\beta(t)-\beta(\tau)] \end{bmatrix} \tag{102}$$

However, the δ-function is zero except when $t = \tau$, and when $t = \tau$, the matrix in the last line in Eq. (102) becomes the identity. Thus, it appears that

$$E\left\{\begin{bmatrix} n_\parallel(t) \\ n_\perp(t) \end{bmatrix}[n_\parallel(\tau) \quad n_\perp(\tau)]\middle|\, \beta(t), \beta(\tau)\right\} = \delta(t-\tau)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{103}$$

and, consequently,

$$E\left\{\begin{bmatrix} n_\parallel(t) \\ n_\perp(t) \end{bmatrix}[n_\parallel(\tau) \quad n_\perp(\tau)]\right\}$$

$$= E\left\{E\left\{\begin{bmatrix} n_\parallel(t) \\ n_\perp(t) \end{bmatrix}[n_\parallel(\tau) \quad n_\perp(\tau)]\middle|\, \beta(t), \beta(\tau)\right\}\right\}$$

$$= \delta(t-\tau)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{104}$$

Thus, the noise force vector $\left[\begin{smallmatrix} n_\parallel(t) \\ n_\perp(t) \end{smallmatrix}\right]$ apparently has the same mean and covariance as white noise. Combining Eqs. (97) and (98), one may write

$$\begin{bmatrix} \dot{V}(t) \\ V\dot{\beta}(t) \end{bmatrix} = \begin{bmatrix} n_\parallel(t) \\ n_\perp(t) \end{bmatrix} \tag{105}$$

At first glance, Eq. (105) appears to be equivalent to what one would obtain by making Eq. (88) stochastic directly, by putting a white-noise force vector on the right-hand side of Eq. (88). Let us explore this further. Let $n_1(t)$ and $n_2(t)$ be two independent Gaussian white noises, each of unity power density. Then

$$E\left\{\begin{bmatrix} n_1(t) \\ n_2(t) \end{bmatrix}\right\} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{106}$$

$$E\left\{\begin{bmatrix} n_1(t) \\ n_2(t) \end{bmatrix}[n_1(\tau) \quad n_2(\tau)]\right\} = \delta(t-\tau)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{107}$$

Now consider the stochastic differential equation

$$\begin{bmatrix} \dot{V}(t) \\ V\dot{\beta}(t) \end{bmatrix} = \begin{bmatrix} n_1(t) \\ n_2(t) \end{bmatrix} \tag{108}$$

The question is, is the process generated by Eq. (108) different from the process generated by Eq. (105)? At first glance, comparing Eq. (101) to Eq. (106) and Eq. (104) to Eq. (107), one is tempted to conclude that Eqs. (105) and (108) generate the same process. In fact, the two processes are quite different.

Introduce the two Wiener processes

$$w_1(t) = \int_0^t n_1(\tau)\, d\tau \tag{109}$$

$$w_2(t) = \int_0^t n_2(\tau)\, d\tau \tag{110}$$

The Ito interpretation of Eq. (108) is

$$\begin{bmatrix} dV(t) \\ d\beta(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1/V(t) \end{bmatrix} \begin{bmatrix} dw_1(t) \\ dw_2(t) \end{bmatrix} \tag{111}$$

The forward Kolmogorov equation corresponding to Eq. (111) is

$$\frac{\partial p}{\partial t} = \frac{1}{2} \frac{\partial^2 p}{\partial V^2} + \frac{1}{2V^2} \frac{\partial^2 p}{\partial \beta^2} \tag{112}$$

The two Ito equations (95) and (111) are clearly different. Further, the Kolmogorov equations (96) and (112) do not have the same solution, i.e., the transition density for the process described by Eq. (95) is different from the transition density for the process described by Eq. (111). If two processes have different transition densities, they are different processes.

Why, then, do Eqs. (105) and (108) appear to be so similar? The safest answer is that the manipulations in Eqs. (101)–(104) are not only nonrigorous, but they are probably meaningless. Another answer is contained in the following plausibility argument based on the $\dot{V}$ equation alone.

For the solution $V(t)$ to Eq. (105), we have in mind exactly the same random process as the $V(t)$ component of the Ito solution of Eq. (95). Since this is the magnitude of the velocity along the flight path, it can never be negative. In fact, one can view the $[1/V(t)]\, dt$ Ito correction term in Eq. (95) as being the force which keeps $V(t)$ always nonnegative, since the expected value of the second term in Eq. (95) is zero. Thus, $n_{\|}$ in Eq. (105) must somehow be correlated with V.

On the other hand, for the solution $V(t)$ of Eq. (108), we have in mind exactly the same random process as the $V(t)$ component of the Ito solution of Eq. (111). But this can be written explicitly as

$$V(t) = V(0) + w_1(t) \tag{113}$$

Since $w_1(t)$ has a Gaussian distribution, there is nothing to prevent $V(t)$ here from being negative at certain times. In fact, as soon as one realizes this, one realizes that for this reason, both Eq. (108) and Eq. (11) are physically meaningless.

The main purpose of this example was to illustrate the kind of paradox one can create for oneself by trying to make direct calculations involving white noise. In any case of doubt in a modeling problem, the safe thing to do is to look at both the Ito and the Stratonovich forms of the equations, and make sure they both have a meaningful interpretation.

The ultimate objective of setting up a mathematical model is to get the predicted output of the model to be an acceptable approximation to the actually observed output of the physical system one is trying to model. This is really the only criterion by which one can judge the correctness of a model.

8. CONCLUSION

In this report, the problem of modeling stochastic nonlinear dynamic systems has been discussed. The various mathematical pitfalls and paradoxes that exist were illustrated by examples. It was asserted that, once the engineer understands the mathematics, he should adopt a phenomenological approach for applying them to real-world problems.

REFERENCES

1. M. C. Wang and G. E. Uhlenbeck, "On the theory of the Brownian motion II," *Rev. Mod. Phys.* **17**:323–342 (1945); reproduced in the book, *Selected Papers on Noise and Stochastic Processes*, N. Wax, ed. (Dover, New York, 1954).
2. J. F. Barrett, "Application of Kolmogorov's equations to randomly disturbed automatic control systems," in: *Automatic and Remote Control* (Proceedings of the First International Congress of IFAC) (Butterworths, London, 1961), pp. 724–733.
3. J. L. Doob, *Stochastic Processes* (John Wiley and Sons, New York, 1953).
4. A. V. Skorokhod, *Studies in the Theory of Random Processes* (Addison-Wesley, Reading, Mass., 1965).
5. R. L. Stratonovich, "A new representation for stochastic integrals and equations," *SIAM J. Control* **4**:362–371 (1966).
6. A. H. Gray and T. K. Caughey, "A controversy in problems involving random parametric excitation," *J. Math. Phys.* **44**(3):288–296 (1965).
7. S. R. McReynolds, "A new approach to stochastic calculus," paper presented at the Seminar on Guidance Theory and Trajectory Analysis, NASA Electronics Research Center, May 1967.
8. E. Wong and M. Zakai, "On the relation between ordinary and stochastic differential equations," *Intern. J. Eng. Sci.* **3**:213–229 (1965).
9. T. Kailath, "The Ito stochastic integral," talk presented to the Los Angeles chapter of the IEEE Professional Group on Information Theory, University of Southern California, February 1968.