Discrete Approximations for Markov-Chain Filters

by

Nigel John Newton

A thesis submitted for the Degree of
Doctor of Philosophy of the University of London
and for the
Diploma of Membership of
the Imperial College of Science and Technology

September 1983

Department of Electrical Engineering
Imperial College of Science and Technology
University of London

Abstract

It is well-known that the conditional distribution of a Markov-chain signal process based on observations of a "signal-plus-white-noise" type obeys a vector stochastic differential equation. The usual Bayes estimates of functions of the signal can be found from this distribution, pointwise in time. Unfortunately, the differential equation mentioned above has, in general, no closed form solution, and so it is necessary to use some form of approximation scheme to produce a practical filter.

We consider discrete approximation schemes based on regular partitions of a finite time-interval and show that, under the condition that the distribution of the observations-process is Wiener-measure (in R^d), the discrete-observations-conditional distribution of the approximation error (suitably normalised) can at best converge to a normal distribution with zero mean and given covariance matrix. For the more practical case, where the distribution of the observations-process is only absolutely continuous with respect to Wiener-measure (i.e. the signal-plus-white-noise case), a similar result holds.

We demonstrate the existence of schemes that are efficient, in that their normalised error sequences converge to this best limit, and schemes that possess the highest possible order of convergence (i.e. first-order) but that are not efficient. The latter methods have limit conditional distributions with non-zero means.

We investigate various practical approximation schemes, in particular a class of schemes based on a "Taylor-series type" expansion of the conditional distribution of the signal about the points of the partition.

A parallel set of conditional-distribution-limit-results apply to the normalised errors in the approximation of the Bayes estimates of functions of the signal process when these approximations are calculated from the corresponding approximate distributions of the signal-process. It is these approximate estimates that form the "output" of any practical filter.

Contents

Abstract . . . 3

Notation . . . 6

Acknowledgements . . . 9

Chapter 1 Introduction

1.1 Optimal filtering . . . 13
1.2 Filters for Markov chains . . . 14
Theorem 1.1 . . . 15
1.3 Discrete solutions of stochastic differential equations . . . 18
1.4 On the results of chapters 2–4 . . . 21

Chapter 2 Some Tools for Handling Discrete Approximation Schemes

2.1 Introduction . . . 23
2.2 A special notation used in Taylor expansions of the solutions of bilinear s.d.e.s . . . 23
Definition 2.1 . . . 24
2.3 A Taylor expansion for representing the solutions of bilinear s.d.e.s . . . 25
Theorem 2.1 . . . 25
2.4 A theorem for bounding the solutions of stochastic difference equations . . . 28
Definition 2.2 . . . 28
Lemma 2.1 . . . 29
Theorem 2.2 . . . 32
2.5 A family of discrete approximation schemes . . . 34
Definition 2.3 . . . 35
Theorem 2.3 . . . 35
2.6 Generalisation of results . . . 39

Chapter 3 The Optimal Asymptotic Behaviour of Approximation Schemes

3.1 Introduction . . . 41
3.2 The equation for the "un-normalised" density . . . 42
3.3 An absolutely continuous transformation of measure . . . 43
Lemma 3.1 . . . 44
3.4 The maximum order of convergence of discrete methods . . . 47
Definition 3.1 . . . 47
Lemma 3.2 . . . 47
Lemma 3.3 . . . 49
Lemma 3.4 . . . 50
3.5 An optimal limit-distribution result for normalised error sequences . . . 52
Theorem 3.1 . . . 53
3.6 Limit-distributions for normalised errors in approximations of Bayes estimates . . . 64
Theorem 3.2 . . . 65

Chapter 4 Some Approximation Schemes

4.1 Introduction . . . 71
4.2 Asymptotic properties of approximation schemes . . . 71

Definition 4.1 71
Lemma 4.1 73
4.3 A paradigmatic efficient scheme 75
Lemma 4.2 77
Definition 4.2 78
Theorem 4.1 78
4.4 A comparison of approximation schemes 85

Chapter 5 Further Considerations
5.1 Properties on a finite partition 92
5.2 Computation 92
5.3 Irregular partitions 94

References 96

Notation

Abbreviations:

a.s.	almost surely
o.d.e.	ordinary differential equation
s.d.e.	stochastic differential equation
w.r.t.	with respect to

Set theoretic symbols:

$\emptyset$	null set
$\in$	"belongs to"
$\subseteq$	"is contained in"
$\cup$	union
$\cap$	intersection
$\exists$	"there exists"
$\forall$	"for all"

Matrix and vector symbols:

I	identity matrix
$b^T, (B^T)$	transpose of a vector (matrix)
$\langle x\,,\, y\rangle$	inner product of two vectors
$\|x\|$	norm of a vector or matrix (Euclidean unless otherwise stated)
$\mathrm{diag}(x_1, \ldots, x_m)$	a diagonal matrix with elements $x_1, \ldots, x_m$
$x_i, (x^{(i)})$	the i^{th} element of a vector; the second form is used where a subscript has another meaning, for instance the i^{th} element of a process $x_t^{(i)}$.

Probability theoretic symbols:

$(\Omega, \mathcal{F}, P)$	a probability space
ω	the generic element of Ω
$\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{F}, \mathcal{G}, \mathcal{P}$	sigma-fields of subsets of Ω
$\sigma(\cdot)$	the sigma-field generated by a system of sets or a random variable
$\mathcal{F}^X$	$\sigma(X)$; where X is a random variable or process
$\mathcal{F}_t^X$	$\sigma(X_S\,;\, s \leq t)$; where X is a random process
P, P_1, P_2	probability measures
E, E_1, E_2	mathematical expectation with respect to P, P_1, P_2
$P(A \mid \mathcal{F})$	$\mathcal{F}$-conditional probability of an event A
$E(X \mid \mathcal{F})$	$\mathcal{F}$-conditional mean of a random variable X

$P_1 \ll P_2$	measure P_1 is *absolutely continuous* with respect to P_2
$P_1 \sim P_2$	P_1 and P_2 are *mutually absolutely continuous*
(a.s.)(P)	*almost surely* with respect to P; (P) is omitted where the measure is obvious
$L_p((\Omega, \mathcal{F}, P))$	space of random variables on Ω whose p^{th} powers are integrable
$\Rightarrow$	weak convergence (convergence in distribution)
$N(m, V)$	the multivariate normal distribution with mean m and covariance V
$n(0, V)$	the multivariate normal density with mean m and covariance V
I_A	the *indicator function* of a set A

Special notation:

$\mathcal{P}_N$	a partition sigma-field (see (3.2.4))
$\hat{q}$	an approximation to the random variable q
$\bar{q}$	$E_2(q \mid \mathcal{P}_N)$ (see (3.4.23))
$\tilde{q}$	$q - E_2(q \mid \mathcal{P}_N)$ (see (3.4.23))
$O_{(\infty)}(\cdot)$	an order of convergence (see definition 2.2)
$O_p(\cdot)$	an order of convergence (see definition 2.2)

M, α, β, $l(\alpha)$, $n(\alpha)$, $-\alpha$, $\alpha-$, ν, $l(S)$, $B(S)$, $A_\alpha(F, G_1, \ldots, G_d)$, $I_\alpha^W(s,t)$, $I_\alpha^W(X,s,t)$, $I_\alpha^W(nh,(n+1)h)$, $I_\alpha^W(q,nh,(n+1)h)$ — see definition 2.1

Other symbols:

$:=$	"defined equal to"
$\approx$	"approximately equal to"
R^d	the space of *real* d-vectors
$\mathcal{N}$	the set of natural numbers

Acknowledgements

I would like to express my many thanks to all who have helped and encouraged me in the work which led to this thesis: most of all, to Dr. J. M. C. Clark my supervisor for his invaluable help and guidance in many discussions during the course of the work, and for the many comments and suggestions he made during the writing of the manuscript. I am also most grateful to Dr. J. E. Marshall* for his understanding and encouragement of my wish to study this mathematical branch of engineering during my undergraduate days.

I am greatly indebted to the Science and Engineering Research Council for their financial support during my period as a student.

Finally I would like to thank Mr. John Peverett for the painstaking work of typing the final version for formatting with TEX.

* Department of Mathematics, University of Bath.

To my parents

Chapter 1

Introduction

1.1 Optimal filtering.

This thesis is concerned with the numerical realisation of a class of optimal filters that could, in principle, be programmed on a digital computer or microprocessor system. By "filter" we mean a recipe by which we can estimate a time-varying parameter or signal, from past and present observations of some related, noise-perturbed signal. For example, we might wish to estimate the "state" of a dynamical system from noise-perturbed observations of the output.

The simplest example of an optimal filter is the Kalman filter (see [AR1]) for linear systems driven by Gaussian white noise.

Much literature exists on the more general theory of filtering. So we shall not go into the details here; the main tools used in that theory and in this thesis are results of *probability theory* and *stochastic calculus*. The reader should be familiar with such notions and results from probability theory as: *probability spaces*, probability as a normalised σ-additive set function, *random variables* as measurable functions, *mathematical expectation* and *conditional expectation, the Radon-Nikodym theorem, Fubini's theorem, almost sure convergence, L_p-convergence, convegence in probability, weak convergence, the Borel-Cantelli lemmas, Hölders's inequality, distributions, moments,* and *the central limit theorem.* There are many good texts in probabilty theory, for example Billingsley [BI1] and Kingman and Taylor [KT1].

Filtering theory for continuous-time signals draws on the theory of continuous-time stochastic processes and the integral calculus appropriate to such processes. The reader should be familiar with such concepts and results as: *stochastic processes* and their *finite-dimensional distributions*, the *Markov property, adapted processes, martingales, Brownian motion, the Wiener integral, stochastic differential equations*, the existence and uniqueness of their solutions, *Itô's rule and the Girsanov theorem.*

A good introductory text to stochastic calculus with practical motivation is Arnold [AR1]. For a more complete (and abstract) handling of the subject see, for example, Ikeda and Watanabe [IW1], Gikhman and Skorokhod [GS1] or Stroock and Varadhan [SV1].

Finally, for a thorough treatment of optimal filtering theory, the reader should consult Liptser and Shiryayev [LS1] or Kallianpur [KA1]. Both these books include sizeable sections on the underlying results of stochastic calculus.

Filtering theory is largely concerned with a recursive formula for the *conditional a-posteriori statistical distribution* for the signal based on prior observations; the usual *Bayes estimates* can be found from this distribution (see for example Larsen [LA1]). For a finite-state signal the "output" of the recursive formula would be a vector of conditional probabilities, and for a continuous-state signal the output might be a conditional density (assuming a density exists). For a discrete-time system the recursive formula would be a difference equation, but for a continuous-time signal it would be a differential or integral equation. (Since we idealise the integral of physical "wide-band" noise by a *Brownian motion* process, these differential equations would be *stochastic differential equations* (s.d.e.s) and cannot be interpreted in the usual Riemann-Stieltjes sense.)

The simplest and earliest example of an optimal filter of this type is the well-known *Kalman filter.* This is a recursive formula for the conditional mean and conditional covariance of the state of a linear system driven by wide-band Gaussian noise, based on observations that are a linear combination of the signal and independent Gaussian wide-band noise.

Versions of the Kalman filtering equations exist for discrete-time and continuous-time systems. The equations are directly applicable for continuous-state signals because they are finite-dimensional; in fact the linearity of the problem ensures that the a-posteriori distribution is Gaussian and can be parameterised by a finite number of statistics (the mean and the covariance).

One of the practical problems with optimal filtering for arbitrary *Markov signals* is that the a-posteriori distributions cannot always be characterised by a finite set of statistics, and so the filtering equations are infinite-dimensional. There is a fair amount of work on special cases for which finite-dimensional filters exist (see for example Beneš [BE1], Brockett [BR1]), but we shall not be concerned with that here.

In this thesis we restrict our attention to the filtering equations appropriate to *Markov-chain signals* because Markov diffusion processes can usually be approximated with arbitrary accuracy by such chains (see Kushner [KU1]).

1.2 Filters for Markov chains.

Suppose we have a signal (s_t) that we wish to estimate from observations of a "signal-plus-wide-band-noise" type; i.e. we observe (Y_t), which is given by

$$Y_t = h(s_t) + n_t; \qquad h(s_t), n_t \in R^d. \tag{1.2.1}$$

h is a bounded, nonlinear function and (n_t) is wide-band Gaussian noise, statistically independent of (s_t). (By "wide-band", we mean that the bandwidth of the noise is much greater than the bandwidth of the signal (s_t).)

The optimal filter in this case would depend in a very complicated way on the nature of (s_t) and (n_t); so we examine first the optimal filter for the case where the signal is a *finite-state Markov chain* and the noise is idealised white noise. We note that there are versions of this optimal filter that represent "nearly optimal" filters for the real signal (s_t) and real noise process (n_t), provided that these can be approximated by a finite-state Markov chain and idealised white noise respectively. These notions are made clearer below.

We represent idealised white noise in its integral form; such a process is called a *Brownian motion* (β_t) [AR1]

$$\beta_t \approx \int_0^t n_s ds. \tag{1.2.2}$$

The optimal estimates of the signal are found by a *Bayes estimation* technique; we use our knowledge of the physical nature of the signal to decide on an *a-priori probability distribution*, then at each time we use the observations up to that time to form an *a-posteriori distribution* for the signal. We use this distribution to find estimates of the signal that optimise various error criteria, such as *least squares.*

An optimal filter for the idealised case is given by the following theorem.

Theorem 1.1. *Let* $(X_t,\ t \in [0, T])$ *be a homogeneous finite-state Markov-chain process defined on some probability space* $(\Omega, \mathcal{F}, P_1)$*, with generator* A *and a-priori distribution* p_0*:*

$$X_t \in \mathcal{S} := \{a_1, a_2, \ldots, a_m\} \qquad t \in [0, T],$$
$$A \in R^{m \times m},$$
$$p_0 := \begin{bmatrix} P_1(X_0 = a_1) \\ \cdot \\ \cdot \\ \cdot \\ P_1(X_0 = a_m) \end{bmatrix}.$$

Let $(W_t,\ t \in [0, T])$ *be the following "observations" process:*

$$W_t := \int_0^t h(X_s) ds + \beta_t, \tag{1.2.3}$$

where h *is a bounded nonlinear function on* $\mathcal{S}$*;*

$$h = \begin{bmatrix} h_1 \\ h_2 \\ \cdot \\ \cdot \\ \cdot \\ h_d \end{bmatrix} : \mathcal{S} \to R^d$$

and (β_t) *is a d-dimensional Brownian motion process, independent of* (X_t)*.*

(i) The a-posteriori distribution for X_t,

$$p_t := \begin{bmatrix} P_1(X_t = a_1 \mid W_s,\ s \in [0,T]) \\ \cdot \\ \cdot \\ \cdot \\ P_1(X_t = a_m \mid W_s,\ s \in [0,T]) \end{bmatrix},$$

is the unique strong solution of the following s.d.e.:

$$p_t = p_0 + \int_0^t A p_s d s + \sum_{i=1}^{d} \int_0^t (B_i - (b_i^T p_s) I) p_s (d W_s^{(i)} - (b_i^T p_s) d s) \qquad t \in [0,T], \tag{1.2.4}$$

where

$$B_i = \operatorname{diag}\{h_i(a_1), h_i(a_2), \ldots, h_i(a_m)\} \qquad i = 1,2,\ldots,d;$$
$$b_i^T = [h_i(a_1)\ h_i(a_2) \ldots\ h_i(a_m)] \qquad i = 1,2,\ldots,d.$$

(ii) If we define an "un-normalised" a-posteriori distribution q_t for X_t by

$$q_t := \exp\Big(\int_0^t \sum_{j=1}^{m} p_s^{(j)} h(a_j)^T d W_s - \frac{1}{2} \int_0^t \| \sum_{j=1}^{m} p_s^{(j)} h(a_j) \|^2 d s \Big) p_t, \tag{1.2.5}$$

then q_t is the unique strong solution of the following bilinear Itô s.d.e.

$$q_0 = p_0$$
$$q_t = q_0 + \int_0^t A q_s d s + \sum_{i=1}^{d} \int_0^t B_i q_s d W_s^{(i)} \qquad t \in [0,T]. \tag{1.2.6}$$

NOTE 1. The idea in (1.2.3) is that

$$W_t \approx \int_0^t Y_s d s.$$

The stochastic integral on the right-hand side of (1.2.4) is to be interpreted in the sense of Itô (see for example Arnold [AR1]).

NOTE 2. Clearly the a-posteriori distribution p_t can be found from the unnormalised version by the formula

$$p_t = \Big(\sum_{j=1}^{m} q_t^{(j)} \Big)^{-1} q_t, \tag{1.2.7}$$

i.e. it is sufficient to solve the simpler s.d.e. (1.2.6) to find p_t.

PROOF OF (i). See, for example, Davis and Marcus [DM1].

PROOF OF (ii). Application of Itô's rule to (1.2.5) yields the result (1.2.6). ∎

Part (i) of theorem 1.1 was first proposed by Wonham [WO1], it is a special case of the *Fujisaki-Kallianpur-Kunita formula* [FKK1]. Part (ii) gives a representation of the a-posteriori distribution due in its general form to Zakai [ZA1].

From the a-posteriori distribution for X_t we can calculate optimal estimates of functions of X_t; for instance if f is some function on the state-space $\mathcal{S}$,

$$f : \mathcal{S} \to R^n$$

such that

$$E_1 \| f(X_t) \|^2 < \infty$$

then the optimal (in the "least-squares" sense) estimate of $f(X_t)$, based on the observations $(W_s,\ s \in [0,t])$ is given by the conditional mean,

$$\begin{aligned} \widehat{f(X_t)} &:= E_1(f(X_t) \mid W_s,\ s \in [0,t]) \\ &= \sum_{j=1}^{m} f(a_j) p_t^{(j)} \\ &= \Big(\sum_{j=1}^{m} q_t^{(j)} \Big)^{-1} \sum_{j=1}^{m} f(a_j) q_t^{(j)}. \end{aligned} \tag{1.2.8}$$

Let us return now to the original problem; that is, the determination of an (almost) optimal filter for the "real" process (s_t) with physical wide-band noise (n_t) in the observations, rather than the idealised problem of theorem 1.1. Suppose first that the signal (s_t) *is* a Markov-chain but that the noise in the observations is not idealised white noise but some physical wide-band noise. We require a formula (hopefully a recursive formula) for calculating an a-posteriori distribution for s_t from each sample path $(Y_s\,;\ s \le t)$ (or $(\int_0^s Y_u du\,;\ s \le t)$), i.e. we require a functional F from the space of continuous R^d-valued functions into R^m,

$$F : C^d[0,T] \to R^m$$

$$\begin{bmatrix} p_t^{(1)} \\ \cdot \\ \cdot \\ \cdot \\ p_t^{(m)} \end{bmatrix} = \begin{bmatrix} P_1(s_t = a_1 \mid \int_0^s Y_u du\,;\ s \le t) \\ \cdot \quad \cdot \\ \cdot \quad \cdot \\ \cdot \quad \cdot \\ P_1(s_t = a_m \mid \int_0^s Y_u du\,;\ s \le t) \end{bmatrix}$$

$$= F\Big(\int_0^s Y_u du\,;\ s \le t \Big).$$

It is plausible that if the distribution of the integral noise process, $(\int_0^t n_s ds)$, is close in some sense to the distribution of the Brownian motion (β_t) (i.e. *Wiener measure*), then the optimal filter for s_t is close to that obtained by formula (1.2.4). Unfortunately a full analysis of this question appears to be very difficult, but one problem is that the solution of (1.2.4) is only defined almost surely uniquely with respect to Wiener measure on $C^d[0,T]$; in particular it prescribes no "output" (a-posteriori distribution) for the set of "inputs" (elements of $C^d[0,T]$) with bounded variation, to which the "real" observations $(\int_0^t Y_s ds)$ belong. This particular problem is overcome by replacing formula (1.2.4) by a *statistically robust* version; by this we mean a functional that defines an "output" for *every* element of $C^d[0,T]$, that coincides with the solution of (1.2.4) almost surely (Wiener measure) and that is continuous in some sense. This problem of *robustness* in optimal filtering equations is dealt with by Clark [CL1], Davis [DA1] and others.

In this thesis we shall be considering discrete approximations for equation (1.2.6); the domain of these approximations extends naturally to all elements of $C^d[0,T]$ and the methods are inherently robust. Some of the methods considered, the Mil'shtein scheme [MI1] and the paradigm (definition 4.2) for example, are probably *uniformly robust* in the time discretisation parameter, which means that the robust approximations obtained by these methods converge to robust exact solutions uniformly ('sup' norm) on compact sets of input paths, but I have not investigated this further. An example of a method which is not uniformly robust is the simple Euler method (see Clark [CL1]). The question of uniform robustness for discrete approximation schemes has been treated in detail by Talay [TA1] for a class of s.d.e.s more general than those considered here.

Next, suppose that the signal we wish to estimate, (s_t), is a (continuous-state) Markov diffusion process. Rather than using the version of the Fujisaki-Kallianpur-Kunita formula appropriate to diffusions, which is an infinite-dimensional (partial) differential equation, we use the results of Kushner in [KU1] and assume that there exists some approximating Markov chain. We solve the filtering equations for this chain. To solve the partial differential equation for the diffusion numerically we would need to use a more complex approximation procedure involving discretisation of the space parameter. Bennaton, in [BE2] and [BE3], has proved results on the order of convergence of Galerkin methods used to solve this problem, which parallel the results of Clark [CL1].

This thesis is primarily concerned then with discrete-time methods for solving equations (1.2.4) to (1.2.8). Given such an approximation method, a filter could be constructed with the form shown in diagram 1.1.

1.3 Discrete solutions of stochastic differential equations.

Stochastic differential equations can be "solved" by finite approximation schemes in much the same way as ordinary differential equations. A first step is to find a "Taylor-type" expansion of the exact solution about each point of a discretisation mesh of the time interval, and to approximate terms in this expansion, thereby forming a one-step approximation scheme. Such an approach gives familiar schemes for ordinary differential equations, the simplest of which is the Euler scheme (see for example Blum [BL1]).

For stochastic differential equations the Taylor expansions differ in form from their o.d.e. counterparts because of the differences between the rules of stochastic calculus and ordinary calculus (for instance, Itô's rule).

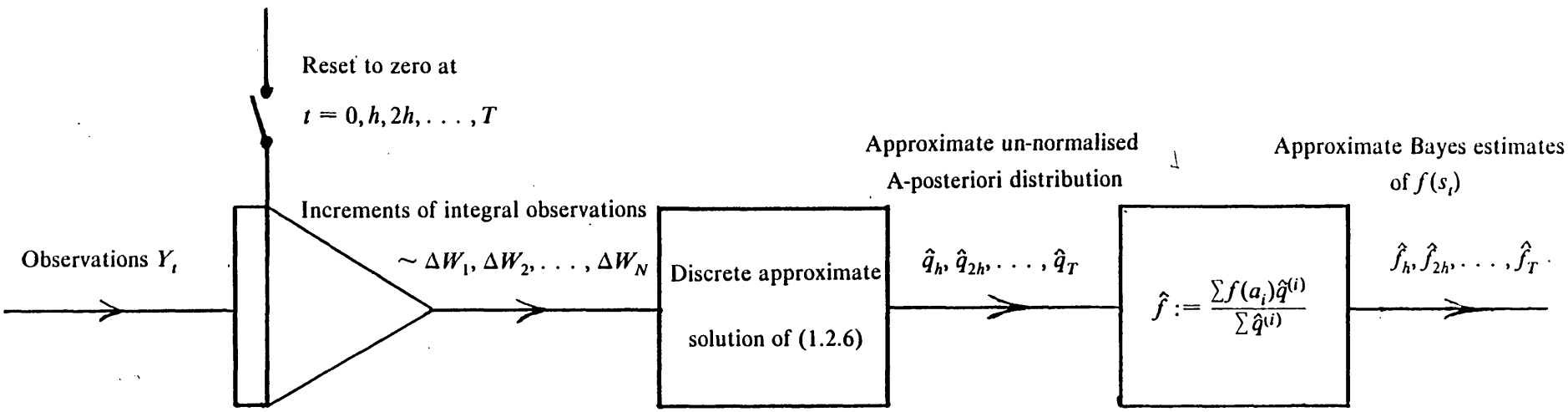

Reset to zero at
$t = 0, h, 2h, \ldots, T$
Observations Y_t
Increments of integral observations
$\sim \Delta W_1, \Delta W_2, \ldots, \Delta W_N$
Discrete approximate
solution of (1.2.6)
Approximate un-normalised
A-posteriori distribution
$\hat{q}_h, \hat{q}_{2h}, \ldots, \hat{q}_T$
$\hat{f} := \frac{\sum f(a_i)\hat{q}^{(i)}}{\sum \hat{q}^{(i)}}$
Approximate Bayes estimates
of $f(s_t)$
$\hat{f}_h, \hat{f}_{2h}, \ldots, \hat{f}_T$

Diagram 1.1

Wagner and Platen [WP1] and Platen [P1] have derived a general Taylor formula for multidimensional s.d.e.s with time-varying drift and diffusion coefficients. They define a family of one-step approximation schemes based on this formula. It is their approach which is adopted in chapters 2 and 3 of this thesis. I refer to their schemes as the *$k/2^{th}$-order Taylor schemes* (see definition 2.3).

The $1/2^{th}$-order Taylor scheme is the stochastic equivalent of the Cauchy-Euler scheme. Though first proposed for use with stochastic equations by Maruyama in [MA1]; it is analysed for time-varying, single-dimensional diffusions by Mil'shtein in [MI1]. Mil'shtein also considered a first-order scheme, which is equivalent to the 1^{st}-order Taylor scheme. He further defined a method with order h^4 for the mean-squared error in one step, (h is the step length). Applied to a bilinear equation such as (1.2.6), this method has a global error (*root*-mean-squared error over an interval) of order 1; it is similar to the $3/2^{th}$-order Taylor scheme but lacks the term in h^2.

Rao, Borwankar and Ramkrishna [RBR1] develop an order 2 scheme, which is equivalent to the 2^{nd}-order Taylor scheme. Rümelin [RU1] considers a general class of Runge-Kutta type schemes, which match terms in the Taylor expansions but are easier to implement than the Taylor schemes if the drift and diffusion coefficients are nonlinear, as derivatives of the coefficients are not directly evaluated.

The above authors are concerned mainly with simulation of the solutions of s.d.e.s. For this, in a scheme such as the $3/2^{th}$-order Taylor scheme, one would use random number generators with appropriate distributions and correlations to simulate the terms required. All the above authors produce neat reductions of their methods that require the minimum number of generators. For the $1/2^{th}$ and 1^{st}-order Taylor schemes one clearly needs to simulate at each step the increment in the random driving term, ΔW: a zero-mean normal random variable with variance h, if the driving term is a Brownian motion. For the $3/2^{th}$-order scheme, Wagner and Platen show that the simulation can be achieved by simulating an extra random variable; in the case of a Brownian driving process this is also normally distributed with mean zero but with variance $h^3/3$. Unfortunately it is correlated with the first random variable, the covariance being $h^2/2$.

Mil'shtein's method with mean-square error in one step of order h^4 also uses these two random variables as a basis for simulation. Rao, Borwankar and Ramkrishna use a further random variable to simulate all the terms needed in their 2^{nd}-order scheme.

We are not concerned here with simulation but with the "pathwise" solution of equations (1.2.4) and (1.2.6) given realisations of the driving term (W_t). More specifically, we require our approximate solutions to be measurable with respect to the driving term at the partition points. Because of this we cannot achieve arbitrary orders of convergence; the maximum order of convergence of discrete approximations to the solution of equation (1.2.6) is in fact 1; this is well-known but is proved here (lemma 3.4) for completeness.

A detailed treatment of the maximum order of convergence of "pathwise" approximate solutions of s.d.e.s is given in Clark and Cameron in [CC1].

1.4 On the results of chapters 2–4.

In this section we outline the results contained in the next three chapters, putting them in the context of the work of other authors.

Chapter 2 contains two theorems which are used extensively in chapters 3 and 4: the first, theorem 2.1, presents a general Taylor expansion of the solution of bilinear s.d.e.s; it is a corollary of theorem 1 in [WP1]. It is used as a first step in the comparison of the solutions of s.d.e.s and their approximations. Theorem 2.2 is a tool for bounding the L_p-norms of the solutions of stochastic difference equations, given the L_p-properties of the driving terms; it is used to bound errors in approximations. For example, in the proof of theorem 3.1 a process, $(N\tilde{q}_{nh}, n = 0, 1, \ldots, N)$, which obeys a rather complicated difference equation, (3.5.10), is approximated by another process, $(\chi_n, n = 0, 1, \ldots, N)$, which obeys the simpler equation (3.5.17); theorem 2.2 is used to show that all the L_p-norms of the error process, $(N\tilde{q}_{nh} - \chi_n, n = 0, 1, \ldots, N)$, converge to zero. Theorem 2.2 is an adaptation of part of theorem 2 in [WP1], which bounds the L_2-norm only; the reason for having the stronger result here is that by obtaining an order of convergence in terms of the L_p-norms for $p > 2$ we are able to use the moment form of the Borel-Cantelli lemma to obtain *almost sure* convergence results.

I have used the notation "$O_p(\frac{1}{N^r})$" to indicate an order of convergence in terms of the L_p-norm, and "$O_{(\infty)}(\frac{1}{N^r})$" to indicate an order for all the L_p-norms (see definition 2.2). I have also borrowed the notation for representing the solutions of s.d.e.s in Taylor series form, devised by Wagner and Platen in [WP1], as this seems a natural and compact, if at first sight daunting, way of handling these expressions; it is explained fully in section 2.2.

Section 2.5 deals with the family of approximation schemes for s.d.e.s proposed by Wagner and Platen; these are the natural one-step schemes. Theorem 2.3, which is a corollary of theorems 2.1 and 2.2, gives their L_p-properties when applied to bilinear equations.

The main results of chapter 3 are the limit-distribution theorems for normalised error sequences (theorems 3.1 and 3.2). These give the optimal "partition-conditional" behaviour of the errors in "partition-measurable" approximation schemes for the bilinear equation (1.2.6) and the corresponding Bayes estimates found from this equation. Theorem 3.1 is essentially the same result as theorem 2 of Clark in [CL2] but the proof is presented here in terms of the discrete approximation results of chapter 2 rather than Clark's approximating o.d.e. approach. To accommodate the non-zero drift component in the driving term of equation (1.2.6) we use the same Girsanov transformation of measure as Clark; this enables a simpler analysis of Wiener-process-driven equations to be extended to equation (1.2.6). The transformation is dealt with in section 3.3.

