THE REPRESENTATION OF NON-LINEAR STOCHASTIC SYSTEMS WITH APPLICATIONS TO FILTERING

by

J. M. C. CLARK

A thesis submitted for the degree of Ph.D. in Engineering

Electrical Engineering Department,
Imperial College,
LONDON, S.W.7.

APRIL, 1966.

ABSTRACT

Physical processes can often be described by a system of ordinary differential equations excited by random disturbances with short correlation times. It is shown in this thesis that such processes can be approximated, in the sense that the second moment of error is small, by Markov diffusion processes of the same dimension. In the approximation the random disturbances are characterised by a matrix which is an integral of their cross-correlation function, but which is not in general the cross-spectral density. If this characteristic matrix is symmetric, the Stratonovich stochastic differential equation of the diffusion approximation is similar in form to the ordinary differential equation of the physical process.

In computer simulations the ordinary differential equation of a physical process can be used as the programming model if the characteristic matrices of the disturbances and the computer noise source are both symmetric and of the same rank.

With the aid of diffusion approximations, much of the theory of the filtering of diffusion processes can be applied to problems in the filtering of physical processes.

ACKNOWLEDGMENTS

I should like to thank Professor Westcott, my supervisor, for his encouragement and understanding, and Dr. Florentin for promoting my interest in the subject of this thesis. I have benefited from many illuminating discussions with Dr. Plimmer, of the Royal Aircraft Establishment, Mr. Mayne and Mr. Cumming.

The work that led to this thesis was supported by the Ministry of Aviation.

CONTENTS

INTRODUCTION

In the analysis and design of engineering systems with random disturbances, it is often convenient to model the processes involved by diffusion processes; that is, Markov processes that are continuous functions of time. The generality of diffusion processes, combined with their simplicity, has led to their use in many different fields of investigation, and in particular in those concerned with stochastic control theory and filtering theory. A diffusion process, however, if chosen to model the behaviour of a process over long periods of time, will not necessarily reproduce the behaviour of the process over very short periods of time; for instance, the realisations of many physical processes have an inherent smoothness, whereas those of a diffusion process are not even differentiable. For systems in which the disturbances are additive this difference between the detailed behaviours of an actual process and its diffusion model can often be ignored, but it becomes important if the disturbances are non-additive. The purpose of this thesis is to consider how, and with what justification, the processes that actually occur in engineering systems can be modelled by diffusion processes.

We shall restrict our attention to the type of engineering process that has the structure of a system of differential equations

$$\dot{X} = f(X,t) + F(X,t)y, \qquad (1)$$

where y is a vector of random disturbances, the periods over which each disturbance is correlated being short compared with the time-constants of the system. A more precise description is given in Section 1.5. The functions f and F will be assumed to be known. Processes that can be modelled by X(t) we shall call "physical" processes. Just how the description (1) is obtained for any particular process we shall not go into, but often the difficulty is not so much how to obtain such a system of equations, but how to obtain a system of reasonably low dimension. The important thing to note is that physical processes as defined by (1) are sufficiently smooth to be differentiable at least once. In Chapter 2 we shall show that a physical process can be approximated, in the sense that the second moment of the error is small, by a diffusion process of the same dimension, and we shall point out which of the statistical parameters of the random disturbances it is necessary to know in order for the diffusion approximation to be determined. Approximations of this sort have also been the concern of Stratonovich [20], Wong and Zakai [21,22], Astrom [23], and Ariaratnam and Graefe [27]. Their work will be discussed in Chapter 2.

Since the development of the theory of Brownian motion, it has been recognised that for many purposes the random disturbances of a physical process can be regarded as "white noise" if they are additive; that is, if in (1) F is independent of X. Then (1) becomes

$$\dot{x} = f(x,t) + F(t)n \qquad (2)$$

where n is a "white noise" process with a correlation function that is a delta function. The process x is, in fact, a diffusion process. This close relation between the description (2) of the diffusion process and the description (1) of the process it approximates is part of the attractiveness of diffusion processes. However, as the rate of change of a diffusion process is infinite with probability 1, equation (2) is not a differential equation in the normal sense, and care has to be taken in its interpretation. This is particularly so if in (2) F depends on x, for then the interpretation is not unique. One precise definition of a stochastic differential equation for diffusion processes has been given by Ito [16]. This is written symbolically in the form

$$dx = g(x,t)dt + G(x,t)dw \qquad (3)$$

where $w(t)$ is a vector Wiener process. Another definition has been given by Stratonovich [7]. The relations between these two forms of stochastic differential equation and also the Fokker-Planck equation of the diffusion process are given in Section 1.4. If G is independent of x, the two forms coincide and the corresponding equation (2) can be formally obtained by dividing (3) by dt and by taking $\frac{dw}{dt}$ to be white noise. In Chapter 2 the results on the diffusion approximation of physical processes will be derived and presented in terms of stochastic differential equations. A further point considered in Chapter 2 is which of the two forms of stochastic differential equation is the most suitable for engineering purposes. The properties of these two forms, considered as equations for diffusion processes, are well-known and are summarised in Section 1.4: briefly, the Ito stochastic differential equation is closely related to the corresponding equations for the probability densities and moments of the diffusion process it describes; the calculus of Stratonovich stochastic differential equations is like the ordinary differential calculus, whereas that of the Ito stochastic differential equations is not. These properties have to be taken into account, but in the context of the diffusion approximation of physical processes an important consideration in selecting the best form is whether or not the stochastic differential equation of the diffusion approximation can be obtained from the ordinary differential equation of the physical process by simply replacing the random disturbances by white noise, as we have seen can be done if the disturbances are additive. It will be shown that, although neither the Ito form nor the Stratonovich form satisfies this requirement in general, the Stratonovich form does so if a particular integral of the matrix correlation function of the random disturbances is symmetric. This matrix integral will be called the "characteristic matrix" of the random disturbances. Its symmetry for some common forms of noise process is studied in Chapter 4.

Finally two general problems are considered to which the approximation results of Chapter 2 can be usefully applied. The first problem is the simulation of one physical process by another, where one is concerned with designing the simulating process so that its statistical characteristics are similar to those of the original process. The following questions arise: Can the ordinary differential equations describing the two processes be taken to be of the same form? What are the statistical parameters of the random disturbances and the generated noise that have to be matched? As we shall see in Chapter 3, the answers depend on the characteristic matrices of the random disturbances and noise processes involved.

The second problem we consider is the filtering of a physical process from continuous observations of a related process. We shall suppose that the filter is to provide an estimate of the present state of a message process and that this estimate is to be the expectation of the message process conditional on the observation process. This formulation is the natural one to consider in problems concerning the control of a physical process: very often the state of the process cannot be measured exactly and only some of the statistical parameters of the noise in the process are known, and an estimate of the present state of the process is required on which to base the control. The closely related problem of the filtering of diffusion processes has been studied by Stratonovich, Kushner and others, and we shall summarise the fundamental work of these authors in Chapter 5. As we shall see, it is relatively straightforward to derive a diffusion filtering algorithm for a diffusion process; that is, a set of differential equations generating an estimate, which is itself part of a diffusion process. However, the input and output of an actual filter - the observation process and the estimate - have finite rates of change and so cannot be diffusion processes. It seems more realistic to take them to be physical processes and to base the design of a filter on a physical filtering algorithm for a physical process. In Chapter 6 we shall be concerned with the problem of deriving such an algorithm from the known results of the filtering of diffusion processes.

CHAPTER 1

BASIC CONCEPTS AND PRELIMINARY DEFINITIONS

1.1 Introduction

In order to make the later chapters more concise, it is convenient to begin by defining the more frequently used terms and concepts. This is the purpose of this chapter. In sections 1.3 and 1.4 we give the definitions of the stochastic integral and the stochastic differential equation and summarize their properties and their relation with the diffusion process. In Section 1.5 we introduce a mathematical description of a random disturbance which includes several commonly used models of noise.

1.2 Notation

Throughout this work equations are written in vector form. There is no distinction in notation between scalar, vector and matrix quantities. The transpose of a matrix is denoted by the superfix T. The Euclidean norm is denoted by $|.|$; that is, if A is scalar, $|A|$ is the modulus of A; if B is a vector, $|B|^2 = \sum_i B_i^2$, B_i being the i:th component of B, and if C is a matrix, $|C|^2 = \sum_{ij} C_{ij}^2 = \text{trace}\,(CC^T)$, C_{ij} being the ij:th element of C.

Partial derivatives are sometimes denoted by subscripts: suppose x is a vector, f is a vector function of x and A is a matrix, then f_x denotes the matrix with ij:th element $\frac{\partial}{\partial x_j} f_i$, and $\langle f_{xx}, A \rangle$ denotes the vector with i:th component $\sum_{jk} A_{jk} \frac{\partial^2}{\partial x_j \partial x_k} f_i$. The rate of change of x(t) with respect to time is denoted by $\dot{x}(t)$.

The expectation, or mean, of a random variable X is denoted by EX or E[X]. The conditional expectation of X given another random variable Y is E[X|Y].

1.3 Basic concepts

All random variables will be assumed to be functions of a basic random parameter γ taking values in a sample space, though we shall not usually indicate this dependence on γ. The phrase 'with probability 1' and the abbreviation 'a.c' (almost certainly) are to be interpreted as: 'for all values of γ, except for a set of values that are taken with zero probability'.

Convergence in the mean. Of the several ways of describing the convergence of a sequence of random variables we shall only need the following: if $X_n, X_{n+1}, \dots$ is a sequence of random variables of finite second moment, then X_n is said to converge to X in the mean as $n \to \infty$ if

$$E\,|X_n - X|^2 \longrightarrow 0\,. \qquad (1.3.1)$$

Following on from this, if we are considering the approximation of one random variable by another or, more generally, one stochastic process $X(t,\gamma)$ by another process $Y(t,\gamma)$ over a period of time $0 \leq t \leq T$ we shall say that $Y(t)$ is a <u>δ-approximation</u> of $X(t)$ if

$$\sup_{0\leq t\leq T} E|Y(t) - X(t)|^2 \leq \delta^2. \qquad (1.3.2)$$

<u>The Wiener process</u> (or Brownian motion process). We shall denote by $w(t,\gamma)$ a vector process that is Gaussianly distributed for each value of t, and has independent increments. Also $w(o) = 0$, and for $s,\ t \geq 0$

$$E[w(t)-w(s)] = 0, \quad E[(w(t)-w(s))(w(t)-w(s))^T] = I\,|t-s|$$

where I is a unit matrix. This is one of the definitions of the Wiener process. Such a process is a.c. continuous (see [12] p. 97).

<u>Stochastic integrals</u>. Suppose $f(t,\gamma)$ is a stochastic process such that it is measurable with respect to t and γ and for fixed t it is a B-measurable function of $[w(s\,,\gamma),\ 0 \leq s \leq t]$ (see Ito [16] Section 1); then we shall say that $f(t,\gamma)$ is <u>admissible</u>. Roughly this means that $f(t,\gamma)$ is a non-pathological function of t and γ

and is independent of future increments of the Wiener process. If $a(t)$ and $F(t)$ are admissible processes (we omit the argument γ) then Ito [15] shows that integrals of the form

$$\int_0^t a(s)ds \ , \qquad \int_0^t F(s)dw(s)$$

can be defined in such a way that these are also admissible processes and are a.c. continuous. The value of the first integral for fixed γ can be taken to be the time integral of $a(s)$, but the corresponding value of the second integral cannot be taken to be the Stieltjes integral of $F(s)$ with respect to $w(s)$, as $w(s)$ is (with probability 1) of unbounded variation. Instead it is defined as the limit (for almost all t, γ) of a sequence of finite sums

$$\sum_i F(t_i)[w(t_{i+1})-w(t_i)] \qquad (1.3.3)$$

where $0 = t_o \leq t_1 \leq \ldots \leq t_n = t$. It is important to note that, in these sums, $F(t_i)$ is evaluated at the beginning of the corresponding partition interval $[t_i, t_{i+1}]$. If F is evaluated at $(1-\alpha)t_i + \alpha t_{i+1}$ for some fixed α $(\neq 0)$ the resulting stochastic integral is not the same as the Ito integral. However, the integral is well defined and there is a simple formula relating it to the Ito integral

[19]. The integral corresponding to $\alpha = \frac{1}{2}$, that is, where F is evaluated at the centre of the partition interval, has been studied by Stratonovich [7] who showed that it can be transformed or differentiated as though it were an ordinary integral; this is not so for the Ito integral, which transforms according to Ito's formula (see (1.3.9)). Ito's stochastic integrals have the following properties: if a(t), b(t), F(t) and G(t) are admissible processes with integrable second moments then

$$\left.\begin{aligned} &E\left[\int_0^t a(s)ds\right] = \int_0^t Ea(s)ds\,, \\ &E\left[\int_0^t a(s)ds\left(\int_0^t b(r)dr\right)^T\right] = \int_0^t\int_0^t E[a(s)b(r)^T]dsdr, \\ &E\left[\int_0^t F(s)dw(s)\right] = 0\,, \\ &E\left[\int_0^t F(s)dw(s)\left(\int_0^t G(r)dw(r)\right)^T\right] = \int_0^t E[F(s)G(s)^T]ds. \end{aligned}\right\} \quad (1.3.4)$$

<u>Integral processes</u>. Suppose a process x(t) can be expressed by the integral equation, for $0 \leq t \leq T$,

$$x(t) = \int_0^t a(s)ds + \int_0^t F(s)dw(s) \qquad (1.3.5)$$

where a(s) and F(s) are admissible processes and

$$\int_0^T |a(s)|\,ds < \infty \text{ a.c.}, \quad \int_0^T |F(s)|^2 ds < \infty \text{ a.c.}. \qquad (1.3.6)^*$$

* In particular this is satisfied if a(s) and F(s) are a.c. continuous.

then we shall call $x(t)$ an <u>integral</u> process. We can write (1.3.5) symbolically in stochastic differential form:

$$dx(s) = a(s)ds + F(s)dw(s). \tag{1.3.7}$$

We define the 'mixed' stochastic integral $\int_0^t G(s)dx(s)$ to be the sum of integrals obtained by replacing dx in accordance with (1.3.7).

For each integral process the functions $a(t)$ and $F(t)$ are unique (for almost all t and γ). We need to refer to them separately. It is convenient to introduce the following notation: we call $a(t)$ the <u>drift</u> of $x(t)$ and denote it by $\underline{d}[x(t)]$; $F(t)$ is called the <u>dispersion</u> of $x(t)$ and $w(t)^T$, and is denoted by $D[x(t):w(t)^T]$. This operator is defined for general arguments as follows: if $x(t)$ and $y(t)$ are scalar integral processes and $F_i(t)$ is the coefficient of $dw_i(t)$ in the expression (1.3.7) for $dx(t)$ and $G_i(t)$ is the corresponding coefficient in $dy(t)$, then the dispersion of $x(t)$ and $y(t)$ is

$$D[x(t):y(t)] = \sum_i F_i(t)G_i(t). \tag{1.3.8}$$

The two arguments of the dispersion operator can be taken to be matrix functions, provided that their matrix product is meaningful. The special case $D[x(t):x(t)^T]$ ($= F(t)F(t)^T$) we shall simply call the dispersion of $x(t)$.

In the literature the terms 'drift' and 'dispersion', which we shall only use in the sense we have given, are sometimes used to refer to determinate functions that are more often called 'drift coefficient' and 'diffusion coefficient' (see the next section).

Ito's formula [16]. Suppose $q(x,t)$ is a scalar function with continuous second order partial derivatives in x and t. Ito's formula states that if $x(t)$ is a vector integral process, then $q(x(t),t)$ is also an integral process and its stochastic differential (in the notation of section 1.2)

$$dq(x(t),t) = (q_t + q_x a + \tfrac{1}{2}\langle q_{xx}, FF^T\rangle)dt + q_x F dw(t), \quad (1.3.9)$$

where $a(t)$ and $F(t)$ are as in (1.3.7) and are evaluated at t, and where the partial derivatives of q are evaluated at $(x(t),t)$. Note that (1.3.9) is different from the ordinary formula for the differential of a determinate function in that it contains the term $\frac{1}{2}\langle q_{xx}, FF^T\rangle dt$. If $x(t)$ has zero dispersion ($F = 0$) then this term is zero (in which case q need only have continuous first order partial derivatives) and (1.3.9) is the same as the ordinary formula.

1.4 Diffusion processes and stochastic differential equations.

Suppose now we have an integral equation

$$x(t) = \int_0^t a(x(s),s)ds + \int_0^t F(x(s),s)dw(s), \quad (1.4.1)$$

which, written as a stochastic differential equation, is

$$dx(t) = a(x(t),t)dt + F(x(t),t)dw(t), \ x(o) = 0. \quad (1.4.2)$$

It can be shown that (Doob [12] p.273 or Ito [16]) if $a(x,t)$ and $F(x,t)$ are measurable, are bounded functions of t and satisfy

$$|a(x,t) - a(y,t)| \leq K|x-y|, \quad |F(x,t)-F(y,t)| \leq K|x-y|$$

for all x and y and for $0 \leq t \leq T$, K being constant, then there is a unique solution to (1.4.1) and it is a Markov process of diffusion type with drift coefficient $a(x,t)$ and diffusion coefficient $F(x,t)F(x,t)^T$; that is, as $h \longrightarrow 0+$,

$$\left.\begin{aligned} &E[x(t+h) - x(t)|x(t) = x] = a(x,t)h + o(h), \\ &E[(x(t+h)-x(t))(x(t+h)-x(t))^T|x(t)=x] = F(x,t)F(x,t)^T h + o(h) \end{aligned}\right\} \quad (1.4.3)$$

and, with suitable regularity conditions, the probability density $p(y,s;\ x,t)$ of $x(t)$ being x, given that $x(s) = y$, satisfies the Fokker-Planck equation

$$\frac{\partial p}{\partial t} = - \sum_i \frac{\partial}{\partial x_i}[a_i(x,t)p] + \tfrac{1}{2} \sum_{ijk} \frac{\partial^2}{\partial x_i \partial x_j}[F_{ik}(x,t)F_{jk}(x,t)p] \, . \quad (1.4.4)$$

The diffusion process $x(t)$ can also be described by the Stratonovich stochastic differential equation

$$\bar{d}x(t) = f(x(t),t)dt + F(x(t),t)\bar{d}w(t) \qquad (1.4.5)$$

where the bar over the differentials denotes that the integrals in the underlying integral equation are to be interpreted as Stratonovich integrals, which we described in the previous section. The relation between the Ito form (1.4.2) and the Stratonovich form (1.4.5) is that [7]

$$f_i(x,t) = a_i(x,t) - \tfrac{1}{2} \sum_{jk} F_{kj}(x,t)\frac{\partial}{\partial x_k}F_{ij}(x,t) \; . \qquad (1.4.6)$$

Another form of this relation, which we shall use, is

$$f(x(t),t) = a(x(t),t) - \tfrac{1}{2}D[F(x(t),t):w(t)] \qquad (1.4.7)$$

where $D[\,.\,:\,.\,]$ is the dispersion operator defined in the previous section.

Stratonovich's representation of a diffusion process has the advantage that the stochastic differential of a function of the process can be obtained by the normal formula for ordinary differentials; Ito's formula involves second-order derivatives of the function. However, Ito's representation has the advantage that the equations for the moments of the process can be expressed more simply in terms of it. For this reason and because we want to make use of the well developed theory of the Ito stochastic integral we shall use the Ito representation in the theoretical part of this work. We shall present some of the results in Stratonovich form for purposes of comparison.

Stochastic differential equations can also be interpreted as the limits of difference equations; we demonstrate this with a simple example in Appendix A.

1.5 Physical processes and random disturbances

Consider the stochastic process $X(t)$ described by the vector differential equation

$$\dot{X}(t) = f(X(t),t) + F(X(t),t)y(t), \quad X(0) = 0 \qquad (1.5.1)$$

where $y(t)$ is a random disturbance. The main purpose of the present work is to show that under certain conditions such a process, even though it is differentiable, can be approximated by a diffusion process. We shall call this process a <u>physical</u> process.

The random disturbance $y(t)$ we shall take to be described by the integral equation

$$y(t) = \int_0^t g(t,u)\,dw(u) \qquad (1.5.2)$$

where $w(t)$ is the vector Wiener process and $g(t,u)$ is a matrix. Such a random disturbance is Gaussian and "depends only on the past". Many of the random phenomena that occur in engineering, such as thermal noise in electrical circuits, some types of vacuum tube noise, and rocket engine noise, have these properties and can be approximated by a description of the form (1.5.2).

