

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

**A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600**



Order Number 9518422

Nonlinear filtering of stochastic dynamical systems

O'Donnell, Brian Edmund, Ph.D.

Iowa State University, 1994

U·M·I

300 N. Zeeb Rd.
Ann Arbor, MI 48106



Nonlinear filtering of stochastic dynamical systems

by

Brian Edmund O'Donnell

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Department: Mathematics
Major: Applied Mathematics

Approved:

Signature was redacted for privacy.

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1994

TABLE OF CONTENTS

	Page
CHAPTER I - INTRODUCTION	1
CHAPTER II - REVIEW	10
1. Semimartingales and Related Processes	10
1.1 General stochastic processes	11
1.2 Martingales	14
1.3 Integrable variation, optional, predictable processes	17
1.4 Square integrable martingales and quadratic variation processes	22
2. Martingale Calculus and Related Results	25
2.1 Integration w.r.t. a martingale	25
2.2 Integral representations and change of measure	28
3. Markov Processes and Convergence	31
3.1 Definitions and examples	32
3.2 Semigroup theory	35
3.3 Weak convergence	37
CHAPTER III - EQUATIONS OF NONLINEAR FILTERING	41
1. Innovations Approach	41
1.1 Innovations process	43
1.2 Gain computation	45
2. Reference Probability Approach	46
2.1 Kallianpur-Striebel formula	47
2.2 Zakai equation	49

3.	Conditional Densities for Markov Signals	51
3.1	Finite state signal	52
3.2	Computational considerations	54
3.3	Diffusion signal	56
CHAPTER IV - APPROXIMATE FILTERS FOR DIFFUSIONS		58
1.	Finite State Diffusion Approximations	59
1.1	Markov chain approximations	59
1.2	Continuous time, discrete state Markov approximations	60
1.3	Example	68
1.4	Truncation of processes	72
2.	Filter Approximations	73
2.1	Examples	74
2.2	Convergence	76
3.	Computational Considerations	80
3.1	Computational algorithms	81
3.2	Algorithm comparison	84
REFERENCES		90
APPENDIX		92

CHAPTER I - INTRODUCTION

Some of the earliest work in stochastic filtering began in the 1940's when N. Wiener studied the following system. Over a given time interval $[0, T]$ a signal process $\{s_t\}$ is transmitted. But, at any time $t \in [0, T]$ only $z_t = s_t + n_t$ can be observed where the process $\{n_t\}$ is some noise process. ($\{n_t\}$ is often modeled as white noise, i.e., a process where $E\{x_{t+\tau}x_t\} = \delta(\tau)$ for all $t \geq 0$.) He wanted to compute \bar{s}_t , the optimal mean squared estimate of s_t given the observations $\{z_s: s \leq t\}$. Assuming $\{s_t\}$ is a stationary process uncorrelated with $\{n_t\}$, Wiener solved this filtering problem using spectral techniques (see [19]).

Using standard linear systems techniques (see [19] and [20]) the above system can be realized in state space form as:

$$(1) \quad \begin{aligned} dx_t &= Ax_t dt + b dv_t \\ dy_t &= cx_t dt + dw_t \end{aligned}$$

for an appropriate m -dimensional state process $\{x_t\}$, and appropriate $m \times m$, $m \times 1$, $1 \times m$ matrices A, b, c . In this realization v_t and w_t are standard Brownian motion processes, $s_t = cx_t$ and symbolically $z_t = y'_t$. In the early 1960's R.E. Kalman developed an algorithm for computing the estimation process $\{\bar{s}_t\}$ using this state space model (again, see [19]).

It was soon realized that the Kalman filter had two major

advantages over the Wiener filter.

- i) The Kalman filter can handle more diverse systems than the Wiener filter. The Kalman filter can be applied to general time-variant linear systems not just stationary time-invariant linear systems.
- ii) The Kalman filter is a recursive algorithm ideally suited to the use of digital computers.

The following simple discrete time example (adapted from [10]) illustrates the basic ideas behind Kalman's algorithm without using much probabilistic machinery. Consider the system:

$$\begin{aligned}x_k &= x \\ \Delta y_k &= x_k + \Delta w_k\end{aligned}$$

for $k=1,2,\dots$ where $x, \Delta w_1, \Delta w_2, \dots$ are iid, $N(0,1)$ random variables defined on a probability space (Ω, \mathcal{F}, P) . For this system the optimal mean squared filter $\bar{x}_k = E\{x_k | \Delta y_1, \dots, \Delta y_k\}$ is the least squares estimate of x in the linear subspace $\mathcal{L}(\Delta y_1, \dots, \Delta y_k)$ of the Hilbert space H of all mean zero, normal random variables on (Ω, \mathcal{F}, P) with inner product $\langle u, v \rangle = E\{uv\}$. Denote the projection of $u \in H$ onto $\mathcal{L}(\Delta y_1, \dots, \Delta y_k)$ by $P_k(u)$. Then, to compute \bar{x}_k we proceed in three steps. First, we construct the "innovations process" $\{\Delta v_k\}$. Define $\Delta v_1 = \Delta y_1$ and $\Delta v_k = \Delta y_k - P_{k-1}(\Delta y_k) = \Delta y_k - P_{k-1}(x) = \Delta y_k - \bar{x}_{k-1}$. Of course,

$\{\Delta v_1, \dots, \Delta v_k\}$ is an orthogonal basis for $\mathcal{L}(\Delta y_1, \dots, \Delta y_k)$.

Second, we represent \bar{x}_k as a linear combination of

$\Delta v_1, \dots, \Delta v_k$. There exist constants a_1, \dots, a_k such that

$\bar{x}_k = \sum_{i=1}^k a_i \Delta v_i$. Third, we compute the "gain" constants a_1, \dots, a_k .

For this, note that

$$\begin{aligned} E\{x \Delta v_1\} &= E\{x^2 + x \Delta w_1 - x \bar{x}_{1-1}\} = E\{x(x - \bar{x}_{1-1})\}, \\ E\{\bar{x}_1 \Delta v_1\} &= E\left\{\sum_{j=1}^1 a_j \Delta v_j \Delta v_1\right\} = a_1 E\{\Delta v_1^2\}. \end{aligned}$$

But,

$$\begin{aligned} E\{x \Delta v_1\} &= E\{\bar{x}_1 \Delta v_1\}, \\ E\{x(x - \bar{x}_{1-1})\} &= E\{(x - \bar{x}_{1-1})^2\}, \\ E\{\Delta v_1^2\} &= E\{(x + \Delta w_1 - \bar{x}_{1-1})^2\} = E\{(x - \bar{x}_{1-1})^2\} + 1. \end{aligned}$$

So, denoting $e_{i-1}^2 = E\{(x - \bar{x}_{i-1})^2\}$, we have $a_1 = \frac{e_{1-1}^2}{e_{1-1}^2 + 1}$.

We can then compute the estimation process recursively as follows. Let $\bar{x}_0 = 0$ and $\Delta \bar{x}_k = \bar{x}_k - \bar{x}_{k-1}$. We get

$$\Delta \bar{x}_k = a_k \Delta v_k = \frac{e_{k-1}^2}{e_{k-1}^2 + 1} (\Delta y_k - \bar{x}_{k-1}).$$

Also, the error process $\{e_k^2\}$ can be computed recursively with $e_0^2 = 1$ and

$$\begin{aligned} e_k^2 &= E\{(x - \bar{x}_k)^2\} \\ &= E\{x(x - \bar{x}_k)\} \\ &= E\{x(x - \bar{x}_{k-1}) - a_k x \Delta v_k\} \end{aligned}$$

$$= E\{(x - \bar{x}_{k-1})^2\} - a_k E\{(x - x_{k-1})^2\} = \frac{e_{k-1}^2}{e_{k-1}^2 + 1}.$$

During the last 30 years much research has been devoted to developing algorithms as computationally practical as Kalman's for the nonlinear filtering problem. Initial efforts were primarily concerned with extending Kalman's ideas to the nonlinear generalization of system (1). Specifically, suppose we are given the partially observed system

$$(2) \quad \begin{aligned} dx_t &= f(x_t)dt + \sigma(x_t)dv_t \\ dy_t &= h(x_t)dt + dw_t \end{aligned}$$

where $f: \mathbb{R}^m \rightarrow \mathbb{R}^m$, $\sigma: \mathbb{R}^m \rightarrow \mathbb{R}^m \times \mathbb{R}^m$, $h: \mathbb{R}^m \rightarrow \mathbb{R}^r$, and v_t, w_t are m, r -dimensional standard Brownian motions w.r.t. a filtration $\{\mathcal{F}_t\}$ on (Ω, \mathcal{F}, P) . We want to find the optimal mean squared estimate of $s_t = \phi(x_t)$ given the observation σ -algebra $Y_t = \sigma\{y_s: s \leq t\}$, i.e., for a given bounded continuous function $\phi: \mathbb{R}^m \rightarrow \mathbb{R}$, find $\bar{\phi}_t = E\{\phi(x_t) | Y_t\}$. More recent efforts concern solving the above nonlinear filtering problem when the state process $\{x_t\}$ is something other than a diffusion process or when the noise process $\{w_t\}$ is something other than standard Brownian motion. The purpose of this paper is to unify many of these nonlinear filtering results.

One can see from some of the terminology already used that much probabilistic machinery will be needed. So, in Chapter II we begin with a review of background material on

stochastic processes. First, we recall the terminology and properties associated with the various classes of processes that arise in filtering theory. Noting that systems (1) and (2) are described in terms of stochastic differential equations, it seems reasonable that next we should review the martingale calculus which is very important in filtering theory. Last, we review the theory of Markov processes, since most often the partially observed system will be Markov.

In Chapter III we derive the main equations of nonlinear filtering. One of the contributions of this paper is the unification of several approaches which leads to the derivation of the filtering equations for quite general systems. The only restriction on the signal process will be that it is a square integrable semimartingale. Also, the noise process in the observation can be either standard Brownian motion or a martingale derived from a conditional Poisson process.

In this chapter we first use the innovations procedure to derive the Kushner-Stratonovich equation, which is a stochastic differential equation for computing the conditional expectation $E\{\phi(x_t)|Y_t\}$. This procedure is analogous to the procedure used above in the simple discrete time example:

- i) We construct an innovations process with orthogonal increments;

- ii) We represent the estimation process in terms of a stochastic integral w.r.t. the innovations process;
- iii) We compute the gain or integrand process of the stochastic integral.

The Kushner-Stratonovich equation can be cumbersome, so next we use the reference probability approach to develop a somewhat simpler equation. The keys to this approach are:

- i) There exists a probability measure Q with $P \ll Q$ on \mathcal{F} , such that under Q the processes x_t, y_t , and $R_t = E^Q \left\{ \frac{dP}{dQ} \middle| \mathcal{F}_t \right\}$ have "nice" properties and under Q the signal and observation processes are independent;
- ii) The filtering problem can be solved in terms of a conditional expectation w.r.t. Q .

We can illustrate points i) and ii) above using our simple discrete time example.