The limit-distribution results of theorem 3.1 are extended from approximations of the solution of (1.2.6) to the corresponding approximate estimates obtained by the Bayes formula (1.2.8) in theorem 3.2. The technique used is the production of a Taylor expansion of the non-linear estimation function implicit in (1.2.8) (given explicitly, Ψ_f, in (3.6.1)) about the point q_T, followed by the bounding of the second and higher-order terms in this expansion; it is what Billingsley in [BI1] p.340 calls the *delta-method.*

Chapter 4 begins by introducing some efficiency concepts for discrete approximate solutions of s.d.e.s; in particular "first-order asymptotic efficiency" (see definition 4.1). A first-order asymptotically efficient scheme is one with the optimal properties discussed in chapter 3. In section 4.3 we

introduce as a paradigm a scheme based on (but distinct from) the one-step schemes of Wagner and Platen. It contains the fewest terms needed for first-order asymptotic efficiency. The paradigm converges with order 1, of course, as does the scheme first proposed by Mil'shtein in [MI1], but the latter is shown, in a corollary to theorem 4.1, not to be first-order asymptotically efficient. Theorem 4.1 also demonstrates the existence of many other approximation schemes which possess the highest order of convergence, 1, but which are not first-order asymptotically efficient. Some of these schemes are shown to possess the optimal limit-distribution for normalised errors given in theorem 3.1, but with non-zero means; they can be thought of as being "asymptotically biased".

Various approximation schemes are compared in section 4.4, in particular the Euler and Mil'shtein schemes and the paradigm, as these contain the fewest terms needed for convergence, first-order convergence and first-order asymptotic efficiency respectively. We also mention briefly the Runge-Kutta schemes of Rümelin.

As far as I am aware the work presented in chapter 4 is original.

Chapter 5 suggests some further areas for future research.

Chapter 2

Some Tools for Handling Discrete Approximation Schemes

2.1 Introduction.

Several results concerning the convergence of discrete approximate solutions of stochastic differential equations are formulated in this chapter. These are the main tools to be used in chapters 3 and 4.

Section 2.3 presents a Taylor expansion representation of the solutions of bilinear stochastic differential equations. Theorem 2.1 is a special case of theorem 1 in [WP1], which gives such an expansion for a wider class of s.d.e.s. It is formulated in terms of a special notation, which is defined in section 2.2. The idea behind these expansions is that truncation of the "remainder" term yields a family of convergent stochastic difference equations that can be used to approximate the solutions of the differential equations.

A theorem for determining the order of convergence of such difference equations is presented in section 2.4; it is based on theorem 2 in [WP1]. In section 2.5 we use the results of the previous two sections to define and investigate the convergence of a class of discrete approximation schemes for bilinear equations, introduced by Wagner and Platen in [WP1]. We call these the Taylor schemes. The order of covergence is given and two counter-examples are used to disprove higher-order convergence in the $1/2^{th}$ and 1^{st} order schemes.

In section 2.6 we mention briefly a generalisation of the results of sections 2.3 and 2.5 from time-invariant bilinear equations to a wider class of s.d.e.s.

2.2 A special notation used in Taylor expansions of the solutions of bilinear stochastic differential equations.

I have adopted the notation used by Wagner and Platen in [WP1] for expressing Taylor expansions of the solutions of stochastic differential equations. The use of a special notation has two advantages: first it provides a very compact way of writing down the multi-term expressions that occur frequently in this type of work, and secondly it facilitates the expression of results that are true for more than one type of expansion (see, for example, theorem 2.1). Its disadvantage, of course, is that the reader needs to become familiar with it before he can interpret the results.

The notation is used in the proofs of results in chapters 3 and 4 but not in their statements. It is used extensively in this chapter both in the statement of results and in their proofs.

The notation introduced in [WP1] is for use with arbitrary nonlinear stochastic differential equations; I use here a version suitable for bilinear equations. Consider the following bilinear equation, defined on some underlying probability space $(\Omega, \mathcal{F}, P)$:

$$\begin{aligned} X_0 &= x, \\ dX_t &= FX_t\,dt + \sum_{i=1}^{d} G_i X_t\,dW_t^{(i)} \qquad t \in [0,T], \end{aligned} \tag{2.2.1}$$

where $X_t \in R^m$; $F, G_1, \ldots, G_d$ are real $m \times m$ matrices and

$$\left(W_t = \begin{bmatrix} W_t^{(1)} \\ \cdot \\ \cdot \\ \cdot \\ W_t^{(d)} \end{bmatrix} \quad t \in [O,T]\right)$$

is a d-dimensional Wiener process.

We make the following defintions.

Definition 2.1. M is the set of indices given by

$$M := \{v, (j_1, j_2, \ldots, j_k) : j_l \in \{0,1,\ldots,d\}; l = 1,2,\ldots,k; k = 1,2,\ldots\}$$

(v is the empty index, containing no terms).

$l(\alpha)$ is the "length" of the index α:

$$\begin{aligned} l(v) &:= 0 \\ l((j_1, j_2, \ldots, j_k)) &:= k. \end{aligned}$$

$n(\alpha)$ is the multiplicity of the term 0 in the index α:

$$\begin{aligned} n(v) &:= 0 \\ n((j_1, j_2, \ldots, j_k)) &:= \text{cardinal}\{j_i = 0; i = 1,2,\ldots,k\}. \end{aligned}$$

$-\alpha$ and $\alpha -$ are the following shortened indices:

$$\begin{aligned} -v = v- = -(j_1) = (j_1)- &:= v, \\ -(j_1, j_2, \ldots, j_k) &:= (j_2, j_3, \ldots, j_k) \qquad k \geq 2, \\ (j_1, j_2, \ldots, j_k)- &:= (j_1, j_2, \ldots, j_{k-1}) \qquad k \geq 2. \end{aligned}$$

$l(S)$ is the supremum of the lengths of indices in a subset S of M:

$$l(S) := \sup_{\alpha \in S} l(\alpha) \qquad S \subseteq M.$$

$B(S)$ is a "remainder set" (see theorem 2.1 for motivation):

$$B(S) := \{\alpha \in M - S : -\alpha \in S\}.$$

$A_\alpha(F, G_1, \ldots, G_d)$ is the indexed matrix product given by

$$\begin{aligned}
&A_\nu(F, G_1, \ldots, G_d) := I, \\
&A_\alpha(F, G_1, \ldots, G_d) := A_{-\alpha}(F, G_1, \ldots, G_d)F \qquad \text{if } \alpha = (0, j_2, \ldots, j_k), \\
&A_\alpha(F, G_1, \ldots, G_d) := A_{-\alpha}(F, G_1, \ldots, G_d)G_{j_1} \qquad \text{if } \alpha = (j_1, j_2, \ldots, j_k) \quad j_1 \neq 0.
\end{aligned}$$

$I_\alpha^W(X, s, t)$ is the indexed integral given by

$$\begin{aligned}
&I_\nu^W(X, s, t) := X_t, \\
&I_\alpha^W(X, s, t) := \int_s^t I_{\alpha-}^W(X, s, u)\, du \qquad \text{if } \alpha = (j_1, j_2, \ldots, 0), \\
&I_\alpha^W(X, s, t) := \int_s^t I_{\alpha-}^W(X, s, u)\, dW_u^{(j_k)} \qquad \text{if } \alpha = (j_1, j_2, \ldots, j_k) \quad j_k \neq 0.
\end{aligned}$$

$I_\alpha^W(s, t)$ is the indexed integral given by

$$\begin{aligned}
&I_\nu^W(s, t) := 1, \\
&I_\alpha^W(s, t) := \int_s^t I_{\alpha-}^W(s, u)\, du \qquad \text{if } \alpha = (j_1, j_2, \ldots, 0), \\
&I_\alpha^W(s, t) := \int_s^t I_{\alpha-}^W(s, u)\, dW_u^{(j_k)} \qquad \text{if } \alpha = (j_1, j_2, \ldots, j_k) \quad j_k \neq 0.
\end{aligned}$$

NOTE. $I_\alpha^W(X, s, t)$ is an m-vector, $I_\alpha^W(s, t)$ is a scalar.

2.3 A Taylor expansion for representing the solutions of bilinear s.d.e.s.

To study the convergence of stochastic difference equations, defined on some partition of the time interval $[0, T]$, to a bilinear stochastic differential equation such as (2.2.1) we would like an expansion of the solution of the latter about the points of the partition. The following theorem, which is a special case of theorem 1 in [WP1], provides such a result. The proof, modified for the bilinear case, is relatively short, so it is included.

Theorem 2.1. *For any subset $S \subseteq M$ with the following properties:*

$$\text{(i) } S \neq \emptyset,$$

(ii) $l(S) < \infty$,

(iii) $-\alpha \in S \qquad \forall\, \alpha \in S$,

the solution of equation (2.2.1) has the representation

$$X_0 = x,$$

$$X_t = \sum_{\alpha \in S} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(s,t) X_s + \sum_{\alpha \in B(S)} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(X, s, t)$$

$$0 \leq s \leq t \leq T. \qquad (2.3.1)$$

PROOF. We prove first that for $\alpha = (j_1, j_2, \ldots, j_k)$, $k \geq 1$

$$I_\alpha^W(X,s,t) = I_\alpha^W(s,t) X_s + I_{(0,\alpha)}^W(FX,s,t) + \sum_{i=1}^{d} I_{(i,\alpha)}^W(G_i X, s, t), \qquad (2.3.2)$$

where

$$(i,\alpha) := (i, j_1, j_2, \ldots, j_k) \qquad i = 0, 1, \ldots, d.$$

Now

$$\begin{aligned} I_{(0,\alpha)}^W(FX,s,t) + \sum_{i=1}^{d} I_{(i,\alpha)}^W(G_i X, s, t) &= \int_s^t \int_s^{t_k} \cdots \int_s^{t_1} F X_u \, du \, dW_{t_1}^{(j_1)} \cdots dW_{t_k}^{(j_k)} \\ &\quad + \sum_{i=1}^{d} \int_s^t \int_s^{t_k} \cdots \int_s^{t_1} G_i X_u \, dW_u^{(i)} \, dW_{t_1}^{(j_1)} \cdots dW_{t_k}^{(j_k)} \\ &= \int_s^t \int_s^{t_k} \cdots \int_s^{t_2} (X_{t_1} - X_s) \, dW_{t_1}^{(j_1)} \cdots dW_{t_k}^{(j_k)} \\ &= I_\alpha^W(X,s,t) - I_\alpha^W(s,t) X_s, \end{aligned}$$

where "$dW_t^{(0)}$" means "dt". This proves (2.3.2).

We now prove (2.3.1) by induction with repect to $l(\alpha)$. It is evident that $S_0 (:= \{v\})$ satisfies conditions (i)–(iii), and

$$B(S_0) = \{(0), (1), (2), \ldots, (d)\}.$$

Now

$$\sum_{\alpha \in S_0} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(s,t) = I$$

and

$$\sum_{\alpha \in B(S_0)} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(X,s,t) = \int_s^t F X_u \, du + \sum_{i=1}^{d} \int_s^t G_i X_u \, dW_u^{(i)}$$

$$= X_t - X_s.$$

This proves (2.3.1) for all S satisfying (i)–(iii) with $l(S) = 0$. Suppose now that (2.3.1) is true for all S that satisfy conditions (i)–(iii) with $l(S) = k \geq 0$. Consider a set S_{k+1} with $l(S_{k+1}) = k + 1$ that also satisfies (i)–(iii). Let

$$S_k := \{\alpha \in S_{k+1} : l(\alpha) \leq k\}.$$

Clearly S_k satisfies (i)–(iii) and $l(S_k) = k$; so

$$X_t = \sum_{\alpha \in S_k} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(s,t) X_s + \sum_{\alpha \in B(S_k)} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(X, s, t).$$

Now $S_{k+1} - S_k \subseteq B(S_k)$; so

$$\sum_{\alpha \in B(S_k)} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(X, s, t) = \sum_{\alpha \in S_{k+1} - S_k} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(X, s, t) + \sum_{\alpha \in B(S_k) - (S_{k+1} - S_k)} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(X, s, t).$$

We apply (2.3.2) to obtain

$$\sum_{\alpha \in S_{k+1} - S_k} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(X, s, t) = \sum_{\alpha \in S_{k+1} - S_k} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(s,t) X_s + \sum_{\alpha \in S_{k+1} - S_k} \Big(A_{(0,\alpha)}(F, G_1, \ldots, G_d) I_{(0,\alpha)}^W(X, s, t) + \sum_{i=1}^{d} A_{(i,\alpha)}(F, G_1, \ldots, G_d) I_{(i,\alpha)}^W(X, s, t) \Big).$$

So
$$X_t = \sum_{\alpha \in S_{k+1}} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(s,t) X_s + \sum_{\alpha \in B} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(X, s, t), \qquad (2.3.3)$$

where

$$B := (B(S_k) - (S_{k+1} - S_k)) \cup \{(i,\alpha) : i = 0, 1, \ldots, d \; ; \; \alpha \in (S_{k+1} - S_k)\}.$$

It is easy to show that $B = B(S_{k+1})$; so (2.3.1) is true for S_{k+1} and also, therefore, for all S that fulfil (i)–(iii). ∎

As an example of this theorem, consider the case where

$$S := \{v, (0), (1), (2), \ldots, (d)\};$$

in which case

$$B(S) = \{(0,0), (1,0), \ldots, (d,0), (0,1), \ldots, (d,1), \ldots, (d,d)\}.$$

Now

$$X_0 = x,$$

$$X_t = \sum_{\alpha \in S} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(s,t) X_s + \sum_{\alpha \in B(S)} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(X, s, t)$$

$$= \Big(I + F(t-s) + \sum_{i=1}^{d} G_i(W_t^{(i)} - W_s^{(i)})\Big) X_s + R_{s,t}, \tag{2.3.4}$$

where $R_{s,t}$ is the remainder term

$$\sum_{\alpha \in B(S)} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(X, s, t).$$

A frequently met discrete approximation scheme for solving s.d.e.s is the "stochastic Euler method", see [MA1]. In the present case this method is given, for the partition $(0, t_1, t_2, \ldots, T)$, by

$$\hat{X}_0 := x$$

$$\hat{X}_{t_{n+1}} := \Big(I + F(t_{n+1} - t_n) + \sum_{i=1}^{d} G_i(W_{t_{n+1}}^{(i)} - W_{t_n}^{(i)})\Big) \hat{X}_{t_n}. \tag{2.3.5}$$

By comparing (2.3.5) and (2.3.4) we obtain a difference equation for the approximation error:

$$X_0 - \hat{X}_0 = 0,$$

$$X_{t_{n+1}} - \hat{X}_{t_{n+1}} = \Big(I + F(t_{n+1} - t_n) + \sum_{i=1}^{d} G_i(W_{t_{n+1}}^{(i)} - W_{t_n}^{(i)})\Big)(X_{t_n} - \hat{X}_{t_n}) + R_{t_n, t_{n+1}}. \tag{2.3.6}$$

The next two sections describe methods of bounding and proving convergence of solutions of stochastic difference equations such as (2.3.6).

2.4 A theorem for bounding the solutions of stochastic difference equations.

We consider first how to bound **terms** in the difference equations obtained by theorem 2.1; beginning with a definition.

Definition 2.2.

(i) We say that a sequence of discrete-parameter random processes

$$((X_{n,N} \in R^m : n = 0, 1, \ldots, N) : N = 1, 2, \ldots)$$

is *of uniform L_p-order r*,and we write

$$X_{n,N} = O_p(1/N^r)$$

if

$$\sup_{n \leq N} E \| N^r X_{n,N} \|^p < \infty \qquad \text{for some } p > 0.$$

(ii) We say that a sequence of random variables $(X_N \in R^m : N = 1, 2, \ldots)$ is *of L_p-order r*, and we also write

$$X_N = O_p(1/N^r)$$

if

$$\sup_N E \| N^r X_N \|^p < \infty \qquad \text{for some } p > 0.$$

NOTE 1. We shall frequently write $O_p(h^r)$ for a uniform-L_p-order-r sequence where we are considering approximation schemes for solving s.d.e.s that are based on regular partitions of the time interval $[0, T] : (0, h, 2h, \ldots, Nh = T)$.

NOTE 2. An indexed process of uniform L_p-order r for all $p > 0$ will be called *of uniform $L_{(\infty)}$ order r*, written

$$X_{n,N} = O_{(\infty)}(1/N^r) \qquad (\text{or } O_{(\infty)}(h^r))$$

The brackets are to distinguish $L_{(\infty)}$-order-r convergence from L_∞-order-r convergence: by

$$X_{n,N} = O_\infty(1/N^r)$$

we would mean

$$\sup_{n \leq N} (\limsup_{p \to \infty} (E \| N^r X_{n,N} \|^p)^{1/p}) < \infty.$$

NOTE 3. A sequence of random variables that is of L_p-order r where $rp > 1$ converges to zero almost surely since

$$\sum_{N=1}^{\infty} E \| X_N \|^p \leq K \sum_{N=1}^{\infty} \frac{1}{N^{rp}} \qquad \text{where } K \geq 0$$
$$< \infty;$$

the moment form of the Borel-Cantelli lemma shows that

$$X_N \to 0 \qquad \text{(a.s.)}.$$

The following lemma puts bounds on the terms in the Taylor expansions of bilinear s.d.e.s introduced in theorem 2.1. It prepares the way for theorem 2.2, which bounds the solutions of stochastic difference equations of the type obeyed by the errors in discrete approximation schemes. It is an extension of a lemma in [WP1], which gives a similar result for second moments only.

Lemma 2.1. *Let $(X_t, t \in [0, T])$ be the solution of the bilinear s.d.e. (2.2.1). For all multiple indices $\alpha \in M$:*

(i)

$$\sup_{0\le s<t\le T} E\left((t-s)^{-(l(\alpha)+n(\alpha))/2}\|E\left(I_\alpha^W(X,s,t)\mid \mathcal{G}\right)\|\right)^p<\infty \tag{2.4.1}$$

for all $p>0$ *and all sub-σ-fields* $\mathcal{G}\subseteq\mathcal{F}$;

(ii)

$$\sup_{0\le s<t\le T} E\left((t-s)^{-(l(\alpha)+n(\alpha))/2}|E\left(I_\alpha^W(s,t)\mid \mathcal{G}\right)|\right)^p<\infty \tag{2.4.2}$$

for all $p>0$ *and all sub-σ-fields* $\mathcal{G}\subseteq\mathcal{F}$;

(iii) If $h:=T/N$ *then the indexed sequences*

$$\left(\left(E\left(I_\alpha^W(X,nh,(n+1)h)\mid\mathcal{G}\right);\ n=0,1,\ldots,N-1\right)N=1,2,\ldots\right)$$

and

$$\left(\left(E\left(I_\alpha^W(nh,(n+1)h)\mid\mathcal{G}\right);\ n=0,1,\ldots,N-1\right)N=1,2,\ldots\right)$$

are of uniform $L_{(\infty)}$ *order* $(l(\alpha)+n(\alpha))/2$ *for all* $\mathcal{G}\subseteq\mathcal{F}$;

(iv)

$$\left.\begin{aligned}E\left(I_\alpha^W(X,s,t)\mid\mathcal{F}_s^W\right)&=0\\ E\left(I_\alpha^W(s,t)\mid\mathcal{F}_s^W\right)&=0\end{aligned}\right\}\quad \text{if } l(\alpha)\neq n(\alpha). \tag{2.4.3}$$

PROOF. It follows by Jensen's inequality that to prove (i) it is sufficient to prove that

$$\sup_{0\le s<t\le T} E\left((t-s)^{-(l(\alpha)+n(\alpha))/2}\|I_\alpha^W(X,s,t)\|\right)^{2m}<\infty\quad\forall\, m=1,2,\ldots \tag{2.4.4}$$

We prove this by induction with respect to $l(\alpha)$.

Since the coefficients in (2.2.1) satisfy the conditions of an existence and uniqueness theorem (theorem 7.1.2 [AR1]) and since the initial condition is finite

$$\sup_{0\le t\le T} E\|X_t\|^{2m}<\infty\quad\forall\, m=1,2,\ldots$$

i.e. (2.4.4) is true for α with length $l(\alpha)=0$; suppose now that it is true for all α with length $l(\alpha)=k\ge 0$. Let $\beta:=(j_1,j_2,\ldots,j_{k+1})$, then $l(\beta)=k+1$.

If $j_{k+1}=0$

$$I_\beta^W(X,s,t)=\int_s^t I_{\beta-}^W(X,s,u)\,du.$$

So for $s<t$

$$\begin{aligned}\left(E\|I_\beta^W(X,s,t)\|^{2m}\right)^{1/2m}&\le\int_s^t\left(E\|I_{\beta-}^W(X,s,u)\|^{2m}\right)^{1/2m}du\\ &\le(t-s)\sup_{s<u\le t}\left(E\|I_{\beta-}^W(X,s,u)\|^{2m}\right)^{1/2m}\end{aligned}$$

and

$$E\left((t-s)^{-(l(\beta)+n(\beta))/2}\|I_\beta^W(X,s,t)\|\right)^{2m} \leq \sup_{s<u\leq t} E\left((t-s)^{-(l(\beta-)+n(\beta-))/2}\|I_{\beta-}^W(X,s,u)\|\right)^{2m}$$
$$\leq \sup_{s<u\leq t} E\left((u-s)^{-(l(\beta-)+n(\beta-))/2}\|I_{\beta-}^W(X,s,u)\|\right)^{2m};$$

so (2.4.4) is true for β.

If $j_{k+1} = 1,2,\ldots,d$

$$I_\beta^W(X,s,t) = \int_s^t I_{\beta-}^W(X,s,u)\, dW_u^{(j_{k+1})};$$

so, by a result on the moments of stochastic integrals (lemma 4.12 in [LS1]),

$$E\|I_\beta^W(X,s,t)\|^{2m} \leq K(t-s)^{m-1}\int_s^t E\|I_{\beta-}^W(X,s,u)\|^{2m}\, du, \qquad \text{where } K>0$$
$$\leq K(t-s)^m \sup_{s<u\leq t} E\|I_{\beta-}^W(X,s,u)\|^{2m} \quad \forall\ s<t$$

and so

$$E\left((t-s)^{-(l(\beta)+n(\beta))/2}\|I_\beta^W(X,s,t)\|\right)^{2m} \leq K \sup_{s<u\leq t} E\left((t-s)^{-(l(\beta-)+n(\beta-))/2}\|I_{\beta-}^W(X,s,u)\|\right)^{2m}$$
$$\leq K \sup_{s<u\leq t} E\left((u-s)^{-(l(\beta-)+n(\beta-))/2}\|I_{\beta-}^W(X,s,u)\|\right)^{2m}.$$

So (2.4.4) is true for β.

This completes the proof of (i). The proof of (ii) is identical. (iii) follows directly from (i) and (ii) by setting $s = nh$ and $t = (n+1)h$. We prove (iv) by induction with respect to $l(\alpha)$ also.

Now $l(v) = n(v)$;so (2.4.3) is true for α with $l(\alpha) = 0$. Suppose now that it is true for α with $l(\alpha) = k \geq 0$. Let $\beta = (j_1, j_2, \ldots, j_{k+1})$, then $l(\beta) = k+1$.

If $j_{k+1} = 0$

$$I_\beta^W(X,s,t) = \int_s^t I_{\beta-}^W(X,s,u)\, du.$$

(2.4.3) is true for β since it is true for $\beta-$ and if $l(\beta-) = n(\beta-)$ then $l(\beta) = n(\beta)$.

If $j_{k+1} = 1,2,\ldots,d$

$$I_\beta^W(X,s,t) = \int_s^t I_{\beta-}^W(X,s,u)\, dW_u^{(j_{k+1})}$$

and since

$$\int_s^t E\|I_{\beta-}^W(X,s,u)\|^2\, du < \infty$$

(2.4.3) is true for β. The same argument applies to $I_\beta^W(s,t)$. ∎

The following theorem can be thought of as a version of the Gronwall lemma suitable for use with stochastic difference equations. The proof is a generalisation of part of the proof of theorem 2 in [WP1].

Theorem 2.2. *Let* $(Y_n \in R^{m\times m} : n = 1,2,\ldots,N)$ *and* $(d_n \in R^m : n = 1,2,\ldots,N)$ *be sequences of matrix-valued and vector-valued discrete-parameter processes respectively, on some underlying probability space* $(\Omega, \mathcal{F}, P)$ *(See note below). Let*

$$\begin{aligned} \mathcal{F}_0 &:= \{\emptyset, \Omega\} \\ \mathcal{F}_n &:= \sigma(Y_k, d_k : k \leq n) \qquad n = 1,2,\ldots,N. \end{aligned}$$

Suppose also that (Y_n) *and* (d_n) *have the following properties:*

(i) $Y_n = O_{(\infty)}(1/N^{1/2})$

(ii) Y_n *is independent of* $\mathcal{F}_{n-1}$ $\forall\ n = 1,2,\ldots,N$

(iii) $\|E\,Y_n\| \leq K_1 1/N \qquad K_1 \geq 0$

(iv) $d_n = O_{(\infty)}(1/N^{r+1/2}) \qquad r \geq 0$

(v) $E\,(d_n \mid \mathcal{F}_{n-1}) = O_{(\infty)}(1/N^{r+1})$

The sequence of processes $(X_n \in R^m : n = 0,1,\ldots,N)$*, defined by*

$$\begin{aligned} X_0 &:= 0 \\ X_{n+1} &:= (I + Y_{n+1})X_n + d_{n+1} \qquad n = 0,1,\ldots,N-1 \end{aligned} \tag{2.4.5}$$

is of uniform $L_{(\infty)}$*- order r, i.e.*

$$\sup_{n \leq N} E\,\|N^r X_n\|^p < \infty \qquad \forall\ p > 0\,. \tag{2.4.6}$$

NOTE. $(X_n), (Y_n)$ and (d_n) are sequences of processes, but we write X_n for $X_{n;N}$ etc.

PROOF. If we multiply both sides of (2.4.5) by N^r we see that it is sufficient to prove the result for $r = 0$. We begin by using induction to prove the following inequality:

$$\begin{aligned} \|X_{n+1}\|^{2i} \leq\ & (1 + Z_{n+1}(i))\|X_n\|^{2i} + 2i\langle Y_{n+1}X_n, X_n\rangle \|X_n\|^{2(i-1)} \\ & + 2i\langle d_{n+1}, X_n\rangle \|X_n\|^{2(i-1)} + r_{n+1}(i), \end{aligned} \tag{2.4.7}$$

where $(Z_n(i))$ and $(r_n(i))$ are non-negative scalar sequences of processes, both of uniform $L_{(\infty)}$- order 1, i.e.

$$\begin{aligned} Z_n(i) &= O_{(\infty)}(1/N) \quad \forall\ i = 1,2,\ldots \\ r_n(i) &= O_{(\infty)}(1/N) \quad \forall\ i = 1,2,\ldots \end{aligned} \tag{2.4.8}$$

In addition, $Z_n(i)$ is independent of $\mathcal{F}_{n-1}\ \forall n = 1,2,\ldots,N$. We obtain, by squaring (2.4.5),

$$\|X_{n+1}\|^2 \leq (1 + 2\|Y_{n+1}\|^2)\|X_n\|^2 + 2\langle Y_{n+1}X_n, X_n\rangle + 2\langle d_{n+1}, X_n\rangle + 2\|d_{n+1}\|^2; \tag{2.4.9}$$

so (2.4.7) is true for $i = 1$. Suppose now that it is true for i; we multiply (2.4.7) by (2.4.9) to obtain

$$\begin{aligned}\|X_{n+1}\|^{2(i+1)} \le\ & (1 + 2(2i+1)\|Y_{n+1}\|^2 + 4i\|Y_{n+1}\|^3 + Z_{n+1}(i)(1 + 2\|Y_{n+1}\| + 2\|Y_{n+1}\|^2))\|X_n\|^{2(i+1)} \\ & + 2(i+1)\langle Y_{n+1}X_n, X_n\rangle\|X_n\|^{2i} + 2(i+1)\langle d_{n+1}, X_n\rangle\|X_n\|^{2i} + 2r_{n+1}(i)\|d_{n+1}\|^2 \\ & + 2r_{n+1}(i)\ \|d_{n+1}\|\ \|X_n\| + r_{n+1}(i)(1 + 2\|Y_{n+1}\| + 2\|Y_{n+1}\|^2)\|X_n\|^2 \\ & + 4i\|d_{n+1}\|^3\|X_n\|^{2i-1} + 2\|d_{n+1}\|^2(1 + Z_{n+1}(i) + 2i\|Y_{n+1}\| + 2i)\|X_n\|^{2i} \\ & + \|d_{n+1}\|(2Z_{n+1}(i) + 8i\|Y_{n+1}\| + 4i\|Y_{n+1}\|^2)\|X_n\|^{2i+1} \end{aligned} \tag{2.4.10}$$

We reduce the last five terms on the right side of (2.4.10) by using the following inequality. If $(\xi_n : n = 1, 2, \ldots, N)$ is a non-negative scalar sequence of processes of uniform $L_{(\infty)}$-order 1 then

$$\xi_n\|X_n\|^s \le \frac{1}{N}\|X_n\|^t + O_{(\infty)}(1/N) \qquad \forall\ 0 < s < t. \tag{2.4.11}$$

This follows directly from Young's inequality:

$$ab \le \frac{1}{p}|a|^p + \frac{1}{q}|b|^q,$$

where $p > 1$, $q > 1$ and $1/p + 1/q = 1$; if a is taken as $\|X_n\|^s$, b as $N\xi_n$, p as t/s and q as $t/(t-s)$. For example, the seventh term on the right-hand side of (2.4.10) is bounded as follows:

$$4i\|d_{n+1}\|^3\|X_n\|^{2i-1} \le \frac{1}{N}\|X_n\|^{2(i+1)} + O_{(\infty)}(1/N).$$

The other terms are bounded in a similar way to yield

$$\begin{aligned}\|X_{n+1}\|^{2(i+1)} \le\ & (1 + Z_{n+1}(i+1))\|X_n\|^{2(i+1)} + 2(i+1)\langle Y_{n+1}X_n, X_n\rangle\|X_n\|^{2i} \\ & + 2(i+1)\langle d_{n+1}, X_n\rangle\|X_n\|^{2i} + r_{n+1}(i+1),\end{aligned}$$

where

$$Z_{n+1}(i+1) := 2(2i+1)\|Y_{n+1}\|^2 + 4i\|Y_{n+1}\|^3 + Z_{n+1}(i)(1 + 2\|Y_{n+1}\| + 2\|Y_{n+1}\|^2) + 5/N$$

and

$$r_{n+1}(i+1) = 2r_{n+1}(i)\|d_{n+1}\|^2 + O_{(\infty)}(1/N).$$

This proves (2.4.7) for $i + 1$ and hence for all natural i.