We have chosen to describe random disturbances in this way not because it is easy to express any particular process in the integral form of (1.5.2) - in general it would not be - but because several broad classes of interesting processes are included in this description; we now list those.

1) Approximately white stationary noise. Suppose a scalar Gaussian stationary process has a continuous power spectral density $P(\omega)$, which satisfies a "realisability" condition of Paley and Wiener (see [9] p.175)

$$\int_{-\infty}^{\infty} \frac{\log P(\omega)}{1+\omega^2}\, d\omega < \infty \ ;$$

then the process can be represented as

$$\int_{-\infty}^{t} g(t-u)\,dw(u) \ . \qquad (1.5.3)$$

If t is positive and g(t-u) decays away sufficiently rapidly for increasing (t-u), that is, the correlation time of the process is sufficiently small, then we can ignore that part of the process produced by integration over negative values of time and the representation (1.5.3) reduces to a special case of (1.5.2).

2) Linear diffusion processes. The model used by Uhlenbeck and Ornstein to describe the rate of change of Brownian motion (see [26] p.93) is a first-order linear diffusion process. Suppose

$$dy = Cydt + Hdw \ , \quad y(o) = 0$$

where y and w are vectors and C and H constant matrices. We can verify that

$$y(t) = \int_0^t e^{C(t-u)}Hdw(u)$$

where the exponential is a matrix exponential. This is of the form (1.5.2).

3) Piecewise constant processes. Consider the process that is constant over intervals of length $\frac{1}{\alpha}$, the values it takes in the intervals being independent and Gaussian. Such a process might be used in a computer to simulate white noise. If in (1.5.2),

$$g(t,u) = \alpha I, \ \frac{n-1}{\alpha} < u \leq \frac{n}{\alpha} \leq t < \frac{n+1}{\alpha}, \ n=1,2,\ldots$$
$$= 0 \quad \text{otherwise},$$

then $y(t)$ is a piecewise constant process of this sort. More general piecewise constant processes, which are correlated over more than one interval, can also be represented by (1.5.2), as we shall see in Chapter 4.

In the next chapter, we shall suppose $y(t)$ to be a.c. piecewise continuous; in Appendix B we give a condition on g that implies this. Otherwise we do not impose any restriction on the analytical properties of the sample paths of $y(t)$; for instance, the diffusion process 2) is not differentiable, whereas the piecewise constant process 3) is, trivially, infinitely differentiable in each interval.

CHAPTER 2

THE APPROXIMATION OF PHYSICAL PROCESSES BY DIFFUSION PROCESSES

2.1 Introduction

In this chapter we consider the approximation of a physical process by a diffusion process over a finite interval of time. We formulate this as a convergence problem by taking the random disturbance of the physical process to be a member of a family of random disturbances described parametrically by a variable α, which can be regarded as an "upper frequency". We can then study the asymptotic behaviour of the physical process as $\alpha \longrightarrow \infty$. The main result, which is given in the next section, can be summarised as follows. The physical process is described by

$$\dot{X}(t) = f(X(t),t) + F(X(t),t)y(t), \quad X(0) = 0 \qquad (2.1.1)$$

where $y(t)$ is a random disturbance of the sort considered in Section 1.5. Considering $y(t)$ as a function of its "upper frequency" α, we show that $X(t)$ converges in the mean to a particular diffusion process $x(t)$ as $\alpha \longrightarrow \infty$, if, among other conditions,

a) for positive t,

$$\lim_{\alpha\to\infty} \frac{1}{t}\int_0^t \int_0^s E[y(s)y(r)^T]\,dr\,ds = A \qquad (2.1.2)$$

where A is a constant matrix, and

b) the integral of $y(t)$ converges in the mean to $Bw(t)$ (these two conditions imply $BB^T = A + A^T$). The diffusion process $x(t)$ is given by the Stratonovich differential equation

$$\bar{d}x_i(t) = f_i dt + \frac{1}{2}\sum_{jmn} F_{jn}\frac{\partial}{\partial x_j} F_{im}(A_{mn}-A_{nm})dt + \sum_j \overset{F_{ij}}{B_{ij}}\,\bar{d}w_j(t),$$

$$x_i(0) = 0\ , \qquad (2.1.3)$$

where f and F are evaluated at $(x(t),t)$. The rate of convergence is such that $x(t)$ is an $O(\alpha^{-\frac{1}{2}})$-approximation to $X(t)$.

Note the following points about the approximating process $x(t)$:

1) $x(t)$ is the same dimension as $X(t)$;

2) the random disturbance $y(t)$ is characterised in the approximation not only by the limit of its integral; that is, $Bw(t)$, but also by the matrix $(A - A^T)$, which is unrelated to $Bw(t)$;

3) if either $F(x,t)$ is independent of x, or the matrix A is symmetric, then the Stratonovich differential equation of the approximation could have been obtained

by replacing the term $y(t)dt$ in (2.1.1) (in differential form) by its formal limit $B\bar{d}w(t)$.

Both the second and third points indicate the importance of knowing the symmetry of the matrix A. In Chapter 4 we consider the symmetry of A for some common forms of random disturbance . As we shall see in the next chapter, if we are only interested in the probability laws of $X(t)$ and not in its functional dependence on the hypothetical process $w(t)$, then the random disturbance is completely characterised in the approximation by A, whether it is symmetric or not. For this reason we shall call A the <u>characteristic matrix</u> of $y(t)$.

If $x(t)$ is a $O(\alpha^{-\frac{1}{2}})$-approximation to two different physical processes $X(t)$ and $X'(t)$ then it is clear that $X(t)$ and $X'(t)$ are $O(\alpha^{-\frac{1}{2}})$-approximations of each other. Relations between their coefficients can readily be obtained by equating corresponding coefficients of dt and $\bar{d}w(t)$ in the two versions of (2.1.3).

In Section 2.4 we compare the results obtained in this chapter with similar results obtained by quite different methods by Stratonovich [20], Wong and Zakai [21,22], Astrom [23] and Ariaratnam and Graefe [27].

2.2 A theorem on the limiting form of a physical process

Suppose $X(t)$ is a vector physical process satisfying for f in $[0,T]$ the equation

$$X(t) = f(X(t),t) + F(X(t),t)g(\alpha,f), \quad X(0) = 0 \tag{2.2.1}$$

where F is a matrix and f and y are vectors, the dimension of y being N. In $[0,T]$ $y(\alpha,t)$ is an a.c. piecewise continuous (that is, with probability 1, $y(\alpha,t)$ is continuous in $[0,T]$ apart from a finite number of finite jumps) random disturbance of the form

$$y(\alpha,t) = \int_0^t g(\alpha,t,s)dw(s) \tag{2.2.2}$$

where $w(t)$ is a vector Wiener process and $g(\alpha,t,s)$ is a measurable matrix function of t and s such that for all positive α

$$\begin{aligned} |g(\alpha,t,s)| &\leq \alpha c e^{-\alpha(t-s)} \quad \text{for } t \geq s \\ &= 0 \text{ otherwise,} \end{aligned} \tag{2.2.3}$$

and

$$\int_s^{\infty} g(\alpha,t,s)dt = B \text{ for all } s \tag{2.2.4}$$

where c is a positive constant and B a constant matrix. Furthermore, suppose that, as α increases,

$$\int_0^t \int_0^s E[y(\alpha,s)y(\alpha,r)^T]drds = At + O(\alpha^{-1}) \tag{2.2.5}$$

uniformly for t in $[0,T]$, where A is a constant matrix and $O(\alpha^{-1})$ denotes a matrix with components of order α^{-1}.

The parameter α can be regarded as a representative upper frequency of $y(\alpha,t)$. We shall call the matrix A the <u>characteristic matrix</u> of $y(\alpha,t)$. For approximately stationary random disturbances, with weighting functions of the form $g(\alpha,t-s)$, the double integral in (2.2.5) can be simplified to a single integral of the correlation function of $y(\alpha,t)$ (see Section 4.1). It will generally be helpful to think of A as the cross-correlation of $y(\alpha,t)$ with the integral of its past. As we shall see in Lemma 2.1, the conditions (2.2.3,4) imply that as α increases the integral $\int_0^t y(\alpha,s)ds$ converges in the mean to $Bw(t)$, so in this sense $y(x,t)$ tends to the "white noise" $Bw(t)$; we shall also see that the matrices A and B are connected by the relation $A + A^T = BB^T$. Let $F_{im}(x,t)$ be the im:th component of the matrix function $F(x,t)$, $F_m(x,t)$ the m:th column of $F(x,t)$, and $Q_{mn}(x,t)$ the column <u>vector</u> with i:th component $\sum_j F_{jn}(x,t)\frac{\partial}{\partial x_j}F_{im}(x,t)$. Suppose the vector functions $f(x,t)$, $F_m(x,t)$ and $Q_{mn}(x,t)$ satisfy the following conditions for all x and all t in $[0,T]$. In these conditions K is a positive constant, the arguments of the functions are x and t, and partial derivatives with respect to x and t are denoted by the suffices x and t

f is continuous with respect to t, continuously differentiable with respect to x, and $|f_x| \leq K$. (2.2.7)

F_m and Q_{mn} are continuously differentiable with respect to both x and t and $|F_{m,x}| \leq K$, $|Q_{mn,x}| \leq K$. (2.2.8)

Either

the functions $|f|$, $|F_m|$, $|Q_{mn}|$, $|F_{mn,t}|$ and $|Q_{mn,t}|$ are all bounded by K, (2.2.9a)

or

$|F_{m,t}| \leq K + K|x|$, $|Q_{mn,t}| \leq K + K|x|$ and for all α, $E|X(t)|^4 \leq K$. (2.2.9b)

We have given the alternative conditions (2.2.9a,b) to broaden the class of physical processes we can consider. Each condition by itself is restrictive in some sense; (2.2.9a) excludes from our consideration linear systems with stochastic coefficients, whereas (2.2.9b) implies that the fourth moment of X(t) is bounded <u>uniformly</u> for increasing α, a condition which, though plausible, is difficult to verify.

<u>Theorem 2.1</u> If conditions (2.2.1-8) hold and either

(2.2.9a) or (2.2.9b) holds, then for increasing α

$$E|X(t) - x(t)|^2 = O(\alpha^{-1}) \tag{2.2.10}$$

uniformly for t in $[0,T]$, where x(t) is the unique diffusion process satisfying the Ito stochastic differential equation

$$\begin{aligned} dx(t) = f(x(t),t)dt + \sum_{mn} Q_{mn}(x(t),t)A_{mn}dt \\ + F(x(t),t)Bdw(t), \quad x(0) = 0 \end{aligned} \tag{2.2.11}$$

and also satisfying the Stratonovich stochastic differential equation

$$\begin{aligned} \bar{d}x(t) = f(x(t),t)dt + \tfrac{1}{2} \sum_{mn} Q_{mn}(x(t),t)[A_{mn}-A_{nm}]dt \\ + F(x(t),t)B\bar{d}w(t), \end{aligned} \tag{2.2.12}$$

where A_{mn} is the mn:th component of A.

<u>Corollary</u> If the condition (2.2.5) defining the characteristic matrix A is replaced by the condition

$$\int_r^{\infty} \int_r^t g(\alpha,t,r)g(\alpha,s,r)^T dsdt = A \tag{2.2.13}$$

then (2.2.10) still holds.

Theorem 2.1 asserts that the physical process $X(t)$, considered as a function of α, tends, as α increases, to a particular diffusion process, and that this diffusion process is an $O(\alpha^{-\frac{1}{2}})$ - approximation to $X(t)$. Furthermore the relation between $X(t)$, or rather $y(\alpha,t)$, and α, can be quite arbitrary within the the constraints (2.2.2,4,5), and $X(t)$ will still converge to the same diffusion process.

In practice we may be concerned with finding a diffusion approximation for a particular physical process, and then it is more natural to think of α as a property* of the disturbance $y(t)$ rather than $y(t)$ as a function of α; that is, the inequality (2.2.3) becomes a definition of α. Similarly we can think of the conditions (2.2.4) and (2.2.13) of the corollary as defining the matrices B and A for a particular disturbance $y(t)$, and hence defining a possible diffusion approximation to $X(t)$. As we see from the corollary, if α, B and A can be meaningfully defined in this way, $x(t)$, as given by (2.2.11), is an $O(\alpha^{-\frac{1}{2}})$ - approximation to $X(t)$.

* In keeping with this interpretation we shall generally write $y(t)$ for $y(\alpha,t)$ and $g(t,s)$ for $g(\alpha,t,s)$.

We note the following points about the approximating process $x(t)$.

1) $x(t)$ is of the same dimension as $X(t)$.

2) If either $F(x,t)$ is independent of x, or A is symmetric, then the Stratonovich stochastic differential equation of the approximation $x(t)$ can be symbolically identified with the ordinary differential equation of $X(t)$ by replacing $y(t)$ in this equation by $B\,\frac{\bar{d}w(t)}{dt}$, which can be regarded as a "white noise" approximation to $y(t)$.

3) The drift and diffusion coefficients of $x(t)$ are respectively

$$f(x,t) + \sum_{mn} Q_{mn}(x,t)A_{mn}, \quad F(x,t)BB^T F(x,t)^T.$$

As we shall see in Lemma 2.1,

$$A + A^T = BB^T.$$

If A is symmetric (that is, $A = \frac{1}{2}BB^T$) or $F(x,t)$ is independent of x (when $Q_{mn}(x,t) = 0$), the only statistical parameters of the N-dimensional random disturbance $y(t)$ that have to be specified in order to determine the drift and diffusion coefficients of $x(t)$ are the $\frac{1}{2}N(N + 1)$ components of BB^T. This symmetric matrix is the intensity coefficient of $y(t)$. In general, however, the characteristic matrix has to be specified, and this has N^2 components.

2.3. Proof of Theorem 2.1

First we explain three analytical techniques that will be used repeatedly in the proof of Theorem 2.1 and we also list some required standard moment inequalities.

1) Equating random processes with integral processes.

The type of random processes we are considering are those of the form

$$Z(t) = \int_0^t g(t,r)Q(r)dw(r) \tag{2.3.1}$$

where $g(t,r)$ is a bounded measurable matrix function and $Q(r)$ is a matrix admissible random process of finite second moment. An example of such a process is $y(t)$ as defined by (2.2.1). Though $Z(t)$ is a stochastic integral it may not be an integral process in the sense of (1.3.5) because the integrand in (2.3.1) depends on t. However, the related process

$$Z(t:s) = \int_0^s g(t,r)\, Q(r)dw(r) \tag{2.3.2}$$

is an integral process in s for fixed t, and clearly

$$Z(t) = Z(t:t).$$

The advantage of equating $Z(t)$ with $Z(t:t)$ is that the Ito calculus is applicable to $Z(t:s)$ whereas it may not be to

$Z(t)$. For example, we can expand $E[Z(t)Z(t)^T]$ as follows:

$$E[Z(t)Z(t)^T] = E[Z(t:t)Z(t:t)^T]$$

which by (1.3.4)

$$= \int_0^t g(t,r)E[Q(r)Q(r)^T]g(t,r)^T dr \quad (2.3.3)$$

2) Changing the order of integration of iterated stochastic integrals. In Lemma B1, Appendix B, it is shown that if $g(s,r)$ and $Q(r)$ are as above then with probability 1

$$\int_0^t \int_0^s g(s,r)Q(r)dw(r)ds = \int_0^t \int_r^t g(s,r)Q(r)dsdw(r). \quad (2.3.4)$$

We note that the right-hand integral is of the form (2.3.1) (with $g(t,r)$ replaced by $\int_r^t g(s,r)ds$). We shall often use this technique in combination with a formula such as (2.3.3) to obtain exact expressions for the second moments of iterated stochastic integrals.

3) Special forms of Ito's formula. Let $U(t)$ and $V(t)$ be matrix integral processes such that the matrix product $U(t)V(t)$ is meaningful. In Ito's formula (1.3.9), let $x_1(t)$ be $U_{ij}(t)$, $x_2(t)$ be $V_{jk}(t)$ and $q(x(t))$ the scalar product $x_1(t)x_2(t)$, which is clearly a twice-differentiable function of $x_1(t)$ and $x_2(t)$. (1.3.9) becomes:

$$d[U_{ij}(t)V_{jk}(t)] = U_{ij}(t)dV_{jk}(t) + dU_{ij}(t)V_{jk}(t)$$
$$+ D[U_{ij}(t):V_{jk}(t)]dt$$

where $D[U_{ij}(t):V_{jk}(t)]$ denotes the cross-dispersion of $U_{ij}(t)$ and $V_{jk}(t)$ (see (1.3.8). By summing over j we obtain the corresponding formula for the ik:th component of $U(t)V(t)$. So the corresponding formula, in integral form, for the matrix $U(t)V(t)$ is

$$U(t)V(t) - U(0)V(0) = \int_0^t [U(s)dV(s) + dU(s)V(s) + D[U(s):V(s)]ds] \tag{2.3.5}$$

We shall come across products where one of the processes, say $U(t)$, has zero dispersion and is of the form $\int_0^t W(s)ds$, Then (2.3.5) can be written as

$$\int_0^t U(s)dV(s) = U(t)V(t) - \int_0^t W(s)V(s)ds \tag{2.3.6}$$

which can be considered as a formula for stochastic integration by parts.

It is convenient to have a formula for the higher-order moments of some processes. In (1.3.9) suppose $x(t)$ is $\int_0^t F(s)dw(s)$, where x, F and W are scaler; suppose $q(x(t) = x(t)^n$. Then on taking the expectation of (1.3.9) we have

$$Ex(t)^n = \tfrac{1}{2}n(n-1) \int_0^t E[x(s)^{n-2}F(s)^2]ds \tag{2.3.7}$$

Some standard inequalities. We shall make use of the following inequalities. Scaler versions of the first two are given in [14] p. 155. We give equivalent inequalities for matrix variables; these follow directly from the scalar versions by norm inequalities $|X + Y| \leq |X| + |Y|$, and $|XY| \leq |X| . |Y|$.

1) The Holder inequality. If X, Y and Z are (matrix) random variables and r, s, and t are positive numbers such that $\frac{1}{r} + \frac{1}{s} + \frac{1}{t} = 1$, then

$$E\,|XYZ| \leq (E|X|^r)^{\frac{1}{r}}(E|Y|^s)^{\frac{1}{s}}(E|Z|^t)^{\frac{1}{t}} \qquad (2.3.8)$$

A similar inequality holds for double products (set $Z = 1$, $t = \infty$).

2) If $X_1, \ldots, X_n$ are (matrix) random variables then

$$E|X_1 + \quad tX_n|^r \leq n^{r-1}(E|X_1|^r + \quad + E|X_n|^r)$$

for $r \geq 1$

(2.3.9)

3) If X(s) is a (matrix) random process

$$E\,\left|\int_0^t X(s)ds\right|^2 \leq t\,\int_0^t E|X(s)|^2 ds \qquad (2.3.10)$$

This can be proved as follows. Assume X(s) is scalar and the right-hand side of (2.3.10) is finite. Then

$$t\int_0^t EX(s)^2 ds = \int_0^t\int_0^t \tfrac{1}{2}[EX(s)^2 + EX(s')^2]ds\,ds'$$

which, by the basic inequality $\frac{1}{2}(A^2 + B^2) \geq AB$,

$$\geq \int_0^t \int_0^t E|X(s)X(s')|dsds'$$

$$\geq \int_0^t \int_0^t E[X(s)X(s')]dsds'$$

As $X(s)X(s')$ is integrable in s, s' and in expectation, we can change the order of integration and expectation (Fubini's Theorem, [14] p. 136). The last integral becomes $E(\int_0^t X(s)ds)^2$ and (2.3.10) follows. We get (2.3.10) for matrix X(t) directly from the definition of $|X(t)|^2$ as $\sum_{ij} X_{ij}(t)^2$. We note that (2.3.10) can also be considered as an integral version of the previous inequality for r=2.

In the following two lemmas we obtain asymptotic bounds for miscellaneous functions of y(t) and X(t) that we will require in the proof of Theorem 2.1.