- i) Suppose under P the joint pdf of $x, \Delta y_1, \dots, \Delta y_n$ is given as $f(x, \Delta y_1, \dots, \Delta y_n) = g(x) \exp \left(-\frac{1}{2} \sum_{k=1}^n (\Delta y_k - x)^2 \right)$ where g is the $N(0,1)$ density function. Let Q be the probability measure with $\frac{dQ}{dP} = R_n^{-1}$ and

$$R_n = \exp \left(\sum_{k=1}^n x \Delta y_k - \frac{1}{2} \sum_{k=1}^n x^2 \right) \text{ for all } k=1, \dots, n.$$

Then,

- ▶ under Q $\{\Delta y_k\}$ is a sequence of iid $N(0,1)$ random variables independent of x ;
- ▶ the marginal distribution of x is the same under Q as under P ;
- ▶ if $\mathcal{F}_k = \sigma\{x, \Delta y_1, \dots, \Delta y_n\}$ then

$$R_k = E^Q \left\{ \frac{dP}{dQ} \middle| \mathcal{F}_k \right\} = \exp \left(\sum_{i=1}^k x \Delta y_i - \frac{1}{2} \sum_{i=1}^k x^2 \right) \text{ for all } k=1, \dots, n.$$

ii) It is easy to verify that for $\phi \in C_b(\mathbb{R})$

$$E^P\{\phi(x) | \Delta y_1, \dots, \Delta y_k\} = \frac{E^Q\{\phi(x) R_k | \Delta y_1, \dots, \Delta y_k\}}{E^Q\{R_k | \Delta y_1, \dots, \Delta y_k\}}$$

for all $k=1, \dots, n$.

The Zakai stochastic differential equation computes the continuous time analog of $E^Q\{\phi(x) R_k | \Delta y_1, \dots, \Delta y_k\}$. This equation is usually easier to work with than the Kushner-Stratonovich equation.

When the state process is Markov, both the Kushner-Stratonovich equation and the Zakai equation give recursive equations for computing the filtered process. Unfortunately,

these equations are usually infinite dimensional. In the last section of Chapter III we consider important exceptions when the Zakai equation gives computationally useful finite dimensional recursive methods for deriving the conditional densities of the filtered process.

When the state process is a diffusion, equations for computing the conditional densities can be developed also, but these equations are infinite dimensional. So, over the last 15 years various numerical methods for approximating the conditional densities have been developed. Although each of these approximation methods were developed by different individuals and use different techniques, there is a common thread running through several of these methods. The main contribution of this paper is the development of a general theory for approximating the optimal filter of a partially observed system when the signal is a diffusion process. This general theory includes several previously developed approximation methods as special cases.

In Chapter IV we present our approximation theory. The theory consists of three main steps.

- i) We approximate the diffusion with a finite state process. We show how to construct finite state processes that will converge to the diffusion signal in distribution.

- ii) We construct filters (conditional expectations) for the finite state approximating processes that converge to the desired filter (conditional expectation) for the diffusion. We show that the approximate filters converge to the desired filter in $L^1(P)$ uniformly on $[0, T]$.
- iii) We develop numerical algorithms for computing the conditional densities of the approximate filters.

The numerical algorithms, of course, depend on the diffusion approximation used and on the way the approximate filter is constructed. But, little research has been conducted studying the effect diffusion and filter approximation choices have on the numerical algorithms. So, we conclude by comparing the efficiency of several possible algorithms using computer generated simulations.

CHAPTER II - REVIEW

In this chapter we review the definitions and properties of the stochastic processes involved in nonlinear filtering. We divide the review into three sections. First, we classify the processes related to semimartingales, since the most general filtering equations are developed for semimartingales. Next, we review the martingale calculus and related results that will be needed to develop the filtering equations. Then, in the last section we examine Markov processes which are the processes of interest in the fourth chapter of this paper. The theorems, propositions, etc., in this review are stated without proofs. These results are all well known and at the beginning of each section references will be listed where these results can be found.

1. Semimartingales and Related Processes

We begin with the basic definitions related to general stochastic processes. Then, we define the concept of a martingale and look at properties and examples of martingales. The semimartingales involved in filtering consist of a martingale plus a predictable process of integrable variation. So, the third subsection examines predictable, optional, and

integrable variation processes. Finally, we define square integrable martingales and quadratic variation processes. These processes are fundamental in the martingale calculus.

The development of the material in this section follows the expositions in references [1] and [2] where proofs of the stated results can be found. We work on a given probability space (Ω, \mathcal{F}, P) . Time varies on an interval I where $I = [0, \infty)$, $[0, \infty]$, or $[0, T]$ for some finite time T . Also, in subsections 1.3 and 1.4 we assume a given filtration $\{\mathcal{F}_t\}$ that is right continuous and complete.

1.1 General stochastic processes

Let E be a complete separable metric space with Borel σ -field $\mathcal{B}(E)$. A (*stochastic*) *process* is a map $x: I \times \Omega \rightarrow E$ such that $x_t(\cdot): \Omega \rightarrow E$ is a random variable for each $t \in I$. A process x is *measurable* if the map x is measurable when $I \times \Omega$ is given the product σ -field $\mathcal{B}(I) \times \mathcal{F}$ where $\mathcal{B}(I)$ is the Borel σ -field on I . A process x is (*right or left*) *continuous* if the paths $t \rightarrow x_t(\omega)$ are (right or left) continuous. In stochastic analysis generally we are able to establish properties of a process (e.g., continuity) only almost surely (a.s.). So we define the following criterion for identifying processes. Process y is a *modification* of x if for each $t \in I$, $x_t = y_t$ a.s. A stronger criterion for identification is: x and y are

indistinguishable if $x_t = y_t$ for all t a.s. A process x that is indistinguishable from the zero process is said to be evanescent. Since we will be primarily concerned with right continuous processes, the following result should be noted.

Proposition 1.1.1 Suppose x and y are right continuous processes and y is a modification of x . Then x and y are indistinguishable.

We also need measurability conditions on a process x that varies with time. A family of σ -fields $\{\mathcal{F}_t\}$ indexed by I is a filtration of (Ω, \mathcal{F}) if each $\mathcal{F}_t \subset \mathcal{F}$ and $\mathcal{F}_s \subset \mathcal{F}_t$ when $s \leq t$. Usually we will assume a filtration $\{\mathcal{F}_t\}$ satisfies the following two conditions. (i) A filtration $\{\mathcal{F}_t\}$ is right continuous if $\mathcal{F}_{t+} = \mathcal{F}_t$ for all $t \in I$ ($\mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s$ and $\mathcal{F}_{t-} = \bigvee_{s<t} \mathcal{F}_s$). (ii) A filtration $\{\mathcal{F}_t\}$ is complete if \mathcal{F}_0 contains all the P -null sets of \mathcal{F} . A process x is adapted to $\{\mathcal{F}_t\}$ if for each $t \in I$, x_t is \mathcal{F}_t measurable. A process x is progressive if for each $t \in I$ the map $x: [0, t] \times \Omega \rightarrow E$ is measurable w.r.t. the product σ -field $\mathcal{B}([0, t]) \times \mathcal{F}_t$. Clearly a progressive process is adapted, but the converse is not always true. However, we do have the following:

Proposition 1.1.2 If x is adapted and right continuous, then x is progressive.

We will want to talk about the value of a process at random times τ as well as fixed times t . A random variable $\tau: \Omega \rightarrow I$ is a *stopping time* w.r.t. $\{\mathcal{F}_t\}$ if $\{\tau \leq t\} \in \mathcal{F}_t$ for each $t \in I$. Of course, fixed times are stopping times. Also, if $\{\mathcal{F}_t\}$ is complete and right continuous and x is a progressive process, then for each $\Gamma \in \mathcal{B}(E)$ $\tau_\Gamma = \inf\{t: x_t \in \Gamma\}$ is a stopping time. The σ -field $\{\mathcal{F}_\tau\}$ of events occurring by time τ is defined:

$$\mathcal{F}_\tau = \{A \in \mathcal{F}: A \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ for each } t \in I\}.$$

Some basic properties of stopping times follow. Assume τ and σ are $\{\mathcal{F}_t\}$ stopping times, then

- i) $\sigma \vee \tau, \sigma \wedge \tau$ are stopping times
- ii) If $A \in \mathcal{F}_0$, then $A \cap \{\sigma \leq \tau\} \in \mathcal{F}_\tau$
- iii) If $\sigma \leq \tau$ a.s. then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$
- iv) $\{\sigma = \tau\}, \{\sigma < \tau\}, \{\sigma > \tau\}$ belong to both \mathcal{F}_σ and \mathcal{F}_τ .

Finally, we have the following:

Lemma 1.1.3 If τ is a stopping time and x is progressive, then $x_\tau \equiv x_{\tau(\omega)}(\omega)$ is \mathcal{F}_τ measurable and the stopped process $x_t^\tau \equiv x_{t \wedge \tau}$ is progressive.

1.2 Martingales

A real-valued stochastic process x is a *supermartingale* (resp. *submartingale*) w.r.t. the filtration $\{\mathcal{F}_t\}$ if:

- i) x is adapted to $\{\mathcal{F}_t\}$
- ii) $E\{|x_t|\} < \infty$ for all $t \in I$
- iii) $E\{x_t | \mathcal{F}_s\} \leq x_s$ a.s. for $s \leq t$ (resp. $E\{x_t | \mathcal{F}_s\} \geq x_s$ a.s. for $s \leq t$).

A process x that is both a submartingale and a supermartingale is called a *martingale*.

Lemma 1.2.1 If x is an $\{\mathcal{F}_t\}$ -martingale (resp. submartingale) and ϕ is a convex (resp. convex increasing) function from \mathbb{R} into \mathbb{R} with $\phi(x_t)$ integrable for each $t \in I$, then $\{\phi(x_t)\}$ is an $\{\mathcal{F}_t\}$ -submartingale.

There are two canonical examples of continuous time martingales that play an important role in this paper.

The primary example of a continuous martingale is the *Wiener or Brownian motion process*:

Definition 1.2.2 A stochastic process w , adapted to $\{\mathcal{F}_t\}$, is a Wiener process w.r.t. $\{\mathcal{F}_t\}$ if:

- i) $w_0 = 0$

- ii) $w_t - w_s$ is independent of \mathcal{F}_s for all $t > s$
- iii) $w_t - w_s \sim N(0, \sigma^2(t-s))$ for all $t > s$.

A Wiener process w with $\sigma^2=1$ is called *standard Brownian motion (SBM)*.

The primary example of a martingale with jumps is related to the *Poisson process*:

Definition 1.2.3 A stochastic process N , adapted to $\{\mathcal{F}_t\}$, is an $\{\mathcal{F}_t\}$ -Poisson process with intensity $\lambda(t)$, where λ is a positive measurable function on I , if:

- i) $N_0 = 0$
- ii) $N_t - N_s$ is independent of \mathcal{F}_s for all $t > s$
- iii) $N_t - N_s \sim \text{POI}(\int_s^t \lambda(u) du)$ for all $t > s$.

A Poisson process N with $\lambda(t) = 1$ is called *standard Poisson*.

Proposition 1.2.4 If N is an $\{\mathcal{F}_t\}$ -Poisson process with integrable intensity $\lambda(t)$, then $N_t - \int_0^t \lambda(s) ds$ is an $\{\mathcal{F}_t\}$ martingale.