We take the mean of both sides of (2.4.7) and use properties (ii) and (iii) of γ_{n+1} to obtain

$$\begin{aligned}E\|X_{n+1}\|^{2i} \le\ & (1 + K_4(i)/N)E\|X_n\|^{2i} + 2iE\langle (K_1/N)X_n, X_n\rangle\|X_n\|^{2(i-1)} \\ & + 2iE\langle E(d_{n+1} \mid \mathcal{I}_n), X_n\rangle\|X_n\|^{2(i-1)} + K_5(i)/N.\end{aligned} \tag{2.4.12}$$

Now

$$\begin{aligned}E\langle (K_1/N)X_n, X_n\rangle\|X_n\|^{2(i-1)} &\le (K_1/N)E\|X_n\|^{2i} \\ E\langle E(d_{n+1} \mid \mathcal{I}_n), X_n\rangle\|X_n\|^{2(i-1)} &\le E\|E(d_{n+1} \mid \mathcal{I}_n)\|\|X_n\|^{2i-1} \\ &\le (K_6(i)/N)E\|X_n\|^{2i} + K_7(i)/N,\end{aligned}$$

where the last step uses property (v) and inequality (2.4.11) with $t = 2i$. So from (2.4.12)

$$E\,\|X_{n+1}\|^{2i} \le (1 + K_8(i)/N)E\,\|X_n\|^{2i} + K_9(i)/N,$$

where $K_8(i)$ and $K_9(i)$ are positive constants, dependent on i but not on n or N. Gronwall's inequality shows that

$$\sup_{n \ge N} E\,\|X_n\|^{2i} < \infty \qquad i = 1,2,\ldots;$$

the remaining moments are bounded using Hölder's inequality. So

$$X_n = O_{(\infty)}(1/N^0).$$

∎

In the deterministic version of theorem 2.2, we would require that Y_n and d_n be of order 1 and $r+1$ respectively to obtain convergence of order r for (X_n); this case is simply Gronwall's inequality. In the random case, however, we have weakened these two conditions by requiring that only the conditional means of Y_n and d_n be of orders 1 and $r+1$ respectively, see conditions (iii) and (v) of the theorem. Conditions (i) and (iv) show that we need only reduced order for the "difference" terms, $Y_n - E(Y_n)$ and $d_n - E(d_n \mid \mathcal{I}_{n-1})$.

One can think of this as follows: at time nh the next increment of the solution to equation (2.4.5) is determined by the "future" terms Y_{n+1} and d_{n+1}; we require that the means of these terms conditioned on the present obey the normal conditions of the Gronwall lemma; the difference terms, $Y_{n+1} - E(Y_{n+1})$ and $d_{n+1} - E(d_{n+1} \mid \mathcal{I}_n)$, which are "orthogonal to the present", may be of order $1/2$ less than the terms Y_{n+1} and d_{n+1} themselves. This is a result of adding large numbers of uncorrelated random variables. (c.f. the central limit theorem).

2.5 A family of discrete approximation schemes.

For any subset $S \subseteq M$ that satisfies the conditions of theorem 2.1 we can define a one-step discrete approximation scheme by truncating the remainder term from (2.3.1):

$$\hat{X}_0 := x$$

$$\hat{X}_{n+1} := \sum_{\alpha \in S} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(nh,(n+1)h)\hat{X}_n \qquad n = 0,1,\ldots,N-1.$$

A very useful subclass of these schemes, in that it prescribes the terms that are needed for different orders of convergence, was proposed by Wagner and Platen in [WP1]; it is the class of schemes resulting from the following definition of the sets S_k:

$$S_k := \{\alpha \in M : l(\alpha) + n(\alpha) \le k \text{ OR } l(\alpha) = n(\alpha) = \tfrac{1}{2}(k+1)\}, \qquad (2.5.1)$$

(i.e. sets that contain all indices shorter than $k+1$ but including the "all zeros" index of length $k+1$.).

Such subsets of M clearly satisfy the conditions of theorem 2.1. So

$$X_{(n+1)h} = \sum_{\alpha \in S_k} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(nh, (n+1)h) X_{nh} + \sum_{\alpha \in B(S_k)} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(X, nh, (n+1)h). \tag{2.5.2}$$

Definition 2.3. Consider the class of discrete approximation schemes given by

$$\hat{X}_0(k) := x,$$
$$\hat{X}_{n+1}(k) := \sum_{\alpha \in S_k} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(nh, (n+1)h) \hat{X}_n(k) \qquad n = 0, 1, \ldots, N-1, \tag{2.5.3}$$

where the S_k are given in (2.5.1) for $k = 1, 2, \ldots$. These schemes will be called the *$k/2^{th}$-order Taylor schemes*. They were compared with the schemes proposed in [MA1], [MI1] and [RBR1] in chapter 1.

Theorem 2.3. *The $k/2^{th}$ order Taylor scheme for solving equations such as (2.2.1) converges with order $k/2$.*

PROOF. We substract (2.5.3) from (2.5.2):

$$\begin{aligned} X_0 - \hat{X}_0(k) &= 0 \\ X_{(n+1)h} - \hat{X}_{n+1}(k) &= (I + Y_{n+1}(k))(X_{nh} - \hat{X}_{n+1}(k)) + d_{n+1}(k), \end{aligned} \tag{2.5.4}$$

where

$$Y_{n+1}(k) := \sum_{\alpha \in S_k - \{v\}} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(nh, (n+1)h),$$
$$d_{n+1}(k) := \sum_{\alpha \in B(S_k)} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(X, nh, (n+1)h).$$

Part (iii) of lemma 2.1 shows that

$$Y_n(k) = O_{(\infty)}(h^{1/2})$$
$$d_n(k) = O_{(\infty)}(h^{\frac{k+1}{2}}).$$

Now, $Y_n(k)$ is independent of $\mathcal{I}_{(n-1)h}$ and by lemma 2.1 part (iv), bearing in mind (2.5.1),

$$E(Y_n) = O_{(\infty)}(h),$$
$$E(d_n \mid \mathcal{I}_{(n-1)h}) = O_{(\infty)}(h^{\frac{k+2}{2}}).$$

We can now apply theorem 2.2 to equation (2.5.4) to yield the desired result,

$$(X_{nh} - X_n(k)) = O_{(\infty)}(h^{k/2}).$$ ∎

This is the result given, in terms of the L_2-norm, in theorem 2 of [WP1] extended to all the L_p-norms.

The $k/2^{th}$-order Taylor schemes are given explicitly below for $k = 1, 2$ and 3:$\Delta W_{n+1}^{(i)}$ means $W_{(n+1)h}^{(i)} - W_{nh}^{(i)}$.

$$\hat{X}_0(1) = \hat{X}_0(2) = \hat{X}_0(3) = x,$$

$$\hat{X}_{n+1}(1) = \Big(I + Fh + \sum_{i=1}^{d} G_i \Delta W_{n+1}^{(i)}\Big)\hat{X}_n(1),$$

$$\hat{X}_{n+1}(2) = \Big(I + Fh + \sum_{i=1}^{d} G_i \Delta W_{n+1}^{(i)} + \sum_{i,j=1}^{d} G_i G_j \int_{nh}^{(n+1)h} \int_{nh}^{t} dW_s^{(j)} dW_t^{(i)}\Big)\hat{X}_n(2),$$

$$\begin{aligned}\hat{X}_{n+1}(3) = \Big(I + Fh + \sum_{i=1}^{d} G_i \Delta W_{n+1}^{(i)} + \sum_{i,j=1}^{d} G_i G_j \int_{nh}^{(n+1)h} \int_{nh}^{t} dW_s^{(j)} dW_t^{(i)} \\ + \sum_{i=1}^{d} FG_i \int_{nh}^{(n+1)h} \int_{nh}^{t} dW_s^{(i)} dt + \sum_{i=1}^{d} G_i F \int_{nh}^{(n+1)h} \int_{nh}^{t} ds\, dW_t^{(i)} \\ + \sum_{i,j,k=1}^{d} G_i G_j G_k \int_{nh}^{(n+1)h} \int_{nh}^{t} \int_{nh}^{s} dW_u^{(k)} dW_s^{(j)} dW_t^{(i)} + \frac{1}{2}F^2h^2\Big)\hat{X}_n(3).\end{aligned} \tag{2.5.5}$$

It is fairly easy to show that the terms prescribed by the Taylor schemes (at least for $k = 1$ and 2) are necessary as well as sufficient for the given orders of convergence. In fact if we perturb the terms in the $1/2^{th}$-order scheme by δF, $\delta G_1, \ldots, \delta G_d$ then we shall have convergence of order $1/2$ to the solution of the the wrong equation; i.e. the scheme given by

$$\hat{X}_0 := x,$$

$$\hat{X}_{n+1} := \Big(I + (F + \delta F)h + \sum_{i=1}^{d} (G_i + \delta G_i)\Delta W_{n+1}^{(i)}\Big)\hat{X}_n, \tag{2.5.6}$$

will converge with order $1/2$ to the solution of the following equation:

$$X_0' = x$$

$$dX_t' = (F + \delta F)X_t' dt + \sum_{i=1}^{d} (G_i + \delta G_i)X_t' dW_t^{(i)}.$$

Clearly the scheme given by (2.5.6) does not in general converge to the solution of (2.2.1).

The argument for the necessity of all the terms in the 1^{st}-order scheme is not as trivial. Consider the scheme given by

$$\hat{X}_0 := x,$$

$$\hat{X}_{n+1} := \Big(I + Fh + \sum_{i=1}^{d} G_i \Delta W_{n+1}^{(i)} + \sum_{i,j=1}^{d} (G_i G_j + \delta G_{i,j}) \int_{nh}^{(n+1)h} \int_{nh}^{t} dW_s^{(j)} dW_t^{(i)}\Big) \hat{X}_n. \tag{2.5.7}$$

We subtract this from the S_2 expansion for the solution of (2.2.1), given in (2.5.2).

$$X_0 - \hat{X}_0 = 0,$$

$$\begin{aligned} X_{(n+1)h} - \hat{X}_{n+1} = \Big(I + Fh + \sum_{i=1}^{d} G_i \Delta W_{n+1}^{(i)} + \sum_{i,j=1}^{d} (G_i G_j + \delta G_{i,j}) \int_{nh}^{(n+1)h} \int_{nh}^{t} dW_s^{(j)} dW_t^{(i)}\Big)(X_{nh} - \hat{X}_n) \\ + \sum_{\alpha \in B(S_2)} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(X, nh, (n+1)h) \\ - \sum_{i,j=1}^{d} \delta G_{i,j} \int_{nh}^{(n+1)h} \int_{nh}^{t} dW_s^{(j)} dW_t^{(i)} X_{nh}. \end{aligned}$$

If we apply lemma 2.1 and theorem 2.2 to this difference equation we see that

$$X_{nh} - \hat{X}_n = O_{(\infty)}(h^{1/2}).$$

Now

$$N^{1/2}(X_{(n+1)h} - \hat{X}_{n+1}) = \Big(I + Fh + \sum_{i=1}^{d} G_i \Delta W_{n+1}^{(i)}\Big) N^{1/2}(X_{nh} - \hat{X}_n) + d_{n+1}(1) + d_{n+1}(2), \tag{2.5.8}$$

where

$$\begin{aligned} d_{n+1}(1) := N^{1/2} \sum_{i,j=1}^{d} (G_i G_j + \delta G_{i,j}) \int_{nh}^{(n+1)h} \int_{nh}^{t} dW_s^{(j)} dW_t^{(i)} (X_{nh} - \hat{X}_n) \\ + N^{1/2} \sum_{\alpha \in B(S_2)} A_\alpha(F, G_1, \ldots, G_d) I_\alpha^W(X, nh, (n+1)h), \end{aligned}$$

$$d_{n+1}(2) := -N^{1/2} \sum_{i,j=1}^{d} \delta G_{i,j} \int_{nh}^{(n+1)h} \int_{nh}^{t} dW_s^{(j)} dW_t^{(i)} X_{nh}.$$

Let the sequence $(\xi_n; n = 0, 1, \ldots, N)$ be given by

$$\xi_0 := 0,$$

$$\xi_{n+1} := \Big(I + Fh + \sum_{i=1}^{d} G_i \Delta W_{n+1}^{(i)}\Big)\xi_n + d_{n+1}(2). \tag{2.5.9}$$

We obtain by subtraction

$$N^{1/2}(X_{(n+1)h} - \hat{X}_{n+1}) - \xi_{n+1} = \Big(I + Fh + \sum_{i=1}^{d} G_i \Delta W_{n+1}^{(i)}\Big)(N^{1/2}(X_{nh} - \hat{X}_n) - \xi_n) + d_{n+1}(1).$$

Again we apply lemma 2.1 and theorem 2.2 to show that

$$N^{1/2}(X_{nh} - \hat{X}_n) - \xi_n = O_{(\infty)}(h^{1/2}).$$

Now

$$E\,\|\xi_{n+1}\|^2 = E\,\|\xi_n\|^2 + E\,\|Y_{n+1}\xi_n\|^2 + E\,\|d_{n+1}(2)\|^2 + 2E\,\langle \xi_n,\, Y_{n+1}\xi_n\rangle + 2E\,\langle \xi_n,\, d_{n+1}(2)\rangle + 2E\,\langle Y_{n+1}\xi_n,\, d_{n+1}(2)\rangle,$$

where

$$Y_{n+1} := Fh + \sum_{i=1}^{d} G_i \Delta W_{n+1}^{(i)}.$$

Now

$$E\,\|d_{n+1}(2)\|^2 = NE \sum_{i,j,k,l=1}^{d} \langle \delta G_{i,j} X_{nh},\, \delta G_{k,l} X_{nh}\rangle \int_{nh}^{(n+1)h} \int_{nh}^{t} dW_s^{(j)}\, dW_t^{(i)} \int_{nh}^{(n+1)h} \int_{nh}^{t} dW_s^{(l)}\, dW_t^{(k)}$$

$$= 2T \sum_{i=1}^{d} E\,\|\delta G_{i,i} X_{nh}\|^2 h + \frac{T}{2} \sum_{\substack{i,j=1 \\ i \neq j}} E\,\|\delta G_{i,j} X_{nh}\|^2 h,$$

and

$$E\,\langle \xi_n,\, Y_{n+1}\xi_n\rangle \geq -KhE\,\|\xi_n\|^2 \qquad K > 0,$$
$$E\,\langle \xi_n,\, d_{n+1}(2)\rangle = 0,$$
$$E\,\langle Y_{n+1}\xi_n,\, d_{n+1}(2)\rangle = 0.$$

The last equality holds because $\Delta W_{n+1}^{(k)}$ and

$$\int_{nh}^{(n+1)h} \int_{nh}^{t} dW_s^{(i)}\, dW_t^{(j)}$$

are uncorrelated for all $i, j, k = 1, 2, \ldots, d$. We have

$$E\,\|\xi_{n+1}\|^2 \geq (1 - Kh)E\,\|\xi_n\|^2 + E\,\|d_{n+1}(2)\|^2;$$

so

$$E\|\xi_N\|^2 \geq \sum_{n=1}^{N}(1-Kh)^{N-n}E\|d_n(2)\|^2$$

$$= \sum_{n=1}^{N}(1-Kh)^{N-n}\Big(2T\sum_{i=1}^{d}E\|\delta G_{i,i}X_{nh}\|^2 + \frac{T}{2}\sum_{i,j=1}^{d}E\|\delta G_{i,j}X_{nh}\|^2\, h$$

$$\to \int_0^T \exp(-K(T-s))\Big(2T\sum_{i=1}^{d}E\|\delta G_{i,i}X_s\|^2 + \frac{T}{2}\sum_{\substack{i,j=1\\ i\neq j}}E\|\delta G_{i,j}X_s\|^2\Big)\,ds$$

> 0 if any $\delta G_{i,j}$ is nonsingular and $X_s \not\equiv 0$.

So, under these conditions,

$$\liminf_N E\|N^{1/2}(X_T - \hat{X}_N)\|^2 > 0.$$

and for $\delta > 0$, $E\|N^{1/2+\delta}(X_T - \hat{X}_N)\|^2$ diverges, which shows that the scheme given by (2.5.7) does not converge with order $1/2 + \delta$ for any $\delta > 0$ (if any $\delta G_{i,j}$ is nonsingular and $X_S \not\equiv 0$).

In the next two chapters we shall be concerned with partition σ-field measurable approximate solutions to bilinear equations, and with this condition the order of convergence is restricted. The $1/2^{th}$-order Taylor scheme uses only $\Delta W_1, \Delta W_2, \ldots, \Delta W_N$ to form an approximation to X_T and is therefore partition-σ-field measurable. The 1^{st} and higher-order Taylor schemes are, in general, not partition σ-field measurable, and so the maximum order of convergence is $1/2$. An important exception to this is the case when all the G_i matrices commute in the sense that $G_iG_j = G_jG_i \;\forall\, i,j = 1,2,\ldots,d$; in which case the 1^{st} order Taylor scheme is partition-σ-field measurable. This is in fact the case in equation (1.2.6).

2.6 Generalisations of results.

Sections 2.3 and 2.5 contain results on the expansion of the solutions of bilinear stochastic differential equations with time-invariant drift and diffusion coefficients, and the convergence of specific approximation schemes to these solutions. These results can easily be generalised to cover a wider class of s.d.e.s, but I have not included the proofs in their most general form since that is not the purpose of this thesis. However, I mention here a generalisation of theorems 2.1 and 2.3.

Consider the s.d.e.

$$\begin{aligned} X_0 &= \xi(\omega) \\ dX_t &= a(t,X_t)\,dt + b(t,X_t)dW_t \qquad t \in [0,T] \end{aligned} \tag{2.6.1}$$

where $X_t \in R^m$, $(W_t \in R^d,\ t \in [0,T])$ is a d-dimensional Wiener process,

$a : [0,T] \times R^m \to R^m$ is a continuous function and

$b : [0,T] \times R^m \to R^{m\times d}$ is a continuous function.

We assume that ξ, a and b fulfil the conditions of an existence and uniqueness theorem such as theorem 6.2.2 in [AR1].

Theorem 2.1 is a special case of theorem 1 in [WP1]; the latter defines Taylor expansions of the solutions of equations such as (2.6.1). To achieve this a differentiability condition on a and b is introduced (see [WP1] theorem 1 (iv)).

Theorem 2 in [WP1] is a similar result to theorem 2.3; it gives the order of convergence of the Taylor schemes when applied to an equation such as (2.6.1), but only in terms of the L_2-norm. It depends on a square integrability condition ([WP1] theorem 2 (iii) and (iv)). By strengthening this condition to p^{th}-power integrability one can easily prove the appropriate order of convergence of the Taylor schemes in terms of the L_p-norm.

Chapter 3

The Optimal Asymptotic Behaviour of Approximation Schemes.

3.1 Introduction.

In this chapter we shall look at the maximum rate of convergence of discrete approximate solutions to the Markov-chain filtering problem, introduced in section 1.2. This will be dealt with in two parts: first the convergence of discrete approximate solutions to the bilinear s.d.e. (1.2.6), and secondly the convergence of the corresponding Bayes estimates obtained by formula (1.2.8).

The main results of this chapter are in sections 3.5 and 3.6. We show in theorem 3.1 that the normalised error in any partition-σ-field measurable approximate solution to (1.2.6), at best converges in distribution to a normal random variable with zero-mean and given covariance matrix. This result is modelled on, and is in part a corollary of, theorem 2 of Clark in [CL2], but the proof presented here is more direct in that it uses the discrete-approximation results of chapter 2 rather than the approximating o.d.e. approach used by Clark. A similar result is obtained in section 3.6 concerning the convergence of the normalised error in the corresponding approximate Bayes estimates of functions of the signal process.

The rate of convergence of discrete approximate solutions to (1.2.6) is limited by the requirement that such solutions are partition-σ-field measurable. This problem is introduced by Clark and Cameron in [CC1]. It is fairly well-known that, in terms of the L_2-norm of the error at the partition points, in general the maximum order of convergence that can be achieved under this restriction is 1. A lower maximum order of convergence (1/2) results if we use interpolation of the approximate solution between the partition points, and look at the supremum of the L_2-norm of the error over the whole time interval.

Rootzén in [RO1] proves a weak-convergence result for approximations to stochastic integrals considered as processes. His results reflect the lower order of convergence of such functional approximations.

The maximum order of convergence would be only 1/2 if the diffusion coefficients in (1.2.6), $B_1, B_2, \ldots, B_d$, did not commute i.e. if $B_iB_j \neq B_jB_i$ for some $i,j \in \{1,2,\ldots,d\}$ (see [CC1]). In the present case these coefficients are diagonal matrices and do commute.

In section 3.3 we introduce an absolutely continuous transformation of probability measure which enables us to use the tools developed in chapter 2 for the analysis of the bilinear s.d.e. (1.2.6). This is a standard Girsanov transformation and yields a new probability measure P_2, with respect to which the observations process $(W_t,\ t \in [0,T])$ is a Brownian motion.

The best "approximation scheme" for solving (1.2.6) in terms of the L_2-norm under P_2 is the partition-σ-field-conditional mean under P_2. This scheme also yields, by substitution into (1.2.8), asymptotically optimal approximate Bayes estimates of functions of the signal-process.

Section 3.4 deals with a few preparatory lemmas, which are used in the proof of theorem 3.1. Lemma 3.4 derives the maximum uniform $L_{(\infty)}$-order of convergence of discrete approximate solutions to (1.2.6).

3.2 The equation for the un-normalised conditional density.

We shall start this chapter by considering discrete approximate solutions to the bilinear s.d.e. (1.2.6), which occurs in the optimal filtering equations for a Markov chain. I have re-written this equation below for convenience; all random elements are defined on the probability space $(\Omega, \mathcal{F}, P_1)$.

$$q_0 = p_0,$$

$$dq_t = Aq_t dt + \sum_{i=1}^{d} B_i q_t dW_t^{(i)} \qquad t \in [0,T], \tag{3.2.1}$$

$$dW_t^{(i)} = h_i(X_t)\,dt + d\beta_t^{(i)} \qquad i = 1,2,\ldots,d \quad t \in [0,T]. \tag{3.2.2}$$

$(X_t,\ t \in [0,T])$ is the Markov-chain signal process that we are trying to estimate, $(\beta_t,\ t \in [0,T])$ is a d-dimensional Brownian motion process and h is an "observations" function:

$$h : R^m \to R^d.$$

As was mentioned in chapter 1, we are interested in methods of solving (3.2.1) that depend on the values of the driving term $(W_t,\ t \in [0,T])$ at the points of regular partitions of the time interval. We can formalise this as follows.

Let Π be the class of *regular partitions* of the interval $[0,T]$:

$$\Pi := \{\pi_N = (0,h,2h,\ldots,Nh) : Nh = T\ ;\ N = 1,2,\ldots\}. \tag{3.2.3}$$

We denote by $\mathcal{M}$ the class of approximation methods of interest. Any element of $\mathcal{M}$ is a discrete-parameter process on $(\Omega, \mathcal{F}, P_1)$:

$$\mu \in \mathcal{M}$$
$$\mu : (\Omega, \mathcal{F}, P_1) \times \mathcal{N} \to R^m$$

$\mu(\cdot, N)$ is an approximation to q_T. μ is measurable with respect to the *partition-σ-field* $\mathcal{P}_N$, given by

$$\mathcal{P}_N := \sigma\{W_h, W_{2h}, \ldots, W_{Nh}\} \qquad N = 1,2,\ldots, \tag{3.2.4}$$

i.e. $\mu(\cdot, N)$ is $\mathcal{P}_N$-measurable for $N = 1,2,\ldots$.

In order to compare different methods belonging to $\mathcal{M}$ we must use some measure of the approximation error $q_T - \mu(\cdot, N)$. One such measure is the conditional L_2-norm,

$$\|q_T - \mu(\cdot, N)\|_{L_2} = \Big(E_1\Big(\sum_{i=1}^{m} (q_T^{(i)} - \mu^{(i)}(\cdot, N))^2 \mid \mathcal{P}_N\Big)\Big)^{1/2}. \tag{3.2.5}$$

Here E_1 denotes taking mathematical expectation with respect to P_1.

It is natural to use the ρ_N-*conditional* L_p-norms of the error because these yield a "measure-of-error" for (almost) all realisations of the observations $(W_h, W_{2h}, \ldots, W_{Nh})$. Loosely speaking, the randomness of these norms is the observed randomness, i.e. that which after observation of $(W_h, W_{2h}, \ldots, W_{Nh})$ is no longer random.

It is well-known that the scheme which (almost surely) minimises the norm given in (3.2.5) in $\mathcal{M}$ is the ρ_N-conditional mean $E_1(q_T \mid \rho_N)$. Sections 3.4 and 3.5 are devoted to analysing the convergence of this optimal scheme. This is achieved by first analysing the convergence of the conditional mean under a new probability measure, which will be introduced in the next section.

3.3 An absolutely continuous transformation of measure.

We would like to be able to use the tools developed in chapter 2 for the analysis of bilinear s.d.e.s driven by Wiener processes, to analyse equation (3.2.1). The problem, of course, is that the driving term in (3.2.1) (W_t) is a *drifting* Wiener process. This problem can be overcome by performing calculations in terms of a new probability measure P_2, under which (W_t) is a Wiener process, and then translating the results into corresponding results under the original measure P_1. The new measure is defined by a standard Girsanov transformation.

Consider the following, in which $(X_t), (B_t), h(\cdot)$ and $(a_i\ ,\ i = 1,2,\ldots,m)$ are as given in theorem 1.1.

$$\sup_{t\in[0,T]} \|h(X_t)\| \leq \max_{i=1,2,\ldots,m} \|h(a_i)\| < \infty;$$

so

$$P_1\Big(\int_0^T \|h(X_t)\|^2 dt < \infty\Big) = 1 \tag{3.3.1}$$

and

$$E_1\Big(\exp\Big(\frac{1}{2}\int_0^T \|h(X_t)\|^2 dt\Big)\Big) < \infty. \tag{3.3.2}$$

We define the pair $(\mu_t, \mathcal{F}_t^{\beta,X})$ by the following:

$$\mu_t := \exp\Big(\int_0^t -h(X_s)^T d\beta_s - \frac{1}{2}\int_0^t \|h(X_s)\|^2 ds\Big),$$

$$\mathcal{F}_t^{\beta,X} := \sigma(\beta_s, X_s\,;\ s \leq t). \tag{3.3.3}$$

A lemma due to Novikov (see theorem 6.1 in [LS1]) shows that $(\mu_t, \mathcal{F}_t^{\beta,X})$ is a martingale under P_1.

We define the new measure P_2 by,

$$P_2(A) := \int_A \mu_T dP_1 \quad \forall\, A \in \mathcal{F}. \tag{3.3.4}$$

P_2 is a probability measure since $P_2(\Omega) = E_1\mu_T = 1$.

The multidimensional version of the Girsanov theorem (see theorem 6.4 in [LS1]) shows that (W_t) is a d-dimensional Wiener process under P_2; furthermore with this probability it is independent of (X_t) (see lemma 3 in [DM1]).

Clearly

$$\mu_t > 0 \qquad \forall\, t \in [0,T] \qquad \text{(a.s.)}(P_1)$$

so P_1 and P_2 are mutually absolutely continuous, and any property that holds *almost surely* under one measure also holds *almost surely* under the other. The boundedness of $h(X_t)$ also implies that

$$\begin{aligned} M_T &:= \frac{1}{\mu_T} \\ &= \exp\Big(\int_0^T h(X_s)^T d\,W_s - \frac{1}{2}\int_0^T \|h(X_s)\|^2 d\,s\Big) \end{aligned} \tag{3.3.5}$$

has moments of all orders under P_2. But

$$dP_1 = M_T dP_2 \tag{3.3.6}$$

and it follows from Hölder's inequality that if $0 < \delta \leq p$,

$$\begin{aligned} E_1\|X_{n,N}\|^{p-\delta} &= E_2 M_T \|X_{n,N}\|^{p-\delta} \\ &\leq (E_2 M_T^{\frac{p}{\delta}})^{\frac{\delta}{p}} (E_2\|X_{n,N}\|^p)^{\frac{p-\delta}{p}}; \end{aligned} \tag{3.3.7}$$

consequently if the sequence $(X_{n,N})$ is of uniform L_p-order r under P_2 then it is also of uniform $L_{p-\delta}$-order r under P_1. In particular the $L_{(\infty)}$-uniform order of $(X_{n,N})$ is the same for both P_1 and P_2.

In sections 3.5 and 3.6 we shall be interested in the weak convergence of ρ_N-conditional distributions of suitably normalised sequences of errors, and in the convergence of their ρ_N-conditional moments. The analysis will be done initially in terms of P_2; the following lemma will be used to obtain corresponding results under P_1.

Lemma 3.1. *Let $(X_N \in R^d : N = 1,2,\ldots)$ be a process on $(\Omega,\mathcal{F})$ such that X_N is $\mathcal{F}_T^W$-measurable for all N. Let P_1 and P_2 be the equivalent probability measures introduced above.*
(i) If, with probability one, (P_1 or P_2), the sequence $(P_2(X_N \in \cdot \mid \rho_N)\,;\, N = 1,2,\ldots)$ of conditional distributions on $(R^d,\mathcal{B}^d)$ converges weakly, i.e. if

$$P_2(X_N \in \cdot \mid \rho_N) \Rightarrow \Pi(\cdot) \quad \text{(a.s.)}, \qquad N \to \infty \tag{3.3.10}$$

where Π is some $\mathcal{F}_T^W$-conditional distribution on $(R^d,\mathcal{B}^d)$, then

$$P_1(X_{N_n} \in \cdot \mid \rho_{N_n}) \Rightarrow \Pi(\cdot) \quad \text{(a.s.)}, \qquad n \to \infty \tag{3.3.11}$$

where $(N_1,N_2,\ldots)$ is any sequence of natural numbers with the following "refining" property:

$$N_n \longrightarrow \infty \text{ and } N_n/N_m \text{ is a natural number if } n \geq m. \tag{3.3.12}$$

(The idea here is that $\mathcal{P}_{N_n} \supseteq \mathcal{P}_{N_m}$*).*

(ii) If for some $p > 0$

$$P_2(\limsup_N E_2(\|X_N\|^{2p} \mid \mathcal{P}_N) < \infty) = 1$$

and

$$E_2(\|X_N\|^p \mid \mathcal{P}_N) \to \xi \quad \text{(a.s.)}, \tag{3.3.13}$$

where ξ *is some* $\mathcal{F}_T^W$*-measurable random variable, then*

$$E_1(\|X_{N_n}\|^p \mid \mathcal{P}_{N_n}) \to \xi \quad \text{(a.s.)}, \quad n \to \infty \tag{3.3.14}$$

where $(N_1, N_2, \ldots)$ *has the property (3.3.12).*

PROOF. Let M_T be defined by (3.3.5). We have already shown that all the moments of M_T exist under both measures.

