BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 81, Number 5, September 1975

BOOK REVIEWS

Stochastic differential equations: theory and applications, by L. Arnold, Wiley-Interscience, New York, 1974, xvi+228 pp., $17.95

Stochastic differential systems. I: *Filtering and control, a function space approach*, by A. V. Balakrishnan, Lecture Notes in Economics and Mathematical Systems, vol. 84, Springer-Verlag, Berlin, 1973, v+252 pp., $8.20

The description of Brownian motion given by Einstein is based on an idealization—that increments of a particle's position over disjoint time intervals are independent random quantities. This leads to a probabilistic model, the Wiener process, in which a random trajectory, although continuous, is nowhere differentiable. In spite of this physically rather unrealistic feature of the theory, certain formal properties of the nonexistent derivative suggest that the derivative be included in some fashion in mathematical models of physical processes affected by noise. The relevance of Brownian motion is best described heuristically: since the formal derivative of a trajectory would be a limit of Brownian increments, the random values at different times would be stochastically independent. Such a trajectory would represent fluctuations uncorrelated in time, and hence the trajectory would serve as a graph of noise, for example, in electromagnetic transmission problems or electrical system problems. Moreover, the Fourier transform of such a trajectory would be a random function of the frequency with constant variance for all frequencies. That is, the derivative would be a uniform superposition of frequencies and thus would represent "white" noise.

The work of Wiener, Langevin, K. Ito, and others has shown that certain integrals of the derivative may be defined rigorously. Wiener found that definite integrals of white noise weighted by a fixed square-integrable function of the time parameter exist as random functions. Ito extended the definition and developed an elegant theory for integrals where the weight function varies with the trajectory under the restriction that the weight for a fixed time is a function only of the trajectory history up to that time.

The first half of each book under review presents a portion of the Ito theory and the second half includes treatments of two problems associated with mathematical models given by the system:

$$X(t) = \int_0^t A(s)X(s)\,ds + \int_0^t B(s)\,dW(s), \tag{1}$$

$$Y(t) = \int_0^t C(s)X(s)\,ds + \int_0^t D(s)\,dW(s). \tag{2}$$

Here, $W(s)$ denotes a vector whose components are scalar-valued independent Wiener processes and A, B, C, D, are given matrix-valued functions of the time parameter. Equations such as (1) are referred to as stochastic differential equations (of Ito type), the term "differential" being used precisely, that is, referring to entities which are defined by their integrals. The random solution of equation (1) then determines the random process $Y(t)$ in equation (2).

Of course, the above system is rather trivially coupled. However, an important mathematical problem is suggested by the modeling situation. The random function $X(t)$ corresponds to the actual state of some physical process; $Y(t)$ corresponds to the observation of the state which, according to (2), includes additional noise which may even be independent of the noise in (1) if one has $D(s)B^*(s) \equiv 0$. The problem of finding the best estimate for $X(t)$ using knowledge of $Y(s)$, $0 \leqq s \leqq t$, is the optimal filtering problem. A solution of this problem which is satisfactory for applications was given for a more general situation in 1960 by R. S. Bucy and R. E. Kalman.

The authors' different motivations for producing these books are revealed by the different manner in which they treat the filtering problem. L. Arnold, in Chapter 12, gives a brief survey of recent work on filtering for a model in which (1) is replaced by a nonlinear Ito type equation in $X(t)$. Then, as an example, he discusses the special features of the solution for the system given by equations (1) and (2): the optimal estimate for $X(t)$ is a linear integral operator evaluated on $Y(s)$ and the estimate may be alternately expressed as an integral of previous estimates and past observations. The reader is referred to various references for proofs of the theorems.

Balakrishnan, in Chapter VI, gives a detailed, self-contained derivation of the Bucy-Kalman filter, and the derivation is based on the following nice idea. If the process $X(t)$ in equation (2) is replaced by the difference between itself and the optimal estimate, the resulting process, $Y(t)$, is a martingale. In addition, a second martingale arises if an appropriate integral of past optimal estimates is subtracted from the estimate. Then this latter martingale can be optimally estimated in an explicit manner from the first martingale. This new estimate is a linear stochastic integral of the first martingale over its past and the nonrandom weight function is explicit. Since the desired estimate is characterized as being the conditional expectation of $X(t)$ relative to all past events determined by the observation process, it suffices to check that the past events determined by the first martingale agree with those of the observation process. This is carried out and one recovers the optimal estimate for $X(t)$. This derivation illustrates why the theory of Ito integrals is so useful for the investigation of such systems. The random processes which arise are Markovian and are amenable to the techniques of analysis developed for such processes. In addition, since equations (1) and (2) are linear in $X(t)$, the processes are jointly Gaussian and one may say quite a bit about each path space measure. Balakrishnan confines his attention to such systems and obtains (known) results such as the absolute continuity of the processes, after a linear normalization, relative to the path space measure for the Wiener process; also, in Chapters IV and V, expressions for the Radon-Nikodym derivative are obtained. In Chapter VIII these results are applied to the problem of identifying the matrices $A(s)$, $B(s)$ in equation (1) if only $C(s)$, $D(s)$ are known. Under some minor restrictions on the relation of the matrices, a procedure for the estimation based on maximizing the expression for the Radon-Nikodym derivative is established. Throughout these notes only the minimal amount of background material is introduced which is necessary for the formulation and solution of the particular filtering and control problems considered. The notes are mathematically self-contained and are tightly organized so that the material in the first six chapters is applied in the remaining two chapters.