Also, we will be concerned with processes that are conditionally Poisson. Suppose $\lambda_t = \lambda(x_t)$ is a positive measurable process depending on the measurable process x with

$E\{\int_0^t \lambda_u du\} < \infty$ a.s. Let $X = \sigma(x_t : t \in I)$ be contained in \mathcal{F}_0 . Then N is a conditional Poisson process (CPP) with intensity λ_t if, given \mathcal{F}_0 , N_t is an $\{\mathcal{F}_t\}$ -Poisson process with intensity λ_t . Again, $N_t - \int_0^t \lambda_s ds$ is an $\{\mathcal{F}_t\}$ -martingale.

Throughout most of the paper we will assume the processes of interest are *cadlag* (right continuous processes with limits on the left). The following two results show that this assumption is not very restrictive.

Proposition 1.2.5 Suppose the filtration $\{\mathcal{F}_t\}$ is right continuous and that x is a supermartingale. A right continuous modification of x exists iff $t-E\{x_t\}$ is right continuous in t .

Proposition 1.2.6 If x is a right continuous supermartingale, then x has a *cadlag* modification.

Also, for much of the paper we will consider only a finite time interval $I = [0, T]$. In this case, the following theorems will apply to the stopped process x^T .

Theorem 1.2.7 Suppose x is a right continuous supermartingale on $I = [0, \infty)$. If $\sup_{t \in I} E\{x_t^-\} < \infty$, then x_t converges to an integrable random variable x_∞ a.s. as $t \rightarrow \infty$.

Theorem 1.2.8

- i) If x is a uniformly integrable martingale on $I = [0, \infty)$, then there exists an integrable random variable x_∞ such that $x_t = E\{x_\infty | \mathcal{F}_t\}$ a.s. and $x_t \rightarrow x_\infty$ a.s. and in L^1 .
- ii) If $x_t = E\{x_\infty | \mathcal{F}_t\}$ for some integrable, $\mathcal{F}_\infty = \bigvee_t \mathcal{F}_t$ measurable random variable, then x is uniformly integrable.

Finally, we state two important properties of martingales.

Theorem 1.2.9 (Doob's Inequality) If x is a right continuous submartingale and $J=[u,v] \subset I$, then $x_v \in L^p$ implies

$$\left\| \sup_{t \in J} |x_t| \right\|_p \leq \frac{p}{p-1} \|x_v\|_p \text{ for } 1 < p < \infty.$$

Theorem 1.2.10 (Optional Sampling) Suppose x is a right continuous $\{\mathcal{F}_t\}$ -supermartingale and σ, τ are $\{\mathcal{F}_t\}$ stopping times with $\sigma \leq \tau$ a.s. for $t \in I = [0, \infty]$. Then x_σ, x_τ are integrable and $x_\sigma \geq E\{x_\tau | \mathcal{F}_\sigma\}$ a.s.

1.3 Integrable variation, optional, predictable processes

A process $\{A_t\}$ is *raw increasing* if the paths of A are

almost surely nonnegative, right continuous, increasing functions of t and $A_t \in L^1$ for $t \in I$. If A_∞ (which necessarily exists) is also in L^1 , we call $\{A_t\}$ *integrable raw increasing*. Furthermore, if $A_0 = 0$ a.s. and A is adapted to $\{\mathcal{F}_t\}$ we say A is *increasing* (resp. *integrable increasing*). A process $\{B_t\}$ which is the difference of two raw increasing processes is a *bounded variation process*, and if $\{B_t\}$ is the difference of two integrable increasing processes, $\{B_t\}$ is a process of *integrable variation*. Again, since $I = [0, T]$ for most of this paper, the main distinction between bounded and integrable variation processes is whether the process is adapted or not.

In filtering the most general processes of interest are *semimartingales*. A semimartingale is a process that can be decomposed as the sum of a uniformly integrable martingale and a process of integrable variation. (Actually, we are defining a *special semimartingale* here but usually $I = [0, T]$ so, we do not need the more general semimartingale definition.) Clearly, a semimartingale is the difference of two submartingales. In the following we will see that a submartingale is also a particular type of semimartingale given by its Doob-Meyer decomposition.

As motivation we first consider how to decompose a discrete time submartingale $\{x_k\}$. Let $A_0 = 0$ and $\Delta A_k = A_{k+1} - A_k = E\{x_{k+1} | \mathcal{F}_k\} - x_k$. Also, let $m_0 = x_0$ and $\Delta m_k = m_{k+1} - m_k = x_{k+1} - E\{x_{k+1} | \mathcal{F}_k\}$.

Note that $x_k = m_k + A_k$. Since x is a submartingale, $\{A_k\}$ is increasing and predictable (A_{k+1} is \mathcal{F}_k -measurable). Also, $\{m_k\}$ is a martingale. So, we have decomposed x as the sum of a martingale and a predictable integrable increasing process. Such a decomposition exists for continuous time submartingales also, symbolically take $dA_t = E\{x_t | \mathcal{F}_{t-}\} - x_{t-}$ and $dm_t = x_t - E\{x_t | \mathcal{F}_{t-}\}$. To make this rigorous many details need to be worked out, particularly regarding the continuous time version of a predictable process.

The *predictable σ -field* Σ_p of the filtration $\{\mathcal{F}_t\}$ is generated by all left continuous $\{\mathcal{F}_t\}$ -adapted processes. The *optional σ -field* Σ_o of the filtration $\{\mathcal{F}_t\}$ is generated by all cadlag $\{\mathcal{F}_t\}$ -adapted processes. A Σ_p -measurable process is a *predictable process* and a Σ_o -measurable process is an *optional process*. Denote $\mathcal{B}(I) \times \mathcal{F}$ as Σ . We say a Σ -measurable process x is *bounded* (denoted $x \in \mathcal{B}^\infty(\Sigma)$) if $\sup_{t \in I} |x_t| \in L^\infty$. Bounded predictable or optional processes are similarly defined (respectively denoted $x \in \mathcal{B}^\infty(\Sigma_p)$ or $x \in \mathcal{B}^\infty(\Sigma_o)$).

Theorem 1.3.1 (Projection Theorem) There are unique linear projections $\pi_p: \mathcal{B}^\infty(\Sigma) \rightarrow \mathcal{B}^\infty(\Sigma_p)$ and $\pi_o: \mathcal{B}^\infty(\Sigma) \rightarrow \mathcal{B}^\infty(\Sigma_o)$ such that for any $x \in \mathcal{B}^\infty(\Sigma)$:

- i) $\pi_p(x)_\tau \chi(\tau < \infty) = E\{x_\tau \chi(\tau < \infty) | \mathcal{F}_\tau\}$ for all stopping times τ

- ii) $\pi_0(x)_\tau \chi(\tau < \infty) = E\{x_\tau \chi(\tau < \infty) | \mathcal{F}_{\tau-}\}$ for all predictable stopping times τ .

We call $\pi_p(x)$ the *predictable projection* of x and $\pi_0(x)$ the *optional projection* of x . Clearly, continuous adapted processes are predictable. Poisson processes are examples of non-predictable processes. Since we will be concerned mainly with cadlag processes, the optional processes will be the adapted processes.

Before we state the Doob-Meyer theorem for continuous time submartingales we need one more definition. A right continuous, uniformly integrable submartingale x is of *class D* if the family $\{x_\tau : \tau \text{ a finite stopping time}\}$ is uniformly integrable. Using the optional sampling theorem it is easy to see that uniformly integrable martingales are of class D. Also, positive submartingales x with $x_t \leq E\{x_\infty | \mathcal{F}_t\}$ on $t \in I$ for some integrable random variable x_∞ are of class D.

Theorem 1.3.2 (Doob-Meyer) Suppose x is a class D submartingale. Then there exists a unique predictable integrable increasing process A and a unique uniformly integrable martingale m such that $x_t = m_t + A_t$.

The following useful results follow from the Doob-Meyer

theorem. Considering our symbolic definition of dA_t above, the first result seems reasonable.

Proposition 1.3.3 Let the class D submartingale x have a Doob-Meyer decomposition $x=m+A$. Then $\pi_p(x)_t=x_{t-}$ iff A is continuous.

For the other result we need the following definition.

Suppose $\{x_t\}$ is an integrable raw increasing process. Then there is a unique predictable integrable increasing process π_p^*x such that

$$E\left\{\int_0^\infty \pi_p(H)_s dx_s\right\} = E\left\{\int_0^\infty H_s d(\pi_p^*x_s)\right\} \text{ for all } H \in \mathcal{B}^*(\Sigma).$$

The process $\{\pi_p^*x_t\}$ is called the *dual predictable projection* of x . Note that π_p^*x , unlike $\pi_p(x)$, is necessarily increasing.

Proposition 1.3.4 The Doob-Meyer decomposition $x=m+A$ for an integrable increasing process x has $m_t=x_t-\pi_p^*x_t$ and $A_t=\pi_p^*x_t$.

If $\{N_t\}$ is a Poisson process with intensity $\lambda(t)$ then,

$\pi_p^*N_t = \int_0^t \lambda(s) ds$ (Proposition 1.3.4) and $\pi_p(N)_t = N_{t-}$ (Proposition 1.3.3) since $\int_0^t \lambda(s) ds$ is continuous.

1.4 Square integrable martingales and quadratic variation processes

For $1 \leq p < \infty$, a process x is L^p -bounded if $\sup_{t \in I} \|x_t\|_p < \infty$. We denote the vector space of L^p -bounded cadlag martingales by \mathcal{H}^p . We say x is a *square integrable martingale* if $x \in \mathcal{H}^2$. Note that a square integrable martingale x is uniformly integrable, so there exists $x_\infty \in L^2$ such that $x_t = E\{x_\infty | \mathcal{F}_t\}$ for all $t \in I$.

Define a bilinear functional on $\mathcal{H}^2 \times \mathcal{H}^2$ by $(x, y) = E\{x_\infty y_\infty\}$. This functional is an inner product giving \mathcal{H}^2 a Hilbert space structure. The subspace $\mathcal{H}_c^2 = \{x \in \mathcal{H}^2 : x \text{ is continuous and } x_0 = 0\}$ is closed in \mathcal{H}^2 . So $\mathcal{H}^2 = \mathcal{H}_c^2 \oplus \mathcal{H}_d^2$ where $\mathcal{H}_d^2 = (\mathcal{H}_c^2)^\perp$ is called the space of *purely discontinuous martingales*.

To rigorously establish the structure of \mathcal{H}_d^2 goes beyond what we will need. Intuitively though, the elements of \mathcal{H}_d^2 are martingales of the form $A_t - \pi_p^* A_t$ where A_t is an L^2 -bounded process that is constant between jump times (jump times may be random or fixed, but if all jump times are fixed then $\pi_p^* A_t = 0$). For example, if N_t is a Poisson process with intensity $\lambda(t)$, then $N_t - \pi_p^* N_t$ with $\pi_p^* N_t = \int_0^t \lambda(s) ds$ is in \mathcal{H}_d^2 . Also, a process A_t that jumps according to a simple random walk at fixed time increments is in \mathcal{H}_d^2 with compensator $\pi_p^* A_t = 0$. We will need the following result about elements in \mathcal{H}_d^2 .