The strict positivity of M_T allows us to define

$$M_N := E_2(M_T \mid \mathcal{P}_N) \quad \text{for } N = 1, 2, \ldots, \infty, \tag{3.3.15}$$

with $M_N(\omega) > 0$ for all $\omega \in \Omega$, where

$$\mathcal{P}_\infty := \sigma\Big(\bigcup_{n=1}^{\infty} \mathcal{P}_{N_n}\Big).$$

It is easy to show that

$$\mathcal{P}_\infty = \mathcal{F}_T^W.$$

Since, for a sequence $(N_1, N_2, \ldots)$ with the refining property (3.3.12),

$$\mathcal{P}_{N_n} \supseteq \mathcal{P}_{N_m} \quad \text{if } n \geq m$$

it follows that $(M_{N_n}, \mathcal{P}_{N_n})$ is a martingale under P_2 and a theorem of Lévy (theorem 1.5 in [LS1]) shows that

$$M_{N_n} \to M_\infty = M_T \quad \text{(a.s.)}, \tag{3.3.16}$$

and

$$E_2((M_\infty - M_{N_n})^2 \mid \mathcal{P}_{N_n}) = E_2(M_\infty^2 \mid \mathcal{P}_{N_n}) - M_{N_n}^2 \to 0 \quad \text{(a.s.)}. \tag{3.3.17}$$

To prove (i) we note that for a set $S \in \mathcal{P}_{N_n}$ and a vector $c \in R^d$

$$\int_S M_{N_n} E_1(\exp(ic^T X_{N_n}) \mid \rho_{N_n}) dP_2 = \int_S M_T E_1(\exp(ic^T X_{N_n}) \mid \rho_{N_n}) dP_2$$

$$= \int_S E_1(\exp(ic^T X_{N_n}) \mid \rho_{N_n}) dP_1$$

$$= \int_S \exp(ic^T X_{N_n}) dP_1$$

$$= \int_S \exp(ic^T X_{N_n}) M_T dP_2$$

$$= \int_S E_2(M_T \exp(ic^T X_{N_n}) \mid \rho_{N_n}) dP_2$$

and so

$$\begin{aligned} M_{N_n} E_1(\exp(ic^T X_{N_n}) \mid \rho_{N_n}) &= E_2(M_T \exp(ic^T X_{N_n}) \mid \rho_{N_n}) \\ &= M_{N_n} E_2(\exp(ic^T X_{N_n}) \mid \rho_{N_n}) \\ &\quad + E_2((M_T - M_{N_n}) \exp(ic^T X_{N_n}) \mid \rho_{N_n}). \end{aligned} \tag{3.3.18}$$

Now

$$\begin{aligned} |E_2((M_T - M_{N_n}) \exp(ic^T X_{N_n}) \mid \rho_{N_n})| &= |E_2((M_\infty - M_{N_n}) \exp(ic^T X_{N_n}) \mid \rho_{N_n})| \\ &\leq E_2(|M_\infty - M_{N_n}| \mid \rho_{N_n}) \\ &\to 0 \quad \text{(a.s.)} \quad \text{by (3.3.17),} \end{aligned} \tag{3.3.19}$$

and since, by (3.3.16)

$$P_2(\liminf_n M_{N_n} > 0) = 1,$$

(3.3.18) and (3.3.19) show that

$$E_1(\exp(ic^T X_{N_n}) \mid \rho_{N_n}) - E_2(\exp(ic^T X_{N_n}) \mid \rho_{N_n}) \to 0 \quad \text{(a.s.)}.$$

This shows that, with probability one, the ρ_{N_n}-conditional characteristic-functions of X_{N_n} under the measure P_1 converge to the $\mathcal{I}_T^W$-conditional characteristic function corresponding to $\Pi(\cdot)$ at all points c with rational co-ordinates. (3.3.11) follows by the continuity properties of characteristic-functions.

The proof of (ii) is similar. We note that for $S \in \rho_{N_n}$

$$\int_S E_1(\|X_{N_n}\|^p \mid \rho_{N_n}) M_{N_n} dP_2 = \int_S E_2(M_T \|X_{N_n}\|^p \mid \rho_{N_n}) dP_2,$$

and so

$$\begin{aligned} M_{N_n} E_1(\|X_{N_n}\|^p \mid \rho_{N_n}) &= M_{N_n} E_2(\|X_{N_n}\|^p \mid \rho_{N_n}) + E_2((M_T - M_{N_n}) \|X_{N_n}\|^p \mid \rho_{N_n}) \\ &= M_{N_n} E_2(\|X_{N_n}\|^p \mid \rho_{N_n}) + E_2((M_\infty - M_{N_n}) \|X_{N_n}\|^p \mid \rho_{N_n}), \end{aligned}$$

We apply Hölder's inequality to the second term on the right-hand side of this expression and use (3.3.17) to prove part (ii) of the lemma. ∎

3.4 The maximum order of convergence of discrete methods.

In this section we shall look at the maximum order of convergence that can be achieved by partition-σ-field-measurable approximate solutions to equation (3.2.1).

Under the measure P_2 this equation is a Wiener-process-driven bilinear s.d.e. of the type dealt with in chapter 2. We noted in section 2.5 that if the diffusion coefficients in an equation of this type commute then the 1^{st} order Taylor approximation scheme introduced in definition 2.3 yields a partition-σ-field-measurable approximate solution. Since the diffusion coefficients $B_1, B_2, \ldots, B_d$ of equation (3.2.1) *do* commute the maximum $L_{(\infty)}$-order of convergence in this case is at least 1.

It is well-known that the maximum L_2-order of convergence of approximations to this equation is 1 and that this order is achieved by the best (L_2) scheme, the conditional mean $E_2(q_T \mid \rho_N)$ (see [CC1]). In fact the conditional means under either measure P_1 or P_2, $E_1(q_T \mid \rho_N)$ and $E_2(q_T \mid \rho_N)$ both achieve the maximum L_p-order of convergence for all p under their respective measures; this maximum order being 1. This is proved in lemma 3.4.

The following two lemmas will be used in the proofs of lemma 3.4 and theorem 3.1. The first lists some properties of the *Brownian-bridge process.*

Definition 3.1. The following process $(\widetilde{W}_t)$ defined on the probabilty space $(\Omega, \mathcal{F}, P_2)$ is called a *Brownian-bridge process*:

$$\widetilde{W}_t := W_t - \left(W_{nh} + \frac{(W_{(n+1)h} - W_{nh})}{h}(t - nh)\right). \tag{3.4.1}$$

Lemma 3.2. *The following properties relating to the Brownian-bridge process hold:*
(i)

$$W_{nh} + \frac{(W_{(n+1)h} - W_{nh})}{h}(t - nh) \text{ is a version of } E_2(W_t \mid \rho_N), \tag{3.4.2}$$

(ii) $(\widetilde{W}_t, t \in [0, T])$ *is Gaussian with zero-mean, and covariance matrix given by*

$$E_2 \widetilde{W}_t \widetilde{W}_s^T = \begin{cases} \frac{1}{h}((n+1)h - t)(s - nh) I & nh \le s \le t \le (n+1)h \\ 0 & \text{otherwise.} \end{cases} \tag{3.4.3}$$

(iii) The following σ*-fields are independent:*

$$\sigma(\widetilde{W}_s, s \in [0, h]), \sigma(\widetilde{W}_s, s \in [h, 2h]), \ldots, \sigma(\widetilde{W}_s, s \in [(N-1)h, Nh]),$$
$$\sigma(W_h), \sigma(W_{2h} - W_h), \ldots, \sigma(W_{Nh} - W_{(N-1)h}).$$

(iv) The discrete-time process

$$\left(N \int_{nh}^{(n+1)h} \widetilde{W}_t \, dt, n = 0, 1, \ldots, N-1\right)$$

is normally-distributed with zero mean, and covariance matrix given by

$$E_2\Big(N\int_{nh}^{(n+1)h}\widetilde{W}_t dt\Big)\Big(N\int_{mh}^{(m+1)h}\widetilde{W}_t dt\Big)^T = \frac{1}{12}T^2hI \qquad \text{if } m = n,$$

$$0 \qquad \text{otherwise.} \tag{3.4.4}$$

PROOF OF (i). We shall denote by ΔW_{n+1} the quantity $W_{(n+1)h} - W_{nh}$. By standard properties of the Wiener process, for $t \in [nh, (n+1)h]$, the following $(N+1)d$-vector has the given distribution:

$$[\Delta W_1^T \ldots \Delta W_n^T \, \Delta W_{n+2}^T \ldots \Delta W_N^T (W_t - W_{nh})^T \, (W_{(n+1)h} - W_t)^T]^T \sim N(0, V), \tag{3.4.5}$$

where

$$V = \operatorname{diag}\{\underbrace{h, \ldots, h}_{(N-1)d \text{ terms}}, \underbrace{(t-nh), \ldots, (t-nh)}_{d \text{ terms}}, \underbrace{((n+1)h-t), \ldots, ((n+1)h-t)}_{d \text{ terms}}\}.$$

Now

$$E_2(W_t \mid \rho_N) = \frac{\int_{R^d} xn(0,V)([\Delta W_1^T \ldots \Delta W_n^T \, \Delta W_{n+2}^T \ldots \Delta W_N^T (x - W_{nh})^T \, (W_{(n+1)h} - x)^T]^T)\, dx}{\int_{R^d} n(0,V)([\Delta W_1^T \ldots \Delta W_n^T \, \Delta W_{n+2}^T \ldots \Delta W_N^T (x - W_{nh})^T \, (W_{(n+1)h} - x)^T]^T)\, dx}, \tag{3.4.6}$$

where $n(0, V)$ is the normal density with mean 0 and covariance V. Evaluating this expression we obtain the result (i).

PROOF OF (ii). From (3.4.1), for $t \in [nh, (n+1)h]$,

$$\widetilde{W}_t = \frac{((n+1)h - t)}{h}(W_t - W_{nh}) - \frac{(t-nh)}{h}(W_{(n+1)h} - W_t). \tag{3.4.7}$$

The properties listed follow directly from the properties of the Wiener process.

PROOF OF (iii). Since the increments of a Wiener process are independent it is sufficient to prove that $\sigma(\widetilde{W}_s, s \in [0,h])$ and W_h are independent. For $t \in [0,h]$,

$$\begin{aligned} E_2(W_h \widetilde{W}_t^T) &= E_2\Big(W_h\Big(\frac{(h-t)}{h}W_t - \frac{t}{h}(W_h - W_t)\Big)\Big) \\ &= 0. \end{aligned}$$

Since W_h and $(\widetilde{W}_s, s \in [0,h])$ are jointly-Gaussian they are independent.

PROOF OF (iv). That the process

$$\Big(N\int_{nh}^{(n+1)h}\widetilde{W}_t dt\,,\ n = 0, 1, \ldots, N-1\Big)$$

is normally-distributed with zero mean is obvious. Now

$$E_2\Big(N\int_{nh}^{(n+1)h}\widetilde{W}_t dt\Big)\Big(N\int_{mh}^{(m+1)h}\widetilde{W}_t dt\Big)^T = N^2 E_2\int_{nh}^{(n+1)h}\int_{mh}^{(m+1)h}\widetilde{W}_t\widetilde{W}_s dt\, ds. \tag{3.4.8}$$

The result follows from (3.4.3). ∎

Lemma 3.3. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$ be independent sub-σ-fields of $\mathcal{F}$. For any integrable random variable $X \in R^m$.*

$$E(E(X \mid \mathcal{A} \vee \mathcal{B}) \mid \mathcal{A} \vee \mathcal{C}) = E(E(X \mid \mathcal{A} \vee \mathcal{C}) \mid \mathcal{A} \vee \mathcal{B}) = E(X \mid \mathcal{A}). \tag{3.4.9}$$

PROOF. Let P_A, P_B and P_C be the restrictions of P to $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ respectively. We consider the product space

$$(\overline{\Omega}, \overline{\mathcal{F}}, \overline{P}) = (\Omega \times \Omega \times \Omega, \mathcal{A} \times \mathcal{B} \times \mathcal{C}, P_A \times P_B \times P_C) \tag{3.4.10}$$

and the mapping

$$T : (\Omega, \mathcal{F}, P) \to (\overline{\Omega}, \overline{\mathcal{F}}, \overline{P}),$$

given by

$$T(\omega) := (\omega, \omega, \omega). \tag{3.4.11}$$

It is easy to show that T is measurable and that

$$P(T^{-1}(\overline{F})) = \overline{P}(\overline{F}) \qquad \text{for all sets } \overline{F} \in \overline{\mathcal{F}}. \tag{3.4.12}$$

This relies on the independence of $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$.

Consider the class of subsets of $\overline{\Omega}$

$$\tau := \{A \times B \times \Omega : A \in \mathcal{A}, B \in \mathcal{B}\}.$$

This is a *π-system* which generates the σ-field

$$\{\overline{F} = \overline{G} \times \Omega : \overline{G} \in \mathcal{A} \times \mathcal{B}\};$$

also $T^{-1}(\tau)$ is a *π-system* which generates $\mathcal{A} \vee \mathcal{B}$, and so

$$T^{-1}\{\overline{F} = \overline{G} \times \Omega : \overline{G} \in \mathcal{A} \times \mathcal{B}\} = \mathcal{A} \vee \mathcal{B}. \tag{3.4.13}$$

Since T is 1–1 and measurable, any integrable random variable, X, on Ω can be represented as the composition with T of an integrable random variable, Y, on $\overline{\Omega}$:

$$X(\omega) = Y(T(\omega)). \tag{3.4.14}$$

It is easy to show, by using (3.4.12) and (3.4.13), that

$$E(X \mid \mathcal{A} \vee \mathcal{B}) = Y_C(T) \qquad \text{(a.s.)} \tag{3.4.15}$$

where

$$Y_C(\omega_A, \omega_B, \omega_C) := \int_{\Omega} Y(\omega_A, \omega_B, \omega_C) d P_C(\omega_C).$$

Similar arguments show that

$$E(E(X \mid \mathcal{A} \vee \mathcal{B}) \mid \mathcal{A} \vee \mathcal{C}) = (Y_C)_B(T) \quad \text{(a.s.)}, \tag{3.4.16}$$
$$E(E(X \mid \mathcal{A} \vee \mathcal{C}) \mid \mathcal{A} \vee \mathcal{B}) = (Y_B)_C(T) \quad \text{(a.s.)}, \tag{3.4.17}$$

and

$$E(X \mid \mathcal{A}) = Y_{BC}(T) \quad \text{(a.s.)}, \tag{3.4.18}$$

where

$$(Y_C)_B := \int_\Omega \int_\Omega Y \, dP_C(\omega_C) \, dP_B(\omega_B),$$

$$(Y_B)_C := \int_\Omega \int_\Omega Y \, dP_B(\omega_B) \, dP_C(\omega_C)$$

and

$$Y_{BC} := \int_{\Omega \times \Omega} Y \, d(P_B \times P_C)(\omega_B, \omega_C).$$

(3.4.9) follows by Fubini's theorem. ∎

NOTE. This will be used for the space $(\Omega, \mathcal{I}, P_2)$ with

$$\mathcal{A} := \sigma(W_h, W_{2h}, \ldots, W_{nh})$$
$$\mathcal{B} := \sigma(\widetilde{W}_t, t \leq nh)$$
$$\mathcal{C} := \sigma(W_{(n+1)h}, W_{(n+2)h}, \ldots, W_{Nh})$$

to show that, for an integrable random variable X,

$$E_2(E_2(X \mid \mathcal{I}_{nh}) \mid \rho_N) = E_2(E_2(X \mid \rho_N) \mid \mathcal{I}_{nh}). \tag{3.4.19}$$

Next we prove the uniform-$L_{(\infty)}$-order-1 convergence of the conditional mean sequences

$$((E_1(q_{nh} \mid \rho_N); n = 0, 1, \ldots, N); N = 1, 2, \ldots)$$

and

$$((E_2(q_{nh} \mid \rho_N); n = 0, 1, \ldots, N); N = 1, 2, \ldots).$$

Lemma 3.4. *The errors in the conditional mean sequences above are of uniform $L_{(\infty)}$-order 1, i.e.*

$$(q_{nh} - E_1(q_{nh} \mid \rho_N)) = O_{(\infty)}(h), \tag{3.4.21}$$
$$(q_{nh} - E_2(q_{nh} \mid \rho_N)) = O_{(\infty)}(h). \tag{3.4.22}$$

PROOF. We prove (3.4.22) first. All random elements are defined on the space $(\Omega, \mathcal{I}, P_2)$. We expand q about the points of the partition up to terms of order 1 using theorem 2.1. For brevity we shall use

the following contracted notation. For an integrable random variable X:

$$\begin{aligned} &\overline{X} \text{ denotes } E_2(X \mid \rho_N), \\ &\widetilde{X} \text{ denotes } X - \overline{X}. \end{aligned} \tag{3.4.23}$$

(This is consistent with the notation $\widetilde{W}_t$ used in lemma 3.2 as $\widetilde{W}_t$ is a version of $W_t - \overline{W}_t$.
For $\alpha \in M$ (see section 2.2)

$$\begin{aligned} A_\alpha &:= A_\alpha(A, B_1, B_2, \ldots, B_d), \\ I_\alpha &:= I_\alpha^W(nh, (n+1)h), \\ I_\alpha(q) &:= I_\alpha^W(q, nh, (n+1)h). \end{aligned} \tag{3.4.24}$$

We obtain by theorem 2.1

$$q_{(n+1)h} = \sum_{\alpha \in S_2} A_\alpha I_\alpha q_{nh} + \sum_{\alpha \in B(S_2)} A_\alpha I_\alpha(q), \tag{3.4.25}$$

where S_2 and $B(S_2)$ are the sets introduced in section 2.5:

$$\begin{aligned} S_2 &= \{\alpha \in M : l(\alpha) + n(\alpha) \leq 2\} \\ &= \{\nu, (0), (i), (i,j) : i,j = 1,2,\ldots,d\}, \\ B(S_2) &= \{\alpha \in M - S_2 : -\alpha \in S_2\} \\ &= \{(0,0), (0,i), (0,i,j), (i,j,k) : i,j,k = 1,2,\ldots,d\}. \end{aligned}$$

We take the ρ_N-conditional mean of both sides of (3.4.25)

$$\overline{q_{(n+1)h}} = \sum_{\alpha \in S_2} A_\alpha I_\alpha \overline{q_{nh}} + \sum_{\alpha \in B(S_2)} A_\alpha \overline{I_\alpha(q)}. \tag{3.4.26}$$

Here we have used the fact that the I_α are ρ_N-measurable if $\alpha \in S_2$.

Subtraction of (3.4.26) from (3.4.25) yields

$$\begin{aligned} \widetilde{q}_{(n+1)h} &= \sum_{\alpha \in S_2} A_\alpha I_\alpha \widetilde{q}_{nh} + \sum_{\alpha \in B(S_2)} A_\alpha \widetilde{I_\alpha(q)} \\ &= (I + Y_{n+1})\widetilde{q}_{nh} + d_{n+1}, \end{aligned} \tag{3.4.27}$$

where

$$Y_{n+1} := \sum_{\alpha \in S_2} A_\alpha I_\alpha - I$$

and

$$d_{n+1} := \sum_{\alpha \in B(S_2)} A_\alpha \widetilde{I_\alpha(q)}.$$

Now Y_{n+1} is independent of $\mathfrak{I}_{nh}$, lemma 2.1 (iii) shows that it is of uniform $L_{(\infty)}$-order $1/2$ and lemma 2.1 (iii) and (iv) shows that

$$|E_2(Y_{n+1})| \leq Kh, \qquad \text{where } K \geq 0.$$

Lemma 2.1 (iii) shows that d_{n+1} is of uniform $L_{(\infty)}$-order 3/2; now

$$\begin{aligned} E_2(d_{n+1} \mid \mathcal{I}_{nh}) &= \sum_{\alpha \in B(S_2)} A_\alpha E_2(\widetilde{I_\alpha(q)} \mid \mathcal{I}_{nh}) \\ &= \sum_{\alpha \in B(S_2)} A_\alpha \widetilde{E_2(I_\alpha(q) \mid \mathcal{I}_{nh})}, \qquad \text{by lemma 3.3,} \\ &= \sum_{\substack{\alpha \in B(S_2) \\ l(\alpha) = n(\alpha)}} A_\alpha \widetilde{E_2(I_\alpha(q) \mid \mathcal{I}_{nh})}, \qquad \text{by lemma 2.1 (iv),} \\ &= O_{(\infty)}(h^2), \qquad \text{by lemma 2.1 (iii).} \end{aligned}$$

Application of theorem 2.2 to equation (3.4.27) proves (3.4.22). Now

$$\begin{aligned} \|E_1(q_{nh} \mid \rho_N) - E_2(q_{nh} \mid \rho_N)\|^p &= \|E_1(q_{nh} - E_2(q_{nh} \mid \rho_N)) \mid \rho_N)\|^p \\ &\le E_1(\|q_{nh} - E_2(q_{nh} \mid \rho_N)\| \mid \rho_N)^p \\ &\le E_1(\|q_{nh} - E_2(q_{nh} \mid \rho_N)\|^p \mid \rho_N) \quad \forall\, p \ge 1 \end{aligned}$$

by Jensen's inequality. So

$$(E_1 \|q_{nh} - E_1(q_{nh} \mid \rho_N)\|^p)^{1/p} \le 2(E_1 \|q_{nh} - E_2(q_{nh} \mid \rho_N)\|^p)^{1/p} \quad \forall\, p \ge 1$$

and

$$\sup_{n \le N} E_1(N \|q_{nh} - E_1(q_{nh} \mid \rho_N)\|)^p < \infty \quad \forall\, p > 0.$$

This proves (3.4.21). ∎

3.5 An optimal limit-distribution for normalised error sequences.

In the last section we saw that the partition-σ-field-conditional mean of the solution of (3.2.1) under the measure P_2 converges with optimal order in terms of all the L_p-norms. In fact the conditional mean is asymptotically optimal in a stronger sense than the order of convergence as we shall see in this section.

The error in the conditional mean is of uniform $L_{(\infty)}$-order 1, which means in particular that the "normalised" error sequence $(N(q_T - E_2(q_T \mid \rho_N)\,;\ N = 1, 2, \ldots)$ is bounded in L_p for all p. The question arises; does this normalised sequence converge in any sense and if so to what does it converge?

I mentioned in section 3.2 that it is the ρ_N-*conditional* properties of the error that we are interested in since they provide us with a sample-wise analysis of the error; in particular we are interested in the ρ_N-conditional distribution of the error. The following theorem shows that the ρ_N-conditional distribution and momen·ts of the normalised error converge to a normal distribution and moments, parameterised by the sample paths of W_t. It is based on theorem 2 in [CL2] although it is proved using the discrete-approximation results of chapter 2. In [CL2] Clark approximates the solution of a stochastic differential equation by the solution of an ordinary differential equation driven by a piecewise-linear approximation of a Brownian motion process in much the same way as Wong

and Zakai [WZ1]: the difference is that he includes an extra Lie-bracket term, which increases the efficiency of the approximation.

The limit-distribution obtained in theorem 3.1 has the same form as that obtained in [CL2]. In the present case, because of the simplicity of the bilinear equation (3.2.1), we are able to show that all the conditional moments of the normalised error converge to those of the limit-distribution.

Theorem 3.1. *Let* $(V_t, t \in [0,T])$ *be a* d*-dimensional Wiener process on* $(\Omega, \mathcal{F}, P)$, $P = P_1$ *or* P_2, *independent of* $(W_t, t \in [0,T])$ *(see note 1 below).*

(i) Under the measure P_2, *with probability one, the* ρ_N*-conditional distibution of the normalised error* $N(q_T - E_2(q_T \mid \rho_N))$ *converges weakly to the* $\mathcal{F}_T^W$*-conditional distribution of the random variable* z_T, *where* z_T *is given by:*

$$z_T := \frac{T}{\sqrt{12}} \sum_{i=1}^{d} \int_0^T \Phi(s,T)(A B_i - B_i A)\Phi(0,s) q_0 d V_s^{(i)}. \tag{3.5.1}$$

$\Phi(\cdot\,,\cdot\,)$ *is the fundamental solution of equation (3.2.1), i.e.*

$$\begin{aligned} &\Phi(s,s) = I \quad \forall\, s \in [0,T] \\ &d\Phi(s,t) = A\Phi(s,t)dt + \sum_{i=1}^{d} B_i \Phi(s,t) d W_t^{(i)} \qquad 0 \le s \le t \le T. \end{aligned} \tag{3.5.2}$$

Furthermore, all the ρ_N*-conditional moments converge almost surely.*

(ii) Let $(N_1, N_2, \ldots)$ *be any sequence of natural numbers with the refining property (3.3.12), that is:*

$$N_i \longrightarrow \infty \quad \text{and} \quad i \ge j \Rightarrow N_i/N_j \text{ is a natural number.}$$

Under the measure P_1, *with probability 1, the* ρ_{N_i}*-conditional distributions of the normalised errors* $N_i(q_T - E_2(q_T \mid \rho_{N_i}))$ *and* $N_i(q_T - E_1(q_T \mid \rho_{N_i}))$ *converge weakly to the* $\mathcal{F}_T^W$*-conditional distribution of* z_T. *Furthermore the* ρ_{N_i}*-conditional moments converge almost surely.*

(iii) z_T *is the final value of the solution of the following s.d.e. scaled by* T*:*

$$\begin{aligned} &\theta_0 = 0, \\ &d\theta_t = A\theta_t dt + \sum_{i=1}^{d} B_i \theta_t d W_t^{(i)} + \frac{1}{\sqrt{12}} \sum_{i=1}^{d} (A B_i - B_i A) q_t d V_t^{(i)}, \qquad t \in [0,T], \\ &z_T = T\theta_T. \end{aligned} \tag{3.5.3}$$

(see note 3 below)

θ can also be represented as part of the solution of the following autonomous equation:

$$\begin{bmatrix} q_0 \\ \theta_0 \end{bmatrix} = \begin{bmatrix} p_0 \\ 0 \end{bmatrix},$$

$$d\begin{bmatrix} q_t \\ \theta_t \end{bmatrix} = A'\begin{bmatrix} q_t \\ \theta_t \end{bmatrix} dt + \sum_{i=1}^{2d} B_i'\begin{bmatrix} q_t \\ \theta_t \end{bmatrix} dU_t^{(i)} \tag{3.5.4}$$

where

$$A' := \begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix},$$

$$B_i' := \begin{bmatrix} B_i & 0 \\ 0 & B_i \end{bmatrix} \quad \text{for } i = 1,2,\ldots,d;$$

$$\begin{bmatrix} 0 & 0 \\ \frac{1}{\sqrt{12}}(AB_{(i-d)} - B_{(i-d)}A) & 0 \end{bmatrix} \quad \text{for } i = d+1, d+2, \ldots, 2d;$$

$$U_t^{(i)} := \begin{cases} W_t^{(i)} & \text{for } i = 1,2,\ldots,d; \\ V_t^{(i-d)} & \text{for } i = d+1, d+2, \ldots, 2d. \end{cases}$$

$(U_t, t \in [0,T])$ is a $2d$-dimensional Wiener process under P_2.

NOTE 1. $(\Omega, \mathcal{F})$, P_1 and P_2 can be defined so as to support $(V_t, t \in [0,T])$ as follows. Let $(V_t, t \in [0,T]) \in (C^d[0,T], \mathcal{B}^d[0,T], W^d)$ the space of d-dimensional continuous functions on $[0,T]$ with the Borel-field generated by the supremum norm, and d-dimensional Wiener-measure. Set

$$(\Omega', \mathcal{F}', P_1') = (\Omega, \mathcal{F}, P_1) \times (C^d[0,T], \mathcal{B}^d[0,T], W^d)$$
$$(\Omega', \mathcal{F}', P_2') = (\Omega, \mathcal{F}, P_2) \times (C^d[0,T], \mathcal{B}^d[0,T], W^d).$$

Now take as $(\Omega, \mathcal{F}, P_1)$ and $(\Omega, \mathcal{F}, P_2)$ the new spaces $(\Omega', \mathcal{F}', P_1')$ and $(\Omega', \mathcal{F}', P_2')$.

NOTE 2. The $\mathcal{F}_T^W$-conditional distribution of z_T is the same under both measures P_1 and P_2. It is normal with zero mean, and covariance matrix D_T given by

$$D_T := E(z_T z_T^T \mid \mathcal{F}_T^W), \qquad E = E_1 \text{ or } E_2$$
$$= \frac{T^2}{12} \sum_{i=1}^{d} \int_0^T \Phi(s,T)(AB_i - B_iA)\Phi(0,s) q_0 q_0^T \Phi(0,s)^T (AB_i - B_iA)^T \Phi(s,T)^T ds. \tag{3.5.5}$$

NOTE 3. (θ_t) is a natural object to use in describing the convergence of the conditional mean sequence for solving (3.2.1) where the equation is defined on a semi-infinite time interval $[0,\infty)$:

$$q_0 := p_0$$
$$dq_t = Aq_t dt + \sum_{i=1}^{d} B_i q_t dW_t^{(i)} \qquad t \in [0,\infty)$$
$$dW_t = h(X_t)dt + d\beta_t \qquad t \in [0,\infty).$$

The signal process (X_t) is assumed to be defined on $[0,\infty)$. $(\beta_t, t \in [0,\infty))$ is a d-dimensional Wiener process on $(\Omega, \mathcal{F}, P_1)$.

We define a *regular partition* π^h of the interval $[0,\infty)$ by the following:

$$\pi^h := \{0, h, 2h, \ldots, nh, \ldots\}$$

Let a *refining sequence* of such partitions be one with the following property:

$$h_i \to 0 \text{ and } \qquad \pi^{h_i} \subseteq \pi^{h_j} \quad \text{if } i \leq j.$$

We also define the following σ-fields:

$$\mathcal{P}^h := \sigma(W_s, s \in \pi^h)$$

The following is an immediate consequence of theorem 3.1.

For each $t \in [0,\infty)$ and each refining sequence of partitions $(\pi^{h_i}: i = 1,2,\ldots)$ such that

$$t \in \pi^{h_i} \quad \text{for some natural } i,$$

$$P_1\Big(\frac{1}{h_i}(q_t - E_1(q_t \mid \mathcal{P}^{h_i} \cap \mathcal{Y}_t^W)) \in \cdot \mid \mathcal{P}^{h_i} \cap \mathcal{Y}_t^W\Big) \Rightarrow P_1(\theta_t \in \cdot \mid \mathcal{Y}_t^W) \quad \text{(a.s.)}$$
$$i \to \infty$$

and all the moments converge almost surely.

$E_1(q_t \mid \mathcal{P}^{h_i} \cap \mathcal{Y}_t^W)$ is the conditional mean of q_t based on all the information in the discrete observations up to time t. θ_t represents the "past-and-present-conditional" statistical behaviour, in the limit, of the normalised error $\frac{1}{h}(q_t - E_1(q_t \mid \mathcal{P}^{h_i} \cap \mathcal{Y}_t^W))$.