Questions involving the appropriateness of the white noise process for modeling physical noise or the existence of alternate models of noise are not discussed in the notes by Balakrishnan but are considered in the book by Arnold. In Chapter 3, Arnold discusses the interpretation of white noise as a generalized process; that is, a random distribution determined by a certain probability measure on the Schwartz space of tempered distributions in the time parameter. He points out that there is in fact a fairly large class of generalized processes which may serve as models of noise. In a discussion of approximation questions in Chapter 10, it is shown that a limit of solutions for equations where white noise is approximated by slightly correlated noise is not necessarily the solution of the corresponding Ito differential equation. Such a limit, for certain systems, is shown to satisfy another type of stochastic equation, studied by Stratonovich. This disparity is studied in detail. The book presents more of the Ito theory than the notes by Balakrishnan. The Markov nature of solutions for nonlinear versions of equation (1) is discussed in Chapter 9 and stability of such systems is studied in Chapter 11. Typically, a discussion of a topic begins with a quite general formulation, and the known major results are reported with appropriate references. Then the theory is illustrated with examples. This provides the reader with an overview of the subject which is missing in Balakrishnan's notes. As a consequence, much of the latter half of the book is not mathematically self-contained.

A second problem treated by both books arises when equation (1) is modified by adding the expression $\int_0^t F(s)u(s)\,ds$ to the right-hand side of equation (1). Here, $F(s)$ is a matrix-valued function and $u(t)$ is a function of the random trajectory $Y(s)$, $s \leqq t$. This produces a coupled system of stochastic equations; $u(t)$ corresponds to a control which modifies the evolution of the physical process and which, of course, should depend only on the previous observations. The first control problem considered is the determination of a function $u(t)$ which minimizes a given quadratic form involving the trajectory of the physical process and the control function over a fixed time interval. As one might expect, a candidate for an optimal control is constructed using the solution of the associated filtering problem. However, as Balakrishnan mentions in Chapter VII, the optimality of the control over the entire class of admissible control function has not been proven. The notes by Balakrishnan also discuss a steady state control problem where the quadratic form is a time average which is an invariant function of the process and its control over each finite time interval. A third control problem is studied where the quadratic form involves only the final value of the physical process over a fixed bounded time interval. The solution of this problem is applied to a control problem where the physical system $X(t)$ is affected by the sum of two controls. The control functions are regarded as the player strategies in a two person zero sum game and the stochastic system is described as a differential game with imperfect information.

The book by Arnold may serve as a textbook or reference work. Its substantial bibliography contains reference lists for topics such as Markov and diffusion processes, stochastic differential equations, stability, filtering, control, and probability theory. There is a good index and each section of the book is well organized. The book is especially valuable for nonexperts on stochastic differential equations who wish to deal with models for processes affected by noise. One can learn the limitations of the theory as well as recent results on a variety of problems. The notes of Balakrishnan are valuable to anyone who desires to master the mathematical techniques involved in modern stochastic control theory.

VICTOR GOODMAN

Optimal control theory, by L. D. Berkovitz, Applied Mathematical Sciences, 12, Springer-Verlag, New York, 1974, ix+304 pp., $9.50

The term "mathematical theory of optimal control" has come to refer to the optimization of a certain class of functionals of state and control variables for dynamical systems whose evolution with time is described by ordinary differential equations. Such problems are similar to the Bolza problem in classical calculus of variations, with the important difference that inequality constraints may be imposed. A large literature on optimal control theory developed during the 1960's, stimulated by the slightly earlier work of Bellman, Pontryagin, and their associates. Most of the questions with which that literature was concerned have by now been resolved. It is the task of authors of books on control theory to preserve the essential aspects, for those interested in the applications and as a foundation for students entering an area of active current research (e.g. control of systems governed by partial differential equations, control systems with time delays, and stochastic control).

In this book Berkovitz gives a readable account not only of the standard Pontryagin necessary conditions for a minimum but also of the problem of existence. The proof given for the Pontryagin necessary conditions follows Gamkrelidze, SIAM J. Control (1965). Like other proofs, it depends on the idea of convex set of variations (due to McShane in 1939) and the Brouwer fixed point theorem.

The traditional method in calculus of variations for proving existence of a minimum is to show precompactness of minimizing sequences and lower semicontinuity. A nicer technique was found in 1959 by Filippov; it avoids lower semicontinuity but uses a theorem about measurable selections. This