Lemma 2.1 Asymptotic properties of y(t) and related functions. If conditions (2.2.2-5) hold and $0 \leq s \leq t \leq T$, then

$$1) \quad A + A^T = BB^T \tag{2.3.11}$$

and for increasing α

2) $E|y(t)|^2 \leq \frac{c^2}{2}\alpha$ (2.3.12)

3) $E\left|Bw(t) - \int_0^t y(r)dr\right|^2 \leq \frac{c^2}{2}\alpha^{-1}$ (2.3.13)

4) $\int_0^t\int_0^u G(u,r)g(u,r)^T drdu = At + O(\alpha^{-1})$ (2.3.14)

5) $E\left|\int_s^t L_{mn}(u,s)du\right|^4 = O(\alpha^{-2})$ (2.3.15)

6) $E|P_{mn}(t:s)|^6 = O(\alpha^{-6})$, (2.3.16)

where the convergence in the last three expressions is uniform in t and s, and where

$$G(t,r) = \int_t^{\infty} g(u,r)du \,, \qquad (2.3.17)$$

L_{mn} (u,s) is a <u>row</u> vector with i:th component

$$\sum_j \Big[G_{mi}(u,s) \int_0^s g_{nj}(u,r)dw_j(r)$$

$$+ g_{ni}(u,s)\int_0^s G_{mj}G(u,r)dw_j(r) \Big] \qquad (2.3.18)$$

and

$$P_{mn}(t:s) = \int_0^s \int_s^t L_{mn}(r,q)drdw(q) \qquad (2.3.19)$$

Proof. It is convenient to prove (2.3.12,13) first.

Let $y(t:s)$ be the integral process $\int_0^s g(t,r)dw(r)$.

Then in the manner of (2.3.3) we can write

$$\begin{aligned} E|y(t)|^2 &= E|y(t:t)|^2 \\ &= \int_0^t |g(t,r)|^2 dr, \end{aligned}$$

which from (2.2.3)

$$\begin{aligned} &\leq \int_0^t \alpha^2 c^2 e^{-2\alpha(t-r)} dr \\ &\leq \tfrac{1}{2} c^2 \alpha, \end{aligned}$$

and (2.3.12) follows. Now from (2.2.2)

$$\int_0^t y(s)ds = \int_0^t \int_0^s g(s,r)dw(r)ds.$$

As $|g(s,r)|^2$ is exponentially bounded we can change the order of integration and write

$$\int_0^t y(s)ds = \int_0^t \int_r^t g(s,r)dsdw(r).$$

Let

$$Y(t) = Bw(t) - \int_0^t y(s)ds \qquad (2.3.20)$$

From (2.2.4) it follows that

$$\int_0^t \int_r^\infty g(s,r)dsdw(r) = \int_0^t Bdw(r) = Bw(t)$$

Therefore

$$Y(t) = \int_0^t G(t,r)\, dw(r)$$

where $G(t,r)$ is given by (2.3.17). Note that

$$|G(t,r)| \leq \int_t^\infty |g(s,r)|\, ds$$

$$\leq c\, e^{-\alpha(t-r)} \qquad (2.3.21)$$

Therefore

$$E\,|Y(t)|^2 = \int_0^t |G(t,r)|^2 dr$$

$$\leq \tfrac{1}{2} c^2 \alpha^{-1}$$

and (2.3.13) follows. (2.3.13) implies that the second

moments of $\int_0^t y(s)ds$ tend, as α increases, to the corresponding moments of $Bw(t)$. So, as $\alpha \to \infty$

$$E\left[\int_0^t y(s)ds \int_0^t y(r)^T dr \right] \longrightarrow BB^T t.$$

But by the normal rules of calculus

$$\int_0^t y(s)ds \int_0^t y(r)^T dr = \int_0^t \int_0^s y(s)y(r)^T drds + \int_0^t \int_0^r y(s)y(r)^T dsdr$$

and the limit of the expectation of the right-hand side of this equation is from (2.2.5) evidently $(A + A^T)t$, so

$$BB^T t = (A + A^T)t$$

and (2.3.11) follows.

We now prove the remaining equations (2.3.14 - 16). In the remainder of this section it will be understood that order-of-magnitude terms such as $O(\alpha^{-1})$ are bounded independantly of any time parameter. If we expand the second moment of $y(t)$ in (2.2.5) in the manner of (2.3.3) we find

$$\int_0^t \int_0^s \int_0^r g(s,q)g(r,q)^T dq dr ds = A(t) + O(\alpha^{-1}),$$

which by a change in the order of integration (g is exponentially bounded)

$$= \int_0^t \int_0^s g(s,q) \int_q^s g(r,q) dr dq ds$$

Now from (2.2.4) and (2.3.17)

$$G(s,q)g(s,q)^T = Bg(s,q)^T - \int_0^s g(r,q) dr g(s,q)^T$$

and so from the last two equations

$$\int_0^t \int_0^s G(s,q)g(s,q)^T dq ds = B \int_0^t \int_0^s g(s,q) dq ds$$

$$- A^T t + O(\alpha^{-1}).$$

Again from (2.2.4)

$$\int_0^t \int_0^s g(s,q)^T dq ds = \int_0^t B^T dq - \int_0^t \int_t^\infty g(s,q)^T ds dq.$$

But

$$\int_0^t \int_t^\infty g(s,q)^T ds dq \leq \int_0^t \int_t^\infty \alpha\, c e^{-\alpha(s-q)} ds dq$$

$$\leq c\, \alpha^{-1}$$

and so

$$\int_0^t \int_0^s G(s,q)g(s,q)^T dq ds \quad = \quad BB^T t \; - \; A^T t \; + \; O(\alpha^{-1})$$

$$= \; At \; + \; O(\alpha^{-1})$$

by (2.3.11), and (2.3.14) follows.

Let $J(r,q)$ be the matrix with ij:th component

$$J_{ij}(r,q) \; = \; \int_s^t [G_{mi}(u,r)g_{nj}(u,q) \; + \; G_{mj}(u,q)g_{ni}(u,r)]du$$

Note that $J(r,q)$ in also a function of s and t, but for the remainder of this proof we shall take s and t to be fixed. Let

$$\beta[\,\cdot\,] = e^{-\alpha(\cdot)},$$

If follows from the bounding inequalities (2.2.3) and (2.3.21) on g and G that

$$J(r,q) \quad \leq \quad 2\,\alpha c^2 \int_s^t \beta(2u - r - q)du$$

$$\leq \; c^2\beta(2s - r - q) \qquad (2.3.22)$$

Now let

$$R_{ij}(u:r) = \int_0^r J_{ij}(u,q)dw_j(q)$$

$$S_{ij}(r) = \int_0^r R_{ij}(q:q)dw_i(q)$$

For fixed u, s and t $R_{ij}(u:r)$ and $S_{ij}(r)$ are integral processes in r. By a change in the order of integration (which is permissible as g and G are bounded) we find that

$$\int_s^t L_{mni}(u,s)du = \sum_j R_{ij}(s:s)$$

where $L_{mni}(u,s)$ is given by (2.3.18). Furthermore

$$P_{mn}(t:s) = \sum_{ij} S_{ij}(s).$$

So by the standard inequality (2.3.9)

$$E \left| \int_s^t L_{mn}(u,s)du \right|^4 \leq N^6 \sum_{ij} ER_{ij}(s:s)^4, \quad (2.3.23)$$

$$E\, P_{mn}(t:s)^6 \leq N^{10} \sum_{ij} E\, S_{ij}(s)^6 \quad (2.3.24)$$

We now consider the order of magnitude of the moments

$ER_{ij}(s:s)^4$ and $ES_{ij}(s)^6$. In the following proof, C_1, C_2,.., will denote constants independent of α and any time parameter. We note in passing that J_{ij} satisfies the inequality (2.3.22) for J, because $|J_{ij}|^2 \leq |J|^2$. As the same moment inequalities will be shown to hold for all i and j, we shall simplify the notation by writing R for R_{ij}, S for S_{ij} and J for J_{ij}.

From Ito's formula for second moments (1.3.4)

$$E\,R(u:r)^2 = \int_0^r J(u,q)^2 dq$$

which from (2.3.22)

$$\leq c^4 \int_0^r \beta(4s - 2u - 2q)dq$$

$$\leq C_1\,\alpha^{-1}\beta(4s - 2u - 2r) \qquad (2.3.25)$$

Now R(u:r) is Gaussian, so for positive integer n

$$E\,R(u:r)^{2n} = (2n-1)(E\,R(u:r)^2)^n$$

$$\leq C_2\,\alpha^{-n}\beta(4ns - 2nu - 2nr) \qquad (2.3.26)$$

In particular $E\,R(s:s)^4 \leq C_2\alpha^{-2}$ (for all i and j) and so (2.3.15) follows from (2.3.23). Now consider the moments of $S(u)$; we have

$$E\, S(u)^2 = \int_0^u E\, R(q:q)^2 dq$$

which by (2.3.25)

$$\leq C_1 \alpha^{-1} \int_0^u \beta(4s - 4q)dq$$

$$\leq C_3 \alpha^{-2} \beta(4s - 4u). \tag{2.3.27}$$

Expanding $S(u)^4$ by Ito's formula (see (2.3.7)) and taking the expectation we have

$$E\, S(u)^4 = 6 \int_0^u E[S(r)^2 R(r:r)^2]dr$$

which by further expansion of the integrand

$$= 6 \int_0^u \int_0^r [M_1 + 4M_2 + M_3]dqdr \tag{2.3.28}$$

where

$$M_1 = E\, [R(q:q)^2 R(r:q)^2]$$

$$M_2 = E\, [S(q)R(q:q)R(r:q)J(r,q)]$$

$$M_3 \;=\; E\,[\,S(q)^2 J(r,q)^2\,].$$

We note that J is determinate and unaffected by the expectation operator. From the Holder inequality (2.3.8) it follows that

$$M_1 \;\leq\; (E\,R(q:q)^4)^{\frac{1}{2}}(E\,R(r:q)^4)^{\frac{1}{2}}$$

which by (2.3.26)

$$\leq\; C_4\alpha^{-2}\beta[\,8s - 2r - 6q\,].$$

So

$$\int_o^u\int_o^r M_1\,dq\,dr \;\leq\; C_4\alpha^{-2}\int_o^u\int_o^r \beta\,[\,8s - 2r - 6q\,]dq\,dr$$

$$\leq\; C_4\alpha^{-4} \quad \text{for } u \leq s.$$

The integrals of M_2 and M_3 can be bounded similarly:

$$M_2 \;\leq\; (E\,S(q)^2)^{\frac{1}{2}}\,(E\,R(q:q)^4)^{\frac{1}{4}}(E\,R(r:q)^4)^{\frac{1}{4}}\;|\,J(r,q)|$$

which by (2.3.22) (2.3.25) and (2.3.27),

$$\leq\; C_5\alpha^{-2}\,\beta(6s - 2r - 4q)$$

and, as before

$$\int_0^u \int_0^r M_2 dq dr \leq C_6 \alpha^{-4}.$$

Similarly

$$\int_0^u \int_0^r M_3 dq dr \leq C_7 \alpha^{-4}$$

So it follows from (2.3.28) that for $u \leq s$,

$$E\ S(u)^4 \leq C_8 \alpha^{-4} \tag{2.3.29}$$

Analysing $E\ S(s)^6$ in a similar fashion, we have

$$E\ S(s)^6 = 15 \int_0^s E[S(r)^4 R(r:r)^2] dr$$

$$= 15 \int_0^s \int_0^r (6N_1 + 8N_2 + N_3) dq dr \tag{2.3.30}$$

where

$$N_1 = E[S(q)^2 R(q:q)^2 R(r:q)^2]$$

$$N_2 = E[S(q)^3 R(q:q) R(r:q)\ J(r,q)]$$

$$N_3 = E[S(q)^4\ J(r,q)^2]$$

By making use of the Holder inequality together with the inequalities already obtained we see that

$$N_1 \leq (ES(q)^4)^{\frac{1}{2}}(R(q:q)^8)^{\frac{1}{4}} (R(r:q)^8)^{\frac{1}{4}}$$

$$\leq C_9\alpha^{-4}\beta(8s - 2r - 6q)$$

$$N_2 \leq (E\ S(q)^4)^{\frac{3}{4}} (E\ R(q:q)^{16})^{\frac{1}{16}} (E\ R(r:q)^{16})^{\frac{1}{16}} /J(r,q)/$$

$$\leq C_{10}\alpha^{-4}\beta(6s - 2r - 4q)$$

and

$$N_3 \leq C_{11}\alpha^{-4}\beta(4s - 2r - 2q)$$

It then follows from (2.3.30) that $E\ S(s)^6 = O(\alpha^{-6})$ and then (2.3.16) follows from (2.3.24).

<u>Lemma 2.2</u> Bounds of some function moments of $X(t)$.
Let $I(x,t)$ be any of the following set of vector functions of x and t:

$$f,\ F_m,\ F_{m,t},\ Q_{mn} \text{ and } Q_{mn,t},$$

and let $J(x,t)$ be any of the set

$$F_{m,x}f,\ Q_{mn,x}f \text{ and } Q_{mn,x}F_p$$

If (2.2.7,8) hold and either (2.2.9a) or (2.2.9b) holds for some constant K' and for t in $[0,T]$,

$$E|I(X(t),t)|^4 \leq K', \quad E|J(X(t),t)|^4 \leq K', \quad (2.3.31)$$

for any I and J.

Proof. If (2.2.7,8,9a) hold, then it follows immediately that

$$|I(x,t)| \leq K , \qquad |J(x,t)| \leq K^2,$$

for all x, and so (2.3.31) holds for $K' = \max(K^4, K^8)$. Suppose (2.2.7,8,9b) hold. Then the three functions f, F_m and Q_{mn} are continuously differentiable in x and the derivative is bounded by K. So by the mean value theorem, f, F_m and Q_{mn} satisfy the inequality

$$|I(x,t) - I(0,t)| \leq K|x|$$

and hence the inequality

$$|I(x,t)| \leq K_1 + K|x|$$

where $K_1 = \max(K, |f(o,t)|, |F_m(o,t)|, |Q_{mn}(o,t)|)$

By (2.2.9b) $F_{m,t}$ and $Q_{mn,t}$ also satisfy the last inequality. Any J will clearly satisfy

$$|J(x,t)| \leq K(K_1 + K|x|)$$

Now $E|X(t)|^4$ is bounded by K so for any I, by the inequality (2.3.9)

$$\begin{aligned} E|I(X(t),t)^4| &\leq 8(K_1{}^4 + K^4 E|X(t)|^4) \\ &\leq 8(K_1{}^4 + K^5). \end{aligned}$$

Similarly $E\,|J(X(t),t)|^4 \leq 8K^4(K_1{}^4 + K^5)$, and (2.3.31) follows.

Proof of Theorem 2.1 Let us express $X(t)$ and $x(t)$ as defined by (2.2.1) and (2.2.11) in integral form:

$$X(t) = \int_0^t f(X(s),s)ds + \int_0^t F(X(s),s)y(s)ds \tag{2.3.32}$$

$$x(t) = \int_0^t f(x(s),s)ds + \int_0^t \sum_{mn} Q_{mn}(x(s),s)A_{mn}ds + \int_0^t F(x(s),s)Bdw(s) \tag{2.3.33}$$

The proof splits naturally into two parts. In the first we transform $X(t)$ by the Ito calculus into an integral equation of the form:

$$X(t) = \int_0^t f(X(s),s)ds + \int_0^t \sum_{mn} Q_{mn}(X(s),s)A_{mn}ds + \int_0^t F(X(s),s)Bdw(s) + R(t). \tag{2.3.34}$$

The remainder $R(t)$ is further expanded by the techniques described at the beginning of this section to a sum of terms for which we can obtain sufficiently accurate estimates of their orders of magnitude. The orders of magnitude of these terms have been calculated in Lemmas 2.1 and 2.2. In this way it is shown that $E|R(t)|^2$ is $O(\alpha^{-1})$. In the second part of the proof we show by a well-known method that the solution of (2.3.34) tends to the solution of (2.3.33) as $\alpha \rightarrow \infty$, and this completes the proof.

First we note the following properties of $x(t)$. By (2.2.7) $f(x,t)$ is continuously differentiable in x and its derivatives are bounded by K, so by the Mean-value Theorem, $f(x,t)$ satisfies the Lipschitz condition

$$|f(x,t) - f(x',t)| \leq K|x-x'| \qquad (2.3.35)$$

for all x, x' and all t in $[0,T]$. Similarly $Q_{mn}(x,t)$ and $F_m(x,t)$ satisfy the same condition. This, together with the continuity in t of these functions, imply the existance and uniqueness (with probability 1) of $x(t)$ as defined by (2.2.11) (see [12] p.277). Furthermore (2.2.12) is indeed the Stratonovich differential equation of $x(t)$; for the correction term that has to be added to (2.2.11) to give the Stratonovich form is $-\frac{1}{2}\sum_{mnj} Q_{mn} B_{mj} B_{nj}$ (see [1.4.6]), but from Lemma 2.1 $\sum_j B_{mj} B_{nj} = A_{mn} + A_{nm}$ and (2.2.12) follows.

We shall make use of the fact that $X(t)$, $F_m(X(t),t)$ and $Q_{mn}(X(t),t)$ are integral processes of zero dispersion. This can be shown as follows. Because $y(t)$ is a.c piecewise continuous in $[0,T]$, $y(t) \leq h$, where h is an a.c. finite random variable. This condition, the a.c. piecewise continuity of $y(t)$, and the conditions (2.2.7,8) imply that, with probability 1, the function $[f(x,t) + F(x,t)\, y(t)]$ is piecewise continuous in t and has continuous derivatives in x that are bounded by $K(1 + Nh)$. Hence by the Mean-value Theorem this function satisfies a Lipschitz condition of the form (2.3.35), with $K(1 + Nh)$ as the constant of proportionality. The usual existence and uniqueness theorem (see [18] Chapter 2) can then be applied to the ordinary differential equation (2.2.1) and so, with probability 1, the solution $X(t)$ is determined uniquely and is continuous in t. Moreover $X(t)$ is the limit, with probability 1, of a sequence $[X^n(t)]$ generated by Picard iteration with

$$X^{n+1}(t) = \int_0^t [f(X^n(s),s) + F(X^n(s),y(s)]ds,$$

$$X^o(t) = 0.$$

Because f and F are continuous functions, and $y(t)$ is both a Borel-measurable function of $[w(s),\ s \leq t]$ for each t and a measurable function of t and γ, each $X^n(t)$ is an admissible process as defined in Section 1.3. Thus $X^{\infty}(t)$ is an admissible process. As we need only determine X(t) with probability 1 we can take it to be $X^{\infty}(t)$. By (2.2.1) $\dot{X}(t)$ is an admissible process and so X(t) is an integral process of zero dispersion. As $F_m(x,t)$ and $Q_{mn}(x,t)$ are continuously differentiable processes in x, $\dot{F}_m(X(t),t)$ and $\dot{Q}_{mn}(X(t),t)$ are admissible processes and therefore $F_m(X(t),t)$ and $Q_{mn}(X(t),t)$ are integral processes of zero dispersion.