PROOF OF THEOREM 3.1. As in the proof of lemma 3.4 we shall use here the following shortened notation. For an integrable random variable X,

$$\begin{aligned} &\overline{X} \text{ denotes } E_2(X \mid \mathcal{P}_N), \\ &\tilde{X} \text{ denotes } X - \overline{X}. \end{aligned} \tag{3.5.6}$$

For $\alpha \in M$,

$$\begin{aligned} A_\alpha &:= A_\alpha(A, B_1, \ldots, B_d), \\ I_\alpha &:= I_\alpha^W(nh, (n+1)h), \\ I_\alpha(q) &:= I_\alpha^W(q, nh, (n+1)h). \end{aligned} \tag{3.5.7}$$

SKETCH OF PROOF. The equation numbers used here refer to the equations in the detailed proof which follows. To prove (i) we start by writing down the $3/2^{th}$-order Taylor expansion of the solution of (3.2.1). We then take the $\mathcal{P}_N$-conditional mean of both sides and subtract to obtain a difference equation for $N\tilde{q}_{nh}$, (3.5.10). The terms on the right-hand side of (3.5.10) are then sorted to obtain a form suitable for the application of theorem 2.2, (3.5.16). All the driving terms except one, $d_{n+1}(2)$ defined in (3.5.13), are of insignificant order in that they produce components of $N\tilde{q}_{nh}$ that converge to zero. The component that does not converge to zero is χ_n, defined in (3.5.17); $N\tilde{q}_{nh} - \chi_n$ is of uniform $L_{(\infty)}$-order $1/2$.

Using the properties of the Brownian-bridge process given in lemma 3.2 we see that χ_N has the same ρ_N-conditional distribution as the variable ξ_N, defined in (3.5.20). Now we show that $z_T - \xi_N$ is of $L_{(\infty)}$-order $1/2$, (3.5.22); so the moment form of the Borel-Cantelli lemma shows that all the ρ_N-conditional moments of $N\tilde{q}_T - \chi_N$ and $\xi_N - z_T$ converge to zero almost surely. This proves (i).

Result (ii) follows by application of lemma 3.1. To prove (iii) we show that the lower m components of the $1/2^{th}$-order Taylor approximate solution of (3.5.4) differs from ξ_n by an amount which is of uniform $L_{(\infty)}$-order $1/2$.

PROOF OF (i). All probabalistic operations are with respect to P_2 unless otherwise stated. We begin by writing down the $3/2^{th}$-order Taylor expansion for q; theorem 2.1 shows that

$$\begin{aligned} q_{(n+1)h} &= \sum_{\alpha \in S_3} A_\alpha I_\alpha q_{nh} + \sum_{\alpha \in B(S_3)} A_\alpha I_\alpha(q) \\ &= \sum_{\alpha \in S_3} A_\alpha \bar{I}_\alpha q_{nh} + \sum_{\alpha \in S_3} A_\alpha \tilde{I}_\alpha q_{nh} + \sum_{\alpha \in B(S_3)} A_\alpha I_\alpha(q), \end{aligned} \tag{3.5.8}$$

where

$$\begin{aligned} S_3 &= \{\alpha \in M : l(\alpha) + n(\alpha) \leq 3 \text{ OR } l(\alpha) = n(\alpha) = 2\} \\ &= \{v, (0), (i), (i,j), (0,i), (i,0), (0,0), (i,j,k) : i,j,k = 1,2,\ldots,d\}, \\ B(S_3) &= \{\alpha \in M - S_3 : -\alpha \in S_3\} \\ &= \{(0,i,j), (0,0,i), (0,i,0), (0,0,0), (i,0,j), (i,j,0), (i,0,0), (0,i,j,k), \\ &\quad (i,j,k,l) : i,j,k,l = 1,2,\ldots,d\}. \end{aligned}$$

We take the ρ_N-conditional mean of both sides to obtain

$$\overline{q_{(n+1)h}} = \sum_{\alpha \in S_3} A_\alpha \bar{I}_\alpha \bar{q}_{nh} + \sum_{\alpha \in S_3} A_\alpha \overline{\tilde{I}_\alpha q_{nh}} + \sum_{\alpha \in B(S_3)} A_\alpha \overline{I_\alpha(q)}. \tag{3.5.9}$$

Subtraction of (3.5.9) from (3.5.8) and multiplication by N yields a difference equation for $N(q_{nh} - E_2(q_{nh} \mid \rho_N))$:

$$\begin{aligned} N\tilde{q}_0 &= 0 \\ N\tilde{q}_{(n+1)h} &= \sum_{\alpha \in S_3} A_\alpha \bar{I}_\alpha N\tilde{q}_{nh} + N \sum_{\alpha \in S_3} A_\alpha \widetilde{\tilde{I}_\alpha q_{nh}} + N \sum_{\alpha \in B(S_3)} A_\alpha \widetilde{I_\alpha(q)} \end{aligned} \tag{3.5.10}$$

Now

$$\sum_{\alpha \in S_3} A_\alpha \bar{I}_\alpha N\tilde{q}_{nh} = (I + Y_{n+1}) N\tilde{q}_{nh} + d_{n+1}(1),$$

where

$$Y_{n+1} := \sum_{i=0}^{d} A_{(i)} \overline{I}_{(i)}$$

$$= A\,h + \sum_{i=1}^{d} B_i \Delta W_{n+1}^{(i)} \tag{3.5.11}$$

and

$$d_{n+1}(1) := \Big(\sum_{i,j=0}^{d} A_{(i,j)} \overline{I_{(i,j)}} + \sum_{i,j,k=1}^{d} A_{(i,j,k)} \overline{I_{(i,j,k)}} \Big) N \tilde{q}_{nh}. \tag{3.5.12}$$

Also

$$N \sum_{\alpha \in S_3} A_\alpha \widetilde{\widetilde{I}_\alpha q_{nh}} = d_{n+1}(2) + d_{n+1}(3),$$

where

$$d_{n+1}(2) := N \sum_{\alpha \in S_3} A_\alpha \widetilde{\widetilde{I}_\alpha \overline{q}_{nh}}$$

$$= N \sum_{i=1}^{d} (A_{(i,0)} \widetilde{I}_{(i,0)} + A_{(0,i)} \widetilde{I}_{(0,i)}) \overline{q}_{nh},$$

$$= N \sum_{i=1}^{d} (A B_i - B_i A) \int_{nh}^{(n+1)h} \widetilde{W}_t^{(i)} d\,t \overline{q}_{nh}, \tag{3.5.13}$$

and

$$d_{n+1}(3) := N \sum_{\alpha \in S_3} A_\alpha \widetilde{\widetilde{I}_\alpha \tilde{q}_{nh}}$$

$$= N \sum_{i=1}^{d} (A_{(i,0)} \widetilde{\widetilde{I}_{(i,0)} \tilde{q}_{nh}} + A_{(0,i)} \widetilde{\widetilde{I}_{(0,i)} \tilde{q}_{nh}}) \tag{3.5.14}$$

(all remaining integrals I_α with $\alpha \in S_3$ are ρ_N-measurable).

We also set

$$d_{n+1}(4) = N \sum_{\alpha \in B(S_3)} A_\alpha \widetilde{I_\alpha(q)}. \tag{3.5.15}$$

We can now re-write equation (3.5.10) as follows:

$$\begin{aligned} N\widetilde{q}_0 &= 0, \\ N\widetilde{q}_{(n+1)h} &= (I + Y_{n+1})N\widetilde{q}_{nh} + d_{n+1}(1) + d_{n+1}(2) + d_{n+1}(3) + d_{n+1}(4). \end{aligned} \tag{3.5.16}$$

We define the sequence $(\chi_n; n = 0, 1, \ldots, N)$ by

$$\begin{aligned} \chi_0 &:= 0, \\ \chi_{n+1} &:= (I + Y_{n+1})\chi_n + d_{n+1}(2) \qquad n = 0, 1, \ldots, N-1. \end{aligned} \tag{3.5.17}$$

Subtraction of (3.5.17) from (3.5.16) yields

$$\begin{aligned} N\widetilde{q}_0 - \chi_0 &= 0, \\ N\widetilde{q}_{(n+1)h} - \chi_{n+1} &= (I + Y_{n+1})(N\widetilde{q}_{nh} - \chi_n) + d_{n+1}(1) + d_{n+1}(3) + d_{n+1}(4). \end{aligned} \tag{3.5.18}$$

Now, lemmas 2.1, 3.3 and 3.4 show that

$$\begin{aligned} Y_{n+1} &= O_{(\infty)}(h^{1/2}) \text{ and is independent of } \mathcal{I}_{nh}, \\ |E(Y_{n+1})| &\leq Kh \qquad \text{where } K \geq 0, \\ d_{n+1}(1) &= O_{(\infty)}(h), \\ E_2(d_{n+1}(1) \mid \mathcal{I}_{nh}) &= O_{(\infty)}(h^2), \\ d_{n+1}(3) &= O_{(\infty)}(h^{3/2}), \\ E_2(d_{n+1}(3) \mid \mathcal{I}_{nh}) &= 0, \\ d_{n+1}(4) &= O_{(\infty)}(h), \end{aligned}$$

and

$$E_2(d_{n+1}(4) \mid \mathcal{I}_{nh}) = O_{(\infty)}(h^2).$$

So, by theorem 2.2,

$$N\widetilde{q}_{nh} - \chi_n = O_{(\infty)}(h^{1/2}). \tag{3.5.19}$$

We define the sequence $(\xi_n; n = 0, 1, \ldots, N)$ by

$$\xi_0 := 0$$

$$\xi_{n+1} := (I + Y_{n+1})\xi_n + \frac{T}{\sqrt{12}} \sum_{i=1}^{d} (AB_i - B_iA)\overline{q}_{nh}\Delta V_{n+1}^{(i)}, \tag{3.5.20}$$

where

$$\Delta V_{n+1}^{(i)} := V_{(n+1)h}^{(i)} - V_{nh}^{(i)}.$$

Now, parts (iii) and (iv) of lemma 3.2 show that ξ_N has the same ρ_N-conditional distribution as χ_N.

By solving (3.5.20), we obtain

$$\xi_N = \frac{T}{\sqrt{12}} \sum_{i=1}^{d} \sum_{n=1}^{N} (I + Y_N) \ldots (I + Y_{n+1})(A B_i - B_i A) \bar{q}_{(n-1)h} \Delta V_n^{(i)}$$

$$= \frac{T}{\sqrt{12}} \sum_{i=1}^{d} \int_0^T f^{(i)}(N, W, q_0)(s) d V_s^{(i)}, \tag{3.5.21}$$

where the $f^{(i)}$ are the piecewise-constant functions

$$f^{(i)}(N, W, q_0)(s) = (I + Y_N) \ldots (I + Y_{n+1})(A B_i - B_i A) \bar{q}_{(n-1)h} \qquad (n-1)h \leq s \leq nh$$
$$= (I + Y_N) \ldots (I + Y_{n+1})(A B_i - B_i A) q_{(n-1)h} + O_{(\infty)}(h) \quad .$$

Consider the s.d.e.

$$X_{nh} = (A B_i - B_i A) q_{(n-1)h}$$

$$d X_t = A X_t d t + \sum_{i=1}^{d} B_i X_t d W_t^{(i)} \qquad t \in [nh, T].$$

This has the solution

$$X_T = \Phi(nh, T)(A B_i - B_i A) q_{(n-1)h} .$$

The $1/2^{th}$ order Taylor approximation for X is given by

$$\hat{X}_T(2) = (I + Y_N) \ldots (I + Y_{n+1})(A B_i - B_i A) q_{(n-1)h};$$

so

$$f^{(i)}(N, W, q_0)(s) - \Phi(nh, T)(A B_i - B_i A)\Phi(0, (n-1)h) q_0 = O_{(\infty)}(h^{1/2}) \quad \forall \quad (n-1)h \leq s \leq nh$$

and

$$f^{(i)}(N, W, q_0)(s) - \Phi(s, T)(A B_i - B_i A)\Phi(0, s) q_0 = O_{(\infty)}(h^{1/2}).$$

So

$$\xi_N - z_T = \frac{T}{\sqrt{12}} \sum_{i=1}^{d} \int_0^T (f^{(i)}(N, W, q_0)(s) - \Phi(s, T)(A B_i - B_i A)\Phi(0, s) q_0) d V_s^{(i)}$$

$$= O_{(\infty)}(h^{1/2}). \tag{3.5.22}$$

It follows by (3.5.19), (3.5.22) and the moment form of the Borel-Cantelli lemma that

$$E_2(\| N \tilde{q}_T - \chi_N \|^p \mid \rho_N) \to 0 \qquad \text{(a.s.)} \qquad \forall\, p > 2, \tag{3.5.23}$$

and

$$E_2(\| \xi_N - z_T \|^p \mid \rho_N) \to 0 \qquad \text{(a.s.)} \qquad \forall\, p > 2. \tag{3.5.24}$$

Since ξ_N and χ_N have identical ρ_N-conditional distibutions,(i) follows.

PROOF OF (ii). Part (i) of this theorem and lemma 3.1 show that under the measure P_1 the $\mathcal{P}_{N_i}$-conditional distribution of $N_i(q_T - E_2(q_T \mid \mathcal{P}_{N_i}))$ converges to the $\mathcal{I}_T^W$-conditional distribution of z_T almost surely, and that the $\mathcal{P}_{N_i}$-conditional moments converge almost surely. In particular

$$E_1(N_i(q_T - E_2(q_T \mid \mathcal{P}_{N_i})) \mid \mathcal{P}_{N_i}) \underset{i \to \infty}{\to} 0 \quad \text{(a.s.)},$$

i.e.

$$N_i(E_1(q_T \mid \mathcal{P}_{N_i}) - E_2(q_T \mid \mathcal{P}_{N_i})) \underset{i \to \infty}{\to} 0 \quad \text{(a.s.)}. \tag{3.5.25}$$

This proves the convergence results for the sequence

$$(N_i(q_T - E_1(q_T \mid \mathcal{P}_{N_i})); \; i = 1, 2, \ldots).$$

PROOF OF (iii). Equation (3.5.4) is a Wiener-process-driven bilinear s.d.e.; so we can use the $1/2^{th}$ order Taylor approximation scheme, given in definition 2.3, to find an approximate solution. This is given by

$$\begin{bmatrix} \hat{q}_0 \\ \hat{\theta}_0 \end{bmatrix} = \begin{bmatrix} p_0 \\ 0 \end{bmatrix},$$

$$\begin{bmatrix} \hat{q}_{n+1} \\ \hat{\theta}_{n+1} \end{bmatrix} = \sum_{\alpha \in S_1} A_\alpha(A', B_1', \ldots, B_{2d}') I_\alpha^U(nh, (n+1)h) \begin{bmatrix} \hat{q}_n \\ \hat{\theta}_n \end{bmatrix},$$

where

$$\begin{aligned} S_1 &= \{\alpha \in M : l(\alpha) + n(\alpha) \leq 1 \text{ OR } l(\alpha) = n(\alpha) = 1\} \\ &= \{v, (0), (1), \ldots, (2d)\}. \end{aligned}$$

So

$$\begin{bmatrix} \hat{q}_{n+1} \\ \hat{\theta}_{n+1} \end{bmatrix} = \left(I + A'h + \sum_{i=1}^{2d} B_i' \Delta U_{n+1}^{(i)}\right) \begin{bmatrix} \hat{q}_n \\ \hat{\theta}_n \end{bmatrix}. \tag{3.5.26}$$

We know from theorem 2.3 that

$$\begin{bmatrix} q_{nh} \\ \theta_{nh} \end{bmatrix} - \begin{bmatrix} \hat{q}_n \\ \hat{\theta}_n \end{bmatrix} = O_{(\infty)}(h^{1/2}). \tag{3.5.27}$$

Now

$$\hat{\theta}_{n+1} = \left(I + Ah + \sum_{i=1}^{d} B_i \Delta W_{n+1}^{(i)}\right) \hat{\theta}_n + \frac{1}{\sqrt{12}} \sum_{i=1}^{d} (AB_i - B_i A) \hat{q}_n \Delta V_{n+1}^{(i)}. \tag{3.5.28}$$

We multiply (3.5.28) by T and subtract from (3.5.20) to obtain

$$(\xi_{n+1} - T\hat{\theta}_{n+1}) = (I + Y_{n+1})(\xi_n - T\hat{\theta}_n) + d_{n+1}(5), \tag{3.5.29}$$

where Y_{n+1} is as defined in (3.5.11) and

$$d_{n+1}(5) := \frac{T}{\sqrt{12}} \sum_{i=1}^{d} (A B_i - B_i A)(\bar{q}_{nh} - \hat{q}_n)\Delta V_{n+1}^{(i)}. \tag{3.5.30}$$

Lemma 3.4 and (3.5.27) show that

$$d_{n+1}(5) = O_{(\infty)}(h)$$

and

$$E_2(d_{n+1}(5) \mid \mathcal{F}_{nh}^U) = 0.$$

Furthermore $Y_{n+1} = O_{(\infty)}(h^{1/2})$ and is independent of $\mathcal{F}_{nh}^U$, and $|E_2(Y_{n+1})| \leq Kh$ for some $K > 0$, (Here $\mathcal{F}_t^U := \sigma(U_s : s \leq t)$). So we can apply theorem 2.2 to obtain

$$\xi_n - T\hat{\theta}_n = O_{(\infty)}(h^{1/2}). \tag{3.5.31}$$

Combining (3.5.22), (3.5.27) and (3.5.31), we have

$$z_T - T\theta_T = O_{(\infty)}(h^{1/2}).$$

Since neither z_T nor $T\theta_T$ is dependent on h

$$z_T = T\theta_T \qquad \text{(a.s.)}$$

∎

There is a much more direct way of proving part (iii) of theorem 3.1 but it relies on a rather subtle result in stochastic differential equation theory: there is a version $\Phi(s,t)$ of the matrix fundamental solution of (3.2.1) that is invertible for all s and t, except perhaps for ω in an exceptional null set that does not depend on s or t. This is a corollary of a theorem on the existence of flows of diffeomorphisms proved in different ways by a number of authors, for example Elworthy [EL1], Malliavin [MA2] and Ikeda and Watanabe [IW1]. One of the simplest statements is theorem 1.2 of Kunita [KU2]. The invertibility of $\Phi(s,t)$ permits the following representation of z_T:

$$z_T = \frac{T}{\sqrt{12}}\Phi(0,T) \sum_{i=1}^{d} \int_0^T \Phi^{-1}(0,s)(A B_i - B_i A)\Phi(0,s) q_0 d V_s^{(i)}. \tag{3.5.32}$$

Now

$$d\Phi(0,T) = A\Phi(0,T)dT + \sum_{i=1}^{d} B_i \Phi(0,T) d W_T^{(i)}$$

and

$$d\Big(\sum_{i=1}^{d} \int_0^T \Phi^{-1}(0,s)(A B_i - B_i A)\Phi(0,s) q_0 d V_s^{(i)}\Big) = \sum_{i=1}^{d} \Phi^{-1}(0,T)(A B_i - B_i A)\Phi(0,T) d V_T^{(i)}.$$

By applying Itô's rule to (3.5.32) we obtain

$$dz_T = Az_T dT + \sum_{i=1}^{d} B_i z_T dW_T^{(i)} + \frac{T}{\sqrt{12}}\Phi(0,T)\sum_{i=1}^{d}\Phi^{-1}(0,T)(AB_i - B_iA)\Phi(0,T)q_0 dV_T^{(i)}$$

$$= Az_T dT + \sum_{i=1}^{d} B_i z_T dW_T^{(i)} + \frac{T}{\sqrt{12}}\sum_{i=1}^{d}(AB_i - B_iA)q_T dV_T^{(i)},$$

which proves (3.5.3).

θ_T, of course, has the physical meaning that it is the limit-in-distribution of the normalised sequence $\frac{1}{h}(q_T - E_2(q_T \mid \mathcal{P}_N))$. If N is sufficiently large, then for almost each sample path of the driving process, the error $(q_T - E_2(q_T \mid \mathcal{P}_N))(W_h(\omega), W_{2h}(\omega), \ldots, W_{Nh}(\omega))$ has approximately the same distribution as $\theta_T(W_t(\omega), t \in [0,T])h$. Loosely speaking, we can think of the randomness in $q_T - E_2(q_T \mid \mathcal{P}_N)$ when parameterised by $(W_h(\omega), W_{2h}(\omega), \ldots, W_{Nh}(\omega))$ as being due to the fact that many sample paths of $(W_t, t \in [0,T])$ pass through the points $W_h(\omega), W_{2h}(\omega), \ldots$ and $W_{Nh}(\omega)$. The randomness in θ_T when parameterised by $(W_t(\omega), t \in [0,T])$ comes from $(V_t, t \in [0,T])$ of course.

We can think of expressions (3.5.1) and (3.5.32) as being convolution integrals. Loosely speaking, if N is sufficiently large then the local error in the conditional-mean scheme accrued over the small interval $[s, s+\delta s]$, $((q_{s+\delta s} - q_s) - (E_2(q_{s+\delta s} \mid \mathcal{P}_N) - E_2(q_s \mid \mathcal{P}_N)))(W_h(\omega), \ldots, W_{Nh}(\omega))$, has approximately the same distribution as

$$\frac{1}{\sqrt{12}}\sum_{i=1}^{d}(AB_i - B_iA)h\delta V_s^{(i)} q_s(W_t(\omega), t \in [0,T]),$$

that is the vector obtained by applying the linear vector-field

$$\frac{1}{\sqrt{12}}\sum_{i=1}^{d}(AB_i - B_iA)h\delta V_s^{(i)}$$

to the solution of (3.2.1) at time s. The distribution of the total error

$$(q_T - E_2(q_T \mid \mathcal{P}_N))(W_h(\omega), \ldots, W_{Nh}(\omega))$$

is found by summing the effects of these vectors at time T; that is, we sum the quantities:

$$\Phi(s,T)(W_t(\omega), t \in [0,T])\frac{1}{\sqrt{12}}\sum_{i=1}^{d}(AB_i - B_iA)h\delta V_s^{(i)} q_s(W_t(\omega), t \in [0,T])$$

over all intervals $[s, s+\delta s]$ in $[0,T]$.

If we apply Itô's rule to equation (3.5.3) we obtain a stochastic differential equation for z_T.

$$z_\epsilon = \epsilon\theta_\epsilon \quad \text{, where } \epsilon > 0,$$

$$dz_T = T\,d\theta_T + \theta_T\,dT$$

$$= \Big(A + \frac{1}{T}I\Big)z_T\,dT + \sum_{i=1}^{d} B_i z_T\,dW_T^{(i)} + \frac{T}{\sqrt{12}}\sum_{i=1}^{d}(AB_i - B_iA)q_T\,dV_T^{(i)}$$

$$T \in (0,\infty). \quad (3.5.33)$$

This equation does not have a unique solution if we include the point $T = 0$.

The convergence of the conditional distribution of the normalised error can be used to investigate the asymptotic behaviour of the "convergence-in-probability" of $E_1(q_T \mid \mathcal{P}_N)$ to q_T:

$$P_1(N_i(q_T - E_1(q_T \mid \mathcal{P}_{N_i})) \in B \mid \mathcal{P}_{N_i}) \to P_1(z_T \in B \mid \mathcal{J}_T^W)$$

for all Borel sets B, with probability 1; in particular

$$P_1(|c^T(q_T - E_1(q_T \mid \mathcal{P}_{N_i}))| \geq \frac{\epsilon}{N_i} \mid \mathcal{P}_{N_i}) \to P_1(|c^T z_T| \geq \epsilon \mid \mathcal{J}_T^W)$$

for all vectors $c \in R^m$ and all positive numbers ϵ, with probability 1. This can be used to estimate the error when N is large.

Since all the positive moments of the normalised error converge we can make similar estimates for large N in terms of much stronger convergence than "in-probability", for instance in terms of the L_p-norms:

$$E_1(|N_i c^T(q_T - E_1(q_T \mid \mathcal{P}_{N_i})|^p \mid \mathcal{P}_{N_i})^{1/p} \to E_1(|c^T z_T|^p \mid \mathcal{J}_T^W)^{1/p} \quad \text{(a.s.)}$$

and so for large N

$$E_1(|c^T(q_T - E_1(q_T \mid \mathcal{P}_N))|^p \mid \mathcal{P}_N)^{1/p} \approx E_1(|c^T z_T|^p \mid \mathcal{J}_T^W)^{1/p}\frac{1}{N}. \quad (3.5.34)$$

Theorem 3.1 is a type of central-limit theorem. Intuitively, we might expect that the normalised error in any $\mathcal{P}_N$-measurable approximation scheme converges in distribution to a normal random variable if it converges at all; in fact if $(\hat{q}_N\ ;\ N = 1,2,\ldots)$ is any $\mathcal{P}_N$-adapted sequence of approximations to q_T the error can be represented as the sum of the error in the conditional mean and a $\mathcal{P}_N$-adapted sequence:

$$q_T - \hat{q}_N = (q_T - E_1(q_T \mid \mathcal{P}_N)) + (E_1(q_T \mid \mathcal{P}_N) - \hat{q}_N).$$

So, if the sequence $N(E_1(q_T \mid \mathcal{P}_N) - \hat{q}_N)$ converges almost surely to some $\mathcal{J}_T^W$-measurable random variable M_T, then the $\mathcal{P}_N$-conditional distribution and moments of the normalised error sequence $N(q_T - \hat{q}_N)$ will converge to the limit-distribution given in theorem 3.1 but with a mean of M_T instead of zero; i.e. the normal distribution $N(M_T, D_T)$. If the sequence $N(E_1(q_T \mid \mathcal{P}_N) - \hat{q}_N)$ does not converge then neither will the $\mathcal{P}_N$-conditional distribution nor the moments of the normalised error $N(q_T - \hat{q}_N)$.

We might expect, therefore, that the limit-distribution given in theorem 3.1 represents the asymptotically maximum rate of convergence in terms of all the L_p-norms ($p \geq 1$); that is, the coefficient on the right hand side of (3.5.34) is minimal. This is certainly the case for the L_{2m}-norms for ($m = 1, 2, \ldots$) as is shown in the following. The L_{2m}-norm of the error in an approximation scheme $\hat{q}_N$ is given by

$$e(\hat{q}_N) = E_1((c^T(q_T - \hat{q}_N))^{2m} \mid \mathcal{P}_N).$$

By differentiating with respect to $\hat{q}$, multiplying by N and equating to zero for a minimum we obtain an equation for the optimal (in the L_{2m} sense) scheme q_N^*:

$$\begin{aligned} 0 &= E_1(N c^T(q_T - q_N^*)^{2m-1} \mid \mathcal{P}_N) \\ &= E_1((N c^T(q_T - E_1(q_T \mid \mathcal{P}_N)) + N c^T(E_1(q_T \mid \mathcal{P}_N) - q_N^*))^{2m-1} \mid \mathcal{P}_N) \\ &= \sum_{r=0}^{2m-1} \frac{(2m-1)!}{(2m-1-r)!r!} E_1((N c^T(q_T - E_1(q_T \mid \mathcal{P}_N)))^{2m-1-r} \mid \mathcal{P}_N)(N c^T(E_1(q_T \mid \mathcal{P}_N) - q_N^*))^r. \end{aligned}$$

But the odd moments of $N c^T(q_T - E_1(q_T \mid \mathcal{P}_N))$ converge along an appropriate subsequence (N_i ; $i = 1, 2, \ldots$) to zero and the even moments are bounded; so

$$\sum_{\substack{r=1 \\ r \text{ odd}}}^{2m-1} \frac{(2m-1)!}{(2m-1-r)!r!} E_1((N_i c^T(q_T - E_1(q_T \mid \mathcal{P}_{N_i})))^{2m-1-r} \mid \mathcal{P}_{N_i})(N_i c^T(E_1(q_T \mid \mathcal{P}_N) - q_{N_i}^*)) \to 0 \quad \text{(a.s.)}$$

and

$$N_i c^T(E_1(q_T \mid \mathcal{P}_{N_i}) - q_{N_i}^*) \to 0 \qquad \text{(a.s.)}.$$

So the optimal scheme in the L_{2m} sense has the same limit-distribution for normalised error as the conditional-mean scheme.

Part (ii) of theorem 3.1 shows that under the "true" measure P_1 the conditional mean $E_1(q_T \mid \mathcal{P}_N)$ has a normalised error that converges on refining partitions to the same limit as the normalised error in $E_2(q_T \mid \mathcal{P}_N)$ in the Wiener measure case. This is to be expected since the increments of the driving term $\Delta W_1, \Delta W_2, \ldots, \Delta W_N$ are given by the sum of a deterministic integral and an increment of a Brownian motion (see equation (3.2.2)). The integral is of uniform $L_{(\infty)}$-order 1 and the increment of Brownian motion is of uniform $L_{(\infty)}$-order 1/2; it seems reasonable, then, that the Brownian motion part of W_t dominates as h decreases.

3.6 Limit-distributions for normalised errors in approximate Bayes estimates.

Hitherto we have been investigating the best convegence of discrete approximate solutions to the bilinear s.d.e. (3.2.1). In this section we shall carry over the main result, theorem 3.1, to the best convergence of approximate Bayes estimates of functions of the original signal process, obtained by formula (1.2.8). It is these approximate Bayes estimates that are the "solution" of the filtering problem. I have restated formula (1.2.8) here for convenience.

Let f be a vector-valued function of the signal process:

$$f : \{a_1, a_2, \ldots, a_m\} \to R^n.$$

The Bayes estimate of $f(X_T)$ is given by

$$E_1(f(X_T) \mid \mathcal{I}_T^W) = \frac{\sum_{i=1}^m f(a_i) q_T^{(i)}}{\sum_{i=1}^m q_T^{(i)}} =: \Psi_f(q_T). \tag{3.6.1}$$

If $\hat{q}_N$ is an approximate solution of (3.2.1) then by "the corresponding approximate Bayes estimate" we mean the quantity

$$\widehat{E_1(f(X_T) \mid \mathcal{I}_T^W)} := \Psi_f(\hat{q}_N) \quad \text{if } P_1\Big(\sum_{i=1}^m \hat{q}_N^{(i)} = 0\Big) = 0. \tag{3.6.2}$$

We would like to examine the convergence of such objects as $\Psi_f(\hat{q}_N)$ (to $\Psi_f(q_T)$) given the convergence properties of $\hat{q}_N$. Once again we shall analyse the convergence of the ρ_N-conditional mean sequence $(E_1(f(X_T) \mid \rho_N); N = 1, 2, \ldots)$ as it possesses the optimal asymptotic properties described in the last section.

Since the function Ψ_f is nonlinear the best approximate solution of (3.2.1) does not in general yield the best approximate Bayes estimate:

$$\text{“}E_1(f(X_T) \mid \rho_N) \neq \Psi_f(E_1(q_T \mid \rho_N)) \quad \text{in general”}. \tag{3.6.3}$$

However, as we shall see, the conditional mean $E_1(q_T \mid \rho_N)$ and indeed $E_2(q_T \mid \rho_N)$ both yield approximate Bayes estimates with optimal properties.

The optimal limit-distribution for the normalised error in approximate Bayes estimates can be found using a method which Billingsley ([BI1], P.340) refers to as the *delta method*. The idea is that Ψ_f is twice-differentiable and, with increasing N, the error in a convergent approximate solution to (3.2.1) decreases so that the corresponding error in the approximate Bayes estimate becomes nearly equal to the error in the approximate solution of (3.2.1) multiplied by the value of the derivative of Ψ_f at the point q_T.