Theorem 1.4.1 Suppose $x \in \mathcal{H}_d^2$ and $y \in \mathcal{H}^2$. Then

$$i) \quad E\{x_\infty y_\infty\} = E\left\{\sum_s \Delta x_s \Delta y_s\right\}$$

$$ii) \quad z_t = x_t y_t - \sum_{s \leq t} \Delta x_s \Delta y_s \text{ is in } \mathcal{H}^1 \text{ with } z_0 = 0.$$

(For a process x , we define $\Delta x_t = x_t - x_{t-}$.)

Now if $x \in \mathcal{H}^2$, x^2 is a class D submartingale and has a Doob-Meyer decomposition $x^2 = m + A$ where m is a uniformly integrable martingale and A is a predictable integrable increasing process. Define the *predictable quadratic variation* of x as $\langle x, x \rangle_t = \langle x \rangle_t = A_t + x_0^2$. This process will be very important when we discuss integration w.r.t. martingales. The terminology is motivated by the following observation:

$$E\{(x_t - x_s)^2 | \mathcal{F}_s\} = E\{x_t^2 - x_s^2 | \mathcal{F}_s\} = E\{(A_t - A_s) | \mathcal{F}_s\} = E\{\langle x \rangle_t - \langle x \rangle_s | \mathcal{F}_s\}.$$

We note that, for standard Brownian motion $\{w_t\}$, $\langle w \rangle_t = t$ since $w_t^2 - t$ is a martingale.

Using Theorem 1.4.1 we also have for $x \in \mathcal{H}^2$

$$x_t^2 - (\langle x^c \rangle_t + \sum_{s \leq t} \Delta x_s^2) = 2x_t^c x_t^d + (x_t^c)^2 - \langle x^c \rangle_t + (x_t^d)^2 - \sum_{s \leq t} \Delta x_s^2$$

is in \mathcal{H}^1 , where $x = x^c + x^d$ is the decomposition of x into its continuous and purely discontinuous parts. So $[x, x]_t = [x]_t = \langle x^c \rangle_t + \sum_{s \leq t} \Delta x_s^2$ is another integrable increasing (not necessarily predictable) process that can be subtracted from x^2 to get a martingale. We call $[x]$ the *optional quadratic variation* of x . For standard Brownian motion $\{w_t\}$, of course, $[w]_t = \langle w \rangle_t = t$. For the standard Poisson martingale $N_t - t = n_t$, $[n]_t = N_t$. Note

that since $[x] - \langle x \rangle$ is a martingale and $[x]$ is increasing, $\langle x \rangle = \pi_p^*[x]$. So, for the standard Poisson martingale $N_t - t = n_t$, the predictable quadratic variation $\langle n \rangle_t = t$ is the same as for standard Brownian motion.

We can now define the *predictable quadratic covariation* and the *optional quadratic covariation* of two processes $x, y \in \mathcal{H}^2$ by

$$\begin{aligned}\langle x, y \rangle &= \frac{1}{2} (\langle x+y \rangle - \langle x \rangle - \langle y \rangle) \\ [x, y] &= \frac{1}{2} ([x+y] - [x] - [y]).\end{aligned}$$

We have the following properties for quadratic covariations.

Lemma 1.4.2 If $x, y \in \mathcal{H}^2$, then

- i) $\langle x, y \rangle$ is the unique predictable integrable variation process with $\langle x, y \rangle_0 = x_0 y_0$ such that $xy - \langle x, y \rangle \in \mathcal{H}^1$
- ii) $[x, y]$ is an integrable variation process with $[x, y]_0 = x_0 y_0$ such that $xy - [x, y] \in \mathcal{H}^1$
- iii) $[x, y]_t = \langle x^c, y^c \rangle_t + \sum_{s \leq t} \Delta x_s \Delta y_s$ for all $t \in I$
- iv) $\langle x, y \rangle_t = 0$ implies $x \perp y$ in the Hilbert space \mathcal{H}^2 .

Finally, we define quadratic covariation processes for semimartingales. We note that a square integrable semimartingale x can be uniquely decomposed as $x = m^c + m^d + A$ where $m^c \in \mathcal{H}_c^2$, $m^d \in \mathcal{H}_d^2$, and A is a predictable integrable variation

process. So define $x^c = m^c$ and the optional quadratic covariation of two such semimartingales as

$$[x, y]_t = \langle x^c, y^c \rangle_t + \sum_{s \leq t} \Delta x_s \Delta y_s.$$

Then, define the predictable quadratic covariation as

$$\langle x, y \rangle_t = \pi_p^*[x, y]_t.$$

2. Martingale Calculus and Related Results

Now we are prepared to define integration w.r.t. a square integrable martingale and state the important differentiation rule for semimartingales. We also state representation and change of measure results for Wiener processes and conditional Poisson processes. It is these results that allow us to derive the filtering equations for the observation processes defined in the next chapter.

Proofs of the results stated in this section can be found in references [1], [2], and [3]. Again, we work on a given probability space (Ω, \mathcal{F}, P) with a complete right continuous filtration $\{\mathcal{F}_t\}$.

2.1 Integration w.r.t. a martingale

In discrete time, integration of a process h w.r.t. a martingale x is defined

$$(h \cdot x)_k = h_0 x_0 + \sum_{i=1}^k h_i (x_i - x_{i-1}).$$

Note that if h_k is \mathcal{F}_{k-1} measurable (h is predictable), then

$$E\{\Delta(h \cdot x)_{k-1} | \mathcal{F}_{k-1}\} = E\{h_k(x_k - x_{k-1}) | \mathcal{F}_{k-1}\} = 0.$$

So the transform $h \cdot h \cdot x$ maps predictable processes to martingales. We would like to retain this property when we define stochastic integrals in continuous time because it is a useful property in many applications. Symbolically, this means we need

$$E\{d(h \cdot x)_t | \mathcal{F}_{t-}\} = E\{h_t dx_t | \mathcal{F}_{t-}\} = 0.$$

So, again we will require that h be predictable. In fact, let $x \in \mathcal{H}^2$ and h be a bounded *simple predictable process* (for a deterministic partition $0 = t_0 < t_1 < \dots < t_n < t_{n+1} = \infty$, h satisfies $h_t = h_{t_i}$ for $t \in (t_i, t_{i+1}]$). Then $(h \cdot x)_t = \int_0^t h_s dx_s$ defined

$$(h \cdot x)_t = h_0 x_0 + \sum_{i=0}^n h_{t_i} (x_{t_{i+1} \wedge t} - x_{t_i \wedge t})$$

is in \mathcal{H}^2 and $E\{(h \cdot x)_\infty^2\} = E\{\int_0^\infty h_s^2 d\langle x \rangle_s\}$. Now let $\mathcal{Q}^2(x)$ denote the space of predictable processes for which

$$\|h\|_x^2 = E\{\int_0^\infty h_s^2 d\langle x \rangle_s\} < \infty.$$

Then we have the following:

Theorem 2.1.1 The map $h \cdot h \cdot x$ defined above that takes bounded simple predictable processes to martingales in \mathcal{H}^2 extends uniquely to a linear isometry, $h \cdot h \cdot x$, taking $\mathcal{Q}^2(x)$ into \mathcal{H}^2 .

Lemma 2.1.2 For $h \in \mathcal{H}^2(x)$ the process $\Delta(h \cdot x)_t$ and $h_t \Delta x_t$ are indistinguishable. In particular, $h \cdot x$ is continuous when x is continuous.

We can also characterize $h \cdot x$ as follows:

Theorem 2.1.3 Let $x \in \mathcal{H}^2$ and $h \in \mathcal{H}^2(x)$.

- (i) The stochastic integral $h \cdot x$ is the unique element of \mathcal{H}^2 such that $E\{(h \cdot x)_{\infty} y_{\infty}\} = E\{\int_0^{\infty} h_s d\langle x, y \rangle_s\}$ for every $y \in \mathcal{H}^2$
- (ii) Denoting the Stieltjes integral $\int_0^t h_s d\langle x, y \rangle_s$ as $h \cdot \langle x, y \rangle_t$, $h \cdot x$ is the unique element of \mathcal{H}^2 such that $\langle h \cdot x, y \rangle = h \cdot \langle x, y \rangle$ for all $y \in \mathcal{H}^2$.

Finally, we state an extension of Ito's formula and the product rule which we will use repeatedly in the next chapter.

Theorem 2.1.4 Suppose x is an \mathbb{R}^n -valued process each of whose components $x(i)$ is a semimartingale. If $\phi: \mathbb{R}^n \rightarrow \mathbb{R}$ is twice continuously differentiable, then

$$\begin{aligned} \phi(x_t) = & \phi(x_0) + \sum_1^n \int_0^t \frac{\partial}{\partial x_1} \phi(x_{s-}) dx_s(i) \\ & + \frac{1}{2} \sum_1^n \sum_1^n \int_0^t \frac{\partial^2}{\partial x_1 \partial x_j} \phi(x_{s-}) d\langle x^c(i), x^c(j) \rangle_s \end{aligned}$$

$$+ \sum_{s \leq t} \left[\phi(x_s) - \phi(x_{s-}) - \sum_{i=1}^n \frac{\partial}{\partial x_i} \phi(x_{s-}) \Delta x_s(i) \right].$$

Corollary 2.1.5 If x and y are semimartingales, then

$$d(xy)_t = x_t dy_t + y_t dx_t + d[x, y]_t.$$

2.2 Integral representations and change of measure

We will repeatedly make use of the following two results that state that a martingale measurable w.r.t. the σ -algebra generated by a SBM or CPP can be represented as a stochastic integral w.r.t. the SBM or CPP. First, we need a couple of definitions.

Definition 2.2.1 The r -dimensional process $\{w_t\} = \{w_t(1), \dots, w_t(r)\}$ where each component is a SBM and $w(i), w(j)$ are independent for all $i \neq j$ is called *r -dimensional standard Brownian motion (SBM)*.

Definition 2.2.2 The r -dimensional process $\{N_t\} = \{N_t(1), \dots, N_t(r)\}$ where the components are CPP's with intensities $\{\lambda_t(1), \dots, \lambda_t(r)\}$ depending on the \mathcal{F}_0 -measurable process $\{x_t\}$ and $N(i), N(j)$ are conditionally (conditioned on \mathcal{F}_0) independent for all $i \neq j$ is called an *r -dimensional conditional Poisson process (CPP)*.