Except where it is ambiguous we shall omit the argument X(t) from functions of X(t). Our main task is to show that the last integral in (2.3.32) converges to the sum of the last two integrals in (2.3.33). Let Y(t) be 'error' process

$$Y(t) = Bw(t) - \int_0^t y(s)ds \qquad (2.3.36)$$

We note the stochastic differential of Y(t)

$$dY(t) = Bdw(t) - y(t)dt$$

Remembering that $F_m(t)$, and hence F(t), is an integral process of zero dispersion we can expand the matrix product F(t) Y(t) by Ito's formula (see (2.3.6)). On rearrangement we have

$$\int_0^t F(s)y(s)ds = \int_0^t F(s)Bdw(s) + \int_0^t dF(s)Y(s) - F(t)Y(t) \qquad (2.3.37)$$

where the m:th column of dF(s)

$$dF_m(s) = (F_{m,x}(s)f(s) + F_{m,t}(s) + \sum_n Q_{mn}(s)y_n(s))\, ds.$$

Later in the proof it will be shown that, of the terms on the right-hand side of (2.3.37), $F(t)\ Y(t)$ vanishes for increasing α and all the terms in $\int_0^t dF(s)Y(s)$ vanish except those of the form

$$\int_0^t Q_{mn}(s)Y_m(s)y_n(s)ds.$$

We shall expand this integral further, but first we expand the product $Y_m(s)\ y_n(s)$. From (2.2.2) we have

$$y_n(s) = \int_0^s \sum_i g_{ni}(s,r)dw_i(r).$$

The definition (2.3.36) of $Y(t)$ coincides with the definition (2.3.20) given in the proof of Lemma 2.1, where it was shown that

$$Y_m(s) = \int_0^s \sum_i G_{mi}\ (s,r)dw_i(r)$$

where the matrix $G(t,r) = \int_t^\infty g(s,r)ds$ for $t \geq r$, and that

$$|G(t,r)| \leq c\, e^{-\alpha(t-r)}$$

Now $Y_m(s)y_n(s) = Y_m(s:s)\, y_n(s:s)$ where $Y_m(s:r)$ and $y_n(s:r)$ are the integral processes in r

$$Y_m(s:r) = \int_0^r \sum_i G_{mi}(s,q)dw_i(q),$$

$$y_n(s:r) = \int_0^r \sum_i g_{ni}(s,q)dw_i(q).$$

Applying Ito's formula in the form (2.3.5), we have

$$\begin{aligned} Y_m(s)y_n(s) &= Y_m(s:s)y_n(s:s) \\ &= \int_0^s \sum_i G_{mi}(s,r)g_{ni}(s,r)dr \\ &+ \int_0^s \sum_i \Big[\sum_j \Big(G_{mi}(s,r) \int_0^r g_{nj}(s,q)dw_j(q) \\ &+ g_{ni}(s,r) \int_0^r G_{mj}(s,q)dw_j(q) \Big) \Big] dw_i(r) \end{aligned}$$

which we shall abbreviate to

$$Y_m(s)y_n(s) = \int_0^s \sum_i G_{mi}(s,r)g_{ni}(s,r)dr + \int_0^s L_{mn}(s,r)dw(r)$$

where $L_{mn}(s,r)$ is a row vector, the i:th component of which is the coefficient of $dw_i(r)$ in the second integral in the preceeding equation. We now compare $\int_o^s Y_m(r)y_n(r)dr$ with the process

$$\int_0^s [A_{mn}dr + \int_r^t L_{mn}(u,r)dudw(r)]$$

The difference is $M_{mn}(s) + P_{mn}(t:s)$, where

$$M_{mn}(s) = \int_0^s [\int_0^r \sum_i G_{mi}(r,q)g_{ni}(r,q)dq - A_{mn}]dr, \tag{2.3.38}$$

$$P_{mn}(t:s) = \int_0^s [\int_0^r L_{mn}(r,q)dw(q)dr - \int_r^t L_{mn}(u,r)dudw(r)] \tag{2.3.39}$$

We note that by changing the order of integration in the first integral in its definition (The second moment of $L_{mn}(r,q)$ is bounded as g and G are bounded) $P_{mn}(t:s)$ can also be expressed as

$$P_{mn}(t:s) = \int_0^s \int_s^t L_{mn}(u,r)dudw(r)$$

and so $P_{mn}(t:t) = P(t:o) = 0$. Also $M(o) = 0$. $M_{mn}(s)$ and $P_{mn}(t:s)$ are integral processes in s. $Q_{mn}(s)$ is an integral process of zero dispersion. So we can expand the product $Q_{mn}(t)\,(P_{mn}(t:t) + M_{mn}(t))$ by Ito's formula in the form (2.3.4). On rearrangement we find that

$$\int_0^t Q_{mn}(s)Y_m(s)y_n(s)ds = \int_0^t Q_{mn}(s)A_{mn}ds$$
$$+ \int_0^t Q_{mn}(s) \int_0^t L_{mn}(u,s)dudw(u)$$
$$+ Q_{mn}(t)\, M_{mn}(t)$$
$$- \int_0^t [M_{mn}(s) + P_{mn}(t:s)]dQ_{mn}(s) \tag{2.3.40}$$

where

$$dQ_{mn}(s) = [Q_{mn,t}(s) + Q_{mn,x}(s) \cdot (f(s) + F(s)y(s))]ds$$

Returning now to the integral $\int_0^t F(s)y(s)ds$, it follows from (2.3.37) and (2.3.40) that

$$\int_0^t F(s)y(s)ds = \sum_{mn} \int_0^t Q_{mn}(s)A_{mn}ds + \int_0^t F(s)Bdw(s) + R(t) \tag{2.3.41}$$

where

$$R(t) = \int_0^t a(s)ds + \int_0^t b(s)dw(s) + c(t)$$

and where

$$a(s) = \sum_m [F_{m,t}(s) + F_{m,x}(s) f(s)] Y_m(s)$$

$$- \sum_{mn} [Q_{mn,t}(s) + Q_{mn,x}(s) (f(s) +$$

$$F(s)y(s))] [M_{mn}(s) + P_{mn}(t:s)],$$

$$b(s) = \sum_{mn} Q_{mn}(s) \int_s^t L_{mn}(u,s)du ,$$

$$c(t) = - F(t) Y(t) + \sum_{mn} Q_{mn}(t)M_{mn}(t)$$

We shall now show $E|R(t)|^2 = O(\alpha^{-1})$. By norm inequalities and the standard inequality (2.3.8) it follows that

$$E|a(s)|^2 \leq (N+N^2) \Big\{ \sum_m E[|F_{m,t}Y_m|^2 + |F_{m,x}fY|^2]$$

$$+ \sum_{mn} E[(|Q_{mn,t}|^2 + |Q_{mn,x}f|^2 + |Q_{mn,x}Fy|^2)$$

$$(M_{mn}^2 + P_{mn}^2)] \Big\}$$

where the terms are evaluated at s. If we apply the Holder inequality (2.3.7) to the individual terms of this equation we have

$$E|a(s)|^2 \leq (N+N^2) \Big\{ \sum_m \big[(E|F_{m,t}|^4 \, EY_m^4)^{\frac{1}{2}} + (E|F_{m,x}f|^4 \, EY_m^4)^{\frac{1}{2}} \big] + \sum_{mn} \big[E|Q_{mn,t}|^2 + E|Q_{mn,x}f|^2 + (E|Q_{mn,x}F|^4 \, E|y|^4)^{\frac{1}{2}} \big] M_{mn}^2 + \sum_{mn} \big[(E|Q_{mn,t}|^4 EP_{mn}^4)^{\frac{1}{2}} + (E|Q_{mn,x}f|^4 \, EP_{mn}^4)^{\frac{1}{2}} + (E|Q_{mn,x}F|^4)^{\frac{1}{2}} (E|y|^{12})^{\frac{1}{6}} (EP_{mn}^6)^{\frac{1}{3}} \big] \Big\}$$

By Lemma 2.2 the fourth moments of all the functions of $X(s)$ occurring in this inequality are bounded as $\alpha \to \infty$. By Lemma 2.1, EY_m^2 is $O(\alpha^{-1})$ and $E|y|^2$ is $O(\alpha)$, where the order terms $O(.)$ are independant of time parameters. But Y_m and y are both Gaussian and so $(EY_m^4)^{\frac{1}{2}}$ is $O(\alpha^{-1})$ and both $(E|y|^4)^{\frac{1}{2}}$ and $(E|y|^{12})^{\frac{1}{6}}$ are $O(\alpha)$. Also by Lemma 2.1, $(EP_{mn}^6)^{\frac{1}{3}}$ and therefore $(EP_{mn}^4)^{\frac{1}{2}}$ are $O(\alpha^{-2})$. It therefore follows that in $[0,T]$

$$E|a(s)|^2 \leq k\,\alpha^{-1} \qquad (2.3.42)$$

for some constant k. From the inequalities (2.3.8,9)

$$E|b(s)|^2 \leq N^2 \sum_{mn}(E|Q_{mn}(s)|^4 \, E\left|\int_s^t L_{mn}(u,s)du\right|^4)^{\frac{1}{2}}.$$

By Lemma 2.2 the first factor in the brackets is bounded, and by Lemma 2.1 the second factor is $O(\alpha^{-2})$. So, for a sufficiently large constant k,

$$E|b(s)|^2 \leq k\,\alpha^{-1}. \qquad (2.3.43)$$

Similarly for sufficiently large k

$$E|c(t)|^2 \leq k\,\alpha^{-1} \qquad (2.3.44)$$

So it follows from the standard inequalities (2.3.8 - 10) that for $0 \leq t \leq T$

$$\begin{aligned} E|R(t)|^2 &\leq 3\Big[\int_0^t (t\,E|a(s)|^2 + E|b(s)|^2)ds \\ &\qquad\qquad + E|c(t)|^2\Big] \\ &\leq 3\,(T^2 + T + 1)k\alpha^{-1} \\ &\leq k_1\alpha^{-1}, \text{ for some constant } k_1 \qquad (2.3.45) \end{aligned}$$

We are now in a position to show that $E|X(t)-x(t)|^2 = O(\alpha^{-1})$. From (2.3.32) and (2.3.41) we see that X(t) can be expressed in the form (2.3.34). We note that, if (2.2.9b) holds, $E|X(t)|^2$ is bounded uniformly in t. If on the other hand (2.2.9a) holds, each of the coefficients f, Q_{mn} and F_m occurring (2.3.34) is bounded uniformly in t by K. So by the standard inequalities (2.3.8,9).

$$E|X(t)|^2 \leq (2+N+N^2) \,[\, t\int_0^t E|f(s)|^2 ds + \sum_{mn} t\int_0^t E|Q_{mn}|^2 |A_{mn}|^2 ds + \sum_m \int_0^t E|F_m(s)|^2 |B|^2 ds + E|R(t)|^2]$$

$$\leq (2+N+N^2)\,(T + |A|T + N|B|)TK^2 + k_1 \alpha^{-1}$$

So in either case $E|X(t)|^2$ is uniformly bounded in t. $E|x(t)|^2$ is also bounded (see [12] p.278). Let

$$\delta X(s) = X(s) - x(s)$$

$$\delta f(s) = f(X(s),s) - f(x(s),s)$$

and so on. As $E|\delta X(s)|^2 \leq 2\,E|X(t)|^2 + 2E|x(t)|^2$, it follows from the remarks above that $E|\delta X(t)|^2$ is bounded uniformly in t by, say, k_2. From (2.3.30,31) we have

$$\delta X(t) = \int_0^t [\,\delta f(s) + \sum_{mn} \delta Q_{mn}(s) A_{mn}\,] ds + \int_0^t \delta F(s) B dw(s) + R(t)$$

So

$$E|\delta X(t)|^2 \leq (N^2+N+2)\,[\,t\int_0^t (E|\delta f(s)|^2 + \sum_{mn} E|\delta Q_{mn}(s)|^2 |A_{mn}|^2) ds +$$

$$+ \sum_m \int_0^t E\,|F_m(s)|^2\,|B|^2 ds \;+\; E|R(t)|^2 \,] \qquad (2.3.46)$$

As f, F_m and Q_{mn} have continuous derivatives in x bounded by K, it follows by the Mean-value Theorem that $|\delta f(s)|$, $|\delta F_m(s)|$ and $|\delta Q_{mn}(s)|$ are all bounded by $K|\delta X(s)|$ and so the second moments of these functions are all bounded by $K^2 E|\delta X(s)|^2$. So from (2.3.46)

$$M(t) \leq h_1 + h_2 \int_0^t M(s)ds \qquad (2.3.47)$$

where $M(t) = E|\delta X(t)|^2$;

$$h_1 = (N^2 + N + 2)k_1 \alpha^{-1},$$

$$h_2 = (N^2 + N + 2)\,[\,T + T|A|^2 + N|B|^2\,]$$

We now show that for t in $[0,T]$

$$M(t) \leq h_1 e^{h_2 t} \qquad (2.3.48)$$

We proceed by induction; suppose that for t in $[0,T]$

$$M(t) \leq h_1 e^{h_2 t} + k_2 \frac{t^n}{n!} \qquad (2.3.49)$$

Then from (2.3.47)

$$M(t) \leq h_1 + h_2 \int_0^t \left(h_1 e^{h_2 s} + k_2 \frac{s^n}{n!}\right) ds$$

$$\leq h_1 e^{h_2 t} + k_2 \frac{t^{n+1}}{(n+1)!}$$

But (2.3.49) holds for $n = o$ as $M(t)$ is bounded by k_2. So (2.3.49) holds for all n and (2.3.48) follows. In particular

$$E|\delta X(t)|^2 \leq h_1 e^{h_2 T} = O(\alpha^{-1}).$$

This is the main result (2.2.10) of Theorem 2.1.

<u>Proof of the corollary.</u> We prove this by showing (2.2.13) implies (2.2.5). Let

$$A'(t,q) = \int_q^t \int_q^s g(s,q)\, g(r,q)^T dr ds$$

By changing the order of integration we find that

$$\int_o^t \int_o^s E[y(s)y(r)^T] dr ds = \int_o^t \int_o^s \int_o^r g(s,q)g(r,q)^T dq dr ds$$

$$= \int_o^t A'(t,q)dq.$$

But from (2.2.13)

$$|A - A'(t,q)| = \left|\int_t^\infty \int_q^s g(s,q)g(r,q)^T dr ds\right|$$

which by (2.2.3)

$$\leq (\alpha c)^2 \int_t^\infty \int_q^s e^{-\alpha(s+r-2q)} dr ds$$

$$\leq c^2 e^{-\alpha(t-q)}$$

So

$$\left| \int_0^t \int_0^s E \left[y(s)y(r)^T \right] dr\,ds - At \right| \leq \int_0^t \left| A'(t,q) - A \right| dq$$

$$\leq c^2 \int_0^t e^{-\alpha(t-q)} dq$$

$$\leq c^2 \alpha^{-1},$$

and (2.2.5) follows.

2.4 Related work

The problem of approximating a physical process by a diffusion process has been considered from different points of view by Stratonovich [20], Wong and Zakai [21,22], and Astrom [23].

In [20] pp.100-103, Stratonovich considers the approximation of a vector process $X(t)$ where (in our notation)

$$\dot{X}(t) = \rho F(X(t),t) , \qquad (2.4.1)$$

and ρ is a small parameter, and $F(x,t)$, for fixed x, is a stationary stochastic process. By expanding the increments and the joint characteristic function of the increments in a power series in ρ, Stratonovich shows that the probability density of $X(t)$ is given by

$$\dot{p} = -\rho \sum_m \frac{\partial}{\partial x_m} \left([EF_m(x,t) + \rho \sum_n \int_0^t K[\frac{\partial}{\partial x_n} F_m(x,t), F_n(x,s)] ds . p \right)$$

$$+\rho^2 \sum_{mn} \frac{\partial^2}{\partial x_m \partial x_n} \left(\int_0^t K[F_m(x,t), F_n(x,s)] ds . p \right) + O(\rho^3), \qquad (2.4.2)$$

where $K[.,.]$ is the covariance of its arguments and x is a fixed parameter. If the $O(\rho^3)$ terms are neglected, (2.4.2) corresponds to the Fokker-Plank equation of a diffusion process. Let us apply this very general result to the special process $X(t)$ considered in Theorem 2.1; we shall assume $y(t)$ is nearly stationary (that is, $g(t,u) = g(t-u)$). The coefficients of p in (2.4.2) become

$$f_m(x,t) + \sum_{nij} \int_0^t \frac{\partial}{\partial x_n} F_{mi}(x,t)F_{nj}(x,s)E[y_i(t)y_j(s)]ds$$

and

$$\frac{1}{2} \sum_{ij} \int_0^t [F_{mi}(x,t)F_{nj}(x,s) + F_{ni}(x,t)F_{mj}(x,s)]E[y_i(t)y_j(s)]ds.$$

Note that we are free to make the second coefficient symmetric in m and n as the operator $\frac{\partial^2}{\partial x_m \partial x_n}$ is symmetric. Now y depends on the parameter α. As this increases, the effective correlation time of y decreases and the limits of the two coefficients are

$$f_m + \sum_{nij} \frac{\partial}{\partial x_n} F_{mi} \cdot F_{nj} A_{ij}$$

and

$$\frac{1}{2} \sum_{ij} F_{mi} F_{mj} (A_{mn} + A_{nm}) ,$$

where the functions are evaluated at (x,t) and the sum in the second coefficient has been rearranged, as y, being nearly stationary, is sufficiently regular for

$\lim_{\alpha \to \infty} \int_0^t E[y_i(t)y_j(s)]ds$ to be constant. But then equation (2.4.2) corresponds to the Fokker-Planck equation of the limiting diffusion process of Theorem 2.1. So the diffusion approximation of Stratonovich and the limiting diffusion process given by Theorem 2.1 are equivalent in the sense that their probability laws are the same.*

In [21] Wong and Zakai consider the convergence of a sequence of scalar physical processes given by

$$\dot{X}_n(t) = f(X_n(t),t) + F(X_n(t),t)y_n(t), \quad X_n(0) = X_o,$$

where $y_n(t)$ is scalar. It is shown that if $y_n(t)$ is piecewise continuous and $\int_o^t y_n(s)ds \longrightarrow w(t)$ a.c. as $n \longrightarrow \infty$, then $X_n(t) \longrightarrow x(t)$ a.c. where $x(t)$ is a diffusion process which would satisfy the Stratonovich stochastic differential equation

$$\bar{d}x(t) = f(x(t),t)dt + F(x(t),t)\bar{d}w, \; x(o) = x_o.$$

We see that Theorem 2.1 also gives this result if $X(t)$ is a scalar process, for then the first order matrix A is trivially symmetric and the correction term in (2.2.12) is zero. The conditions in [21] imposed on $y(t)$ are weaker than those of Theorem 2.1, and enable $y(t)$ to be a quite general non-Gaussian approximation to white noise. Also, no assumption is made about the rate of divergence of $F(X_n,t)y_n$. The conditions of f and F are

* The author first learned of Stratonovich's different approach after the proof of Theorem 2.1.

similar to those of Theorem 2.1, though F is taken to be always positive.

In [22] Wong and Zakai have weakened the conditions on f and F, but have taken y(t) to be a piecewise constant Gaussian process. Also in this paper Wong and Zakai have extended their result to cover a case where $X_n(t)$ is a special vector process; we shall discuss this result again in Chapter 4.

The proofs in [21] and [22] are shorter and more elegant than the proof we have given of Theorem 2.1. It is not yet known, however, if they can be extended so that they also cover the vector case considered in Theorem 2.1.

In [23, Section 8] Astrom considers a particular linear system with stochastic coefficients, which in our notation would be the physical process

$$\dot{X} = -n_1 X + n_2 + Xy_1 + y_2$$

where y_1 and y_2 are stationary random disturbances and n_1 and n_2 are positive constants. He derives the Ito stochastic differential equation of the diffusion limit of this process for the disturbances tending to white noise. The same diffusion limit is given by Theorem 2.1 of this chapter, provided the extra assumption is made that the disturbances tending to white noise have a characteristic matrix that is symmetric (conditions for this are given in Chapter 4).

A second point considered in [23] is the steady-state behaviour of a process. Astrom shows experimentally that the steady-state distribution of a particular physical process approximates the steady-state distribution of its theoretical diffusion limit. Now Theorem 2.1 does not express the steady-state properties of a physical process not only because its proof depends on the interval T being finite but also because the second moment of the process or its diffusion limit may be infinite, in which case the concept of approximation in the mean breaks down. However, Astrom's result, in which the second moment of the diffusion limit is infinite in the steady state, demonstrates that in some cases the physical process may still be approximated by its diffusion limit in the steady-state but in the weaker sense that their distributions approximate each other.

Linear systems with stochastic coefficients have also been considered by Ariaratnam and Graefe [27]. In this paper, the authors have derived Fokker-Planck equations for processes described by two different types of linear differential equations: in the first the coefficients are perturbed by "incremental Brownian motion"; in the second they are perturbed by Gaussian white noise. The derivations are formal, but the results are consistent with the interpretation that the two different types of differential equations are, respectively, Ito and Stratonovich stochastic differential equations.

However, the equations containing Gaussian white noise can also be reasonably interpreted as a limiting form of the ordinary differential equations describing a physical process. The point to be made here is that with such an interpretation the derivation of the Fokker-Planck equation given in [27], depending as it does on the correlation functions of the noise being Dirac delta functions, breaks down.