The following theorem gives a limit-distibution result for the error in approximations of the Bayes estimate given by (3.6.1); the main difficulty in the proof lies in the fact that for small q_T, $1/\sum_{i=1}^m q_T^{(i)}$ can be unboundedly large.

Theorem 3.2. *Let $(\hat{q}_N : N = 1, 2, \ldots)$ be a sequence of ρ_N-measurable approximations to q_T such that for some measure P that is equivalent to P_1 and P_2 in the sense of absolute continuity*

$$P(N(q_T - \hat{q}_N) \in \cdot \mid \rho_N) \Rightarrow P(\xi_T \in \cdot \mid \mathcal{I}_T^W) \quad \text{(a.s.)}$$

and

$$E(\|N(q_T - \hat{q}_N)\|^p \mid \rho_N) \to E(\|\xi_T\|^p \mid \mathcal{I}_T^W) \quad \text{(a.s.)} \quad \forall\, p > 0, \tag{3.6.4}$$

(where ξ_T is some random variable in R^m) then the error in the corresponding approximate Bayes

estimate $\Psi_f(\hat{q}_N)$ *converges as follows:*

$$P(N(\Psi_f(q_T)-\Psi_f(\hat{q}_N))\in\cdot\mid\rho_N)\Rightarrow P(\eta_T\in\cdot\mid\mathcal{J}_T^W)\qquad\text{(a.s.)}$$

and

$$E(\|N(\Psi_f(q_T)-\Psi_f(\hat{q}_N))\|^p\mid\rho_N)\to E(\|\eta_T\|^p\mid\mathcal{J}_T^W)\qquad\text{(a.s.)}\qquad\forall\, p>0,\tag{3.6.5}$$

where

$$\begin{aligned}\eta_T&:=(D\Psi_f)(q_T)\xi_T\\&=\frac{\sum_{j,k=1}^m(f(a_j)-f(a_k))q_T^{(k)}\xi_T^{(j)}}{\left(\sum_{j=1}^m q_T^{(j)}\right)^2}.\end{aligned}\tag{3.6.6}$$

Here, $D\Psi_f$ *means the Jacobian derivative of* Ψ_f.

PROOF. We first establish the existence of the moments of $\frac{1}{\sum_{i=1}^m q_T^{(i)}}$. In the derivation of the Zakai equation (3.2.1) q_T is defined in terms of the normalised conditional distribution of the signal-process p_t by

$$\begin{aligned}q_t&:=\sum_{i=1}^m q_t^{(i)}p_t\\\sum_{i=1}^m q_t^{(i)}&:=\exp\Big(\int_0^t\sum_{i=1}^m p_s^{(i)}h(a_i)^T\,dW_s-\frac{1}{2}\int_0^t\|\sum_{i=1}^m p_s^{(i)}h(a_i)\|^2ds\Big)\end{aligned}\tag{3.6.7}$$

(see equation (1.2.5)). Since

$$p_t^{(i)}\geq 0\qquad\text{(a.s.)}\quad\forall\ i=1,2,\ldots,m\quad t\geq 0$$

and

$$\sum_{i=1}^m p_t^{(i)}=1\qquad\text{(a.s.)}\quad\forall\, t\in[0,T]$$

it follows that

$$q_t^{(i)}\geq 0\qquad\text{(a.s.)}\quad\forall\ i=1,2,\ldots,m\quad t\geq 0$$

and

$$\sum_{i=1}^m q_t^{(i)}>0\qquad\text{(a.s.)}\quad\forall\, t\in[0,T].\tag{3.6.8}$$

Furthermore, if we apply Itô's rule to (3.6.7) we obtain

$$d\Big(\sum_{i=1}^m q_t^{(i)}\Big)^{-1}=\Big(\sum_{i=1}^m q_t^{(i)}\Big)^{-1}\|\sum_{i=1}^m p_t^{(i)}h(a_i)\|^2dt-\Big(\sum_{i=1}^m q_t^{(i)}\Big)^{-1}\sum_{i=1}^m p_t^{(i)}h(a_i)^T\,dW_t.\tag{3.6.9}$$

Now

$$\sup_{\substack{t \\ \omega \in C}} \| \sum_{i=1}^{m} p_t^{(i)} h(a_i) \|^2 < \infty \qquad \text{where } P(C) = 1;$$

so the coefficients of (3.6.9) fulfil the conditions of the existence and uniqueness theorem (see theorem 4.6 in [LS1]), and since

$$\Big(\sum_{i=1}^{m} q_0^{(i)}\Big)^{-1} = 1 \qquad \text{(a.s.)}$$

all the moments of

$$\Big(\sum_{i=1}^{m} q_t^{(i)}\Big)^{-1}$$

exist for all $t \in [0, T]$.

We now define a set S which contains q_T almost surely and on which

$$\Big(\sum_{i=1}^{m} q_t^{(i)}\Big)^{-1}$$

is well defined. Let

$$S := \Big\{x \in R^m : x_i \geq 0\,;\ i = 1,2,\ldots,m : \sum_{i=1}^{m} x_i > 0\Big\}. \tag{3.6.10}$$

S is convex. Consider the following function, defined on S:

$$\Psi(x) := \Big(\sum_{i=1}^{m} x_i\Big)^{-1} x; \tag{3.6.11}$$

clearly

$$\Psi_f(x) = [f(a_1) f(a_2) \ldots f(a_m)]\Psi(x).$$

Let ν be a unit vector in R^m and let α be a positive scalar such that $x, x + \alpha\nu \in S$. We define the function $g(\alpha)$ and produce a Taylor expansion of it about 0.

$$g(\alpha) := \Psi(x + \alpha\nu), \tag{3.6.12}$$

$$\frac{dg}{d\alpha}(\alpha) = \Big(\sum_{i=1}^{m} (x_i + \alpha\nu_i)\Big)^{-1}\nu - \Big(\sum_{i=1}^{m} \nu_i\Big)\Big(\sum_{i=1}^{m} (x_i + \alpha\nu_i)\Big)^{-2}(x + \alpha\nu),$$

$$\frac{d^2g}{d\alpha^2}(\alpha) = -2\Big(\sum_{i=1}^{m} \nu_i\Big)\Big(\sum_{i=1}^{m} (x_i + \alpha\nu_i)\Big)^{-2}\nu + 2\Big(\sum_{i=1}^{m} \nu_i\Big)^2\Big(\sum_{i=1}^{m} (x_i + \alpha\nu_i)\Big)^{-3}(x + \alpha\nu), \tag{3.6.13}$$

and

$$\Psi(x+\alpha v)-\Psi(x)=g(\alpha)-g(0)$$
$$=\frac{dg}{d\alpha}(0)\alpha+\frac{d^2g}{d\alpha^2}(\lambda\alpha)\frac{\alpha^2}{2} \quad \text{for some } 0\le\lambda\le 1. \tag{3.6.14}$$

Now

$$\|\frac{d^2g}{d\alpha^2}(\lambda\alpha)\|^2 \le K_1\Big(\sum_{i=1}^m (x_i+\lambda\alpha v_i)\Big)^{-4}+K_2\Big(\sum_{i=1}^m (x_i+\lambda\alpha v_i)\Big)^{-6}\|x+\lambda\alpha v\|^2$$

and since S is convex $x+\lambda\alpha v\in S$; so

$$\|x+\lambda\alpha v\|^2=\sum_{i=1}^m (x_i+\lambda\alpha v_i)^2$$
$$\le\Big(\sum_{i=1}^m (x_i+\lambda\alpha v_i)\Big)^2.$$

This shows that for a natural number M

$$\|\frac{d^2g}{d\alpha^2}(\lambda\alpha)\|^{2M} \le K_3\Big(\sum_{i=1}^m (x_i+\lambda\alpha v_i)\Big)^{-4M}$$
$$=K_3\Big((1-\lambda)\sum_{i=1}^m x_i+\lambda\sum_{i=1}^m (x_i+\alpha v_i)\Big)^{-4M}$$
$$\le(1-\lambda)K_3\Big(\sum_{i=1}^m x_i\Big)^{-4M}+\lambda K_3\Big(\sum_{i=1}^m (x_i+\alpha v_i)\Big)^{-4M} \tag{3.6.15}$$

This follows since $\frac{1}{x^{4M}}$ is a convex downward function.

Now, $q_T\in S$ (a.s.) and for any ω in the following set

$$A_N:=\{\omega:\hat{q}_N\in S\}$$

we use (3.6.13), (3.6.14) and (3.6.15) to obtain

$$N(\Psi(q_T)-\Psi(\hat{q}_N))=\Big(\sum_{i=1}^m \hat{q}_N^{(i)}\Big)^{-1}\Big(I-\Big(\sum_{i=1}^m \hat{q}_N^{(i)}\Big)^{-1}[\hat{q}_N\,\hat{q}_N\ldots\hat{q}_N]\Big)N(q_T-\hat{q}_N)+R_N, \tag{3.6.16}$$

where

$$\|R_N\|^{2M} \le K_4\Big((1-\lambda)\Big(\sum_{i=1}^m \hat{q}_N^{(i)}\Big)^{-4M}+\lambda\Big(\sum_{i=1}^m q_T^{(i)}\Big)^{-4M}\Big)N\|q_T-\hat{q}_N\|^2.$$

since the moments of

$$\Big(\sum_{i=1}^{m} q_T^{(i)}\Big)^{-1}$$

exist

$$E(\|R_N\|^{2M} I_{A_N} \mid \mathcal{P}_N) \leq \frac{K_5(\omega)}{N} \qquad \text{(a.s.)},$$

where I_{A_N} is the *indicator function* of the set A_N. Since I_{A_N} is $\mathcal{P}_N$-measurable and since for each ω there exists an $N_0(\omega)$ such that

$$N \geq N_0(\omega) \Rightarrow \omega \in A_N$$

it follows that

$$E(\|R_N\|^{2M} \mid \mathcal{P}_N) \to 0 \qquad \text{(a.s.)}. \tag{3.6.17}$$

Now $\hat{q}_N$ is $\mathcal{P}_N$-measurable and converges to q_T almost surely. So it is evident from (3.6.16) and (3.6.17) that the $\mathcal{P}_N$-conditional distribution and moments of $N(\Psi(q_T) - \Psi(\hat{q}_N))$ converge almost surely to the $\mathcal{I}_T^W$-conditional distribution and moments of the following vector:

$$\Big(\sum_{i=1}^{m} q_T^{(i)}\Big)^{-1}\Big(I - \Big(\sum_{i=1}^{m} q_T^{(i)}\Big)^{-1}[q_T\, q_T \ldots q_T]\Big)\xi_T.$$

The theorem follows because

$$\Psi_f(\cdot) = [f(a_1)f(a_2)\ldots f(a_m)]\Psi(\cdot),$$ ∎

Corollary.

(i)

$$P_2(N(\Psi_f(q_T) - \Psi_f(E_2(q_T \mid \mathcal{P}_N))) \in \cdot \mid \mathcal{P}_N) \Rightarrow P_2(\chi_T \in \cdot \mid \mathcal{I}_T^W) \qquad \text{(a.s.)}$$

and

$$E_2(\|N(\Psi_f(q_T) - \Psi_f(E_2(q_T \mid \mathcal{P}_N)))\|^p \mid \mathcal{P}_N) \to E_2(\|\chi_T\|^p \mid \mathcal{I}_T^W) \qquad \text{(a.s.)} \qquad \forall\, p > 0; \tag{3.6.18}$$

and for any sequence $(N_1, N_2, \ldots)$ with the refining property (3.3.12)

$$P_1(N_i(\Psi_f(q_T) - \Psi_f(E_2(q_T \mid \mathcal{P}_{N_i}))) \in \cdot \mid \mathcal{P}_{N_i}) \Rightarrow P_1(\chi_T \in \cdot \mid \mathcal{I}_T^W) \qquad \text{(a.s.)}$$

and

$$E_1(\|N_i(\Psi_f(q_T) - \Psi_f(E_2(q_T \mid \mathcal{P}_{N_i})))\|^p \mid \mathcal{P}_{N_i}) \to E_1(\|\chi_T\|^p \mid \mathcal{I}_T^W) \qquad \text{(a.s.)} \qquad \forall\, p > 0, \tag{3.6.19}$$

where χ_T is given by

$$\chi_T := \frac{\sum_{j,k=1}^{m} (f(a_j) - f(a_k)) q_T^{(k)} z_T^{(j)}}{\big(\sum_{j=1}^{m} q_T^{(j)}\big)^2}. \tag{3.6.20}$$

z_T was given in (3.5.1).

(ii) The limit-distribution for the normalised error in the approximate Bayes estimate given in part (i) of this corollary is optimal in the sense that it is also the limit-distribution for the normalised error in the best approximate Bayes estimate, i.e

$$P_2(N(\Psi_f(q_T) - E_2(\Psi_f(q_T) \mid \mathcal{P}_N)) \in \cdot \mid \mathcal{P}_N) \Rightarrow P_2(\chi_T \in \cdot \mid \mathcal{I}_T^W) \quad \text{(a.s.)}$$

and

$$E_2(\|N(\Psi_f(q_T) - E_2(\Psi_f(q_T) \mid \mathcal{P}_N))\|^p \mid \mathcal{P}_N) \to E_2(\|\chi_T\|^p \mid \mathcal{I}_T^W) \quad \text{(a.s.)} \quad \forall\, p > 0; \tag{3.6.21}$$

and for any sequence $(N_1, N_2, \ldots)$ with the refining property (3.3.12)

$$P_1(N_i(\Psi_f(q_T) - E_1(\Psi_f(q_T) \mid \mathcal{P}_{N_i})) \in \cdot \mid \mathcal{P}_{N_i}) \Rightarrow P_1(\chi_T \in \cdot \mid \mathcal{I}_T^W) \quad \text{(a.s.)}$$

and

$$E_1(\|N_i(\Psi_f(q_T) - E_1(\Psi_f(q_T) \mid \mathcal{P}_{N_i}))\|^p \mid \mathcal{P}_{N_i}) \to E_1(\|\chi_T\|^p \mid \mathcal{I}_T^W) \quad \text{(a.s.)} \quad \forall\, p > 0. \tag{3.6.22}$$

PROOF OF COROLLARY. (i) is a direct result of the theorem and theorem 3.1; (3.6.19) follows by lemma 3.1.

Because of (3.6.18)

$$E_2(N(\Psi_f(q_T) - \Psi_f(E_2(q_T \mid \mathcal{P}_N))) \mid \mathcal{P}_N) \to 0 \quad \text{(a.s.)},$$

i.e.

$$N(E_2(\Psi_f(q_T) \mid \mathcal{P}_N) - \Psi_f(E_2(q_T \mid \mathcal{P}_N))) \to 0 \quad \text{(a.s.)}.$$

which proves (3.6.21). (3.6.22) follows in the same manner from (3.6.19). ∎

The limit-distribution in the corollary, the $\mathcal{I}_T^W$-conditional distribution of χ_T, has the optimal properties of the limit-distribution in theorem 3.1, discussed in the last section. It is a normal distribution with zero mean, and covariance matrix D'_T given by

$$\begin{aligned} D'_T &:= E(\chi_T \chi_T^T \mid \mathcal{I}_T^W) \\ &= (D\Psi)(q_T) E(z_T z_T^T \mid \mathcal{I}_T^W)(D\Psi_f)^T(q_T) \\ &= \frac{T^2}{12} \sum_{i=1}^{d} \int_0^T (D\Psi_f)(q_T)\Phi(s,T)(AB_i - B_iA)\Phi(0,s) q_0 q_0^T \Phi(0,s)^T \\ &\qquad (AB_i - B_iA)^T \Phi(s,T)^T (D\Psi_f)^T(q_T)\, ds. \end{aligned} \tag{3.6.23}$$

It follows from the discussion in the last section that if $(\hat{q}_N; N = 1, 2, \ldots)$ is a $\mathcal{P}_N$-adapted sequence of approximations to q_T with the property that $N(E_1(q_T \mid \mathcal{P}_N) - \hat{q}_N)$ converges almost surely to a random variable M_T, then the $\mathcal{P}_N$-conditional distribution of the normalised error in the corresponding Bayes estimate of $f(X_T)$, $N(\Psi_f(q_T) - \Psi_f(\hat{q}_N))$, converges along an appropriate subsequence to a normal distribution with mean $(D\Psi_f)(q_T)M_T$ and covariance matrix D'_T, i.e.

$$P_1(N_i(\Psi_f(q_T) - \Psi_f(\hat{q}_{N_i}) \in \cdot \mid \mathcal{P}_{N_i}) \Rightarrow N((D\Psi_f)(q_T)M_T, D'_T)(\cdot) \quad \text{(a.s.)}, \tag{3.6.24}$$

where $(N_i : i = 1, 2, \ldots)$ has the property (3.3.12). The moments also converge almost surely. If the sequence $N(E_1(q_T \mid \mathcal{P}_N) - \hat{q}_N)$ does not converge then neither does the $\mathcal{P}_N$-conditional distribution of the normalised error in the Bayes estimates nor its moments.

Chapter 4

Some Approximation Schemes.

4.1 Introduction.

This chapter considers specific finite-difference methods for solving the bilinear s.d.e. (3.2.1), with particular attention given to the asymptotic properties. The main difference between the approximations obtained by the schemes in this chapter and the partition-conditional mean sequence, analysed in chapter 3, is that the former can, in principle, be calculated.

The asymptotic properties that we are mostly concerned with are; *almost sure convergence, asymptotic efficiency,* 1^{st}*-order asymptotic efficiency* and of course *order of convergence.* The efficiency concepts are described in section 4.2: roughly speaking a 1^{st}-order asymptotically efficient sequence of approximations is one with the optimal asymptotic properties of the partition-conditional mean sequence; the asymptotic behaviour of the corresponding approximate Bayes estimates, found by formula (3.6.1), is also optimal.

In section 4.3 we construct a 1^{st} order asymptotically efficient scheme, the "Paradigm": theorem 4.1 shows that the removal of one or more terms from this scheme results in the loss of 1^{st} order asymptotic efficiency. In particular the Euler (see Maruyama [MA1]) and Mil'shtein [MI1] schemes discussed in chapter 1 are not 1^{st} order asymptotically efficient.

Various schemes are compared in section 4.4 with particular mention given to the Euler, Mil'shtein and Paradigm schemes since these contain the minimum number of terms needed for almost sure convergence, 1^{st} order convergence and 1^{st} order asymptotic efficiency respectively. We also look at some of the Runge-Kutta schemes of Rümelin [RU1].

4.2 Asymptotic properties of approximation schemes.

The main properties used in this chapter for comparing approximation schemes for solving equations of the type (3.2.1) are asymptotic properties. Asymptotic analysis of errors does not lead to quantitative statements about the errors of given methods on given partitions but does indicate the type of method that is worth implementing: complex methods with the same asymptotic properties as simpler schemes are probably not worth considering; it would be better to use a simpler scheme with a finer partition to improve accuracy.

Definition 4.1. Let $X(\in R^n)$ be a square-integrable vector-valued random variable on a probability space $(\Omega, \mathcal{J}, P)$. Let $(\mathcal{D}_N \subseteq \mathcal{J} : N = 1, 2, \ldots)$ be an increasing sequence of sub-sigma-fields of $\mathcal{J}$.

(a) $(X_N \in R^n : N = 1,2,\ldots)$ is called *asymptotically efficient* in the class of ρ_N-adapted sequences of approximations to X if (X_N, ρ_N) is adapted and, with probability one,

$$\frac{E((c^T(X - X_N))^2 \mid \rho_N)}{\mathrm{var}(c^T X \mid \rho_N)} \to 1 \qquad \forall\, c \in R^n - \{0\}. \tag{4.2.1}$$

Here

$$\mathrm{var}(\cdot \mid \rho_N) := E((\cdot - E(\cdot \mid \rho_N))^2 \mid \rho_N).$$

(b) $(X_N \in R^n : N = 1,2,\ldots)$ is called *r^{th}-order asymptotically efficient* in the class of ρ_N-adapted sequences of approximations to X if (X_N, ρ_N) is adapted and, with probability one,

$$\frac{E((N^r c^T(X - X_N))^2 \mid \rho_N) + 1}{\mathrm{var}(N^r c^T X \mid \rho_N) + 1} \to 1 \qquad \forall\, c \in R^n. \tag{4.2.2}$$

An asymptotically efficient sequence is one whose error has asymptotically minimal variance, i.e. it is asymptotically optimal in the least squares sense. An asymptotically efficient sequence certainly has the maximum order of convergence. r^{th}-order asymptotic efficiency is weaker than asymptotic efficiency; there are however special cases where it is equivalent. Suppose for instance that, with probability one,

$$\mathrm{var}(N^r c^T X \mid \rho_N) \to \xi_c < \infty \qquad \forall\, c \in R^n.$$

If the sequence $(X_N : N = 1,2,\ldots)$ is r^{th}-order asymptotically efficient then, with probability one,

$$E((N^r c^T(X - X_N))^2 \mid \rho_N) \to \xi_c \qquad \forall\, c \in R^n.$$

If $P(\xi_c > 0) = 1 \;\; \forall\, c \in R^n - \{0\}$ then this is equivalent to asymptotic efficiency, but if there exists a set of non-zero probability on which ξ_c is zero for some $c \neq 0$ then asymptotic efficiency does not follow.

In the case of ρ_N-adapted sequences of approximations to the solution of (3.2.1) we know that the maximum order of convergence is 1. In the next section we shall demonstrate the existence of a first-order asymptotically efficient sequence of approximations to q_T. We know from theorem 3.1 that, with probability one,

$$\mathrm{var}_2(N c^T q_T \mid \rho_N) \to c^T D_T c \qquad \forall\, c \in R^m,$$

where D_T is the following conditional-covariance matrix:

$$D_T = \frac{T^2}{12} \sum_{i=1}^{d} \int_0^T \Phi(s,T)(A B_i - B_i A)\Phi(0,s) q_0 q_0^T \Phi(0,s)^T (A B_i - B_i A)^T \Phi(s,T)^T \, ds. \tag{4.2.3}$$

If D_T is almost surely positive-definite then any first-order asymptotically efficient scheme will also be asymptotically efficient. If, however, there is a set of non-zero probability on which D_T is not positive-definite then asymptotic efficiency does not follow. In effect, on the set on which D_T is not posi-

tive-definite the conditional mean sequence $(E_2(q_T \mid \mathcal{P}_N) : N = 1,2,\ldots)$ converges faster than linearly. It is not a requirement of a first-order asymptotically efficient method that it should "keep up" with this increased rate of convergence over the generic maximum, it is however a requirement of an asymptotically efficient method.

As an example, suppose that the matrices A and B_i $(i = 1,2,\ldots,d)$ in (3.2.1) commute, i.e. $AB_i = B_iA(\forall i = 1,2,\ldots,d)$ (of course $B_iB_j = B_jB_i \forall i,j = 1,2,\ldots,d$). In this case q_T is $\mathcal{P}_N$-measurable for all $N \geq 1$ and there is no maximum order of convergence.

It would be interesting to find conditions under which D_T is almost surely positive-definite but I have not looked at this problem; one would certainly require the system of vectors

$$\{\Phi(s,T)(AB_i - B_iA)\Phi(0,s)q_0 \,;\, i = 1,2,\ldots,d \,;\, s \in [0,T]\}$$

to span R^m. The following lemma establishes a property which, under certain conditions, is equivalent to r^{th}-order asymptotic efficiency.

Lemma 4.1. *Let X be a square-integrable vector-valued random variable on some probability space $(\Omega, \mathcal{F}, P)$,*

$$X \in R^n, \qquad E\|X\|^2 < \infty.$$

Let $(X_N, \mathcal{P}_N \,;\, N = 1,2,\ldots)$ be an adapted sequence of approximations to X. Furthermore suppose that, with probability one,

$$\limsup_N \operatorname{var}(N^r c^T X \mid \mathcal{P}_N) < \infty \qquad \forall\, c \in R^n, \tag{4.2.4}$$

then the following two statements are equivalent:

(a) $(X_N \,;\, N = 1,2,\ldots)$ *is an r^{th}-order asymptotically efficient sequence of approximations to X,*

(b)

$$N^r(E(X \mid \mathcal{P}_N) - X_N) \to 0 \qquad \text{(a.s.)}. \tag{4.2.5}$$

NOTE. Since $\operatorname{var}(N^r c^T X \mid \mathcal{P}_N)$ is continuous in c, (4.2.4) is equivalent to

$$P(\limsup_N \operatorname{var}(N^r c^T X \mid \mathcal{P}_N) < \infty) = 1 \qquad \forall\, c \in R^n$$

(consider c with rational components).

PROOF.

$$E((c^T(X - X_N))^2 \mid \mathcal{P}_N) = \operatorname{var}(c^T X \mid \mathcal{P}_N) + (c^T(E(X \mid \mathcal{P}_N) - X_N))^2;$$

so (a) is equivalent to

$$P\Big(\frac{\operatorname{var}(N^r c^T X \mid \mathcal{P}_N) + (N^r c^T(E(X \mid \mathcal{P}_N) - X_N))^2 + 1}{\operatorname{var}(N^r c^T X \mid \mathcal{P}_N) + 1} \to 1 \qquad \forall\, c \in R^n\Big) = 1.$$

Under condition (4.2.4) this is eqivalent to (b). ∎

Corollary. *If* $(\hat{q}_N : N = 1,2,\ldots)$ *is first-order asymptotically efficient in the class of* ρ_N*-adapted sequences of approximations to* q_T *under the measure* P_2*, then* $(\Psi_f(\hat{q}_N) : N = 1,2,\ldots)$ *is first-order asymptotically efficient in the class of* ρ_N*-adapted sequences of approximations to* $\Psi_f(q_T)$ *under the measure* P_2*. Here* Ψ_f *is the Bayes estimation function defined in (3.6.1). Furthermore if* $(N_i ; i = 1,2,\ldots)$ *is any sequence of natural numbers with the refining property (3.3.12), i.e.*

$$N_i \longrightarrow \infty \quad \text{and} \quad i \geq j \Rightarrow N_i/N_j \text{ is an integer,}$$

then $(\hat{q}_{N_i} : i = 1,2,\ldots)$ *and* $(\Psi_f(\hat{q}_{N_i}) : i = 1,2,\ldots)$ *are first-order asymptotically efficient in the class of* ρ_{N_i}*-adapted sequences of approximations to* q_T *and* $\Psi_f(q_T)$ *respectively under the measure* P_1.

PROOF. By theorems 3.1 and 3.2 the following all hold with probability one:

$$\begin{aligned}
\operatorname{var}_2(N c^T q_T \mid \rho_N) &\to c^T D_T c \qquad \forall\, c \in R^m,\\
\operatorname{var}_2(N c^T \Psi_f(q_T) \mid \rho_N) &\to c^T D'_T c \qquad \forall\, c \in R^n,\\
\operatorname{var}_1(N_i c^T q_T \mid \rho_{N_i}) &\to c^T D_T c \qquad \forall\, c \in R^m,
\end{aligned}$$

and

$$\operatorname{var}_1(N_i c^T \Psi_f(q_T) \mid \rho_{N_i}) \to c^T D'_T c \qquad \forall\, c \in R^n.$$

Here, D_T and D'_T are the conditional-covariance marices given in (3.5.5) and (3.6.23) respectively. This means that condition (4.2.4) of lemma 4.1 is satisfied by q_T and $\Psi_f(q_T)$ under P_2, and under P_1 along the subsequence $(N_i ; i = 1,2,\ldots)$. So we need to prove that

$$N(E_2(\Psi_f(q_T) \mid \rho_N) - \Psi_f(\hat{q}_N)) \to 0 \quad \text{(a.s.)}, \tag{4.2.6}$$

$$N_i(E_1(q_T \mid \rho_{N_i}) - \hat{q}_{N_i}) \to 0 \quad \text{(a.s.)}, \tag{4.2.7}$$

and

$$N_i(E_1(\Psi_f(q_T) \mid \rho_{N_i}) - \Psi_f(\hat{q}_{N_i})) \to 0 \quad \text{(a.s.)}, \tag{4.2.8}$$

given that

$$N(E_2(q_T \mid \rho_N) - \hat{q}_N) \to 0 \quad \text{(a.s.)}. \tag{4.2.9}$$

(4.2.9) shows that under the measure P_2 the ρ_N-conditional distribution and moments of the normalised error $N(q_T - \hat{q}_N)$ converge to the optimal limits given in theorem 3.1 almost surely, i.e.

$$\begin{aligned}
&P_2(N(q_T - \hat{q}_N) \in \cdot \mid \rho_N) \Rightarrow N(0, D_T)(\cdot) \qquad \text{(a.s.)}\\
\text{and} \qquad &E_2(\|N(q_T - \hat{q}_N)\|^p \mid \rho_N) \to \int_{R^m} \|x\|^p N(0, D_T)(dx) \qquad \text{(a.s.)},
\end{aligned} \tag{4.2.10}$$

Theorem 3.2 and lemma 3.1 show that similar results are true for the normalised sequence

$N(\Psi_f(q_T) - \Psi_f(\hat{q}_N))$ under P_2 and the normalised subsequences $N_i(q_T - \hat{q}_{N_i})$ and $N_i(\Psi_f(q_T) - \Psi_f(\hat{q}_{N_i}))$ under P_1, i.e.

$$P_2(N(\Psi_f(q_T)-\Psi_f(\hat{q}_N)) \in \cdot \mid \rho_N) \Rightarrow N(0, D'_T)(\cdot) \quad \text{(a.s.)}, \tag{4.2.11}$$

$$P_1(N_i(q_T - \hat{q}_{N_i}) \in \cdot \mid \rho_{N_i}) \Rightarrow N(0, D_T)(\cdot) \quad \text{(a.s.)} \tag{4.2.12}$$

and

$$P_1(N_i(\Psi_f(q_T) - \Psi_f(\hat{q}_{N_i})) \in \cdot \mid \rho_{N_i}) \Rightarrow N(0, D'_T)(\cdot) \quad \text{(a.s.)}, \tag{4.2.13}$$

where $N(0, D'_T)$ is the optimal limit-distribution for Bayes estimates given in the corollary to theorem 3.2. The moments also converge almost surely. These expressions in turn imply the results (4.2.6), (4.2.7) and (4.2.8). ∎

The above lemma and corollary show that any sequence of approximations to q_T that is first-order asymptotically efficient under P_2 yields a corresponding sequence of approximate Bayes estimates that is also first-order asymptotically efficient under P_2. Appropriate subsequences of approximations to q_T and approximate Bayes estimates are first-order asymptotically efficient under P_1. The normalised error sequences in the approximations to q_T and $\Psi_f(q_T)$ converge to the optimal limits given in theorem 3.1 and the corollary to theorem 3.2 respectively.

In the remainder of this chapter the asymptotic properties of various specific approximation schemes will be analysed. The analysis will be done for approximations to q_T under P_2, the corresponding results for the approximate Bayes estimates under P_2 follow by the corollary to lemma 4.1. Likewise the corresponding results for approximations to q_T and the corresponding approximate Bayes estimates along appropriate subsequences (N_i) under P_1 follow.