Theorem 2.2.3 Suppose $\{w_t\}$ is an r -dimensional SBM and $\mathcal{G}_t = \sigma\{w_s : s \leq t\}$. If $\{m_t\}$ is a square integrable $\{\mathcal{G}_t\}$ -martingale, then there exist \mathcal{G}_t -predictable processes $k(1), \dots, k(r)$ with $E\{\int_0^\infty k_s^2(i) ds\} < \infty$ for all $i=1, \dots, r$ such that

$$m_t = m_0 + \sum_{i=1}^r \int_0^t k_s(i) dw_s(i) \quad \text{a.s.}$$

Theorem 2.2.4 Suppose $\mathcal{G}_t = \sigma\{N_s : s \leq t\}$ where $\{N_t\}$ is an r -dimensional CPP with $\{\mathcal{G}_t\}$ -measurable intensity vector $(\lambda_t(1), \dots, \lambda_t(r))$. If $\{m_t\}$ is a right continuous square integrable $\{\mathcal{G}_t\}$ -martingale, then there exist \mathcal{G}_t -predictable processes $k(1), \dots, k(r)$ with $E\{\int_0^\infty k_s^2(i) \lambda_s(i) ds\} < \infty$ for all $i=1, \dots, r$ such that

$$m_t = m_0 + \sum_{i=1}^r \int_0^t k_s(i) (dN_s(i) - \lambda_s(i) ds) \quad \text{a.s.}$$

Just as important as the above theorems are the following results which describe how standard Brownian or Poisson processes behave under a Girsanov type change of measure.

Theorem 2.2.5 Suppose $\{y_t\}$ is an r -dimensional SBM under P on $[0, T]$ and $h: \Omega \times [0, T] \rightarrow \mathbb{R}^r$ is a measurable process with $\int_0^T |h_s|^2 ds < \infty$ a.s. Define

$$R_t = \exp \left[\int_0^t h_s^* dy_s - \frac{1}{2} \int_0^t |h_s|^2 ds \right]$$

where $*$ is the transpose operation. If $E\{R_T\}=1$ and Q is the probability measure on \mathcal{F} defined by $\frac{dQ}{dP}=R_T$, then $\{R_t\}$ is an $\{\mathcal{F}_t\}$ -martingale under Q and $w_t = y_t - \int_0^t h_s ds$ is an r -dimensional SBM under Q on $[0, T]$.

Theorem 2.2.6 Suppose $\{N_t\}$ is an r -dimensional Poisson process with intensity vector $e=(1, \dots, 1)$ under P on $[0, T]$ and $\lambda: \Omega \times [0, T] \rightarrow \mathbb{R}^r$ is a positive \mathcal{F}_0 -measurable process with $\int_0^T \lambda_s(i) ds < \infty$ a.s. for each $i=1, \dots, r$. Define

$$R_t = \exp \left[\int_0^t (\ln \lambda_s)^* dN_s + \int_0^t e^* (e - \lambda_s) ds \right]$$

where $*$ is again the transpose operation. If $E\{R_T\}=1$ and Q is again the probability measure on \mathcal{F} defined by $\frac{dQ}{dP}=R_T$, then $\{R_t\}$ is an $\{\mathcal{F}_t\}$ -martingale under Q and for each $i=1, \dots, r$

$$n_t = N_t(i) - \int_0^t \lambda_s(i) ds$$

is an $\{\mathcal{F}_t\}$ -martingale under Q on $[0, T]$.

Finally, we will need the following result due to Doléans-Dade.

Theorem 2.2.7 (Doléans-Dade) Suppose $\{m_t\}$ is a semimartingale with $m_0=0$. Then there exists a unique semimartingale $\{R_t\}$

such that $R_t = 1 + \int_0^t R_{s-} dm_s$. Furthermore, R_t is given as

$$R_t = \exp \left[m_t - \frac{1}{2} \langle m^c, m^c \rangle_t \right] \prod_{0 \leq s \leq t} (1 + \Delta m_s) e^{-\Delta m_s}$$

where the infinite product is absolutely convergent a.s.

Our interest in this result is related to the previous two theorems.

- i) Take $m_t = \int_0^t h_s^* dy_s$ and we get that R_t in Theorem 2.2.5 satisfies $R_t = 1 + \int_0^t R_{s-} h_s^* dy_s$.
- ii) Take $m_t = \int_0^t (\lambda_s - e)^* (dN_s - eds)$ and we get that R_t in Theorem 2.2.6 satisfies $R_t = 1 + \int_0^t R_{s-} (\lambda_s - e)^* (dN_s - eds)$.

3. Markov Processes and Convergence

In this section we examine the processes of interest in Chapter IV, Markov processes. First, we review the main definitions and look at examples. Then, we summarize the semigroup theory of Markov processes leading to the representation of a Markov process as a semimartingale. We need this result in Chapter III. Since we are concerned with convergence properties of Markov processes in Chapter IV, in the last subsection we review weak convergence theory and state the main results used in Chapter IV.

The sources used for the material in this section are [8] and [9]. As usual, we work on a probability space (Ω, \mathcal{F}, P) with a complete right continuous filtration $\{\mathcal{F}_t\}$. Also, S will denote a complete separable metric space and $B(S)$ will denote the Banach space of bounded measurable real-valued functions on S . Finally, unless otherwise stated, we assume that all processes are cadlag processes adapted to $\{\mathcal{F}_t\}$.

3.1 Definitions and examples

A function $p(t, x, A)$ defined on $I \times S \times \mathcal{B}(S)$ is a (*time-homogeneous*) *transition function* if

- i) for each $(t, x) \in I \times S$, $p(t, x, \cdot) \in \mathcal{P}(S)$ (the space of probability measures on S)
- ii) for each $x \in S$, $p(0, x, \cdot) = \delta_x$ (unit mass at x)
- iii) for each $A \in \mathcal{B}(S)$, $p(\cdot, \cdot, A) \in B(I \times S)$
- iv) *Chapman-Kolmogorov property*: for $s, t \geq 0$, $x \in S$,

$$A \in \mathcal{B}(S) \text{ we have } p(t+s, x, A) = \int_S p(s, y, A) p(t, x, dy).$$

(In the above, the time index set I can be either $[0, \infty)$ or $\{0, 1, 2, \dots\}$.) Suppose $\{x_t\}$ is an S -valued process on I . Then x is a (*time-homogeneous*) *Markov process* with transition function p if

$$P(x_{t+s} \in A | \mathcal{F}_t) = p(s, x_t, A)$$

for all $s, t \geq 0$ and $A \in \mathcal{B}(S)$, or equivalently, if

$$E\{\phi(x_{t+s}) | \mathcal{F}_t\} = \int_S \phi(y) p(s, x_t, dy)$$

for all $s, t \geq 0$ and $\phi \in B(S)$.

Intuitively, the Markov property says that the state of the process at time $t+s$ depends only on its state at time t and not on the entire history of the process up to time t . Time-homogeneity means that the expected change in the process over a given time interval depends only on the length of the interval. Unless otherwise stated, when we talk about a Markov process we will mean a time homogeneous Markov process. It should also be noted that since S is complete and separable, there exists a Markov process x whose finite-dimensional distributions are uniquely determined by a given transition function and initial distribution (see [8]).

A process x is *strong Markov* if for every $\{\mathcal{F}_t\}$ -stopping time $\tau < \infty$ a.s.

$$P(x_{\tau+t} \in A | \mathcal{F}_\tau) = p(t, x_\tau, A)$$

for all $t \geq 0$ and $A \in \mathcal{B}(S)$, or equivalently, if

$$E\{\phi(x_{\tau+t}) | \mathcal{F}_\tau\} = \int_S \phi(y) p(t, x_\tau, dy)$$

for all $t \geq 0$ and $\phi \in B(S)$. Also, a Markov process is said to be *Feller* if for each $t \geq 0$ and each bounded continuous function ϕ on S , $E_x\{\phi(x_t)\}$ is a continuous function in x on S . Cadlag Feller processes have the strong Markov property. So, verification of the Feller property is a good way to verify that a Markov process is strong Markov.

The first example of interest to us is when both the time interval I and the state space S are discrete. In this case we say a Markov process is a *Markov chain*, and the transition matrix Π is the matrix (possibly infinite dimensional) with entries $\pi(x,y)=p(1,x,y)$ for $x,y \in S$. Note that for a Markov chain $\{x_n\}$ with transition matrix Π and initial distribution vector p_0 , we can compute the distribution vector p_n and the expectation $E\{\phi(x_n)\}$ ($\phi \in B(S)$) for the process at time n by

$$p_n = (\Pi^n)^* p_0 \text{ and } E\{\phi(x_n)\} = p_0^* \Pi^n \phi .$$

(We use the same notation whether we think of ϕ as a function on S or a vector with entries $\phi(x)$, $x \in S$.)

Next, we consider the case where $I=[0,\infty)$ but S is still discrete. Now, we call the Markov process a *discrete state Markov process* and the transition matrix $\Pi(t)$ for each $t \in I$ is the matrix with entries $\pi(t)(x,y)=p(t,x,y)$ for $x,y \in S$. Again, for a given initial distribution vector p_0 , we can compute the distribution vector p_t and the expectation $E\{\phi(x_t)\}$ ($\phi \in B(S)$) for the process at time t by

$$p_t = \Pi^*(t) p_0 \text{ and } E\{\phi(x_t)\} = p_0^* \Pi(t) \phi .$$

Finally, when $I=[0,\infty)$ and $S=\mathbb{R}^m$, the most common Markov processes are diffusion processes. We define diffusions as solutions to stochastic differential equations. Suppose $\{w_t\}$ is m -dimensional SBM. Also, assume $f:\mathbb{R}^m \rightarrow \mathbb{R}^m$ and $\sigma:\mathbb{R}^m \rightarrow \mathbb{R}^m \times \mathbb{R}^m$ are

vector and matrix valued functions respectively that are Lipschitz continuous, i.e., there exists a positive constant L such that

$$|f(x) - f(y)| \leq L|x - y| \text{ and } \|\sigma(x) - \sigma(y)\| \leq L|x - y|$$

for all $x, y \in \mathbb{R}^m$ (say, $\|A\| = \sup_x |Ax|$ is the matrix norm). Then the stochastic differential equation

$$x_t = x_0 + \int_0^t f(x_s) ds + \int_0^t \sigma(x_s) dw_s$$

has a unique (up to indistinguishability) continuous $\{\mathcal{F}_t\}$ -measurable solution on I for each $x_0 \in \mathbb{R}^m$. The solution of the above SDE is called a *diffusion* with *drift* $f(x)$ and *diffusion matrix* $a(x) = \sigma(x)\sigma^*(x)$. It turns out that a diffusion process is a Markov process (actually a strong Markov process).

3.2 Semigroup theory

Suppose $\{x_t\}$ is an S -valued Markov process. For each $t \in I$, define the linear operator $T_t: B(S) \rightarrow B(S)$ for $\phi \in B(S)$ by

$$(T_t \phi)(x) = E_x\{\phi(x_t)\} \text{ for all } x \in S.$$

First, we should check that the map $\phi \rightarrow T_t \phi$ takes bounded functions to bounded functions. In fact, T_t is a *contraction operator* for $t \in I$, i.e., $|(T_t \phi)(x)| \leq \|\phi\|_\infty$ for all $x \in S$. This property is easily verified via Jensen's inequality. Next, the Markov property can be used to show that the family of operators $\{T_t: t \in I\}$ is a *semigroup*, i.e., $T_{s+t} = T_s \circ T_t$ for all $s, t \in I$. Note that if $\{x_t\}$ is a Feller process, then the

operators in $\{T_t\}$ also map bounded continuous functions to bounded continuous functions. The family $\{T_t\}$ is called the *semigroup of conditioned shifts corresponding to $\{x_t\}$* .