It is still possible to give a meaning to "white noise" in this context and to manipulate the correlation functions of the noise as though they were delta functions, provided that we modify the usual definition of a delta function. If y is a stationary random disturbance (see Section 4.1), then

$$\int_0^{\infty} \rho(t)dt = A ,$$

$$\int_{-\infty}^{\infty} \rho(t)dt = A + A^T ,$$

where $\rho(t-s)$ is the matrix correlation function $E[y(t)y(s)^T]$. A is the characteristic matrix of y and $A + A^T$ its intensity coefficient. Any white noise replacement of y must satisfy at least these two relations. Except in the special case where A is symmetric, correlation functions proportional to Dirac delta functions can be chosen to satisfy only the second relation, but not the first, even if the convention is adopted that the integral of the Dirac delta function over the positive real line is half that over the whole real line. Consider instead the correlation function

$$A\,\delta^+(t) \; + \; A^T\,\delta^+(-t)$$

where $\delta^+(t)$ is a "one-sided" delta function with the property:

$$\int_\alpha^\beta \delta^+(t)dt = 1$$

if

$$\alpha \leq 0\,, \quad \beta > 0$$

(for a Dirac delta function α would have to be strictly negative). This correlation function satisfies both relations. By replacing random disturbances by noise defined in terms of this correlation function we can formally manipulate the equations of physical processes in the manner of Ariaratnam and Graefe in [27]. For example, suppose we want an equation for the mean of a scalar physical process

$$\dot{X} = y_1 X + y_2 \qquad X(0) = (0) ,$$

where y_1 and y_2 are stationary random disturbances with a characteristic matrix A. We replace y_1 and y_2 by the noise processes n_1 and n_2 with correlation function $A\,\delta^{+}(t) + A^T\delta^{+}(-t)$. We make the correlation between n_1 and X explicit by expressing X as an integral:

$$\dot{X}(t) = n_1(t) \int_0^t \dot{X}(s)ds + n_2(t)$$

$$= \int_0^t n_1(t)[n_1(s)X(s) + n_2(s)]ds + n_2(t) .$$

If we take expectations of both sides and commute the operations of expectation, differentiation and integration, this equation reduces to

$$\dot{m} = A_{11}m + A_{12} ,$$

where m is the mean of X. The diffusion approximation given by Theorem 2.1 of the physical process has a mean that also satisfies this equation, and this is why the preceding formal derivation is justified. The equations of other moments and the moments of more general physical processes can be similarly derived and justified.

CHAPTER 3

THE SIMULATION OF ONE PHYSICAL PROCESS BY ANOTHER

3.1 Approximation of distributions

A consequence of the diffusion process $x(t)$, defined in Theorem 2.1, being the limit in the mean of the physical process $X(t)$ is that the joint distributions and finite moments of $x(t)$ are also the limits of the corresponding joint distributions and moments of $X(t)$. ([14] pp.157,168). The distributions of $x(t)$ are determined by the drift coefficient

$$f(x,t) + \sum_{mn} Q_{mn}(x,t)A_{mn} \qquad (3.1.1)$$

and the diffusion coefficient

$$F(x,t)(A + A^T)F(x,t)^T , \qquad (3.1.2)$$

in which we have replaced BB^T by $A + A^T$. So we see that in these distributions the random disturbance $y(t)$ is completely characterised by its characteristic matrix A.

Suppose we want to simulate $X(t)$ by another physical process $X'(t)$ so that their distributions are approximately the same. We can do this by so choosing $X'(t)$

that it can be approximated by a diffusion process with the same drift and diffusion coefficients as $x(t)$. To be more precise we shall call $X'(t)$ a <u>δ-simulation</u> of $X(t)$ if $x(t)$ is a <u>δ-approximation</u> of $X(t)$, $x'(t)$ is a δ-approximation of $X'(t)$, and $x(t)$ and $x'(t)$ have the same drift and diffusion coefficients. If $X(t)$ satisfies (2.2.1) and the corresponding equation for $X'(t)$ is denoted by primed coefficients, then it is clear that $X'(t)$ is a δ-simulation of $X(t)$ if α is sufficiently large and

$$f' + \sum_{mn} Q'_{mn} A'_{mn} = f + \sum_{mn} Q_{mn} A_{mn} \quad , \tag{3.1.3}$$

$$F'(A' + A'^T)F'^T = F(A + A^T)F^T \; . \tag{3.1.4}$$

3.2 <u>Simulation with restrictions</u>

Suppose we want to simulate a physical process $X(t)$ on an analogue computer (or a digital computer; see Appendix A). The computer we shall suppose to contain a physical 'white noise' generator with output $z(t)$. It is clearly of interest to know under what circumstances we can simulate $X(t)$ by programming the computer directly from the equation of $X(t)$, and by simulating $y(t)$ by a rescaled version of $z(t)$.

This is the motivation for the following theorem.

<u>Theorem 3.1</u>. Suppose $y(t)$ and $z(t)$ are random disturbances of the same dimension and comparable upper frequencies α, and that their characteristic matrices, A and W, are such that $A + A^T$ and $W + W^T$ are of the same rank. Suppose $X(t)$ is a physical process satisfying the conditions of Theorem 2.1, and

$$\dot{X} = f(X,t) + F(X,t)y(t), \quad X(o) = 0. \qquad (3.2.1.)$$

Then X is $O(\alpha^{-\frac{1}{2}})$-simulated by the physical process X' where

$$\dot{X}' = f + \sum_{mn} Q_{mn}[A-CWC^T]_{mn} + FCz(t), \quad X'(0) = 0. \; (3.2.2)$$

In this equation $[.]_{mn}$ is the mn:th element of its matrix argument; f, Q_{mn} and F are evaluated at (X',t) and C is a constant matrix such that

$$A + A^T = C(W + W^T)C^T. \qquad (3.2.3)$$

<u>Corollary</u>. If 1) F is independent of x, or 2) there is a matrix C_o such that $A = C_oWC_o^T$, then X' satisfies the same equation as X but with y replaced by $C_o z$.

<u>Proof</u>. We note that (3.2.3) is the condition for the congruence of two symmetric matrices. As all symmetric matrices of equal rank are congruent, we can always find a matrix C satisfying (3.2.3). If X satisfies the conditions of Theorem 2.1, so does X'. So both X and X'

can be approximated by diffusion processes. That one simulates the other then follows from the formulas (3.1.3,4). In the corollary, the first condition implies that $Q_{mn} = 0$; a consequence of the second condition is that C, as defined by (3.2.3), can be taken to be C_o; so both conditions imply that the correction term in (3.2.2) is zero, and the corollary follows.

The corollary states that direct simulation of a physical process on a computer is possible if the noise is additive, or if the characteristic matrices of the noise and the computer noise generator satisfy a relation $A = CWC^T$ for some C; this last condition necessarily holds if the characteristic matrices are symmetric and of the same rank, for then they are congruent.

<u>A counter example</u>. The corollary is pointless if direct simulation is always possible. We give here a simple example to show that this is not the case. Consider the scalar process

$$\dot{X} = y_1 X + y_2$$

where the characteristic matrix of the vector random disturbance $(y_1, y_2)^T$ is

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} .$$

As we shall see in the next chapter, y_1 might be a process piecewise constant over intervals of length α^{-1} and y_2 the same but lagging on y_1 by one interval. If z_1 and z_2 are independent processes, the characteristic matrix W of $(z_1, z_2)^T$ is necessarily diagonal. To be a direct simulation of X, X' must satisfy an equation of the form

$$\dot{X}' = y_1' X' + y_2'$$

where the vector $y' = Cz$ for some C. The characteristic matrix of y' is CWC^T, which is necessarily symmetric as W is diagonal, and so we shall take it to be

$$A' = \begin{bmatrix} a & c \\ c & b \end{bmatrix} .$$

The formulas (3.1.3,4) must also be satisfied; these turn out to be

$$ax + c = x \quad ,$$

$$2(ax^2 + 2cx + b) = 2(x^2 + 2x + 1) .$$

There are no values of a, b and c for which these equations are identities in x, so X cannot be directly simulated.

3.3 Congruent characteristic matrices

If we take C_0 to be nonsingular, the second condition in the corollary of Theorem 3.1 reduces to the slightly stronger condition that A and W are congruent. In this section we consider some of the properties that are common to congruent matrices. Note that the equivalence class of all matrices congruent to the characteristic matrix W of a process z can be interpreted as the class of characteristic matrices of all non-singular linear transformations of z; for, if $y = Cz$, the characteristic matrix of y is CWC^T.

The congruence of asymmetric matrices is not a subject considered in the better known books on matrices. However, some of the necessary conditions for congruence can be derived fairly readily. Suppose A is congruent to W; that is, there is a non-singular matrix C such that

$$A = CWC^T \;;$$

then by transposition

$$A^T = CW^TC^T,$$

so for arbitrary λ,

$$A - \lambda A^T = C(W - \lambda W^T)C^T \,,$$

Let us take determinants, then

$$\det\,(A-\lambda A^T) = (\det\,C)^2\det(W-\lambda W^T). \qquad (3.3.1)$$

$\det(C)$ is not zero as C is non-singular. Each side of (3.3.1) is a polynomial of degree at most n, where n is the order of A. It follows that the roots of the equations

$$\det(A-\lambda A^T) = 0, \quad \det(W-\lambda W^T) = 0 \qquad (3.3.2)$$

are the same, and that this is a necessary condition for A and W to be congruent. If A and W are non-singular then the equations (3.3.2) reduce to the characteristic equations

$$\det(AA^{T-1}-\lambda I) = 0, \quad \det(WW^{T-1}-\lambda I) = 0 \qquad (3.3.3)$$

and then the necessary condition is that the eigenvalues of the matrices AA^{T-1} and WW^{T-1} are to be the same. Note that if A is symmetric

$$\det(A-\lambda A^T) = \det(A(1-\lambda)) = (1-\lambda)^n \det A$$

and so all the roots are + 1. Similarly if A is skew-symmetric they are all -1.

CHAPTER 4

THE CHARACTERISTIC MATRICES OF SOME COMMON FORMS OF RANDOM DISTURBANCES

4.1 An interpretation of the characteristic matrix

In the previous two chapters we have seen that symmetry, or asymmetry, is an important property of the characteristic matrix of a random disturbance. In this chapter we derive the conditions for which the characteristic matrices of some common forms of random disturbances are symmetric.

The equation (2.2.5) defining the characteristic matrix can be written as, for $t > 0$

$$A = \lim_{\alpha\to\infty} \frac{1}{t} \int_0^t \int_0^s E[y(s)y(r)^T]dr\, ds. \tag{4.1.1}$$

We now give some simple examples to illustrate the meaning of the symmetry of A. We note immediately from (4.1.1) that if the components of $y(t)$ are independent, A is diagonal and is trivially symmetric. To obtain a process with an asymmetric A, let us consider two scalar processes, y_1 and y_2, which rapidly reach a stationary state; that is,

$$E[y_1(s)y_2(r)] = \rho_{12}(s-r) ,$$

except for values of s and r close to the origin.

For such processes it is unnecessary to carry out the time-averaging as well as the ensemble-averaging in (4.1.1) when calculating A, and we have

$$A_{12} \approx \int_0^s \rho_{12}(s-r)dr \approx \int_{-\infty}^0 \rho_{12}(t)dt \ , \qquad (4.1.2)$$

$$A_{21} \approx \int_0^s \rho_{12}(r-s)dr \approx \int_0^\infty \rho_{12}(t)dt \ . \qquad (4.1.3)$$

It is not difficult to see how it can arise that $A_{12} \neq A_{21}$; for instance, suppose y_1 is a process with autocovariance function $\rho(t)$ and y_2 is the same as y_1, but delayed by δ. Then $\rho_{12}(t) = \rho(t+\delta)$, and as $\rho(t)$ is symmetric about the origin,

$$A_{12} - A_{21} \approx 2\int_0^\delta \rho(t)dt \ ,$$

which, for values of δ comparable to or greater than the effective correlation time of y_1, is finite and positive. Therefore A is asymmetric.

4.2 Piecewise constant Markov processes and linear diffusion processes.

In this section it is convenient to replace condition (2.2.5) by the stronger one (2.2.15)

$$\int_u^\infty g(s,u) \int_u^s g(r,u)^T drds = A \ . \qquad (4.2.1)$$

Consider now the process defined by the difference equation*

$$\left.\begin{aligned} &y(\tfrac{n+1}{\alpha}) = Gy(\tfrac{n}{\alpha}) + \alpha H[\, w(\tfrac{n+1}{\alpha}) - w(\tfrac{n}{\alpha})],\ y(o) = 0\,, \\ &y(t) = y(\tfrac{n}{\alpha})\,, \quad \tfrac{n}{\alpha} \le t < \tfrac{n+1}{\alpha}\,, \quad n = 0,1,2,\ldots \end{aligned}\right\} \quad (4.2.2)$$

where G and H constant matrices and the eigenvalues of G are less than one in absolute value. This process is Gaussian, Markov, and piece-wise constant over intervals of length $\frac{1}{\alpha}$. When considered as a function of α it possesses a characteristic matrix, obtained as a limit as $\alpha \longrightarrow \infty$. We have the following result.

Theorem 4.1. The characteristic matrix of the piece-wise constant process just defined is symmetric if and only if GHH^T is symmetric.

Corollary. The process with independent steps given by $G = 0$ has a symmetric characteristic matrix.

Proof. It can be verified that

$$y(t) = \int_0^t g(s,u)\,dw(u)$$

where

*The parameter α given here will not in general coincide with the upper frequency denoted by α in Theorem 2.1, but will be some multiple of it.

$$\left.\begin{aligned} g(s,u) &= 0, \quad \tfrac{m-1}{\alpha} < u \leq \tfrac{m}{\alpha}, \quad s < \tfrac{m}{\alpha}, \\ &= \alpha H, \quad \text{"} \quad , \tfrac{m}{\alpha} \leq s < \tfrac{m+1}{\alpha}, \\ &= \ldots \\ &= \alpha G^n H, \quad \text{"} \quad , \tfrac{n-m}{\alpha} \leq s < \tfrac{n+m+1}{\alpha}. \end{aligned}\right\} \quad (4.2.3)$$

The first task is to show that g satisfies the bounding and normalising conditions (2.2.3,4). Though $|G|$ is not necessarily less than one, $|G^m| \longrightarrow 0$ exponentially as $m \longrightarrow \infty$ because the eigenvalues of G are less than one in absolute value. So g is exponentially bounded and the following series is absolutely convergent:

$$\int_0^s g(r,u)dr = I + G + \ldots + G^n\alpha(s-s_n)$$

$$= (1-G^n)(1-G)^{-1}H + G^nH\alpha(s-s_n) \qquad (4.2.4)$$

where $\frac{m-1}{\alpha} < u \leq \frac{m}{\alpha}$ and $s_n = \frac{m+n}{\alpha} \leq s < \frac{m+n+1}{\alpha} = s_{n+1}$. So

$$B = \int_0^\infty g(r,u)dr = (1-G)^{-1}H.$$

From (4.2.1) it follows that

$$A = \sum_{n=0}^{\infty} \int_{s_n}^{s_{n+1}} [\alpha G^n H_B{}^T(I-G^n)^T + \alpha^2 G^n HH^T G^{nT}(s-s_n)]ds$$

$$= \sum_{n=0}^{\infty} [G^n HB^T - G^n HB^T G^{nT} + \tfrac{1}{2} G^n HH^T G^{nT}]$$

$$= BB^T + \tfrac{1}{2}\sum_{n=0}^{\infty} G^n HH^T G^{nT} - \sum_{n=0}^{\infty} G^n RG^{nT}, \qquad (4.2.5)$$

where $R = HB^T$. Let $U = \sum_{n=0}^{\infty} G^n R G^{nT}$. The first two terms in (4.2.5) are symmetric, so the symmetry of A is the same as the symmetry of U. U is symmetric if R is symmetric, and as U satisfies the relation

$$U - R = GUG^T, \tag{4.2.6}$$

R is symmetric if U is symmetric. (I-G) is nonsingular, so the symmetry of R is the same as the symmetry of

$$(I-G)R(I-G)^T = HH^T - GHH^T ,$$

and on retracing our steps we find the theorem is proved.

<u>Remarks</u>. As we have already mentioned in Chapter 2, Wong and Zakai [22] have given the stochastic differential equation for the limit of a vector physical process perturbed by piecewise constant noise with independent Gaussian steps. If in (4.2.2) we set G = 0 we get a process very similar to the noise process of Wong and Zakai. Now from the corollary of Theorem 4.1 it follows that the characteristic matrix is symmetric. By making the matrix A in (2.2.11) symmetric, we find that the stochastic differential equation given by Theorem 2.1 for the limit of the corresponding physical process agrees with that of Wong and Zakai.

<u>Theorem 4.2</u>. Let $y(t)$ be the linear diffusion process

$$dy(t) = \alpha Cy(t)dt + \alpha H dw(t) , \quad y(o) = 0 \qquad (4.2.7)$$

where the eigenvalues of the matrix C have negative real parts. Then the characteristic matrix of $y(t)$ is symmetric if and only if CHH^T is symmetric.

<u>Proof</u>. This is similar in nature to that of the previous theorem. We verify that

$$y(t) = \int_o^t \alpha e^{\alpha C(t-u)} H \, dw(u) ,$$

where the exponential is a matrix exponential, and so

$$g(t,u) = \alpha e^{\alpha C(t-u)} H , \qquad t \geq u \qquad (4.2.8)$$

As the real parts of the eigenvalues of C are negative, $|g|$ is exponentially bounded in the form (2.2.3), [25, p.128]. Now

$$\int_u^s \alpha e^{\alpha C(t-u)} H dr = (e^{\alpha C(s-u)} - I)C^{-1}H$$

and $B = -C^{-1}H$. So from (4.2.1), setting s-u to be t,

$$A = \int_o^\infty \alpha e^{\alpha Ct} HB^T (I - e^{\alpha C^T t}) dt$$

$$= BB^T - U$$

where

$$U = \int_o^\infty \alpha e^{\alpha Ct} HB^T e^{\alpha C^T t} dt .$$

A and U have the same symmetry. U is symmetric if HB^T is symmetric. U can be shown [24, p.125] to be a solution of

$$CU + UC^T = -HB^T ,$$

so HB^T is symmetric if U is symmetric. HB^T and CHH^T have the same symmetry and so A and CHH^T have the same symmetry. This is the result we want.

Example. Here is a direct verification of Theorem 4.1 for the following process:

$$y_1(t) = \alpha[w(\tfrac{n+1}{\alpha}) - w(\tfrac{n}{\alpha})] , \qquad \tfrac{n+1}{\alpha} \leq t < \tfrac{n+2}{\alpha} ,$$

$$y_2(t) = \alpha[w(\tfrac{n}{\alpha}) - w(\tfrac{n-1}{\alpha})] ,$$

where $w(t)$ is a scalar Wiener process. $y_1(t)$ is a piecewise constant process with independent steps; $y_2(t)$ is $y_1(t)$ delayed by an interval $\frac{1}{\alpha}$. Cross sections at constant **u** of the weighting functions $g_1(t,u)$ and $g_2(t,u)$ of y_1 and y_2 are illustrated in Figure 4.1.

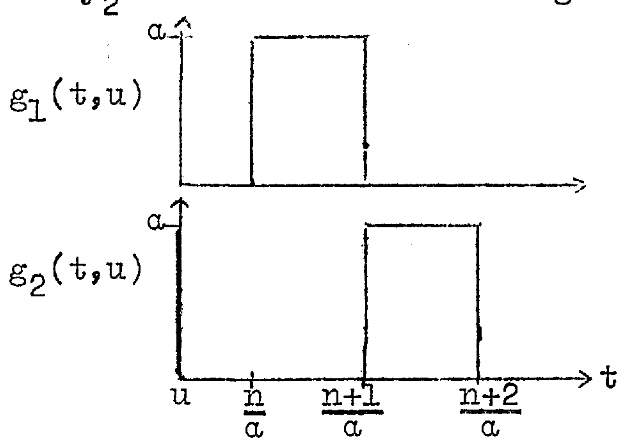

Figure 4.1

The integral of each weighting function with respect to t is clearly 1. So

$$B_1 = B_2 = 1$$

The integral $\int_0^t y_2(s)ds$ is independent of $y_1(t)$ and so

$$A_{12} = 0 .$$

But we know that $A_{21} = B_1B_2 - A_{12}$, so

$$A_{21} = 1.$$

The complete characteristic matrix is $\begin{bmatrix} \frac{1}{2} & 0 \\ 1 & \frac{1}{2} \end{bmatrix}$, which is asymmetric. The joint process $(y_1, y_2)^T$ is a special case of the process considered in Theorem 4.1, with

$$G = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \text{, and } H = \begin{bmatrix} 1 \\ 0 \end{bmatrix} .$$

$GHH^T = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ is asymmetric, which by Theorem 4.1 confirms the asymmetry of the characteristic matrix.