4.3 A paradigmatic efficient scheme.

We shall now look at some specific approximation schemes for solving the differential eqaution (3.2.1), looking in particular at the asymptotic behaviour of the error in the approximate value of q_T and the error in the corresponding approximate Bayes estimates. We begin by investigating one-step methods, i.e. methods in which the approximate value of q at one point of the partition is calculated from the approximate value at the previous point and the increment of the driving-process W.

Under the measure P_2 the ρ_N-conditional mean sequence $(E_2(q_T \mid \rho_N)\,;\, N = 1, 2, \ldots)$ can be "Taylor-expanded" in a similar fashion to that given in chapter 2. In fact, if we define the sets S_k as in (2.5.1),

$$S_k := \{\alpha \in M : l(\alpha) + n(\alpha) \leq k \text{ OR } l(\alpha) = n(\alpha) = \tfrac{1}{2}(k+1)\},$$

then theorem 2.1 shows that

$$q_{(n+1)h} = \sum_{\alpha \in S_k} A_\alpha(A, B_1, \ldots, B_d) I_\alpha^W(nh, (n+1)h) q_{nh}$$

$$+ \sum_{\alpha \in B(S_k)} A_\alpha(A, B_1, \ldots, B_d) I_\alpha^W(q, nh, (n+1)h). \tag{4.3.1}$$

We take the conditional means of both sides of (4.3.1) to obtain

$$\overline{q_{(n+1)h}} = \sum_{\alpha \in S_k} A_\alpha(A, B_1, \ldots, B_d)(\overline{I_\alpha^W(nh,(n+1)h)}\,\overline{q_{nh}} + \overline{\widetilde{I_\alpha^W(nh,(n+1)h)q_{nh}}})$$

$$+ \sum_{\alpha \in B(S_k)} A_\alpha(A, B_1, \ldots, B_d)\overline{I_\alpha^W(q,nh,(n+1)h)}\,. \tag{4.3.2}$$

Here, for an integrable random variable X, $\overline{X}$ denotes $E_2(X \mid \rho_N)$ and $\widetilde{X}$ denotes $X - E_2(X \mid \rho_N)$. The "awkward" term in (4.3.2), $\overline{\widetilde{I_\alpha^W(nh,(n+1)h)q_{nh}}}$, can be eliminated by expanding the integral using the following formula:

$$\int_{nh}^{(n+1)h} I_\beta(nh,s)d\,W_s = \left(\frac{\Delta W_{n+1}}{h}\right) \int_{nh}^{(n+1)h} I_\beta(nh,s)d\,s + \int_{nh}^{(n+1)h} I_\beta(nh,s)d\,\widetilde{W}_s.$$

Then we obtain

$$\widetilde{I_\alpha^W(nh,(n+1)h)} = \sum_{k=0}^{l(\alpha)-n(\alpha)} \left(\frac{\Delta W_{n+1}}{h}\right)^k f_k(\widetilde{W}_t;\ t \in [nh,(n+1)h]),$$

where the functionals f_k are measurable and have zero mean,

$$E_2 f_k(\widetilde{W}_t,\ t \in [nh,(n+1)h]) = 0.$$

So

$$\overline{\widetilde{I_\alpha^W(nh,(n+1)h)q_{nh}}} = \sum_{k=0}^{l(\alpha)-n(\alpha)} \left(\frac{\Delta W_{n+1}}{h}\right)^k \overline{f_k(\widetilde{W}_t, t \in [nh,(n+1)h])q_{nh}}$$

$$= \sum_{k=0}^{l(\alpha)-n(\alpha)} \left(\frac{\Delta W_{n+1}}{h}\right)^k \overline{E_2(f_k(\widetilde{W}_t, t \in [nh,(n+1)h]) \mid \rho_N \vee \mathfrak{I}_{nh}^W)q_{nh}}$$

$$= 0, \qquad \text{by lemmas 3.2 (iii) and 3.3.}$$

So, from (4.3.2),

$$\overline{q_{(n+1)h}} = \sum_{\alpha \in S_k} A_\alpha(A, B_1, \ldots, B_d)\overline{I_\alpha^W(nh,(n+1)h)}\,\overline{q_{nh}}$$

$$+ \sum_{\alpha \in B(S_k)} A_\alpha(A, B_1, \ldots, B_d)\overline{I_\alpha^W(q,nh,(n+1)h)}\,. \tag{4.3.3}$$

This expansion leads naturally to the following family of one-step approximation schemes:

$$\overline{\hat{q}_0}(k) := q_0$$
$$\overline{\hat{q}_{n+1}}(k) := \sum_{\alpha \in S_k} A_\alpha(A, B_1, \ldots, B_d)\overline{I_\alpha^W(nh,(n+1)h)}\,\overline{\hat{q}_n}(k). \tag{4.3.4}$$

These schemes can be thought of as the *smoothed $k/2^{th}$-order Taylor schemes* since they yield approximate solutions that are the ρ_N-conditional means of the Taylor schemes given in definition 2.3; i.e.

$$\overline{\hat{q}_n}(k) = E_2(\hat{q}_n(k) \mid \rho_N). \tag{4.3.5}$$

The following lemma gives some asymptotic properties of the smoothed $k/2^{th}$-order Taylor schemes.

Lemma 4.2. *The difference between the partition-σ-field-conditional mean of the solution of (3.2.1) and the approximate value obtained by the smoothed $k/2^{th}$-order Taylor scheme, defined in (4.3.4), is of uniform $L_{(\infty)}$-order $k/2$, i.e.*

$$\sup_{n \leq N} E_2 \| N^{k/2}(E_2(q_{nh} \mid \rho_N) - \overline{\hat{q}_n}(k)) \|^p < \infty \quad \forall\, p > 0\,. \tag{4.3.6}$$

PROOF. Because of (4.3.5)

$$E_2(q_{nh} \mid \rho_N) - \overline{\hat{q}_n}(k) = E_2(q_{nh} - \hat{q}_n(k) \mid \rho_N),$$

where $\hat{q}_n(k)$ is the approximate solution of (3.2.1) obtained by the $k/2^{th}$-order Taylor scheme, given in definition 2.3. Theorem 2.3 shows that

$$q_{nh} - \hat{q}_n(k) = O_{(\infty)}(h^{k/2});$$

(4.3.6) follows by Jensen's inequality. ∎

The $1/2^{th}$ and 1^{st}-order Taylor schemes (the Euler and Mil'shtein schemes respectively) yield ρ_N-measurable approximate solutions; so they coincide with the smoothed $1/2^{th}$ and 1^{st}-order Taylor schemes. So the latter converge with order $1/2$ and 1 respectively. The question arises: is the smoothed 1^{st}-order Taylor scheme 1^{st}-order asymptotically efficient? In fact it is not; this is proved in part (ii) of theorem 4.1 below. However, for $k \geq 3$ the smoothed $k/2^{th}$-order Taylor schemes are 1^{st}-order asymptotically efficient; this is also shown in the theorem.

Of particular interest is the smoothed $3/2^{th}$-order Taylor scheme, because it represents the minimal expansion needed for 1^{st}-order asymptotic efficiency. I use this scheme as a paradigm; it is given explicitly below.

Definition 4.2. The following approximation scheme for solving bilinear s.d.e.s such as (3.2.1) will be called the *paradigm*:

$$\begin{aligned}
q_0^* &:= q_0,\\
q_{n+1}^* &:= \Big(I + Ah + \sum_{i=1}^{d} B_i \Delta W_{n+1}^{(i)} + \frac{1}{2}\sum_{i=1}^{d} B_i^2 (\Delta W_{n+1}^{(i)\,2} - h)\\
&\quad + \frac{1}{2}\sum_{\substack{i,j=1\\ i\neq j}}^{d} B_i B_j \Delta W_{n+1}^{(i)} \Delta W_{n+1}^{(j)} + \frac{1}{2}A^2h^2 + \frac{1}{2}\sum_{i=1}^{d}(AB_i + B_iA)h\Delta W_{n+1}^{(i)}\\
&\quad + \frac{1}{6}\sum_{i=1}^{d} B_i^3 (\Delta W_{n+1}^{(i)\,3} - 3h\Delta W_{n+1}^{(i)}) + \frac{1}{6}\sum_{\substack{i,j=1\\ i\neq j}}^{d} B_i^2 B_j (\Delta W_{n+1}^{(i)\,2} - h)\Delta W_{n+1}^{(j)}\\
&\quad + \frac{1}{6}\sum_{\substack{i,j,k=1\\ i\neq j, j\neq k, k\neq i}}^{d} B_i B_j B_k \Delta W_{n+1}^{(i)} \Delta W_{n+1}^{(j)} \Delta W_{n+1}^{(k)}\Big) q_n^*. \qquad (4.3.7)
\end{aligned}$$

For the case where the observations process is scalar ($d = 1$) this reduces to

$$\begin{aligned}
q_0^* &:= q_0,\\
q_{n+1}^* &:= \Big(I + Ah + B\Delta W_{n+1} + \frac{1}{2}B^2(\Delta W_{n+1}^2 - h) + \frac{1}{2}A^2h^2 + \frac{1}{2}(AB + BA)h\Delta W_{n+1}\\
&\quad + \frac{1}{6}B^3(\Delta W_{n+1}^3 - 3h\Delta W_{n+1})\Big) q_n^*. \qquad (4.3.8)
\end{aligned}$$

The paradigm is the $3/2^{th}$-order Taylor scheme with the term

$$\sum_{i=1}^{d}(AB_i - B_iA)\int_{nh}^{(n+1)h} \widetilde{W}_t^{(i)}\,dt$$

missing. In the degenerate case, where the matrix A commutes with all the B_i matrices, this term is zero and the paradigm converges with order $3/2$.

Theorem 4.1.

(i) For any $k \geq 3$, the smoothed $k/2^{th}$-order Taylor schemes defined in (4.3.4) yield 1^{st}-order asymptotically efficient approximate solutions to bilinear s.d.e.s of the type (3.2.1). In particular, the paradigm yields a 1^{st}-order asymptotically efficient approximate solution.

(ii) Consider the following approximation scheme for (3.2.1)

$$\hat{q}_0 := q_0,$$

$$\begin{aligned}\hat{q}_{n+1} := \Big(I + Ah + \sum_{i=1}^{d} B_i \Delta W_{n+1}^{(i)} + \frac{1}{2}\sum_{i=1}^{d} B_i^2 (\Delta W_{n+1}^{(i)\,2} - h) + \frac{1}{2}\sum_{\substack{i,j=1 \\ i\neq j}}^{d} B_i B_j \Delta W_{n+1}^{(i)} \Delta W_{n+1}^{(j)} \\ + (\tfrac{1}{2}A^2 + D_0)h^2 + \sum_{i=1}^{d} (\tfrac{1}{2}(AB_i + B_i A) + D_i) h \Delta W_{n+1}^{(i)} \\ + \sum_{i=1}^{d} (\tfrac{1}{6}B_i^3 + F_i)(\Delta W_{n+1}^{(i)\,3} - 3h\Delta W_{n+1}^{(i)}) \\ + \sum_{\substack{i,j=1 \\ i\neq j}}^{d} (\tfrac{1}{6}B_i^2 B_j + F_{i,j})(\Delta W_{n+1}^{(i)\,2} - h)\Delta W_{n+1}^{(j)} \\ + \sum_{\substack{i,j,k=1 \\ i\neq j\, j\neq k, k\neq i}}^{d} (\tfrac{1}{6}B_i B_j B_k + F_{i,j,k})\Delta W_{n+1}^{(i)} \Delta W_{n+1}^{(j)} \Delta W_{n+1}^{(k)}\Big)\hat{q}_n. \end{aligned} \tag{4.3.9}$$

(This is the paradigm with perturbed $3/2^{th}$-order terms.)

Let the process $(\Psi_t,\ t \in [0,T])$ be defined as follows:

$$\begin{bmatrix} q_0 \\ \Psi_0 \end{bmatrix} = \begin{bmatrix} q_0 \\ 0 \end{bmatrix},$$

$$d\begin{bmatrix} q_t \\ \Psi_t \end{bmatrix} = \begin{bmatrix} A & 0 \\ -D_0 & A \end{bmatrix}\begin{bmatrix} q_t \\ \Psi_t \end{bmatrix} dt + \sum_{i=1}^{d} \begin{bmatrix} B_i & 0 \\ -D_i & B_i \end{bmatrix}\begin{bmatrix} q_t \\ \Psi_t \end{bmatrix} dW_t^{(i)} \qquad t \in [0,T]. \tag{4.3.10}$$

Then

(a) The approximation to q_T given by (4.3.9) is 1^{st}-order asymptotically efficient if and only if both of the following conditions are true:

(e1)

$$\Psi_T = 0 \qquad \text{(a.s.)}$$

and

(e2) for each $i,j,k = 1,2,\ldots,d$

$$P_2(F_i q_t = F_{i,j} q_t = F_{i,j,k} q_t = 0 \qquad \forall\, t \in [0,T]) = 1.$$

Also

(b) If (e2) above holds then

$$N(E_2(q_T \mid \rho_N) - \hat{q}_N) \to T\Psi_T \qquad \text{(a.s.)}, \tag{4.3.11}$$

where Ψ_T is given by (4.3.10).

NOTE. Of particular interest are the cases that represent the paradigm with any combination of $3/2^{th}$-order terms removed. For example, when

$$D_0 = -\frac{1}{2}A^2;\ D_1 = -\frac{1}{2}AB_1;\ D_2,\ldots,D_d,\ F_i,\ F_{i,j},\ F_{i,j,k} = 0 \quad \forall\, i,j,k = 1,2,\ldots,d$$

the scheme given in (4.3.9) is the paradigm less the terms $\frac{1}{2}A^2h^2$ and $\frac{1}{2}AB_1h\Delta W_{n+1}^{(1)}$.

Corollary. *Let* $(\chi_t;\ t \in [0,T])$ *be part of the solution of the following s.d.e.:*

$$\begin{bmatrix} q_0 \\ \chi_0 \end{bmatrix} = \begin{bmatrix} q_0 \\ 0 \end{bmatrix},$$

$$d\begin{bmatrix} q_t \\ \chi_t \end{bmatrix} = \begin{bmatrix} A & 0 \\ -A^2/2 & A \end{bmatrix}\begin{bmatrix} q_t \\ \chi_t \end{bmatrix} dt + \sum_{i=1}^{d}\begin{bmatrix} B_i & 0 \\ -(AB_i + B_iA)/2 & B_i \end{bmatrix}\begin{bmatrix} q_t \\ \chi_t \end{bmatrix} dW_t^{(i)}$$

$$t \in [0,T]\,. \tag{4.3.12}$$

The Mil'shtein scheme [MI1] described in chapter 1 yields a 1^{st}*-order asymptotically efficient approximation to* q_T *if and only if both of the following conditions are true*:

$$\chi_T = 0 \qquad \text{(a.s.)}$$

and for each $i,j,k = 1,2,\ldots,d$

$$P_2(B_i^3 q_t = B_i^2 B_j q_t = B_i B_j B_k q_t = 0 \qquad \forall\, t \in [0,T]) = 1\,.$$

PROOF OF THEOREM 4.1 PART (i). It follows from lemma 4.2 and the moment form of the Borel-Cantelli lemma that for $k \geq 3$

$$N(E_2(q_T \mid \rho_N) - \bar{\bar{q}}(k)) \to 0 \qquad \text{(a.s.)}.$$

(i) follows by lemma 4.1.

PROOF OF THEOREM 4.1 PART (ii). We subtract (4.3.9) from (4.3.7) to obtain a difference equation for $N(q_n^* - \hat{q}_n)$. Theorem 2.2 shows that the solution of this equation differs from (ξ_n), the solution of equation (4.3.14), by an amount which is of uniform $L_{(\infty)}$-order $1/2$, i.e.

$$N(q_n^* - \hat{q}_n) - \xi_n = O_{(\infty)}(h^{1/2}), \tag{4.3.13}$$

where $\xi_0 := 0$,

$$\xi_{n+1} := \Big(I + Ah + \sum_{i=1}^{d} B_i \Delta W_{n+1}^{(i)}\Big)\xi_n + d_{n+1}(1) + d_{n+1}(2), \tag{4.3.14}$$

$$d_{n+1}(1) := -N\Big(D_0 h^2 + \sum_{i=1}^{d} D_i h \Delta W_{n+1}^{(i)}\Big)\hat{q}_n(1),$$

$$d_{n+1}(2) := -N\Big(\sum_{i=1}^{d} F_i({\Delta W_{n+1}^{(i)}}^3 - 3h\Delta W_{n+1}^{(i)}) + \sum_{\substack{i,j=1 \\ i\neq j}}^{d} F_{i,j}({\Delta W_{n+1}^{(i)}}^2 - h)\Delta W_{n+1}^{(j)}$$

$$+ \sum_{\substack{i,j,k=1 \\ i\neq j\, j\neq k, k\neq i}}^{d} F_{i,j,k}\Delta W_{n+1}^{(i)}\Delta W_{n+1}^{(j)}\Delta W_{n+1}^{(k)}\Big)\hat{q}_n(1); \tag{4.3.15}$$

$\hat{q}_n(1)$ is the $1/2^{th}$-order Taylor approximation for q_{nh} (see defintion 2.3). We also define (η_n) by the following:

$$\eta_0 := 0,$$

$$\eta_{n+1} := \Big(I + Ah + \sum_{i=1}^{d} B_i \Delta W_{n+1}^{(i)}\Big)\eta_n + d_{n+1}(1). \tag{4.3.16}$$

We prove (b) first. Let $(d_{n+1}(3))$ be the following approximation to $(d_{n+1}(2))$;

$$d_{n+1}(3) := -N\Big(\sum_{i=1}^{d} F_i({\Delta W_{n+1}^{(i)}}^3 - 3h\Delta W_{n+1}^{(i)}) + \sum_{\substack{i,j=1 \\ i\neq j}}^{d} F_{i,j}({\Delta W_{n+1}^{(i)}}^2 - h)\Delta W_{n+1}^{(j)}$$

$$+ \sum_{\substack{i,j,k=1 \\ i\neq j\, j\neq k, k\neq i}}^{d} F_{i,j,k}\Delta W_{n+1}^{(i)}\Delta W_{n+1}^{(j)}\Delta W_{n+1}^{(k)}\Big)(\hat{q}_n(1) - q_{nh}). \tag{4.3.17}$$

Since (e2) is true

$$P_2(d_n(3)) = d_n(2); \quad \forall\, n \leq N;\ N = 1,2,\ldots) = 1.$$

We subtract (4.3.16) from (4.3.14) and use this approximation to obtain

$$\xi_{n+1} - \eta_{n+1} = \Big(I + Ah + \sum_{i=1}^{d} B_i \Delta W_{n+1}^{(i)}\Big)(\xi_n - \eta_n) + d_{n+1}(3) \qquad \text{(a.s.)}.$$

Now $d_{n+1}(3) = O_{(\infty)}(h^{3/2})$ and $E_2(d_{n+1}(3) \mid \mathcal{F}_{nh}^W) = 0$; so theorem 2.2 shows that

$$\xi_n - \eta_n = O_{(\infty)}(h^{1/2}). \tag{4.3.18}$$

The $1/2^{th}$-order Taylor expansion for equation (4.3.10) is given by

$$\begin{bmatrix} \hat{q}_0(1) \\ \hat{\Psi}_0(1) \end{bmatrix} := \begin{bmatrix} q_0 \\ 0 \end{bmatrix},$$

$$d\begin{bmatrix} \hat{q}_{n+1}(1) \\ \hat{\Psi}_{n+1}(1) \end{bmatrix} := \Big(I + \begin{bmatrix} A & 0 \\ -D_0 & A \end{bmatrix} h + \sum_{i=1}^{d} \begin{bmatrix} B_i & 0 \\ -D_i & B_i \end{bmatrix} \Delta W_{n+1}^{(i)}\Big) \begin{bmatrix} \hat{q}_n(1) \\ \hat{\Psi}_n(1) \end{bmatrix} \quad n = 0, 1, \ldots, N-1.$$

Clearly,

$$\eta_n = T\hat{\Psi}_n(1) \quad \forall\ n = 0, 1, \ldots, N, \tag{4.3.19}$$

and by theorem 2.3

$$T\hat{\Psi}_N(1) - T\Psi_T = O_{(\infty)}(h^{1/2}). \tag{4.3.20}$$

We combine (4.3.13), (4.3.18), (4.3.19) and (4.3.20) and apply the moment form of the Borel-Cantelli lemma to show that

$$N(q_N^* - \hat{q}_N) \to T\Psi_T \qquad \text{(a.s.)}.$$

This, together with part (i) of the present theorem and lemma 4.1, proves (b).

Part (b) of the present theorem and lemma 4.1 show that if condition (e2) is true then the 1^{st}-order asymptotic efficiency of the approximation scheme given in (4.3.9) is equivalent to condition (e1). To prove (a) it therefore remains to be shown that if (e2) is not true then the approximation scheme given by (4.3.9) is not 1^{st}-order asymptotically efficient.

Suppose (e2) is not true, ξ_N is an approximation to $N(q_N^* - \hat{q}_N)$; in the following we show that $P_2(\xi_N \to 0) \neq 1$; we do this by first bounding the moments of ξ_N below by the moments of $\xi_N - \eta_N$. We solve (4.3.16) to obtain

$$\eta_N = \sum_{n=1}^{N} \Big(I + Ah + \sum_{i=1}^{d} B_i \Delta W_N^{(i)}\Big) \ldots \Big(I + Ah + \sum_{i=1}^{d} B_i \Delta_{n+1}^{(i)}\Big) d_n(1);$$

similarly, we solve (4.3.14) and (4.3.16) and subtract to obtain

$$\xi_N - \eta_N = \sum_{n=1}^{N} \Big(I + Ah + \sum_{i=1}^{d} B_i \Delta W_N^{(i)}\Big) \ldots \Big(I + Ah + \sum_{i=1}^{d} B_i \Delta W_{n+1}^{(i)}\Big) d_n(2).$$

Now, $d_n(2)$ is uncorrelated with ΔW_m and $d_m(1)$ for all $m, n \leq N$, which means that η_N and $\xi_N - \eta_N$ are uncorrelated. So

$$\begin{aligned} E_2\|\xi_N\|^2 &= E_2\|\eta_N\|^2 + E_2\|\xi_N - \eta_N\|^2 \\ &\geq E_2\|\xi_N - \eta_N\|^2. \end{aligned} \tag{4.3.21}$$

Also, again using the orthogonality of $d_n(2)$ and ΔW_m,

$$E_2\|\xi_{n+1}-\eta_{n+1}\|^2 = E_2\Big\|\Big(I + Ah + \sum_{i=1}^{d} B_i\Delta W_{n+1}^{(i)}\Big)(\xi_n-\eta_n)\Big\|^2 + E_2\|d_{n+1}(2)\|^2$$
$$\geq (1-Kh)E_2\|\xi_n-\eta_n\|^2 + E_2\|d_{n+1}(2)\|^2, \tag{4.3.22}$$

where K is a positive constant depending on the matrices $A, B_1, \ldots, B_d$. Now

$$E_2\|d_{n+1}(2)\|^2 = T^2\Big(6\sum_{i=1}^{d} E_2\|F_i\hat{q}_n(1)\|^2 + 2\sum_{\substack{i,j=1\\ i\neq j}}^{d} E_2\|F_{i,j}\hat{q}_n(1)\|^2$$
$$+ \sum_{\substack{i,j,k=1\\ i\neq j\, j\neq k, k\neq i}}^{d} E_2\|F_{i,j,k}\hat{q}_n(1)\|^2\Big)h. \tag{4.3.23}$$

From (4.3.21), (4.3.22) and (4.3.23) we have the following:

$$E_2\|\xi_N\|^2 \geq \sum_{n=1}^{N}(1-Kh)^{N-n}E_2\|d_n(2)\|^2$$
$$\to T^2\int_0^T \exp(T-s)\Big(6\sum_{i=1}^{d} E_2\|F_iq_s\|^2 + 2\sum_{\substack{i,j=1\\ i\neq j}}^{d} E_2\|F_{i,j}q_s\|^2$$
$$+ \sum_{\substack{i,j,k=1\\ i\neq j\, j\neq k, k\neq i}}^{d} E_2\|F_{i,j,k}q_s\|^2\Big)ds.$$

Since the sample paths of q are continuous, except perhaps for a set of probability 0, condition (e2) is equivalent to the following: for each $i,j,k = 1, \ldots, d$

$$(P_2\times\lambda)(F_iq = F_{i,j}q = F_{i,j,k}q = 0) = T,$$

where λ is Lebesgue measure on [0,T]. So, if (e2) is not true

$$\liminf_N E_2\|\xi_N\|^2 > 0. \tag{4.3.24}$$

Theorem 2.2 shows that (ξ_n) is of uniform $L_{(\infty)}$-order 1; so that, in particular,

$$\sup_N E_2\|\xi_N\|^2 < \infty.$$

So, by the Vallee-Poussin theorem, (ξ_n) is uniformly integrable. It is a property of uniformly integrable

sequences of random variables that their almost-sure convergence to some limit implies the convergence of their first moment to that of the limit (see the corollary to theorem 1.3 in [LS1]). Since the first moment of $\|\xi_N\|^2$ does not converge to zero this implies that

$$P_2(\|\xi_N\| \to 0) \neq 1.$$

It follows by (4.3.13) and the moment form of the Borel-Cantelli lemma that

$$P_2(N(q_N^* - \hat{q}_N) \to 0) \neq 1;$$

so

$$P_2(N(E_2(q_T \mid \rho_N) - \hat{q}_N) \to 0) \neq 1.$$

This, together with lemma 4.1, shows that the approximation given by (4.3.9) is not 1^{st}-order asymptotically efficient. ∎

The conclusion of part (ii) of theorem 4.1 looks at first-sight stronger than necessary; it seems reasonable to prove only that the scheme given in (4.3.9) does not *generally* yield 1^{st}-order asymptotically efficient approximate solutions, i.e. that there exist simple, specific counter-examples for which the scheme does not yield a 1^{st}-order asymptotically efficient sequence of approximations. It seems however that the only counter-examples that simplify the analysis are those in which the matrices $A, B_1, \ldots, B_d$ all commute, e.g. a scalar equation or one with $A = 0$ etc. These, however, are degenerate cases where the maximum order of convergence is infinite and 1^{st}-order asymptotic efficiency is no longer an optimal property. It is not worthwhile to show that a method does not possess a certain property in a case where that property is not very important.

Part (ii)(b) of theorem 4.1 shows that if the "ΔW^3" terms in the approximation scheme are "correct", i.e. if they are the terms prescribed in the paradigm, then the normalised difference between the conditional mean $E_2(q_T \mid \rho_N)$ and the approximate value $\hat{q}_N$ converges almost surely to $T\Psi_T$. This means that the ρ_N-conditional distribution and moments of the normalised error $N(q_T - \hat{q}_N)$ converge almost surely to the optimal distribution and moments given in theorem 3.1 but with the non-zero mean $T\Psi_T$, i.e.

$$P_2(N(q_T - \hat{q}_N) \in \cdot \mid \rho_N) \Rightarrow N(T\Psi_T, D_T) \quad \text{(a.s.)} \tag{4.3.25}$$

and all the conditional moments converge almost surely; a similar result applies for the corresponding Bayes estimates.

The distribution on the right-hand side of (4.3.25) is of course that of the random variable $T\epsilon_T := T(\Psi_T + \theta_T)$ where θ_T is given in (3.5.3). Clearly ϵ_T is given by

$$\begin{aligned}
\begin{bmatrix} q_0 \\ \epsilon_0 \end{bmatrix} &= \begin{bmatrix} q_0 \\ 0 \end{bmatrix}, \\
d\begin{bmatrix} q_t \\ \epsilon_t \end{bmatrix} &= \begin{bmatrix} A & 0 \\ -D_0 & A \end{bmatrix} \begin{bmatrix} q_t \\ \epsilon_t \end{bmatrix} dt + \sum_{i=1}^{d} \begin{bmatrix} B_i & 0 \\ -D_i & B_i \end{bmatrix} \begin{bmatrix} q_t \\ \epsilon_t \end{bmatrix} dW_t^{(i)} \\
&\quad + \frac{1}{\sqrt{12}} \sum_{i=1}^{d} \begin{bmatrix} 0 & 0 \\ (AB_i - B_iA) & 0 \end{bmatrix} \begin{bmatrix} q_t \\ \epsilon_t \end{bmatrix} dV_t^{(i)}.
\end{aligned} \tag{4.3.26}$$

Once again, because there exists a version of the fundamental solution of (3.2.1) which is invertible, $\Phi(s,t)$, ϵ_T has an explicit representation. In fact

$$\epsilon_T = \int_0^T \Phi(s,T)D_0\Phi(0,s)q_0 d\,s + \sum_{i=1}^{d}\int_0^T \Phi(s,T)D_i\Phi(0,s)q_0 d\,W_s^{(i)}$$
$$+ \frac{1}{\sqrt{12}}\sum_{i=1}^{d}\int_0^T \Phi(s,T)(A\,B_i - B_i A)\Phi(0,s)q_0 d\,V_s^{(i)}. \qquad (4.3.27)$$

If one or more of the "ΔW^3" terms in an approximation scheme are "incorrect", i.e. if condition (e2) of theorem 4.1 part (ii) does not hold, then I suspect that neither the ρ_N-conditional distribution of the normalised error nor its moments converge.

4.4 A comparison of some approximation schemes.

In this section various approximation schemes for solving equation (3.2.1) are compared, with particular emphasis given to their asymptotic properties. It is natural to begin by analysing the fundamental one-step schemes defined in (4.3.4). For $k = 1, 2$ and 3 these are the Euler scheme (Maruyama [MA1]), the Mil'shtein scheme (Mil'shtein [MI1]) and the paradigm (Definition 4.2). These have the minimum terms required for almost-sure convergence, 1^{st}-order convergence and 1^{st}-order asymptotic efficiency respectively. We would like to know which of them involves the least amount of computation to achieve a given degree of accuracy. In such comparisons we must take into account the fact that the simpler methods, which require fewer calculations at each step, need a finer partition to achieve a given degree of accuracy than more complex methods.

For high accuracy the first-order schemes are better than the Euler scheme since, roughly speaking, doubling the accuracy-requirement doubles the number of steps needed in the first-order schemes but quadruples the number needed in the Euler scheme.

It seems worthwhile to compare the accuracy of the 1^{st}-order schemes for large N. The error in the paradigm is predominantly caused by the absence of the d non-ρ_N-measurable $3/2^{th}$-order terms $(A\,B_i - B_i A)\int_{nh}^{(n+1)h} \widetilde{W}_t^{(i)} dt$; the error in the Mil'shtein scheme is predominantly caused by the absence of all the $3/2^{th}$-order terms: $1 + d + d^3$ of them. Since these terms are uncorrelated we might expect that if N is large the root-mean-square error in the Mil'shtein scheme is roughly $(K \times (1 + d + d^3)/d)^{1/2}$ times as big as the root-mean-square error in the paradigm for a given step size, where K is some constant depending on the matrices $A, B_1, \ldots, B_d$: if the matrices "nearly commute" then K is large. Since both schemes converge linearly we would require roughly $(K \times (1 + d + d^3)/d)^{1/2}$ times as many steps in the Mil'shtein scheme as in the paradigm for a given degree of accuracy.