Now, we assume $I=[0, \infty)$. We can define the *infinitesimal generator A of $\{T_t\}$ or $\{x_t\}$* as the linear operator satisfying

$$(A\phi)(x) = \lim_{t \downarrow 0} \frac{(T_t \phi)(x) - \phi(x)}{t} \quad \text{for all } \phi \in D_A.$$

Here $D_A = \{\phi \in B(S) : \text{the limit above converges uniformly in } x\}$.

For example, when $\{x_t\}$ is a discrete state Markov process with a differentiable transition matrix $\Pi(t)$, then the generator is the matrix

$$A = \left. \frac{d\Pi(t)}{dt} \right|_{t=0} \quad \text{with } D_A = B(S).$$

Also, when $\{x_t\}$ is a diffusion with drift $f(x)$ and diffusion matrix $a(x)$, then the generator is the differential operator

$$A = \sum_{i=1}^m f_i \frac{\partial}{\partial x_i} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m a_{ij} \frac{\partial^2}{\partial x_i \partial x_j}$$

where D_A contains $C_0^2(S) = \{\phi \in B(\mathbb{R}^m) : \phi \text{ is twice continuously differentiable with compact support}\}$.

The semigroup $\{T_t\}$ corresponding to a Markov process $\{x_t\}$ has the following properties:

Theorem 3.2.1 If $\phi \in D_A$, then

- i) $T_t \phi \in D_A$ for all $t \geq 0$
- ii) A and T_t commute for all $t \geq 0$
- iii) (Kolmogorov's backward equation):

$$\frac{\partial (T_t \phi)(x)}{\partial t} = A(T_t \phi)(x) \text{ for all } x \in S.$$

Using this theorem we can derive:

Proposition 3.2.2 (Dynkin's formula) If $\phi \in D_A$ and $A\phi(x_t)$ is right continuous, then

$$E_x\{\phi(x_t)\} = \phi(x) + E_x\left\{\int_0^t (A\phi)(x_s) ds\right\}.$$

Now, using Dynkin's formula we get a fundamental result we will need in the next chapter.

Theorem 3.2.3 Under the conditions of Proposition 3.2.2,

$$\phi(x_t) - \phi(x_0) - \int_0^t (A\phi)(x_s) ds = m_t \text{ is an } \{\mathcal{F}_t\}\text{-martingale.}$$

3.3 Weak convergence

Let $C_b(S)$ denote the space of real-valued bounded continuous functions on S . Also, let $\mathcal{P}(S)$ denote the family of Borel probability measures on S . For a sequence $\{\mu_n\} \subset \mathcal{P}(S)$, we say $\{\mu_n\}$ converges weakly to $\mu \in \mathcal{P}(S)$ (denoted $\mu_n \rightarrow \mu$) if

$$\lim_{n \rightarrow \infty} \int_S \phi d\mu_n = \int_S \phi d\mu \quad \forall \phi \in C_b(S).$$

If the sequence $\{\mu_n\}$ and μ are measures induced by S -valued random variables $\{x_n\}$ and x , then we say $\{x_n\}$ converges in distribution to x (denoted $x_n \rightarrow x$) if $\mu_n \rightarrow \mu$. We have that $\mu_n \rightarrow \mu$ (or $x_n \rightarrow x$) iff $\{\mu_n\}$ converges to μ in $\mathcal{P}(S)$, where $\mathcal{P}(S)$ is topologized with the Prohorov metric ρ :

$$\rho(\mu, \nu) = \inf\{\epsilon > 0 : \mu(F) \leq \nu(F^\epsilon) + \epsilon \quad \forall \text{ closed } F \subseteq S\}.$$

$$(F^\epsilon = \{x \in S : \inf_{y \in F} d(x, y) < \epsilon\} \text{ where } d \text{ is the metric on } S.)$$

Now, we would like to discuss the standard technique for showing that a sequence $\{x_n\}$ converges in distribution to x . First, if S is complete and separable, then the Prohorov space $\mathcal{P}(S)$ is complete and separable, so that a precompact sequence $\{\mu_n\} \subseteq \mathcal{P}(S)$ has a convergent subsequence. Next, we have that a sequence $\{\mu_n\} \subseteq \mathcal{P}(S)$ is precompact iff $\{\mu_n\}$ is tight, i.e., for each $\epsilon > 0$ there exists a compact set $K_\epsilon \subseteq S$ such that $\inf \mu_n(K_\epsilon) > 1 - \epsilon$. So, to show that $x_n \rightarrow x$, we show that the sequence of induced measures $\{\mu_n\}$ is tight and then show that $x_{n_k} \rightarrow x$ for each convergent subsequence $\{x_{n_k}\}$.

Often in filtering S is the space $D_E(I)$. If E is a complete separable metric space, then $D_E(I)$ is the complete separable metric space of all E -valued right continuous functions on I with left limits topologized with the Skorohod metric. The Skorohod metric is described in [8], but we will

not need to explicitly work with this metric. Likewise, several criteria for tightness of sequences of measures on $D_E(I)$ are given in [8] which we do not explicitly use. However, we do need the following two results from [8].

Theorem 3.3.1 Suppose $\{x_t\}$ is an \mathbb{R}^m -valued diffusion with drift $f(x)$ and diffusion matrix $a(x)$ for $t \in I = [0, T]$. Also, suppose $\{x^N\}$ and $\{B^N\}$ are sequences of $D_{\mathbb{R}^m}(I)$ -valued random variables, and $\{A^N\}$ is a sequence of $m \times m$ nonnegative definite matrices of processes $A^N(i, j)$ taking values in $D_{\mathbb{R}}(I)$. Let $\mathcal{F}_t^N = \sigma(x_s^N, B_s^N, A_s^N; s \leq t)$ and assume the following conditions are satisfied.

- i) The components of $M^N = x^N - B^N$ are $\{\mathcal{F}_t^N\}$ -martingales.
- ii) The components of $M^N (M^N)^* - A^N$ are $\{\mathcal{F}_t^N\}$ -martingales.
- iii) $\lim_{N \rightarrow \infty} E \left\{ \sup_{t \leq T} |x_t^N - x_{t-}^N|^2 \right\} = 0.$
- iv) $\lim_{N \rightarrow \infty} E \left\{ \sup_{t \leq T} |B_t^N - B_{t-}^N|^2 \right\} = 0.$
- v) $\lim_{N \rightarrow \infty} E \left\{ \sup_{t \leq T} |A_t^N(i, j) - A_{t-}^N(i, j)|^2 \right\} = 0$ for $i, j = 1, \dots, m.$
- vi) $\sup_{t \leq T} \left| B_t^N(i) - \int_0^t f_1(x_s^N) ds \right| \rightarrow 0$ in probability for each $i = 1, \dots, m.$

$$\text{vii) } \sup_{t \leq T} \left| A_t^N(i, j) - \int_0^t a_{ij}(x_s^N) ds \right| \xrightarrow{40} 0 \text{ in probability for} \\ \text{each } i, j = 1, \dots, m.$$

Then $x_n \rightarrow x$ provided the distributions of the sequence $\{x_0^N\}$ converge weakly to the distribution of x_0 .

Theorem 3.3.2 (Skorohod Representation) Suppose the sequence $\{x_n\}$ of S -valued random variables converges in distribution to the S -valued random variable x and these random variables induce the measures $\{\mu_n\}, \mu$ on S . Then there exist S -valued random variables $\{y_n\}, y$ defined on a single probability space such that $\{y_n\}, y$ induce the above measures $\{\mu_n\}, \mu$ on S and

$$\lim_{n \rightarrow \infty} y_n \rightarrow y \text{ a.s.}$$

CHAPTER III - EQUATIONS OF NONLINEAR FILTERING

The equations of nonlinear filtering developed here fall into three categories. First, we derive the Kushner-Stratonovich equation for the conditional expectation of a signal process given the information of an observed process. The innovations approach pioneered by Fujisaki, Kallianpur, and Kunita [7] is used in this first section. Then, we use the reference probability approach to derive the Zakai equation for the unnormalized conditional expectation. Finally, we derive equations for the unnormalized conditional densities of Markov signal processes. As we will see, these last equations are the most useful for computational purposes.

1. Innovations Approach

Suppose we are given a partially observed system consisting of a signal process $\{\xi_t\}$ and an observation process $\{y_t\}$ defined on a probability space (Ω, \mathcal{F}, P) and adapted to a complete right continuous filtration $\{\mathcal{F}_t: 0 \leq t \leq T\}$. We wish to develop a recursive algorithm for computing the optimal least squares estimate of ξ_t given the observations in the σ -algebra $Y_t = \sigma\{y_s: 0 \leq s \leq t\}$, i.e., we want to compute $E\{\xi_t | Y_t\}$ via a recursive equation. The equation derived is a generalization of the Kushner-Stratonovich equation which was first developed

for a Markov signal process observed in additive white noise. We combine methods used by Elliott [4] and Wong [5] to develop the equation when the signal and observation process are more general semimartingales.

Specifically, unless otherwise stated, we assume that the signal process $\{\xi_t\}$ is a real-valued semimartingale of the form:

$$\xi_t = \xi_0 + \int_0^t f_s ds + m_t$$

where f is an $\{\mathcal{F}_t\}$ -adapted process with $E\{\int_0^T f_s^2 ds\} < \infty$ and $\{m_t\}$ is a square integrable cadlag $\{\mathcal{F}_t\}$ -martingale on $[0, T]$. Note that there is no loss of generality in assuming that the signal process is one-dimensional since we would simply derive an equation for the conditional expectation of each individual component of a multi-dimensional signal.

Also, we assume that the observation process $\{y_t\}$ is an \mathbb{R}^r -valued semimartingale of the form $y_t = \int_0^t g_s ds + n_t$ where

i) $g_t = h(\xi_t)$ and $n_t = w_t$ is r -dimensional SBM,

or ii) $y_t = N_t$ is an r -dimensional CPP with intensity

vector $\lambda(\xi_t)$, i.e., $g_t = \lambda(\xi_t)$ and $n_t = N_t - \int_0^t \lambda_s ds$

(where $\lambda_s = \lambda(\xi_s)$).

In either case we want g to be an $\{\mathcal{F}_t\}$ -adapted process with $E\{\int_0^T |g_s|^2 ds\} < \infty$. Also, if $y_t = N_t$ we want each component of λ_t to be positive and bounded away from zero so that each component of n_t is a square integrable $\{\mathcal{F}_t\}$ -martingale on $[0, T]$.

Finally, we make the usual assumption that each component of n_t is independent of m_t .

1.1 Innovations process

For an arbitrary cadlag process ϕ we define the optional projection

$$\phi_t^0 = E\{\phi_t | Y_t\}$$

and the predictable projection

$$\phi_t^p = E\{\phi_t | Y_{t-}\}.$$

So, if ϕ is bounded or a positive process, then by the Projection Theorem $\phi^0 = \Pi_0(\phi)$ and $\phi^p = \Pi_p(\phi)$.

The innovations approach to filtering involves expressing ξ^0 as a $\{Y_t\}$ -semimartingale with martingale part a stochastic integral w.r.t. the innovations process defined below.

Lemma 1.1.1 The innovations process $\tilde{n}_t = y_t - \int_0^t g_s^p ds$ is a $\{Y_t\}$ -martingale.