CHAPTER 5

THE FILTERING OF INTEGRAL AND DIFFUSION PROCESSES

5.1 Introduction

Our primary purpose in this chapter is to present some results on the filtering of integral and diffusion processes that are required in the next chapter, which is concerned with physical processes, but some of these results are of independent interest.

Suppose that x and z are nonlinear vector integral processes. We shall call x the message process, z the observation process and the conditional expectation

$$m = E[x(t) \mid z(t') \ , \quad t_0 \leq t' < t]$$

the estimate of x. Then in this chapter we consider two 'filtering' problems that can be roughly stated as:

a) find the form of the stochastic differential equation of m;

b) find functions q, h and H such that $q(\bar{m},t)$ is a reasonable approximation in the mean to m, $\bar{m}$ being the vector solution of

$$d\bar{m} = h(\bar{m},t)dt + H(\bar{m},t)dz \ . \qquad (5.1.1)$$

Equation (5.1.1) could be interpreted as a theoretical filtering algorithm, as it gives a dynamic description of how an approximate estimate is generated continuously as a function of the observation process. However, it is a stochastic differential equation and so cannot be used directly for the design of 'physical' filters; that is, filters that are described by ordinary differential equations. We leave the discussion of this point to the next chapter.

A solution of problem (a), as we shall see, may serve as a starting point for the solution of (b). Problem (a) and the related one of finding a differential equation for p_t, the conditional probability density of the message process, has been studied by several authors for the case where the message and observation processes are joint diffusion processes of the general form

$$dx = a(x,z,t)dt + F(x,z,t)dw \quad , \tag{5.1.2}$$

$$dz = b(x,z,t)dt + G(z,t)dw, \; z(t_o) = 0 \tag{5.1.3}$$

where GG^T is nonsingular. In 1960 Stratonovich [1] obtained a differential equation for p_t for a wide class of Markov processes, including, in particular, diffusion processes for which a, b, F and G are independent of time. In 1964, Kushner [2], following more closely Ito's calculus, derived a stochastic differential equation for p_t (for a, b, F and G independent of z and t). Other derivations have been given by Bucy [3] (for a, b, F and G independent of z and t and $FG^T = 0$) and Wonham [4] (for $FG^T = 0$, a and F independent of z). Wonham [5] has also derived rigorously the corresponding estimate equations for a message process that is a Markov step process. Recently Kushner [6] has derived rigorously a stochastic differential equation for the estimate m (for a, b, F and G independent of z and $FG^T = 0$).

In the derivations of all these authors p_t or m is defined explicitly; that is, the conditional probability density or expectation of the message process for a finite number of observations is first constructed by Bayes formula and then p_t or m, or rather the equation describing p_t or m, is obtained by a limiting process as the number of observations taken into account is increased.

For the case where the message and observation processes are integral processes with a.c. continuous drift and dispersion terms, we can derive the stochastic differential of the estimate in a comparatively simple way in which the taking of limits is avoided, provided that we are prepared to make the plausible assumptions that the estimate is also on integral process with a.c. continuous drift and dispersion terms and that it is an integral functional of the observation process. We give this derivation in Section 5.2: instead of constructing the estimate by Bayes formula we define it implicitly by two of the many identities that conditional expectations necessarily satisfy; these are

$$YE[X|Y] = E[YX|Y] \quad , \tag{5.1.4}$$

$$E[E[X|Y,Z]|Z] = E[X|Z] \tag{5.1.5}$$

where X, Y and Z are random variables. By using other identities it may be possible to weaken the assumptions, though the assumption of a.c. continuity of various terms seems essential to our method.

If the message and observation process are taken to be the diffusion processes considered by Kushner, Bucy or Wonham, the estimate equation obtained in Section 5.2 is the same as the corresponding equations of these authors. However, from a formal comparison made in Section 5.3 it appears to differ in general from the corresponding equation of Stratonovich. In making this comparison we have proceeded on the assumption that the equation for p_t given by Stratonovich in [1] is to be interpreted as a Stratonovich differential equation. This is not stated in [1], but in a later paper [7] Stratonovich explains that some related equations are to be interpreted in this form. However, for two important problem formulations: 1) b is linear in x and z, and G and FG^T are determinate, 2) b and G are independent of z and $FG^T = 0$, the estimate equation given in Section 5.3 agrees with the estimate equation of Stratonovich. It is interesting to note that, on the basis of this comparison, some, but not all, of the estimate equations given by the authors we have already cited appear to be equivalent to Stratonovich's equation, because the corresponding problem formulations come under 1) or 2).

It remains for us to consider problem (b) and look for filtering algorithms. In the last section we restrict the message process to be a diffusion process of small dispersion; that is the 'a priori' covariance of the message process at any time is small. Blagoveshchenskii [8] has shown that such processes can be expanded in a power series of a small parameter and that any first N coefficients of this expansion form a diffusion process. We formally apply his analysis to the estimate equation (5.2.4) and obtain diffusion equations for the first three coefficients of the expansion for the estimate. The sum of the corresponding terms in the expansion form an approximation to the estimate of known order of accuracy. The equations for the coefficients do not form a solution of (b), as they contain non-observable forcing functions, but we can use them as a basis for finding algorithms that do. We suggest two such algorithms and compare one of them with a similar algorithm given by Wonham [4].

5.2 The estimate equation of an integral process

Suppose x and z are integral processes such that for $0 \leq t_o \leq t \leq T$,

$$dx(t) = a(t)dt + F(t)dw(t), \tag{5.2.1}$$

$$dz(t) = b(t)dt + G(t)dw(t), \quad z(t_o) = 0 \tag{5.2.2}$$

and that the drifts a and b, and the dispersions F and G, are a.c. continuous stochastic processes with finite second moments. Let

$$E_t[-] = E[- \mid z(s),\ t_o \leq s \leq t] \ ;$$

so that $m(t) = E_t x(t)$.

Suppose G is determined by past z; that is, $E_t G(t) = G(t)$ with probability 1, and that GG^T is positive definite.

If m is an integral process such that the differential of m is of the form

$$dm(t) = R(t)dt + S(t)dz(t) \tag{5.2.3}$$

where R and S are determined by past z, and that the drift $\underline{d}[m]$ (= R + Sb) and the dispersion $D[m:w^T]$ (= SG) are a.c. continuous processes with finite second moments, then for $t_o \leq t \leq T$,

$$dm = E_t a dt + E_t[x(b-E_t b)^T + FG^T](GG^T)^{-1}(dz - E_t b dt), \quad (5.2.4)$$

all terms being evaluated at t.

Proof. The conditions we have imposed on the coefficients in (5.2.1), (5.2.2) and (5.2.3) imply that the second moments of x, z and m are finite. So, by the Schwarz inequality, $E[x(t)z(t)^T]$ is finite; but then so must be $E_s[x(t)z(t)^T]$ with probability 1 for all s and t. Similarly all the conditional expectations we shall be considering are finite. The two identities, for $t_o \leq s \leq t \leq T$,

$$E_s m(t) = E_s x(t) , \quad (5.2.5)$$

$$E_s[m(t)z(t)^T] = E_s[x(t)z(t)^T] , \quad (5.2.6)$$

which are generalisations of (5.1.4) and (5.1.5), are sufficient to determine R and S. From (5.2.3) it follows that

$$\begin{aligned} E_s m(t) &= E_s[m(s) + \int_s^t (R(r) + S(r)b(r))dr + S(r)G(r)dw(r)] \\ &= m(s) + \int_s^t E_s[R(r) + S(r)b(r)]dr , \end{aligned}$$

as $E_s m(s) = m(s)$ with probability 1, and the increments of w in the interval [s,t] are independent of z before s. Similarly

$$E_s x(t) = m(s) + \int_s^t E_s a(r)dr ,$$

so from (5.2.5)

$$\int_s^t E_s[R(r) - S(r)b(r) - a(r))dr = 0 . \qquad (5.2.7)$$

The a.c. **continuity** of a process with finite expectation implies the a.c. continuity of any of its conditional expectations, and as a and R + Sb are a.c. continuous, so are their conditional expectations. As (5.2.7) is valid for all t in [s,T] the integrand vanishes with probability 1. In particular

$$E_s[R(s) - S(s)b(s) - a(s)] = 0 . \qquad (5.2.8)$$

From (5.2.2), (5.2.3) and Ito's formula for the transformation of integral processes, we have

$$m(t)z(t)^T = m(s)z(s)^T + \int_s^t [R(r)dr + S(r)dz(r))z(r)^T + m(r)dz(r)^T + S(r)G(r)G(r)^T dr] .$$

So

$$E_s[m(t)z(t)^T] = m(s)z(s)^T + \int_s^t E_s[(R(r) + S(r)b(r))z(r)^T + m(r)b(r)^T + S(r)G(r)G(r)^T]dr . \qquad (5.2.9)$$

Similarly

$$E_s[x(t)z(t)^T] = m(s)z(s)^T + \int_s^t E_s[a(r)z(r)^T + x(r)b(r)^T + F(r)G(r)^T]dr . \qquad (5.2.10)$$

We see from the identity (5.2.6) that the integrals in (5.2.9) and (5.2.10) are equal with probability 1. If we equate the integrands for the reasons of continuity we have already given, we have

$$E_s[(R + Sb - a)z^T - (m-x)b^T + SGG^T + FG^T] = 0 , \quad (5.2.11)$$

evaluations being at s. In (5.2.8) and (5.2.11) the terms R, S, z, m and GG^T are determined by past z and so are unaffected by the conditional expectation operator E_s. It follows from these two equations that, evaluations still being at s,

$$S = E_s[x(b-E_s b)^T + FG^T](GG^T)^{-1} ,$$

$$R = E_s a - SE_s b ,$$

and by setting t = s, equation (5.2.4) follows.

<u>Conditional moments</u>. It follows from Ito's formula that any twice continuously differentiable function of an integral process with a.c. continuous drift and dispersion terms is itself an integral process with a.c. continuous drift and dispersion terms. So if h is some power of x, and the appropriate expectations are finite, the equation for the conditional moment $E_t h$ corresponding to (5.2.4) is

$$dE_t h = E_t \underline{d}[h]dt + E_t[h(b-E_t b)^T + D[h:z^T]](GG^T)^{-1}(dz-E_t b dt), \quad (5.2.12)$$

the drift and dispersion terms $\underline{d}[h]$ and $D[h:z^T]$ being given by Ito's formula.

Remarks. The derivation of (5.2.4) does not really depend on our assumption that G is determined by past values of z. If, however, G is not determined by past z, the whole problem can become unrealistic. Suppose, for instance, we make the scalar observation $z(t)$, $t_o \leq t \leq T$; then in principle we can 'observe' the sum

$$\sum_i [z(t_i + \delta) - z(t_i)]^2$$

taken over a partition of $[t_o, T]$ into intervals of length δ. By decreasing δ we can make this sum converge to

$$\int_{t_o}^{T} G(t)^2 dt.$$

If, for sake of argument, G were $\sqrt{x}$, where x is an unknown positive constant, this integral would be $x(T-t_o)$ with probability 1. Then we would have an exact observation of x, and the equations we have been considering would only be meaningful for the trivial case where $m(t) = x(t)$ with probability 1. This unrealistic situation occurs in other estimation problems and is associated, Bartlett ([9] p.244) explains, with the Gaussianness of the Wiener process $w(t)$.

5.3 The estimate in Stratonovich form; a comparison with the results of Stratonovich

If the message and observation processes of the last section are taken to be the diffusion processes considered by Kushner in [2] or [6], by Bucy in [3] or by Wonham in [4], the equation (5.2.3) for the estimate is formally the same as the corresponding equations of these authors. In this section we transform (5.2.3) into a Stratonovich stochastic differential equation and then compare it with a corresponding result of Stratonovich. For message and observation processes that are of the general form (5.1.2,3) but with the coefficients independent of time, Stratonovich [1] gives a differential equation for the conditional probability density of $x(t)$. It is not clear from [1] how this equation is to be interpreted. However, in a later paper [7] Stratonovich explains that some equations in [10], which are special cases of the equations in [1], are to be taken as Stratonovich stochastic differential equations. So we shall proceed on the assumption that the equation in [1] is also of this form.

Stratonovich's equation. This is

$$\bar{d}p_t = - \sum_i \frac{\partial}{\partial x_i} [(a + FG^T W[\bar{d}z - bdt])p_t]_i$$
$$+ \frac{1}{2} \sum_{ij} \frac{\partial^2}{\partial x_i \partial x_j} [(FF^T - FG^T WGF^T)p_t]_{ij} dt \qquad (5.3.1)$$
$$+ [b^T W(\bar{d}z - \tfrac{1}{2}bdt) - E_t[b^T W(\bar{d}z - \tfrac{1}{2}bdt)]]p_t ,$$

where $[-]_i$ is the i:th component of the contained vector and $[-]_{ij}$ the ij:th component of the contained matrix. The functional dependence of a, for instance, is $a(x', z(t), t)$; x' here is simply the parameter of the conditional probability density p_t; $z(t)$ is, however, the stochastic observation process. W is defined to be $(GG^T)^{-1}$. The conditional expectation of $x(t)$ is

$$E_t x = \int x' p_t(x') dx' .$$

If we apply these formulae to (5.3.1) and integrate by parts where necessary we find that

$$\bar{d}E_t x = E_t[a - \tfrac{1}{2}x(b^T Wb - E_t b^T W E_t b) - FG^T Wb]dt$$
$$+ E_t[x(b - E_t b)^T + FG^T]W\bar{d}z \quad . \qquad (5.3.2)$$

It is this equation that we shall compare with (5.2.4) in Stratonovich form.

<u>(5.2.4) in Stratonovich form</u>. If the conditional expectations occurring in (5.2.4) are components of some extended diffusion process we can transform (5.2.4) into

Stratonovich form by means of the formula given in Section 1.4. We assume this transformation is valid. To make the transformation, we insert a correction term $-\frac{1}{2}D[H:z]dt$ into the right member of (5.2.4), H being the coefficient of dz.

First we note that, if for all i the i:th column J_i of a matrix integral process J satisfies (5.2.11), then

$$D[E_tJ:z] = E_t[J(b-E_tb) + D[J:z]] \ . \qquad (5.3.3)$$

For

$$D[E_tJ:z] = \sum_i E_t[J_i:z_i]$$

which by (5.2.11) and the definition of dispersion

$$= \sum_i E_t[J_i(b-E_tb)^T + D[J_i:z^T]]W(GG^T)_i \ ,$$

where $(GG^T)_i$ is the i:th column of GG^T. But $W = (GG^T)^{-1}$, so

$$D[E_tJ:z] = \sum_i E_t[J_i(b_i - E_tv_i) + D[J_i:z_i]] \ ,$$

which is the same as (5.3.3).

We are now in a position to determine $D[H:z]$. Set J to be

$$(x - m)b^TW + FG^TW \ .$$

Then $H = E_tJ$, and by (5.3.3)

$$D[H:z] = D[E_tJ:z]$$

$$= E_t[J(b-E_tb)] + E_tD[J:z] \ . \qquad (5.3.4)$$

It follows from the definition of dispersion that

$$E_tD[J:z] = E_t[D[(x-m):z^T]Wb + (x-m)D[b^TW:z] + D[FG^TW:z]] .$$

But

$$D[x:z^T] = FG^T ,$$

$$D[m:z^T] = E_t[x(b-E_tb)^T + FG^T] ,$$

and so (5.3.4) reduces to

$$D[H:z] = E_t[(xb^T-mb^T-xE_tb^T+2FG^T)W(b-E_tb)] + K. \qquad (5.3.5)$$

where

$$K = E_t[(x-m)D[b^TW:z] + D[FG^TW:z]] . \qquad (5.3.6)$$

So the resulting Stratonovich stochastic differential equation for the estimate m corresponding to (5.2.4) is

$$\bar{d}m = E_t[a - \tfrac{1}{2}x(b^TWb - E_tb^TWE_tb) - FG^TWb]dt + E_t[x(b-E_tb)^T + FG^T]W\bar{d}x - \tfrac{1}{2}Kdt . \qquad (5.3.7)$$

Stratonovich's equation (5.3.2) and our equation (5.3.7) are equivalent only if

$$K = 0.$$

It is not difficult to show that this is so for the following two important special cases:

1) the observation process is linear and the correlation between the "noises" on the observation and message processes is determinate; that is, b is linear in x and z, and G and FG^T are determinate;

2) the observation process has no dynamics and there is no correlation between the "noises" on the observation and message processes; that is, b and G are independent of z and $FG^T = 0$.

5.4 Filtering algorithms

Let us suppose that the message process is a diffusion process, its dispersion is small, and the random spread of its initial value is also small; that is,

$$dx = a(x,t)dt + \rho F(x,t)dw, \quad x(t_o) = c + \rho r, \qquad (5.4.1)$$

where ρ is a small parameter, c is a constant and r a random variable of finite second moment. The message process is evidently a function of ρ; Blagoveshchenskii [8] shows that, if a and F are sufficiently smooth, it can be expanded as a power series in ρ and that any first N coefficients of this series form a diffusion process. If we take continuous observations of z where

$$dz = b(x,t)dt + G(t)dw, \quad z(t_o) = 0 \qquad (5.4.2)$$

and $FG^T = 0$, then the observation process will be a function of ρ and so m, the estimate of x, will be too. One might conjecture that this estimate can also be expanded as a power series in ρ:

$$m = m_o + m_1\rho + \tfrac{1}{2}m_2\rho^2 + \ldots + R_N\rho^N, \qquad (5.4.3)$$

R_N being a remainder with bounded second moment. Clearly the sum of the first N terms of this series form a $O(\rho^N)$- approximation to the estimate. If we can find a set of equations such that the only forcing term is the observation process z and such that some of the generated variables are $O(\rho^N)$- approximations to the terms $m_i\rho^i$ in (5.4.3), then these equations taken together with (5.4.3) constitute a solution to the problem (b) posed in section 5.1.

In the remainder of this section we shall formally derive a set of such equations that will give us an $O(\rho^3)$- approximation to the estimate. The method for obtaining higher order approximations is similar, but then the equations are considerably more complicated. We start by finding the diffusion equations satisfied by the differential coefficients of x with respect to ρ. Let a function of ρ and its first and second differential coefficients, all evaluated at $\rho = 0$, be denoted by the suffices 0, 1 and 2 respectively. Then differentiating

(5.4.1) with respect to ρ we have

$$dx_0 = a_0 dt , \quad x_0(t_o) = c , \tag{5.4.4}$$

$$dx_1 = a_{x0}x_1 dt + F_0 dw , \quad x_1(t_o) = r, \tag{5.4.5}$$

$$dx_2 = \langle a_{xx0}, x_1 x_1^T \rangle dt + a_{x0}x_2 dt + F_{x0}x_1 dw,$$

$$x_2(t_o) = 0 \tag{5.4.6}$$

where the operator $\langle (-)_{xx}, xx^T \rangle$ is the differential operator $\sum_{ij} x_i x_j \frac{\partial^2}{\partial x_i \partial x_j} (-)$, x_i being the i:th component of x. The validity of these equations is proved in [8]. Note that x_0 is a determinate process.

As the observation process depends on x, and x depends on ρ, the expectation operator E_t also depends on ρ. We list some of the more obvious properties of this dependence.

1) If E_t and x are sufficiently smooth functions of ρ it follows from the linearity of E_t that

$$\begin{aligned} (E_t x)_0 &= E_{t0} x_0 , \\ (E_t x)_1 &= E_{t1} x_0 + E_{t0} x_1 \end{aligned} \tag{5.4.7}$$

and so on.

2) If c is a function of time and not of ρ,

$$E_t c = c ,$$

and so from (5.4.7)

$$E_{t0}c = c \ , \quad E_{t1}c = E_{t2}c = 0 \ . \qquad (5.4.8)$$

3) Let us write the a priori expectation operator E_{t_o} as E. Suppose u is part of a diffusion process which is independent of ρ and such that $D[u:z^T] = 0$, then

$$E_{t0}u(t) = Eu(t) \ , \qquad (5.4.9)$$

for, if $\rho = 0$, x is determinate; so the observation process z is independent of u and contributes no new information about u.