For the purposes of solving the filtering problem we would like to minimise the number of "on-

line" calculations, that is calculations that must be made when the filter is operating. The number of "off-line" calculations, those that can be made at the time the filter is designed, is not so important. I have estimated the number of "on-line" scalar multiplications required at each step by the three schemes for the two cases where A is a "full" matrix and where A is "sparse". In both cases I have assumed that the B matrices are diagonal, as is the case in the filtering problem: this greatly simplifies the Mil'shtein and paradigm schemes.

I have based my estimates of the number of calculations required at each step on the following calculation sequences, which seem optimal if one assumes that m (the dimension of q) is much larger than d (the dimension of W). If the Markov chain that we are trying to estimate is an approximation to a diffusion then it is likely that A will be sparse and that $m \gg d$.

Euler scheme

Off-line calculations:

$$E_0 := I + A\,h.$$

On-line calculations at each step:

$$K_1 := \sum_{i=1}^{d} B_i \Delta W_{n+1}^{(i)},$$

$$\hat{q}_{n+1}(E) := (E_0 + K_1)\hat{q}_n(E). \tag{4.4.1}$$

Mil'shtein scheme

Off-line calculations:

$$M_0 := I + \Big(A - \frac{1}{2}\sum_{i=1}^{d} B_i^2\Big)h,$$

$$M_i := \frac{1}{\sqrt{2}} B_i \qquad i = 1, 2, \ldots, d.$$

On-line calculations at each step:

$$K_1 := \sum_{i=1}^{d} M_i \Delta W_{n+1}^{(i)},$$

$$\hat{q}_{n+1}(M) := (M_0 + K_1(\sqrt{2}I + K_1))\hat{q}_n(M). \tag{4.4.2}$$

Paradigm

Off-line calculations:

$$P_0 := I + \Big(A - \frac{1}{2}\sum_{i=1}^{d} B_i^2\Big)h + \frac{1}{2}A^2h^2,$$

$$P_i' := \frac{1}{2}\Big(AB_i + B_iA - B_i^3 - \frac{1}{3}\sum_{\substack{j=1 \\ j\neq i}}^{d} B_j^2 B_i\Big)h \qquad i = 1, 2, \ldots, d,$$

$$P_i := \frac{1}{6^{1/3}} B_i.$$

On-line calculations at each step:

$$K_1 := \sum_{i=1}^{d} P_i \Delta W_{n+1}^{(i)},$$

$$K_2 := \sum_{i=1}^{d} P_i' \Delta W_{n+1}^{(i)},$$

$$\hat{q}_{n+1}(P) := (P_0 + K_2 + K_1(6^{1/3} I + K_1(\frac{3}{6^{1/3}} I + K_1)))\hat{q}_n(P). \tag{4.4.3}$$

The following table lists the number of on-line scalar multiplications required at each step for the three schemes.

Table 4.1.

Number of scalar multiplications in one step.

Scheme	A–full	A–tridiagonal
Euler	$m(m+d)$	$m(d+3)$
Mil'shtein	$m(m+d+1)$	$m(d+4)$
Paradigm	$m(m+(m+1)d+2)$	$m(2d+5)$

$B_1, B_2, \ldots, B_d$ are assumed to be diagonal.

If m is large there seems little point in using the Euler scheme since the Mil'shtein scheme is similar in complexity but converges with twice the order. The paradigm is roughly $d+1$ times as complicated in one step as the Mil'shtein scheme if A is a full matrix, but the Mil'shtein scheme requires about $(K \times (1 + d + d^3)/d)^{1/2}$ times as many steps for the same accuracy, where K depends on $A, B_1, \ldots, B_d$. Whether d is large or small these two schemes seem to be roughly equivalent in overall complexity.

If the matrix A is "sparse" the paradigm is only twice as complicated as the Mil'shtein scheme in one step. It is likely in a lot of cases that the Mil'shtein scheme will require more than twice the number of steps to achieve the same accuracy, particularly if d is large.

I have not had time to carry out any computer simulations of these schemes but A. Estandia has kindly given me a plot from some simulation work he has been doing, which gives a flavour of their performance. This particular example concerns the approximate solution of the following equation:

$$q_0 := \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix},$$

$$dq_t = Aq_t dt + Bq_t dW_t, \tag{4.4.4}$$

where

$$A := \begin{bmatrix} -2 & 1 & 0 & 1 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 1 & 0 & 1 & -2 \end{bmatrix}, \qquad B := \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

(4.4.4) is the equation for the un-normalised conditional distribution of a four-state Markov chain $(X_t,\ t \in [0,T])$ with generator A. The initial distribution q_0 is the steady-state solution of the Kolmogorov forward equation; so we are assuming that X_0 is unknown.

$$X_t \in \{a_1, a_2, a_3, a_4\};$$

$$P(X_0 = a_i) = \frac{1}{4} \text{ for } i = 1,2,3,4;$$

$$W_t = \int_0^t (I_{\{X_t=a_1\}} - I_{\{X_t=a_3\}})dt + \beta_t. \tag{4.4.5}$$

Diagram 4.1 shows the mean-square error in the approximate values of $q_{1.024}$ obtained by the Euler, Mil'shtein and paradigm schemes for $1,2,4,\ldots,2^i,\ldots,1024$ partition points. This is calculated by averaging the squared error over 100 sample paths of simulated Wiener process W, using the approximate value obtained by the paradigm with $N = 1024$ as the reference. More simulations and the justification of such simulations are needed before conclusions can be drawn, but diagram 4.1 has the form that we would expect from the theory. The rate of convergence of the mean-square errors, $1/N$ in the Euler case and $1/N^2$ in the Mil'shtein and paradigm cases, can be clearly seen, as can the "asymptotic bias" of the non-1st-order-asymptotically-efficient Mil'shtein scheme.

Other schemes One problem with Taylor-expansion-type approximation schemes such as those discussed above, when applied to *general* nonlinear s.d.e.s, is that they require the evaluation of derivatives of the drift and diffusion coefficients. For instance, consider the following scalar nonlinear s.d.e:

$$dx_t = a(x_t)dt + b(x_t)dW_t. \tag{4.4.6}$$

The Mil'shtein approximation is given by

$$\hat{x}_{n+1} := \hat{x}_n + (a(\hat{x}_n) - \frac{1}{2}\frac{db}{dx}(\hat{x}_n)b(\hat{x}_n))h + b(\hat{x}_n)\Delta W_{n+1} + \frac{1}{2}\frac{db}{dx}(\hat{x}_n)b(\hat{x}_n)\Delta W_{n+1}^2; \tag{4.4.7}$$

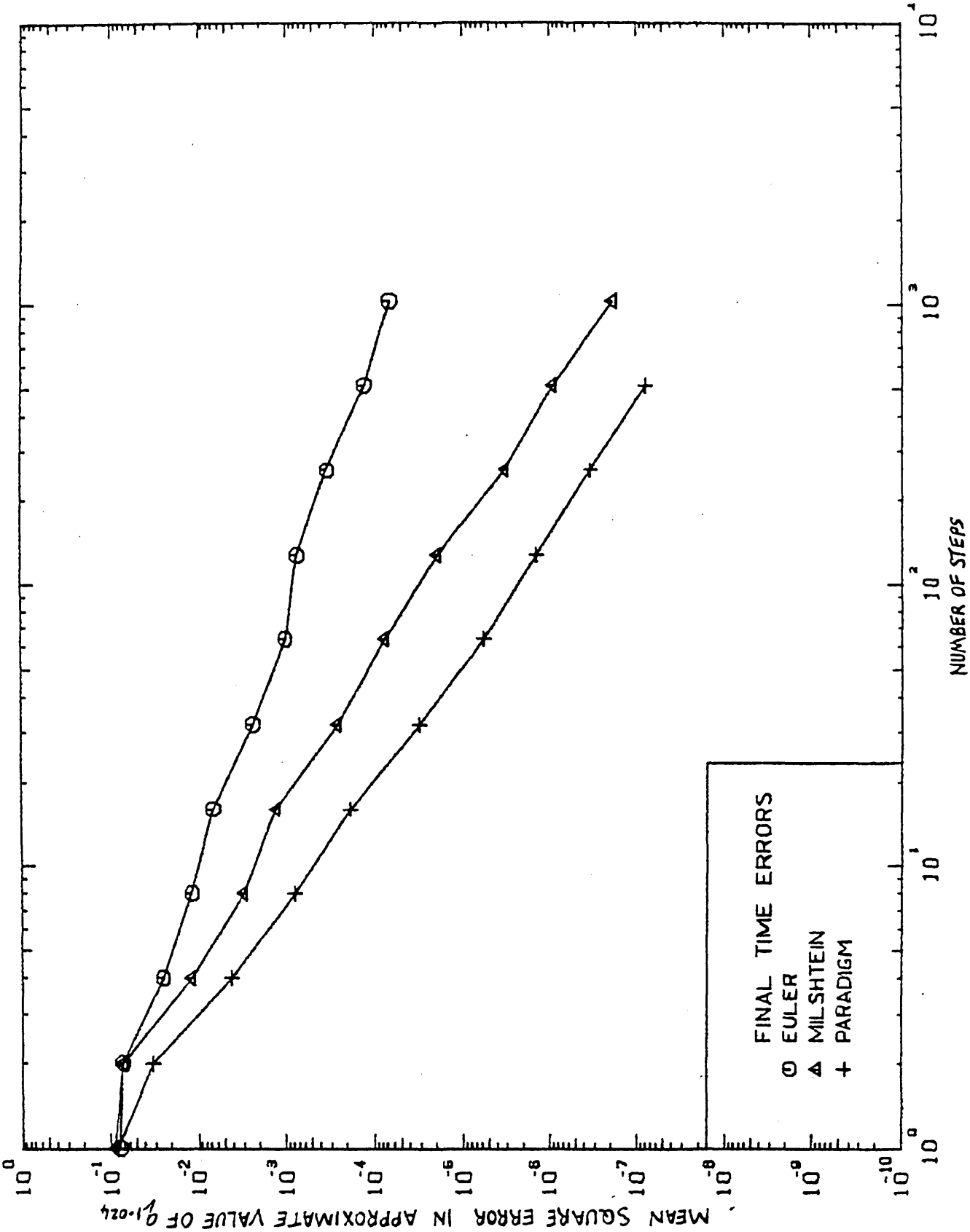
so it is necessary to evaluate the derivative $\frac{db}{dx}$. The higher-order Taylor schemes defined by Wagner and Platen in [WP1] require the evaluation of higher-order derivatives. This problem is overcome in ordinary differential equation theory by using methods of the Runge-Kutta type, which match terms in Taylor expansions without the need to evaluate derivatives. Rümelin in [RU1] considers methods of this type applied to stochastic differential equations. He begins by giving a general explicit Runge-Kutta scheme and then shows that the maximum rate of convergence that can be achieved in one step is h^3 in the sense that

$$E|X_h - \hat{X}_1|^2 \leq Kh^3 \qquad \text{for some positive constant } K.$$

This corresponds to the fact that the maximum order of convergence over an interval is $O_2(h)$, i.e.

$$\sup_{n \leq \frac{N}{N}} E|N(X_{nh} - \hat{X}_n)|^2 < \infty.$$

Diagram 4.1


FINAL TIME ERRORS
EULER
MILSHTEIN
PARADIGM
MEAN SQUARE ERROR IN APPROXIMATE VALUE OF $q_{1.024}$
NUMBER OF STEPS

Rümelin also shows that this maximum order can only be achieved if the coefficients used in the Runge-Kutta scheme are those obtained from the Stratonovich form of the original differential equation.

It seems to me that the best way of choosing a Runge-Kutta scheme is by matching the terms in the Taylor expansions appropriate to stochastic differential equations (formulated by Wagner and Platen in [WP1] and discussed in chapter 2). For instance a minimal Runge-Kutta scheme for almost sure convergence is the Euler scheme, a minimal Runge-Kutta scheme for 1^{st}-order convergence (the maximum order) is the following, which matches terms in the Mil'shtein scheme. I have given this scheme as applied to equation (4.4.6).

$$\begin{aligned} x_n^{(1)} &:= \hat{x}_n + \frac{1}{2}b(\hat{x}_n)\Delta W_{n+1}, \\ \hat{x}_{n+1} &:= \hat{x}_n + c(\hat{x}_n)h + b(x_n^{(1)})\Delta W_{n+1}, \end{aligned} \tag{4.4.8}$$

where c is the following Stratonovich drift-coefficient:

$$c(x) := a(x) - \frac{1}{2}\frac{db}{dx}(x)b(x). \tag{4.4.9}$$

Specific Runge-Kutta schemes designed for use with ordinary differential equations match the terms in Taylor expansions appropriate to ordinary differential equations rather than those appropriate to stochastic differential equations. For instance the following scheme, which matches the terms needed for 2^{nd}-order convergence in o.d.e.s:

$$\begin{aligned} x_n^{(1)} &:= \hat{x}_n + c(\hat{x}_n)h + b(\hat{x}_n)\Delta W_{n+1}, \\ \hat{x}_{n+1} &:= \hat{x}_n + \frac{1}{2}(c(\hat{x}_n) + c(x_n^{(1)}))h + \frac{1}{2}(b(\hat{x}_n) + b(x_n^{(1)}))\Delta W_{n+1}. \end{aligned} \tag{4.4.10}$$

This scheme is sometimes referred to as the Heun scheme (as are others, see Gear [GE1]) it has the following equivalent Taylor expansion:

$$\begin{aligned} \hat{x}_{n+1} = \hat{x}_n &+ c(\hat{x}_n)h + b(\hat{x}_n)\Delta W_{n+1} + \frac{1}{2}\frac{dc}{dx}(\hat{x}_n)c(\hat{x}_n)h^2 \\ &+ \frac{1}{2}\frac{db}{dx}(\hat{x}_n)b(\hat{x}_n)\Delta W_{n+1}^2 + \frac{1}{2}\Big(\frac{dc}{dx}(\hat{x}_n)b(\hat{x}_n) + \frac{db}{dx}(\hat{x}_n)c(\hat{x}_n)\Big)h\Delta W_{n+1} \\ &+ \frac{1}{4}\frac{d^2b}{dx^2}(\hat{x}_n)b^2(\hat{x}_n)\Delta W_{n+1}^3 + O_2(h^2) \end{aligned} \tag{4.4.11}$$

if one assumes various differentiability conditions on b and c. (4.4.11) shows that the scheme given in (4.4.10) achieves the maximum order of convergence as does the simpler scheme (4.4.8) but despite its extra complexity does not match the order-$h^{3/2}$ terms needed for 1^{st}-order asymptotic efficiency.

It seems, therefore, that the usual Runge-Kutta schemes designed for ordinary differential equations are not ideal for use with stochastic differential equations. What one needs is a family of Runge-Kutta schemes designed specially for s.d.e.s such as (4.4.8), but I have not investigated this further.

To solve the Markov-chain filtering problem we wish to solve the *bilinear* s.d.e. (3.2.1), and the evaluation of the derivatives of the drift and diffusion coefficients needed for the Taylor-expansion methods is easy (the first derivatives are constant matrices, the second and higher-order derivatives are zero). So there seems little point in using Runge-Kutta methods in the present case.

It is worth mentioning the well-known Wong-Zakai scheme [WZ1] for approximating the solutions of s.d.e.s by the solutions of ordinary differential equations. For equation (3.2.1) this takes the

form

$$\hat{q}_0 := q_0,$$

$$\dot{\hat{q}}_t(N) = \Big(A - \frac{1}{2}\sum_{i=1}^{d} B_i^2\Big)\hat{q}_t(N) + \sum_{i=1}^{d} B_i \hat{q}_t(N)\dot{\overline{W}}_t \qquad t \in [0,T]. \tag{4.4.12}$$

Here $\overline{W}_t := E_2(q_T \mid \rho_N)$; so

$$\dot{\overline{W}}_t = \frac{\Delta W_{n+1}}{h} \quad \text{if} \quad t \in [nh,(n+1)h].$$

It is distinct from the other schemes discussed in this section because it is not a finite-difference method. Clark in [CL2] has recently shown that this scheme is not (1^{st}-order asymptotically) efficient but a similar scheme with a compensation term is; for equation (3.2.1) this is given by

$$\hat{q}_0 := q_0$$

$$\dot{\hat{q}}_t(N) = \Big(A - \frac{1}{2}\sum_{i=1}^{d} B_i^2 + \frac{h}{12}[B,[B,A]]\Big)\hat{q}_t(N) + \sum_{i=1}^{d} B_i \hat{q}_t(N)\dot{\overline{W}}_t \qquad t \in [0,T], \tag{4.4.13}$$

where $[B,A]$ is the Lie-bracket operation—

$$[B,A] := AB - BA.$$

Chapter 5

Further Considerations

5.1 Properties on a finite partition.

In chapters 3 and 4 we were considering *asymptotic* properties of approximation schemes for the Markov-chain filtering problem. These properties allow us to make such statements as, "method A is more accurate than method B if the number of partition points is sufficiently large; we have not quantified "sufficiently large". For a practical filter design we must also consider the properties of methods on specific finite partitions; it may be that one scheme is asymptotically more accurate than another but that for any number of partition points that can be considered as practical in a given example, the second scheme is the more accurate.

The numerical analyst considers finite-partition properties of schemes such as their *region of absolute stability*; that is the area in the complex plane in which the product of the step-size h and an eigenvalue of a simple test differential equation may be,such that the effect of the error incurred in one step propagates to subsequent steps in a stable (decreasing) fashion (see [GE1]).

As is the case with square-root versions of the Kalman filter ([BI2]) it may be that there are algorithms for solving the Markov-chain filtering problem which are algebraically equivalent to the scheme discussed in chapter 4 but which have better numerical properties. This requires further investigation.

5.2 Computation.

I mentioned in chapter 4 that the filtering problem is rather a special case of solving general stochastic differential equations in that we are interested in minimising the amount of "on-line" computation needed to achieve a given accuracy. Minimising calculations that can be performed "off-line", i.e. at the time the filter is designed, is not as important. For example it is perfectly acceptable when solving equation (1.2.6) to use, say, the term $\exp(A\,h)$ in expressions, as it can be calculated "off-line".

Also, when designing specific filters we can take into account the special nature of equation (1.2.6) to minimise the "on-line" calculations. In particular we can exploit the fact that the B matrices are diagonal and the A-matrix is sparse (possibly tri-diagonal). If A is tri-diagonal (1.2.6) becomes a large system of slightly-coupled equations.

So far we have been considering approximation schemes that first solve the unnormalised bilinear equation (1.2.6) and apply the normalisation at the final time T: with this technique we do not know in advance whether q will become large or small with time;so it would probably be necessary to use some form of floating point arithmetic in our calculations. This can be avoided by normalising at each step to find an approximate value of p, the solution of (1.2.4), which we know must have components that sum to unity. This approximate value can then be used as the "starting-value" for the next step. For instance, instead of using the paradigm to solve (1.2.6) and then normalising at the final time we could use the following scheme, which is the paradigm with normalisation at each step:

$$\hat{p}_0 := p_0,$$

$$\begin{aligned}\hat{\pi}_{n+1} := \Big(I + Ah + \sum_{i=1}^{d} B_i \Delta W_{n+1}^{(i)} + \frac{1}{2}\sum_{i=1}^{d} B_i^2 (\Delta W_{n+1}^{(i)\,2} - h) \\ + \frac{1}{2}\sum_{\substack{i,j=1 \\ i\neq j}}^{d} B_i B_j \Delta W_{n+1}^{(i)} \Delta W_{n+1}^{(j)} + \frac{1}{2}A^2 h^2 + \frac{1}{2}\sum_{i=1}^{d} (AB_i + B_i A) h \Delta W_{n+1}^{(i)} \\ + \frac{1}{6}\sum_{i=1}^{d} B_i^3 (\Delta W_{n+1}^{(i)\,3} - 3h\Delta W_{n+1}^{(i)}) + \frac{1}{6}\sum_{\substack{i,j=1 \\ i\neq j}}^{d} B_i^2 B_j (\Delta W_{n+1}^{(i)\,2} - h)\Delta W_{n+1}^{(j)} \\ + \frac{1}{6}\sum_{\substack{i,j,k=1 \\ i\neq j, j\neq k, k\neq i}}^{d} B_i B_j B_k \Delta W_{n+1}^{(i)} \Delta W_{n+1}^{(j)} \Delta W_{n+1}^{(k)}\Big)\hat{p}_n,\end{aligned}$$

$$\hat{p}_{n+1} := \Big(\sum_{i=1}^{m} \hat{\pi}_{n+1}^{(i)}\Big)^{-1} \hat{\pi}_{n+1}, \tag{5.2.1}$$

$$\widehat{f(X_T)} := \sum_{i=1}^{m} \hat{p}_N^{(i)} f(a_i). \tag{5.2.2}$$

This scheme is analytically identical to the paradigm with normalisation at the final time, i.e.

$$\hat{p}_N = \Big(\sum_{i=1}^{m} q_N^{*(i)}\Big)^{-1} q_N^*; \tag{5.2.3}$$

it is a scheme for solving the nonlinear s.d.e. (1.2.4),and because of the results on limit-distributions in chapters 3 and 4 and (5.2.3) it has errors that converge in distribution in the following manner:

$$P_1(N_t(p_T - \hat{p}_{N_t}) \in \cdot \mid \rho_{N_t}) \Rightarrow N(0, D_T'')(\cdot) \quad \text{(a.s.)}, \tag{5.2.4}$$

where $N(0, D_T'')(\cdot)$ is the multivariate normal distribution with zero mean,and covariance matrix

D''_T, given by

$$D''_T = \frac{T^2}{12}\Big(\sum_{i=1}^{m} q_T^{(i)}\Big)^{-2}(I - [p_T\, p_T \cdots p_T]) \sum_{i=1}^{d}\int_0^T \Phi(s,T)(AB_i - B_iA)\Phi(0,s)$$
$$p_0 p_0^T \Phi(0,s)^T (AB_i - B_iA)^T \Phi(s,T)^T ds (I - [p_T\, p_T \cdots p_T])^T. \tag{5.2.5}$$

The p_{N_t}-conditional moments also converge almost surely.

The computation needed in the scheme given in (5.2.1) can be reduced if instead of evaluating $(\sum_{i=1}^m \hat{\pi}_{n+1}^{(i)})^{-1}$ we approximate this term by $2-\sum_{i=1}^m \hat{\pi}_{n+1}^{(i)}$. The idea here is that for large N, $\sum_{i=1}^m \hat{\pi}_{n+1}^{(i)}$ will be close to 1 and so the second and higher-order terms in the following expansion are small:

$$\Big(1+\Big(\sum_{i=1}^{m}\hat{\pi}_{n+1}^{(i)}-1\Big)\Big)^{-1} = 1-\Big(\sum_{i=1}^{m}\hat{\pi}_{n+1}^{(i)}-1\Big)+\Big(\sum_{i=1}^{m}\hat{\pi}_{n+1}^{(i)}-1\Big)^2-\ldots. \tag{5.2.6}$$

5.3 Irregular partitions.

The methods we looked at in chapter 4 are all based on regular samples of the observations process; i.e. filters that use these methods sample their inputs and make calculations for a new output at regular time-intervals. There are of course other techniques that seem worth investigating: for instance, methods in which calculations are made when the change in the observations-process since the last sample-point first reaches a certain size. For equation (1.2.6) the sample-points $t_0, t_1, t_2, \ldots$ might be given by the following:

$$\begin{aligned} t_0 &:= 0, \\ t_{k+1} &:= \inf\{t > t_k : \|y_t - y_{t_k}\| \geq \epsilon\} \qquad k = 0,1,\ldots. \end{aligned} \tag{5.3.1}$$

Convergence in this case would of course be in terms of ϵ. This technique would presumably increase the accuracy in the estimate of the stochastic integral in (1.2.6) relative to the deterministic integral as more frequent calculations would be made during large changes in the observations-process and vice-versa.

Similarly one might use a compromise with sample-times based on the change in some norm of the observations-process and time:

$$\begin{aligned} t_0 &:= 0, \\ t_{k+1} &:= \inf\{t > t_k : \|(y_t - y_{t_k},\, t - t_k)\| \geq \epsilon\} \qquad k = 0,1,\ldots. \end{aligned} \tag{5.3.2}$$

One practical disadvantage of these techniques is that the amount of calculation at each step would be restricted to that which could be performed within the minimum time between two sample points; the latter would therefore have to be fixed at some positive value.

These then are a few of the areas in which further research would seem worthwhile. The "finite-partition" and computation questions are particularly important as they concern aspects of great significance to the design of practical filters.

References

[AR1] L. Arnold, *Stochastic Differential Equations: Theory and Applications*, Wiley, New York, 1974.

[BE1] V. E. Beneš, Exact finite-dimensional filters for certain diffusions with non-linear drift, *Stochastics*, **5**, 1981, pp. 65–92.

[BE2] J. Bennaton, *Ph.D. Thesis*, Department of Computing and Control, Imperial College, University of London, London, January 1980.

[BE3] J. Bennaton, Discrete-time Galerkin approximations to the non-linear filtering solution, submitted to: *J. Math. Anal. Appl.*

[BI1] P. Billingsley, *Probability and Measure*, Wiley, New York, 1979.

[BI2] G. J. Bierman, *Factorisation Methods for Discrete Sequential Estimation*, Academic Press, New York, 1977.

[BL1] E. K. Blum, *Numerical Analysis and Computation: Theory and Practice*, Addison-Wesley, 1972.

[BR1] R. W. Brockett, Non-linear systems and non-linear estimation theory, in [HW1].

[CC1] J. M. C. Clark and R. J. Cameron, The maximum rate of convergence of discrete approximations for stochastic differential equations, in: *Stochastic Differential Systems*, Ed. B. Grigelionis, (Proceedings of the I.F.I.P. working conference, Vilnius, Lithuania, U.S.S.R., 1978), Springer-Verlag, Berlin, 1980.

[CL1] J. M. C. Clark, The design of robust approximations to the stochastic differential equations of nonlinear filtering, in: *Communication Systems and Random Process Theory*, Ed. J. K. Skwirzynski, N.A.T.O. Advanced Study Institute Series, Sijthoff and Noorshoff, Alphen aan den Rijn, 1978.

[CL2] J. M. C. Clark, An efficient approximation scheme for a class of stochastic differential equations, to appear in: W. H. Fleming (Ed.), *Recent Advances in Filtering and Optimisation*, (I.F.I.P. working conference, Cocoyoc, Feb. 1982), Springer-Verlag.

[DA1] M. H. A. Davis, Pathwise non-linear filtering, in [HW1].

[DM1] M. H. A. Davis and S. I. Marcus, An introduction to non-linear filtering, in [HW1].

[EL1] K. D. Elworthy, *Stochastic Differential Equations on Manifolds*, London Mathematical Society, Lecture Note Series 7D, Cambridge University Press, 1982.

[FKK1] M. Fujisaki, G. Kallianpur and H. Kunita, Stochastic differential equations for the non-linear filtering problem, *Osaka J. Math.*, **9**, 1972, pp. 19–40.

[GE1] C. W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall, New Jersey, 1971.

[GS1] I. I. Gikhman and A. V. Skorokhod, *Stochastic Differential Equations*, Springer-Verlag, New York, 1972.

[HW1] M. Hazewinkel and J. C. Willems, *Stochastic Systems*, N.A.T.O. Advanced Study Institute Series, D. Reidel, Dordrecht, 1981.

[IW1] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, North Holland Mathematical Library 24, North Holland, 1981.

[KA1] G. Kallianpur, *Stochastic Filtering Theory*, Springer-Verlag, New York, 1980.

[KT1] J. F. C. Kingman and S. J. Taylor, *Introduction to Measure and Probability*, Cambridge University Press, 1966.

[KU1] H. J. Kushner, A robust discrete-state approximation to the optimal non-linear filter for a diffusion, *Stochastics*, 3, 1979, pp. 75–83.

[KU2] H. Kunita, On the decomposition of solutions of stochastic differential equations, in: D. Williams (Ed.), *Stochastic Integrals*, (Proc. LMS Durham Symposium, 1980), Lecture Notes in Mathematics 851, Springer-Verlag, Berlin, 1981.

[LA1] H. J. Larson, *Introduction to Probability Theory and Statistical Inference*, Wiley, 1969.

[LS1] R. S. Liptser and A. N. Shiryayev, *Statistics of Random Processes I: General Theory*, Springer-Verlag, New York, 1977.

[MA1] G. Maruyama, Continuous Markov processes and stochastic equations, *Rend. Circ. Mat. Palermo*, **4**, 1955, pp. 48–90.

[MA2] P. Malliavin, *Géométrie Différentielle Stochastique*, Les Presses de l'Université de Montréal, Montréal, 1978.

[MI1] G. N. Mil'shtein, Approximate integration of stochastic differential equations, *Theory of Prob. and its Appl.*, **19**, 1974, pp. 557–562.

[MS1] E. J. McShane, *Stochastic Calculus and Stochastic Models*, Academic Press, New York, 1974.

[PL1] E. Platen, An approximation method for a class of Itô processes, *Lietuvos Matematikos Rinkinys*, **21**, no.1, 1981.

[RBR1] N. J. Rao, J. D. Borwankar and D. Ramkrishna, Numerical solution of Itô differential equations, *SIAM J. Contr.*, **12**, 1974, pp. 124–139.

[RO1] H. Rootzén, Limit-distributions for the error in approximations of stochastic integrals, *Ann. Prob.*, **8**, 1980, pp. 241–251.

[RU1] W. Rümelin, Numerical treatment of stochastic differential equations, to appear in: *SIAM J. Num. Anal.*, 1982.

[SV1] D. W. Stroock and S. R. S. Varadhan, *Multidimensional Diffusion Processes*, Springer-Verlag, New York, 1979.

[TA1] D. Talay, Résolution trajectorielle et analyse numérique des equations différentielles stochastiques, *Stochastics*, **9**, no.4, 1983, pp. 275–306.

[WO1] W. M. Wonham, Some applications of stochastic differential equations to optimal non-linear filtering, *SIAM J. Contr.*, **2**, 1965, pp. 347–369.

[WP1] W. Wagner and E. Platen, Approximation of Itô differential equations, *Internal Publication of the Zentralinstitut für Mathematik und Mechanik der Akademie der Wissenschaffen der D.D.R.*, Berlin, 1978.

[WZ1] E. Wong and M. Zakai, On the convergence of ordinary integrals to stochastic integrals, *Ann. Math. Stat.*, **36**, 1965, pp. 1560–1564.

[ZA1] M. Zakai, On the optimal filtering of diffusion processes, *Z. Wahr. Verw. Geb.*, **11**, 1969, pp. 230–243.