Proof Obviously, \tilde{n}_t is Y_t -measurable for each $t \in [0, T]$ and the integrability of \tilde{n}_t follows from the assumptions made on the observations y_t . Also, for $t > s$

$$\begin{aligned} E\{\tilde{n}_t - \tilde{n}_s | Y_s\} &= E\{n_t - n_s | Y_s\} + E\left\{\int_s^t g_u - g_u^p du \mid Y_s\right\} \\ &= E\{E\{n_t - n_s | \mathcal{F}_s\} | Y_s\} + E\left\{\int_s^t g_u - g_u^0 du \mid Y_s\right\} \\ &= 0 + \int_s^t E\{E\{g_u - g_u^0 | Y_u\} | Y_s\} du = 0. \end{aligned}$$

Note that in the second equality we used the fact that

$\int_s^t g_u^p du = \int_s^t g_u^0 du$ a.s. This result is proved in [4] Lemma 1.1, but since $g_u^p = g_u^0$, except at possibly countably many values of u for each path, the result seems reasonable. ■

The following lemma shows that ξ^0 is a $\{Y_t\}$ -semimartingale.

Lemma 1.1.2 $\tilde{m}_t = \xi_t^0 - \xi_0^0 - \int_0^t f_s^0 ds$ is a $\{Y_t\}$ -martingale.

Proof Same as Lemma 1.1.1. ■

Next we want to show that there exists a $\{Y_t\}$ -predictable \mathbb{R}^r -valued process $\{k_t\}$ with $\tilde{m}_t = \int_0^t k_s^* d\tilde{n}_s$. For case (ii) where y is a CPP the representation follows immediately from Theorem II.2.2.4. For case (i) where the signal is observed in additive white noise a little more work is required. In this case it is easy to show that the innovations process \tilde{n} is SBM w.r.t. $\{Y_t\}$ and, clearly, $\sigma\{\tilde{n}_s : s \leq t\} \subseteq Y_t$. So if the "innovations conjecture" that $Y_t \subseteq \sigma\{\tilde{n}_s : s \leq t\}$ is true, then we can use Theorem II.2.2.3 to obtain the desired representation for \tilde{m}_t . When $r=1$ the innovations conjecture is proved in [6], but even without assuming the validity of the innovations conjecture the representation result still follows as proved in [7].

1.2 Gain computation

For our recursive filtering equation we only have left to describe the gain, the $\{k_t\}$ process. This is accomplished using the product rule, Corollary II.2.1.5. We have

$$\begin{aligned}
 (1) \quad \xi_t y_t &= \int_0^t \xi_{s-} dy_s + \int_0^t y_{s-} d\xi_s + [m, n]_t \\
 &= \int_0^t \xi_{s-} dy_s + \int_0^t y_{s-} d\xi_s \\
 (2) \quad \xi_t^0 y_t &= \int_0^t \xi_{s-}^0 dy_s + \int_0^t y_{s-} d\xi_s^0 + [\tilde{m}, \tilde{n}]_t \\
 &= \int_0^t \xi_{s-}^0 dy_s + \int_0^t y_{s-} d\xi_s^0 + \int_0^t (d[\tilde{n}]_s k_s)
 \end{aligned}$$

where $[\tilde{m}, \tilde{n}]_t$ is the vector with components $[\tilde{m}, \tilde{n}(i)]_t$ and $d[\tilde{n}]_s$ is the diagonal matrix with diagonal entries $d[\tilde{n}(i)]_s$.

Next, note that for an arbitrary process $\{z_t\}$,

$$E\{dz_t | Y_{t-}\} = E\{dz_t^0 | Y_{t-}\}, \text{ so that}$$

$$E\{d(\xi_t^0 y_t) | Y_{t-}\} = E\{d(\xi_t y_t)^0 | Y_{t-}\} = E\{d(\xi_t y_t) | Y_{t-}\}$$

and

$$E\{y_{t-} d\xi_t^0 | Y_{t-}\} = y_{t-} E\{d\xi_t^0 | Y_{t-}\} = y_{t-} E\{d\xi_t | Y_{t-}\} = E\{y_{t-} d\xi_t | Y_{t-}\}.$$

Thus, taking the predictable projections of equations (1) and (2), we get

$$\xi_t^0 - E\{dy_t | Y_{t-}\} + E\{d[\tilde{n}]_t | Y_{t-}\} k_t = E\{\xi_t - dy_t | Y_{t-}\}$$

or

$$\xi_t^0 - g_t^P dt + \gamma_t k_t dt = (\xi_t - g_t)^P dt$$

where γ_t is the diagonal matrix defined by $\frac{E\{d[\tilde{n}]_t | Y_{t-}\}}{dt}$.

In case (i) γ_t is the $r \times r$ identity matrix. In case (ii) γ_t is the diagonal matrix with diagonal entries given by the predictable intensity vector λ_t^P . The gain k_t is substituted in the integral representing the $\{Y_t\}$ -martingale \tilde{m}_t and Lemma 1.1.2 gives us the following result.

Theorem 1.1.3 (Kushner-Stratonovich equation)

$$\xi_t^0 = \xi_0^0 + \int_0^t f_s^0 ds + \int_0^t [(\xi_{s-} - g_s)^P - \xi_{s-}^0 - g_s^P]^* \gamma_s^{-1} (dy_s - g_s^P ds)$$

Finally, in Chapter IV we will assume that the signal $\xi_t = \phi(x_t)$ where $\{x_t\}$ is an S -valued time homogeneous Markov process with generator A and $\phi \in D_A$. Then Theorem 1.1.3 and Theorem II.3.2.3 give us:

Theorem 1.1.4

$$\Pi_0(\phi_t) = \Pi_0(\phi_0) + \int_0^t \Pi_0(A\phi)_s ds + \int_0^t [\Pi_p(\phi_{s-} - g_s) - \Pi_0(\phi_{s-}) \Pi_p(g_s)]^* \gamma_s^{-1} (dy_s - \Pi_p(g_s) ds)$$

2. Reference Probability Approach

The above equations can be simplified if, instead of working with the conditional expectation w.r.t. the

probability measure P , we work with an appropriate conditional expectation w.r.t. a reference probability measure Q . In this section we derive the "unnormalized" conditional expectation equations corresponding to the ordinary conditional expectation equations in Section 1.

The notation and assumptions are the same as in the previous section with the following exceptions. First, we assume that $\{g_t\}$ is a bounded process. This assumption ensures that $E\{R_T\}=1$ in the Girsanov theorems (Theorem II.2.2.5 and Theorem II.2.2.6). Second, we assume the existence of a reference measure Q such that

- i) when y is as in case (i) then $\{y_t\}$ is SBM under Q
- ii) when y as in case (ii) then $\{y_t\}$ is a standard Poisson process under Q .

2.1 Kallianpur-Striebel formula

Assuming the existence of the reference probability measure Q described above, we can use Theorems II.2.2.5 and II.2.2.6 to define the measure P on \mathcal{F} with $\frac{dP}{dQ} = R_T$ such that $y_t = \int_0^t g_s ds + n_t$ where $\{n_t\}$ is either r -dimensional SBM under P or an r -dimensional conditional Poisson martingale under P . So, under P , $\{y_t\}$ is the observation process described in Section 1 while, under Q , $\{y_t\}$ is a standard process (Brownian or Poisson). The following result shows we can formulate the

filtering problem under Q as well as under P .

Theorem 2.1.1 (Kallianpur-Striebel)

For any random

variable ϕ

$$\phi_t^0 = E^P\{\phi|Y_t\} = \frac{E^Q\{\phi R_t|Y_t\}}{E^Q\{R_t|Y_t\}} = \frac{\sigma_t(\phi)}{\sigma_t(1)}$$

Proof Assume $A \in Y_t$. Then

$$E^P\{\chi_A E^P\{\phi|Y_t\}\} = E^Q\{\chi_A R_t E^P\{\phi|Y_t\}\} = E^Q\{\chi_A E^Q\{R_t|Y_t\} E^P\{\phi|Y_t\}\},$$

and

$$E^P\{\chi_A \phi\} = E^Q\{\chi_A R_t \phi\} = E^Q\{\chi_A E^Q\{R_t \phi|Y_t\}\}.$$

Since $E^P\{\chi_A E^P\{\phi|Y_t\}\} = E^P\{\chi_A \phi\}$ and $A \in Y_t$ is arbitrary, the result follows. ■

The process $\sigma_t(\xi_t) = E^Q\{\xi_t R_t|Y_t\}$ is called the unnormalized conditional expectation (UCE) and the Kallianpur-Striebel formula shows us how to compute the ordinary conditional expectation from the UCE.

Our goal then is to derive recursive equations for computing the UCE, but first we note two facts we will need. Let $z_t = y_t - \int_0^t \delta ds$ where for y as in case (i) δ is the r -dimensional zero vector and for y as in case (ii) δ is the r -dimensional vector with each component equal to one. Then from Theorem II.2.2.7,

$$(3) \quad R_t = 1 + \int_0^t R_{s-} (g_s - \delta) * dz_s.$$

Also, denoting $\bar{R}_t = E^Q\{R_t | Y_t\}$, Theorems II.2.2.3 and II.2.2.4 guarantee the existence of a $\{Y_t\}$ -predictable vector process $\{k_t\}$ such that

$$(4) \quad \bar{R}_t = 1 + \int_0^t k_s^* dz_s.$$

2.2 Zakai equation

First, we compute the vector process $\{k_t\}$ in equation

(4). The method is similar to the gain computation in Section

1. Using equations (3) and (4),

$$(5) \quad R_t Y_t = \int_0^t R_{s-} dy_s + \int_0^t y_{s-} dR_s + \int_0^t R_{s-} (d[z]_s (g_s - \delta))$$

and

$$(6) \quad \bar{R}_t Y_t = \int_0^t \bar{R}_{s-} dy_s + \int_0^t y_{s-} d\bar{R}_s + \int_0^t (d[z]_s k_s)$$

where $d[z]_s$ is the diagonal matrix with diagonal entries $d[z(i)]_s$. Taking the predictable projections of equations (5) and (6) w.r.t. Q and simplifying, we get

$$\begin{aligned} & \bar{R}_t E^Q\{dy_t | Y_t\} + E^Q\{d[z]_t k_t | Y_t\} \\ &= E^Q\{R_t dy_t | Y_t\} + E^Q\{R_t d[z]_t (g_t - \delta) | Y_t\}. \end{aligned}$$

or

$$E^Q\{d[z]_t k_t | Y_t\} = E^Q\{R_t d[z]_t (g_t - \delta) | Y_t\}.$$

So, $k_t = E^Q\{R_t (g_t - \delta) | Y_t\}$, but by Theorem 2.1.1

$$k_t = E^Q\{R_t (g_t - \delta) | Y_t\} = \bar{R}_t E^P\{g_t - \delta | Y_t\} = \bar{R}_t (g_t^P - \delta).$$

Thus,

$$(7) \quad \bar{R}_t = 1 + \int_0^t \bar{R}_{s-} (g_s^P - \delta)^* dz_s.$$

Now we can use equation (7) and Theorem 1.1.3 to derive the Zakai equation for the UCE.