We now derive equations satisfied by the differential coefficients of E_t. For a process u as defined above the conditional expectation $E_t u$ satisfies, by (5.2.4),

$$dE_t u = E_t \underline{d}[u]dt + (E_t[ub^T] - E_t u E_t b^T)W(dz - E_t b dt),$$

$$E_{t_o}u(t_o) = Eu(t_o), \qquad (5.4.10)$$

where $W = (GG^T)^{-1}$. In this equation $b(x,t)$, $z(t)$ and the operator E_t are functions of ρ. With the aid of (5.4.4)-(5.4.9) we can show that

$$(E_t[ub^T] - E_t u E_t b^T)_0 = 0 \ ,$$

$$(E_t[ub^T] - E_t u E_t b^T)_1 = (E[ux_1^T] - EuEx_1^T)b_{x0}^T \ ,$$

and

$$(dz - E_t bdt)_0 = (b_0 - E_{t0}b_0)dt + Gdw$$

$$= Gdw \ .$$

If in (5.4.10) we get $\rho = 0$, we obtain an equation that just reinforces our somewhat intuitive argument that $E_{t0}u = Eu$. If we differentiate (5.4.10) for ρ and then set $\rho = 0$, we have, using the above equations

$$dE_{t1}u = E_{t1}\underline{d}[u]dt + (E[ux_1^T] - EuEx_1^T)b_{x0}^T WGdw ,$$

$$E_{t_o1}u(t_o) = 0 . \qquad (5.4.11)$$

The equations for the higher order differential coefficients of E_t can be found in the same manner, but we shall not need them.

<u>The differential coefficients of the estimate</u>. From (5.4.7) and (5.4.9) it follows that the coefficients

$$\left.\begin{aligned} m_0 &= E_{t0}x_0 = c_0 , \\ m_1 &= E_{t1}x_0 + E_{t0}x_1 = Ex_1 , \\ m_2 &= E_{t2}x_0 + 2E_{t1}x_1 + E_{t0}x_2 = 2E_{t1}x_1 + Ex_2 \end{aligned}\right\} (5.4.12)$$

and from (5.4.5), (5.4.6) and (5.4.11) evaluated at $\rho = 0$ that

$$\left.\begin{aligned} &dEx_1 = a_{x0}Ex_1 dt , \qquad Ex_1(t_o) = Er , \\ &dEx_2 = \langle a_{xx0}, E[x_1x_1^T]\rangle dt + a_{x0}Ex_2 dt, \ Ex_2(t_o) = 0 , \\ &dE_{t1}x_1 = a_{x0}E_{t1}x_1 dt + (E[x_1x_1^T] - Ex_1Ex_1^T)b_{x0}^T WGdw , \\ &\qquad E_{t_o1}x_1(t_o) = 0 , \end{aligned}\right\} (5.4.13)$$

and

$$dE[x_1x_1^T] = (a_{x0}E[x_1x_1^T] + E[x_1x_1^T]a_{x0}^T + F_0F_0^T)dt \ .$$

If v is the conditional covariance of x; that is

$$v = E_t[xx^T] - E_txE_tx^T \ ,$$

then we find that

$$v_0 = v_1 = 0, \quad v_2 = 2(E[x_1x_1^T] - Ex_1Ex_1^T) \ . \qquad (5.4.14)$$

We see from the equations (5.4.12) - (5.4.14) that equations for the differential coefficients of m can be conveniently written as

$$\left.\begin{aligned}
m_0 &= x_0, \ (dx_0 = a(x_0,t)dt, \ x_0(t_o) = c), \\
dm_1 &= a_{x0}m_1dt, \quad m_1(t_o) = Er \ , \\
dm_2 &= (a_{x0}m_2 + \langle a_{xx0}, \tfrac{1}{2}v_2 + m_1m_1^T\rangle)dt, \\
&\quad + v_2b_{x0}WGdw, \quad m_2(t_o) = 0 \ ,
\end{aligned}\right\} \qquad (5.4.15)$$

where

$$dv_2 = a_{x0}v_2 + v_2a_{x0}^T + 2F_0F_0^Tdt \ ,$$

$$v_2(t_o) = 2 \ \mathrm{Cov}[r] \ .$$

<u>Filtering algorithms</u>. The equations (5.4.15) cannot be used as they stand as a filtering algorithm for the estimate, for the forcing term dw is not 'observable'. However, we note that this forcing term only occurs in

the equation for m_2, and for an $O(\rho^3)$- approximation of the estimate we need only an $O(\rho)$- approximation to m_2. It is not difficult to show that the 'observable' term $dz - b_0 dt$ is an $O(\rho)$- approximation to dw.[1]

From a physical point of view it would be desirable for ρ only to occur in its natural place as a coefficient of F, for then it would be unnecessary to measure ρ explicitly.

These considerations suggest the following two filtering algorithms; it is not difficult to show with the analysis of this section that both these algorithms give $O(\rho^3)$- approximations to the estimate.[2]

1) The first is for the case where $Ex(t_o) = c$:

$$\left.\begin{aligned} &\dot{x}_0 = a(x_0, t) , \quad x_0(t_o) = c , \\ &\dot{\delta v} = a_{x0}\delta v + \delta v\, a_{x0}^T + \rho^2 F_0 F_0 , \quad \delta v(t_o) = E[\rho^2 r^2] , \\ &\text{and} \\ &d\delta m = a_{x0}\delta m\, dt + \tfrac{1}{2}\langle a_{xx0}, \delta v\rangle dt \\ &\qquad + \delta v\, b_{x0}^T W(dz - b_o dt) , \quad \delta m(t_o) = 0 , \end{aligned}\right\}(5.4.16)$$

where, as before, the suffix 0 indicates evaluations at x_0. $x_0 + \delta m$ is the $O(\rho^3)$- approximation to the estimate. Note that x_0 and δv are determinate and so can be calculated separately from δm, and that the stochastic processes δm and z only occur linearly.

2) A more compact but nonlinear algorithm is

$$\left.\begin{aligned} \dot{\underline{v}} &= \underline{a_x v} + \underline{v a_x^T} + \rho^2 \underline{FF^T} , \quad \underline{v}(t_o) = \mathrm{Cov}[x(t_o)] , \\ d\underline{m} &= \underline{a} dt + \tfrac{1}{2} \langle \underline{a_{xx}}, \underline{v} \rangle dt + \underline{v b_x^T} W (dz - \underline{b} dt) , \\ & \qquad \underline{m}(t_o) = Ex(t_o) , \end{aligned}\right\} \quad (5.4.17)$$

where the underlining indicates evaluations at $\underline{m}$. The estimate of x is $O(\rho^3)$-approximated by $\underline{m}$.

In these two algorithms δv and $\underline{v}$ are $O(\rho^3)$-approximations to the conditional covariance of x, which is itself of order ρ^2. It is interesting to note that the equations defining δv and $\underline{v}$ do not depend on the statistics of the observation process z. This implies that the difference between the covariance of x if no observations are made (or the observation noise is infinite) and the covariance of x conditional on the observations z is of order ρ^3, and can be neglected.

A related algorithm. In any application of these algorithms it would be necessary to analyse more thoroughly the errors involved, and in particular to investigate the dependence of the errors on time. It may be found that the period of time over which the approximations are acceptable can be extended by retaining in the algorithms some of the terms of order ρ^3. This point is illustrated by an algorithm suggested by Wonham [4]. If applied to the process x of this section it generates a $O(\rho^3)$-approximation to the estimate and an approximation to the conditional covariance of possibly greater accuracy. It is similar in form to the algorithm (5.4.17) though it contains terms that are order ρ^4. If the message and observation process are linear and Gaussian, Wonham's algorithm coincides with the exact filtering algorithm of Kalman and Bucy, whereas ours do not. It is reasonable to suppose, therefore, that for nearly linear processes Wonham's algorithm, if compared to ours, generates acceptable approximations over a longer period of time.

The general case. For more complicated message and observation processes or for more accurate approximations the analysis we have given is more suitable for checking the accuracy of postulated algorithms than for deriving them. However, on inspecting (5.4.17) we see that this algorithm could have been derived by discarding from the exact equations for the estimate and conditional covariance all terms of order ρ^3. This is roughly the method used by Wonham in [4], though his criterion for discarding is different. In this way more general filtering algorithms might be generated.

CHAPTER 6

THE FILTERING OF PHYSICAL PROCESSES

6.1 Introduction

In the last chapter we considered the estimation of one diffusion process given a continuous observation of another, and derived a filtering algorithm in the form of a stochastic differential equation. In this chapter we consider the similar problem of estimating a physical process given a continuous observation of a second physical process and we derive a filtering algorithm that is an ordinary differential equation. Calling the message and observation physical processes X and Z respectively, the problem is:

find a differential equation

$$\dot{M} = h(M,t) + H(M,t)\dot{Z} \tag{6.1.1}$$

such that h and H are continuous functions, and a differentiable function $q(M(t),t)$ that is a reasonable approximation in the mean to the estimate of X

$$E[X(t)|Z(s), \quad t_0 \leq s \leq t] .$$

This problem is more realistic than the corresponding problem (b) of the last chapter in that the message process, the observation process and the approximate estimate all now have piecewise continuous velocities, and that (6.1.1), being an ordinary differential equation, can be simulated by a practical filter. In the remainder of this chapter we suppose that X and Z are joint physical processes satisfying the conditions of Theorem 2.1 and such that for $t_o \leq t \leq T$,

$$\dot{X} = f(X,Z,t) + F(X,Z,t)y, \qquad (6.1.2)$$

$$\dot{Z} = g(X,Z,t) + G(Z,t)y, \quad Z(t_o) = 0 \qquad (6.1.3)$$

where GG^T is positive definite and the random disturbance y is such that the characteristic matrix A satisfies the relation $A + A^T = I$. So that we may make use of Theorem 2.1 we suppose that the random initial value $X(t_o)$ has been generated by some physical process. The diffusion processes that are the limit of X and Z for increasing α (see Theorem 2.1) we shall call x and z, respectively.

6.2 Physical processes considered as integral processes.

The message and observation process, being physical processes, are integral processes, but for the same reason their dispersions are zero. The estimate equation (5.2.4) for integral processes was derived under the assumption that the observation was of positive definite dispersion, and so this result cannot be applied directly to X and Z. If the random disturbance y is also an integral process of positive definite dispersion, then $\dot{Z}$ is an integral process of positive definite dispersion. Now, as $\dot{Z}$ is a.c. finite there is no difference between the expectation of X conditional on past values of $\dot{Z}$ (and $Z(t_0)$) and the expectation of X conditional on past values of Z. So we can redefine the observation process to be $\dot{Z}$, and from (5.2.4) obtain an equation for the estimate of X. Unfortunately for our purpose , the dispersion of y will enter this equation as a coefficient. This term is determinate and increases with α. It would be attractive to derive filtering algorithms directly from this estimate equation, but so far the author has been unable to do this.

6.3 A procedure for obtaining practical filtering algorithms

A practical filtering algorithm can be derived by creating a sequence of approximations in the following way:

1) X and Z are approximated by the diffusion processes x and z by means of theorem 2.1;

2) The estimate $E[X(t)|Z(s), t_0 \leq s \leq t]$ is approximated by $E[X(t)|z(s), t_0 \leq s \leq t]$, which in the notation of the last chapter is $E_tX(t)$;

3) $E_t X(t)$ is approximated by $E_t x(t)$;

4) with the methods of Section 5.4 a stochastic differential equation depending on z is derived, from which an approximation to $E_t x(t)$ can be generated;

5) finally, by means of Theorem 2.1, this equation is converted into an ordinary differential equation depending on $\dot{Z}$; this is the filtering algorithm from which the final approximate estimate can be generated.

The second step in this sequence is based on the major assumption that, if two observation processes approximate each other, then the two corresponding conditional expectations of any random variable also approximate each other. So far the author has been unable to find a sufficiently general set of conditions for which this assumption is valid. The problem is analogous to the simpler one of determining the conditions for which $E[U|Y_n]$ converges to $E[U|Y]$ if Y_n converges to Y; U, Y and Y_n being scalar random variables. However, we can prove convergence for the following special cases; the proofs are given in Appendix C: if U, Y and Y_n are either (1) discrete random variables taking a finite number of values independent of n, or

(2) Gaussian variables of positive variance, and if Y_n converges to Y in probability, then $E[U|Y_n]$ converges to $E[U|Y]$ in the mean.

These examples suggest that there is a broad class of processes for which the second step of the sequence is reasonable.

An application. We now follow through this procedure in more detail for message and observation processes of the form: for $t_o \leq t \leq T$.

$$\dot{X} = f(X,t) + \rho F(X,t)y, \quad \mathrm{Cov}[X(t_o)] = O(\rho^2), \tag{6.3.1}$$

$$\dot{Z} = g(X,t) + G(t)y, \quad Z(t_o) = 0 \tag{6.3.2}$$

where ρ is a small parameter and $FG^T = 0$. We are prepared to neglect terms that are either $O(\rho^3)$ or $O(\alpha^{-\frac{1}{2}})$; that is, terms for which the square root of the second moment is of these orders.

It follows from Theorem 2.1 that X and Z can be $O(\alpha^{-\frac{1}{2}})$- approximated by x and z, where

$$dx = [f(x,t) + \rho^2 K(x,t)]dt + \rho F(x,t)dw \ , \tag{6.3.3}$$

$$dz = g(x,t)dt + G(t)dw, \quad z(t_o) = 0 \ , \tag{6.3.4}$$

where the i:th component of the vector K is

$$K_i = \sum_{jmn} \frac{\partial}{\partial x_j}(F_{im})F_{jn}A_{mn} . \qquad (6.3.5)$$

The value of $x(t_o)$ is a $O(\alpha^{-\frac{1}{2}})$- approximation to $X(t_o)$. It follows from this that $E_t x(t)$ is a $O(\alpha^{-\frac{1}{2}})$- approximation of $E_t X(t)$, for by the Holder inequality,

$$\begin{aligned} E\,|\,E_t X(t) - E_t x(t)\,|^2 &\leq E[E_t\,|X(t) - x(t)|^2] \\ &= E\,|\,X(t) - x(t)|^2 = O(\alpha^{-1}) . \end{aligned}$$

We now apply the procedure of Section 5.4 to derive a filtering algorithm, in the form of a stochastic differential equation, for $E_t x$. The only difference between the message process considered in Section 5.4 and the one considered here is that, relative to (5.4.1), (6.3.3) contains the extra term $\rho^2 K dt$. This introduces extra terms into the filtering algorithms. One possible filtering algorithm, which corresponds to (5.4.16), is

$$\dot{x}_0 = f(x_0, t) , \quad x_0(t_o) = EX(t_o) , \qquad (6.3.6)$$

$$\delta\dot{v} = f_{x0}\delta v + \delta v f_{x0}^T + \rho^2 F_0 F_0^T , \quad \delta v(t_o) = \mathrm{Cov}[X(t_o)], \qquad (6.3.7)$$

and

$$\begin{aligned} d\delta m = f_{x0}\delta m dt + \tfrac{1}{2}\langle f_{xx0}, \delta v\rangle dt + \rho^2 K_0 dt \\ + \delta v g_{x0}^T (GG^T)^{-1}(dz - g_0 dt), \quad \delta m(t_o) = 0 \end{aligned} \qquad (6.3.8)$$

where, as before, the suffix 0 indicates evaluations at (x_0, t). $x_0 + \delta m$ is a $O(\rho^3)$- approximation to $E_t x(t)$.

It only remains now to carry out the fifth step of the procedure: the conversion of (6.3.6), (6.3.7) and (6.3.8) into a set of ordinary differential equations. For the particular filtering algorithm we have chosen this is relatively straightforward, for x_0 and δv are determinate functions and so the equation for δm is of the form

$$d\delta m = C_1(t)\delta m dt + C_2(t)dt + C_3(t)(g(x,t)dt + G(t)dw), \tag{6.3.9}$$

in which we have expressed dz explicitly. As we see from (6.3.2) and (6.3.9) δm and x form a joint diffusion process to which we can apply Theorem 2.1 and obtain approximating physical processes. A $O(\alpha^{-\frac{1}{2}})$- approximation to x is of course X; the corresponding $O(\alpha^{-\frac{1}{2}})$- approximation to δm is given by

$$d\delta M = C_1(t)\delta M dt + C_2(t)dt + C_3(t)(g(X,t)dt + G(t)ydt) ,$$

$$\delta M(t_o) = 0 \quad , \tag{6.3.10}$$

which happens to be of the same form as (6.3.9), as the dispersion of δM is determinate. But the term in the brackets is $\dot{Z}dt$ and so we can write

$$\begin{aligned}\dot{\delta M} = f_{x0}\delta M + \tfrac{1}{2}\langle f_{xx0},\delta v\rangle + \rho^2 K_0 \\ + \delta v g_{x0}^T(GG^T)^{-1}(\dot{Z} - g_0), \quad \delta M(t_o) = 0 \ .\end{aligned} \tag{6.3.11}$$

By combining all five of the steps of the procedure, we see that $x_o + \delta M$ is a sufficiently accurate approximation of the estimate of X and that (6.3.6), (6.3.7) and (6.3.11) together form a filtering algorithm that satisfies the conditions we posed in Section 6.1.

The general case. If the message and observation processes are of the more general form (6.1.2) and (6.1.3), or if we require more accurate approximations, the stochastic differential equation generating the approximation to $E_t x$ is generally nonlinear and contains terms depending on z as well as x. However, the processes generated by these equations together with x and z form a joint diffusion process. So we can still apply Theorem 2.1; the resulting ordinary differential equation, in which the forcing functions are $\dot{Z}$ and Z, is the required filtering algorithm. In converting the stochastic equation into an ordinary equation correction terms are generally introduced. This did not occur in the above example as the equations involved are linear in the random variables.

CHAPTER 7

SUMMARY AND CONCLUSIONS

It has been shown in Theorem 2.1 that with certain assumptions a physical process can be approximated by a diffusion process of the same dimension. The most interesting assumption is, roughly, that the matrix integral

$$\frac{1}{t}\int_0^t \int_0^s E[y(s)y(r)^T]drds$$

of the vector random disturbance $y(t)$ of the physical process is approximately equal to a constant matrix A for values of t comparable to the time-constants of the process. This matrix, which has been called the "characteristic matrix" of $y(t)$, plays an important role; it is this matrix rather than the matrix generalisation of the "mean power per unit bandwidth" that characterises the random disturbance in the probability distributions of the diffusion approximation. Moreover, if the characteristic matrix is symmetric, the differential equation obtained from that of a physical process by replacing the disturbances by "white noise" can be precisely interpreted as the Stratonovich stochastic differential equation of a diffusion approximation.

In [20] Stratonovich has demonstrated that the probability density of a physical process with stationary random disturbances of short correlation time approximately satisfies an equation of Fokker-Planck type. Theorem 2.1 is based on different assumptions but is closely related to this result; in particular Stratonovich's result implies that some such term as the characteristic matrix has to be defined. In its formulation Theorem 2.1 is more in line with the theorems on the relation between ordinary and stochastic differential equations given by Wong and Zakai in [21,22]. In these theorems, however, the models chosen for the random disturbances happen to possess symmetric characteristic matrices; in this situation the significance of the characteristic matrix is lost and the proofs in [21,22] do not require its definition.

The somewhat artificial assumption that the disturbances are Gaussian was made in Theorem 2.1 to simplify its proof. Stratonovich in [20] and Wong and Zakai in [21] did not have to make this assumption. This suggests that it is unnecessary and that Theorem 2.1 might also hold for general non-Gaussian disturbances; this would make the result much more useful.

In the simulation of physical processes the symmetry of the characteristic matrix is important. To simulate a process, an analogue computer can be programmed directly from the equations of the process if either the random disturbances of the process are additive or the characteristic matrices of the disturbances and the computer noise source are congruent to each other. This second condition is satisfied if both the characteristic matrices are symmetric, and this is the form of the condition that is of practical interest. If these conditions are not satisfied it is generally necessary to include extra terms in the equations used for programming the computer. Similar problems arise when simulating processes on digital computers; these have been mentioned only briefly in this thesis and have yet to be investigated.

A vector random disturbance has a symmetric characteristic matrix if its components are independent or if its matrix correlation function is an even function of time. Other conditions, for random disturbances that are linear diffusion processes or piecewise constant Gaussian Markov processes, have been derived in Chapter 4.

By approximating physical processes by diffusion processes we can apply to the problem of filtering physical processes some of the results of the theory of filtering diffusion processes, which has been studied by several authors for different formulations. Many of their results can be summarised in one general formula. This is an Ito stochastic differential equation satisfied by the estimate of the message process, which we have taken to be its mean conditional on the past values of the observation process. An apparently novel derivation of this estimate equation, based on the implicit properties of conditional moments, has been given in Section 5.2. There are some minor discrepancies between this equation and the corresponding equation of Stratonovich's [1]. Similar stochastic differential equations can be derived for other conditional moments of the message process, but it is not usually possible to choose from these a set that make up a diffusion filtering algorithm; that is, a finite set of stochastic differential equations describing the estimate, in which the observation process is the only forcing function. However, filtering algorithms giving approximations to the estimate can be derived by perturbation methods; this has been demonstrated in Section 5.4.

We can carry over these results to physical processes by making a series of approximations. The procedure has been discussed in Chapter 6. The steps that have to be taken, and the points that have to be considered, are illustrated in Figure 7.1. The estimate equation given in Section 5.2 applies equally well to message and observation process that are just integral processes; that is, processes that are sums of indefinite ordinary and stochastic integrals. Physical processes are integral processes, and it would be attractive to derive the physical filtering algorithm for these processes directly from the corresponding estimate equation, thus avoiding the chain of approximations required in the derivation we have given. There are certain difficulties to this approach, but it should be a point of further research.

Parts of our derivation of the physical filtering algorithm have not been fully justified; in particular the assumption has been made that, if the physical observation process is approximated by a diffusion observation process, the corresponding conditional means of the message process approximate each other. This is certainly true for some special cases; it would be interesting to know whether it is true more generally.

Altogether, though, the derivation is on a sufficiently sound mathematical basis to make it worthwhile to test the resulting filtering algorithms experimentally. It is only then that their usefulness can be finally judged.

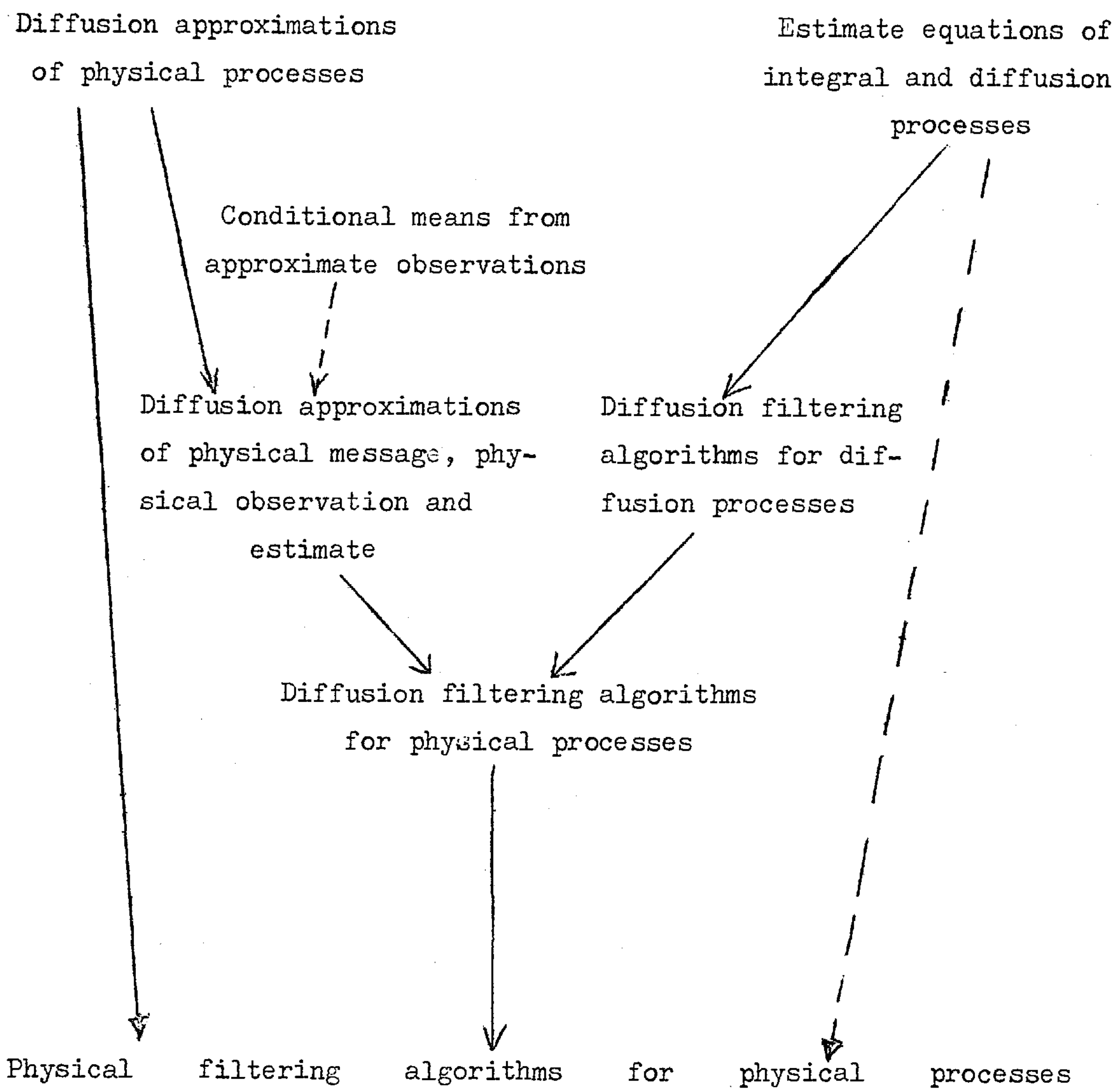

Figure 7.1 The derivation of physical filtering algorithms

REFERENCES

1. R.L. Stratonovich, Conditional Markov processes. Theory of Probability and its Applications, 5 (2), pp.156-178, 1960 (English Trans.).
2. H.J. Kushner, On the differential equations satisfied by conditional probability densities of Markov processes, with applications. Siam J. on Control, 2, No.1, pp.106-119, 1964.
3. R.S. Bucy, Nonlinear filter theory, IEEE Trans. on Automatic Control, AC-10, p.198, April, 1965.
4. W. M. Wonham, Advances in nonlinear filtering, Lecture, M.I.T., Summer Session, 1965.
5. W. M. Wonham, Some applications of stochastic differential equations to optimal nonlinear filtering, J.Siam, series A: Control, 2, No.3, pp.347-369, 1964.
6. H.J. Kushner, Dynamical equations for optimum non-linear filtering, Private communication, Dec.1965.
7. R.L. Stratonovich, A new form of representation of stochastic integrals and equations (in Russian), Vestnik Mosk.Univ. Series 1, Math and Mech.I, pp.3-12, Jan.1964.

8. Yu.N. Blagoveshchenskii, Diffusion processes depending on a small parameter. Theory of Probability and its Applications, 1, No.2, pp.130-140, 1962 (translation).

9. M.S. Bartlett, An Introduction to Stochastic Processes, University Press, Cambridge, 1962.

10. R.L. Stratonovich, On the theory of optimal non-linear filtering of random functions. Theory of Probability and its Applications, 4, pp.223-225, 1959 (English translation).

11. R.E. Kalman, R.S. Bucy. New results in linear filtering and prediction. J.Basic Engineering (ASME Transactions) 83D, pp.95-108, March 1961.

12. J.L. Doob, Stochastic Processes, Wiley, London, 1953.

13. E.B. Dynkin, Markov Processes, Vol.I, Springer Verlag, Berlin, 1965.

14. Michel Loeve, Probability Theory, Van Nostrand, Princeton, second edition, 1960.

15. Kiyosi Ito, Stochastic integrals, Proc.Imp.Acad. Tokyo, 20, pp.519-524, 1944.

16. Kiyosi Ito, Stochastic differential equations on a differentiable manifold, Nagoya Math.J., 1, pp.35-47, 1950.

17. Kiyosi Ito, On stochastic differential equations, Memoirs Amer.Math.Soc. 4, 1951.

18. Solomon Lefschetz, Differential Equations: Geometric Theory, Wiley, New York, 1963.

19. K.J. Astrom, Private communication.

20. R.L. Stratonovich, Topics in the Theory of Random Noise, Vol.I, Gordon and Breach, New York, 1963.

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

22. E. Wong, M. Zakai, On the relation between ordinary and stochastic differential equations, Int.J.Engng. Sce. 3, pp.213-229, 1965.

23. K.J. Astrom, On a first order stochastic differential equation, International J.of Control, 1 (4), April, 1965.

24. R. Bellman, Introduction to Matrix Analysis, McGraw Hill, New York, 1960.

25. F.R. Gantmacher, Matrix Theory, Vols.I,II, Chelsea Publishing, New York, 1960.

26. Nelson Wax (Editor), Selected Papers on Noise and Stochastic processes, Dover Publications, New York, 1954.

27. S.T. Ariaratnam, P.W.U. Graefe, Linear systems with stochastic coefficients II, International J. of Control, 2 (2), 1965.

APPENDIX A: STOCHASTIC DIFFERENCE EQUATIONS AND THE APPROXIMATION OF A PHYSICAL PROCESS

Stochastic differential equations. The Ito and Stratonovich stochastic differential equations can be interpreted as the limits of finite difference equations. To illustrate this we summarise a simple example given by Kushner [2, Appendix 1]. Suppose $x(t) = e^{w(t)}$, where $w(t)$ is a scalar Wiener process. Then over an interval of time $[t, t+\delta t]$ the increment of $x(t)$ is

$$\delta x = x(t+\delta t) - x(t) = e^{w}(\delta w + \tfrac{1}{2}\delta w^2 + \ldots)$$

$$= x(\delta w + \tfrac{1}{2}\delta w^2 + \ldots) . \qquad (A.1)$$

Note the order of magnitude of the terms in (A.1): the means and standard deviations of δw^3 and the higher powers are $o(\delta t)$; the mean of δw^2 is δt and its standard deviation is $\sqrt{3}\delta t$; and the mean of δw is zero and its standard deviation $\sqrt{\delta t}$. This suggests that the solution of the difference equation

$$\delta w = x(\delta w + \tfrac{1}{2}\delta w^2) , \qquad (A.2)$$

which is (A.1) with the $o(\delta t)$ terms neglected, might converge in the mean to $x(t)$ as $\delta t \longrightarrow 0$. This is indeed so. Moreover, on integration, the effect of the random part of δw^2 accumulated over all the intervals vanishes as $\delta t \longrightarrow 0$, and so the solution of the difference equation

$$\delta x = x\delta w + \tfrac{1}{2}x\delta t \qquad (A.3)$$

also converges to $x(t)$.

The Ito equation of $x(t)$, obtained by differentiating $x(t)$ by Ito's formula, is

$$dx = xdw + \tfrac{1}{2}xdt \ , \qquad (A.4)$$

which can be considered to be the limiting form of (A.3).

The backward difference

$$\delta' x = x(t) - x(t-\delta t) = c^{w(t)}\ (1 - e^{-\delta' w})$$

$$= x(\delta' w - \tfrac{1}{2}\delta' w^2 + \ldots) \ ,$$

and so the central difference

$$\bar{\delta}x = \tfrac{1}{2}(\delta x + \delta' x) = x(\bar{\delta}w + o(\delta t)) \ .$$

As for the forward difference equation, the solution of the central difference equation

$$\bar{\delta}x = x\bar{\delta}w \qquad (A.5)$$

converges to $x(t)$ as $\delta t \longrightarrow 0$; this equation matches the Stratonovich equation of x, which is

$$\bar{d}x = x\bar{d}w \ . \qquad (A.6)$$

Approximation of a physical process. Suppose $X(t)$ is the physical process given by

$$\dot{X} = Xy \ , \ X(0) = 0 \ . \tag{A.7}$$

y is a continuous disturbance approximating white noise; that is, the integral of y approximates in the mean a Wiener process $w(t)$. As y is continuous we can integrate (A.7) in the usual way to get the solution

$$X(t) = e^{\int_0^t y(s)ds} .$$

If we make the assumption that y is also Gaussian, then it is not difficult to show that $X(t)$ is approximated in the mean by the diffusion process $x(t) = e^{w(t)}$. For this particular example, and for other scalar examples, if we multiply the differential equation of X by dt and replace ydt by $\bar{d}w$, we obtain the Stratonovich equation of the diffusion approximation x. However, the differential equation of a vector physical process and the Stratonovich equation of its diffusion approximation do not in general have this equivalence (see Chapter 3).

The simulation of X (or its diffusion approximation) on a digital computer. Suppose we want to simulate X (that is, generate a process with similar statistical characteristics) on a digital computer, on which we can generate a sequence of independent Gaussian random variables R_n, each of zero mean and unit variance. There are several ways of doing this. First consider the forward difference formula

$$x_{n+1} - x_n = x_n R_n \sqrt{h} + \tfrac{1}{2} x_n h \ . \qquad (A.8)$$

Now $\sum_{r=0}^{n} R_r \sqrt{h}$ behaves exactly like a Wiener process $w'(nh)$, and it follows from our previous remarks about (A.3) that, as the step length h is made smaller, the solution x_n converges to a process of the form $x^{w'(nh)}$, which is a simulation of x and therefore of X. This is one way of simulating X. Another way is to first simulate X by a continuous process X', which is a known function of the variables R_n, and then generate approximations to this on the computer. For instance, consider

$$\dot{X}' = X' y_{h_o} \qquad (A.9)$$

where y_{h_o} is a step process which takes the value $\frac{R_n}{\sqrt{h_o}}$ in the interval $[nh_o, (n+1)h_o]$. Then X' simulates X, the error depending on the size of h_o (see Chapter 3). For any set of realizations of the R_n, (A.9) is an ordinary differential equation which can be integrated by the normal integration formulas to give an approximation to X'. The accuracy of this approximation depends on the ratio of the integration steplength to h_o and on the order of the integration formula. If the step length is taken to be h_o, we cannot use the first-order forward-difference formula, which is (A.8) without the term $\frac{1}{2}x_n h_o$, as its solution would not converge, for decreasing h_o, to that of (A.8). It seems reasonable to suppose, however, that second and higher order formulas, with step lengths h_o, would give a solution converging to X'.

APPENDIX B: ITERATED STOCHASTIC INTEGRALS

Lemma B1: This is a slight extension of a theorem of Doob's [12] p. 430. Let $g(s,r)$ be a bounded measurable matrix function of s and r in $[0,t]$ Let $Q(r)$ be an admissible matrix random process such that $E|Q(r)|^2$ is bounded in $[0, T]$. Then the processes

$$I(u) = \int_u^t g(s,u)\, Q(u)ds \tag{B.1}$$

$$J(u) = \int_0^u g(u,r)\, Q(r)dw(r) \tag{B.2}$$

are admissible processes in u and have bounded second moments, and

$$\int_0^t \int_r^t g(s,r)Q(r)dsdw(r) = \int_0^t \int_0^s g(s,r)Q(r)dw(r)ds \quad \text{(a.c)} \tag{B.3}$$

Proof. We can suppose without loss of generality that $g(s,r) = 0$ for $r > s$. All the integrations can then be taken over the period $[0,t]$. $g(s,u)$, being measurable and bounded, is integrable, so $\int_0^t g(s,u)ds$ is measurable with respect to u. $Q(u)$ is admissible; that is, it is measurable with respect to u and γ (the basic probability parameter) and for fixed u it is Borel-measurable in

$[w(r),\ o \leq r \leq u]$. So the product of these two terms, which is $I(u)$, is also admissible. That $E|I(u)|^2$ is bounded follows from the boundeness of $E|Q(u)|^2$. For fixed u, $J(u)$ is a well-defined stochastic integral which can be taken to be Borel-measurable in $[w(r),\ o \leq r \leq u]$. Moreover $E|J(u)|^2$ is bounded. To show that $J(u)$ is admissible we have to show it is a measurable function u and γ. As $g(u,r)$ is measurable and bounded it can be expressed as the limit, for almost all u and r, of a sequence of finite sums

$$\sum_i f_i^n(u)\, h_i^n(r)$$

where $f_i^n(u)$ and $h_i^n(r)$ are measurable bounded functions. Let

$$J^n(u) = \sum_i f_i^n(u) \int_o^t h_i^n(r)\, Q(r) dw(r)$$

$J^n(u)$ is clearly measurable in u and γ. Now

$$E|J(u) - J^n(u)|^2$$

$$\leq \int_o^t |g(u,r) - \sum_i f^n(u) h_i^n(r)|^2 E|Q(r)|^2 dr,$$

$$\longrightarrow 0 \qquad (8.5)$$

for almost all u as $n \longrightarrow \infty$.

This implies that $J(u)$ is also measurable in u and so it is admissible.

The left side of (B.3) is $\int_0^t I(r)dw(r)$ and the right side is $\int_0^t J(s)ds$. As $I(r)$ and $J(s)$ are admissible with bounded second moments these integrals are well defined. Let

$$I^n(r) = \sum_i \left(\int_0^t f_i^n(s)ds\right) h_i^n(r)\, Q(r).$$

Clearly

$$\int_0^t I^n(r)dw(r) = \int_0^t J^n(s)ds. \qquad \text{(B.6)}$$

By the inequality (2.3.10)

$$E\left|\int_0^t J(s)ds - \int_0^t J^n(s)ds\right|^2$$

$$\leq t\int_0^t E|J(s) - J^n(s)|^2 ds$$

which by (B.5) tends to zero as $n \longrightarrow \infty$; that is, $\int_0^t J(s)ds = \underset{n \to \infty}{\text{l.i.m}} \int_0^t J^n(s)ds$.

Also

$$E\left|\int_0^t I(r)dw(r) - \int_0^t I^n(r)dw(r)\right|^2$$

$$= \int_0^t E\,|\, I(r) - I^n(r)|^2 dr$$

$$\leq \int_0^t \Big| \int_0^t [\, g(s,r) - \sum_i f_i^n(s) h_i^n(r) \,] ds \Big|^2 \, E|Q(r)|^2 dr$$

$$\longrightarrow 0 \quad \text{as} \quad n \longrightarrow \infty .$$

So $\int_0^t I(r)dwr = \underset{n \to \infty}{\text{l.i.m.}} \int_0^t I^n(r)dw(r)$. As equality in the mean implies equality with probability 1, (B.3) follows from (B.6).

APPENDIC C: THE CONVERGENCE OF CONDITIONAL EXPECTATIONS

Lemma C1: Discrete variables: If U, Y and Y_n are discrete random variables taking only a finite number of values independent of n, and if Y_n converges to Y in probability as n increases, then $E[U|Y_n]$ converges to $E[U|Y]$ in the mean.

Proof. As $E[U|Y_n]$ is bounded uniformly for n by $\max|U|$ we need only prove convergence in probability. Let p_{ijk} be the probability that U, Y and Y_n take the values i, j and k respectively. As the values Y and Y_n take are independent of n, convergence of Y_n to Y in probability implies that p_{ijk} vanishes if $j \neq k$. So considering only those values of the conditional expectation that occur with non-vanishing probability; that is, where $Y_n = Y = h$ and $\sum_i p_{ihh}$ does not vanish, then

$$E[U|Y{=}h] - E[U|Y_n{=}h]$$

$$= \frac{\sum_{ik} i\, p_{ihk}}{\sum_{ik} p_{ihk}} - \frac{\sum_{ij} i\, p_{ijh}}{\sum_{ij} p_{ijh}}$$

$$= \frac{\sum_{ijkm} i(p_{ihk}p_{mjh} - p_{ijh}p_{mhk})}{\sum_{ijkm} p_{ihk}p_{mjh}} .$$

The denominator in this expression does not vanish as n increases, but the numerator does. So those values of the conditional expectations which occur with non-vanishing probability converge. The result we require follows.

<u>Lemma C2</u>: Gaussian variables. If U, Y and Y_n are Gaussian variables of positive variance and Y_n converges to Y in probability for increasing n then $E[U|Y_n]$ converges to $E[U|Y]$ in the mean.

Proof. For Gaussian variables convergence in probability and in the r:th mean for any r are equivalent, so we need only prove convergence in the first mean. As U, Y and Y_n are Gaussian ,

$$E[U|Y_n] = EU + VU(VY_n)^{-1}(Y_n - EY_n) ,$$

$$E[U|Y] = EU + VU(VY)^{-1}(Y - EY)$$

where V is the variance operator. Thus

$$E\left|E[U|Y_n] - E[U|Y]\right| \leq VU(VY)^{-1}(E|Y_n-Y| + |EY_n-EY|)$$

$$+ \delta.VU(E|Y_n| + |EY_n|)$$

where $(VY_n)^{-1} = (VY)^{-1} + \delta$. As Y_n and Y are Gaussian, $E|Y_n - Y|$, $|EY_n-EY|$ and δ all vanish for increasing n, so the right member of the inequality vanishes. Thus $E[U|Y_n]$ converges to $E[U|Y]$ in the first mean.