Theorem 2.2.1 (Zakai)

$$\sigma_t(\xi_t) = \sigma_0(\xi_0) + \int_0^t \sigma_s(f_s) ds + \int_0^t [\sigma_{s-}(\xi_{s-} g_s) - \delta \sigma_{s-}(\xi_{s-})]^* (dy_s - \delta ds)$$

Proof

$$\begin{aligned} \bar{R}_t \xi_t^0 &= \xi_t^0 + \int_0^t \bar{R}_{s-} d\xi_s^0 + \int_0^t \xi_{s-}^0 d\bar{R}_s + \int_0^t \bar{R}_{s-} (g_s^P - \delta)^* d[y]_s \gamma_s^{-1} [(\xi_{s-} g_s)^P - \xi_{s-}^0 g_s^P] \\ &= \xi_t^0 + \int_0^t \bar{R}_{s-} f_s^0 ds + \int_0^t \bar{R}_{s-} [(\xi_{s-} g_s)^P - \xi_{s-}^0 g_s^P]^* \gamma_s^{-1} (dy_s - g_s^P ds) \\ &\quad + \int_0^t \bar{R}_{s-} \xi_{s-}^0 (g_s^P - \delta)^* (dy_s - \delta ds) \\ &\quad + \int_0^t \bar{R}_{s-} (g_s^P - \delta)^* d[y]_s \gamma_s^{-1} [(\xi_{s-} g_s)^P - \xi_{s-}^0 g_s^P] \end{aligned}$$

where again $d[y]_s$ is the diagonal matrix with diagonal entries $d[y(i)]_s$. After some lengthy computations the above equation simplifies to:

$$(8) \quad \bar{R}_t \xi_t^0 = \xi_t^0 + \int_0^t \bar{R}_{s-} f_s^0 ds + \int_0^t \bar{R}_{s-} [(\xi_{s-} g_s)^P - \delta \xi_{s-}^0]^* (dy_s - \delta ds).$$

Finally, we note that $\xi_t^0 = \sigma_0(\xi_0)$ and use the Kallianpur-Striebel formula on $\bar{R}_t \xi_t^0$ and the two integrals in equation (8) to finish the proof. ■

As in Section 1, we restate Theorem 2.2.1 for the case where the signal $\xi_t = \phi(x_t)$ with $\{x_t\}$ an S-valued time homogeneous Markov process with generator A and $\phi \in D_A$.

Theorem 2.2.2

$$\sigma_t(\phi_t) = \sigma_0(\phi_0) + \int_0^t \sigma_s(A\phi)_s ds + \int_0^t [\sigma_{s-}(\phi_{s-}g_s) - \delta\sigma_{s-}(\phi_{s-})]^* (dy_s - \delta ds).$$

3. Conditional Densities for Markov Signals

Although we have derived filtering equations for general semimartingale signals, the most common situation occurs when the signal is a Markov process. Then, Theorems 1.1.4 and 2.2.2 provide recursive equations for computing normalized and unnormalized conditional expectations. Unfortunately, these equations are usually infinite dimensional. There are two notable exceptions, however. When the partially observed system is linear, Theorem 1.1.4 can be used to derive the finite dimensional Kalman-Bucy filter (see [11] p. 68). Also, when the Markov signal has a finite dimensional state space, Theorem 2.2.2 can be used to derive a finite dimensional stochastic differential equation (SDE) for computing the unnormalized conditional density (UCD). Equations for computing the UCD are the focus of this section. First, we derive the finite dimensional SDE described above. Then, we discuss some of the computational aspects of this equation.

Last, we describe conditions under which equations for computing the UCD of a diffusion process can be developed. Throughout this section we use the notation of Section II.3 and assume that the signal $\{x_t\}$ is an S-valued (time homogeneous) Markov process with generator A.

3.1 Finite state signal

Assume $S=\{x_1, \dots, x_N\}$. Initially we suppose the signal is unobservable and denote $P(x_t=x)$ as $p_t(x)$ for all $x \in S$. Then we have the following.

Theorem 3.1.1 (Kolmogorov's forward equation)

The

distribution vector p_t satisfies the initial value problem

$$\frac{dp_t}{dt} = A^* p_t \text{ with initial value } p_0.$$

(As usual A^* is the adjoint matrix of A.)

Proof Recall that for $\phi \in B(S)$, $T_t \phi = \Pi(t) \phi$ where T_t is the semigroup operator and $\Pi(t)$ is the transition matrix corresponding to $\{x_t\}$. So, Kolmogorov's backward equation (Theorem II.3.2.1) can be written $\Pi'(t) \phi = A \Pi(t) \phi$ for all $\phi \in B(S)$. Therefore,

$$(p_t')^* \phi = p_0^* \Pi'(t) \phi = p_0^* A \Pi(t) \phi = p_0^* \Pi(t) A \phi = p_t^* A \phi.$$

Since ϕ is arbitrary, we have $p_t' = A^* p_t$. ■

Now, suppose we partially observe $\{x_t\}$ through $\{y_t\}$ and denote $\rho_t(x) = E^P\{\chi(x_t=x) | Y_t\}$ for all $x \in S$. Also, define the unnormalized conditional density (distribution):

$$q_t(x) = E^Q\{\chi(x_t=x) R_t | Y_t\}$$

where Q and $\{R_t\}$ are as in the previous section.

Theorem 3.1.2 The UCD vector q_t satisfies the stochastic differential equation

$$dq_t = A^* q_t dt + \sum_{i=1}^r (G_i - D) q_{t-} (dy_t(i) - \delta(i) dt)$$

where $q_0 = p_0$, $G_i = \text{diag}(g_1(x_1), \dots, g_1(x_N))$, and D is the $N \times N$ zero matrix when δ is the zero vector or the $N \times N$ identity matrix when δ is a vector of ones.

Proof Since $\sigma_t(\phi_t) = \phi^* q_t$, we simply change notation in Theorem 2.2.2 to get

$$d(\phi^* q_t) = \phi^* A^* q_t dt + \sum_{i=1}^r (\phi^* G_i q_{t-} - \phi^* (D) q_{t-}) (dy_t(i) - \delta(i) dt)$$

where we use the fact that $\sigma_{t-}(\phi_{t-} g_t) = \sigma_{t-}(\phi_{t-} g_{t-})$. Since $\phi \in B(S)$ is arbitrary, the result follows. ■

Of course, the normalized conditional density ρ_t can be computed once q_t is known. Specifically,

$$\rho_t(x) = \frac{q_t(x)}{\sum_{x \in S} q_t(x)}.$$

3.2 Computational considerations

When $y_t = N_t$ is a conditional Poisson process with intensity $g(x_t) = \lambda(x_t)$, the equation in Theorem 3.1.2 becomes

$$q_t = q_0 + \int_0^t A^* q_s ds + \int_0^t \sum_{i=1}^r (G_i - I) q_{s-} (dy_s(i) - ds) .$$

Let $T_0 = 0$ and $T_k = \inf\{t > T_{k-1} : \Delta y_t(i) = 1 \text{ for some } i = 1, \dots, r\}$. Since the components of y have no common jumps a.s., the above SDE can be evaluated as follows.

- i) If T_k is a jump time corresponding to $y(i)$, then $\Delta q_{T_k} = (G_i - I) q_{T_k-}$ or $q_{T_k} = G_i q_{T_k-}$.
- ii) If $T_k < t < T_{k+1}$, then q_t satisfies the initial value problem $\frac{dq_t}{dt} = \left(A^* - \sum_{i=1}^r (G_i - I) \right) q_t$ with initial value q_{T_k} .

When $y_t = \int_0^t h(x_s) + w_t$ where w_t is SBM, the equation in Theorem 3.1.2 becomes

$$q_t = q_0 + \int_0^t A^* q_s ds + \int_0^t \sum_{i=1}^r G_i q_{s-} dy_s(i) .$$

Evaluating this SDE is problematical in that y_t is only a mathematical model for the actual integrable variation observation \tilde{y}_t . So, we must be able to choose a version of $\pi(y) = E\{\phi(x_t) | Y_t\}$ with "nice" properties. Specifically, we want $\pi: C_{\mathbb{R}^r}[0, t] \rightarrow \mathbb{R}$ to be continuous where $C_{\mathbb{R}^r}[0, t]$ is the space

of continuous \mathbf{R}^r -valued functions endowed with the sup norm. Clark in [18] discusses this robustness property in detail and derives an algorithm for computing a robust version of $\rho_t(\mathbf{x}) = \mathbb{E}^P\{\chi(\mathbf{x}_t = \mathbf{x}) | \mathbf{Y}_t\}$. The algorithm involves solving an equation that is a deterministic linear differential equation for each observation path.

We will heuristically develop this algorithm. The complete proof is in [18]. Suppose $\{\xi_k^N\}$ is a Markov chain with transition matrix $\Pi_N = \Pi(\Delta)$ where $\Delta = \frac{T}{N}$ and $\Pi(t)$ is the transition matrix of our finite state Markov process $\{\mathbf{x}_t\}$. Define

$$R_t = \exp \left[\sum_{i=1}^r (G_i Y_t(i) - \frac{1}{2} G_i^2 t) \right]$$

$$G_N(k) = R_{(k+1)\Delta} R_{k\Delta}^{-1}$$

$$\mathbf{r}_k^N = R_{k\Delta}^{-1} \mathbf{q}_k^N$$

where $\mathbf{q}_t^N(\mathbf{x})$ is the UCD of $\mathbf{x}_t^N = \xi_{[t/N]}^N$. In Subsection IV.3.1 it is shown that $\mathbf{q}_{k+1}^N = \Pi_N^* G_N(k) \mathbf{q}_k^N$. In terms of $\{\mathbf{r}_k^N\}$, this equation can be rewritten as

$$\mathbf{r}_{k+1}^N = R_{(k+1)\Delta}^{-1} \Pi_N^* R_{(k+1)\Delta} \mathbf{r}_k^N \quad \text{or as} \quad \frac{\Delta \mathbf{r}_k^N}{\Delta} = R_{(k+1)\Delta}^{-1} \mathbf{A}_N^* R_{(k+1)\Delta} \mathbf{r}_k^N$$

where $A_N^* = (\Pi_N^* - I) \frac{1}{\Delta}$. This is the discrete form of Clark's robust equation:

$$(9) \quad r_t' = R_t^{-1} A^* R_t r_t, \quad r_0 = p_0$$

where $r_t = R_t^{-1} q_t$. So, the conditional density satisfies

$$\rho_t = \frac{R_t r_t}{\sum_{x \in S} R_t(x, x) r_t(x)}.$$

3.3 Diffusion signal

When the signal process is a diffusion satisfying the appropriate conditions, equations analogous to those above for computing the UCD can be obtained. A full account of these results is given by Pardoux in [17]. We will briefly summarize his work.

Again, we begin by assuming the signal $\{x_t\}$ is unobservable and derive Kolmogorov's forward equation. Let $\{x_t\}$ be a diffusion satisfying

$$dx_t = f(x_t) dt + \sigma(x_t) dv_t$$

where the coefficients have the following properties (actually Pardoux allows the diffusion to be time variant).

i) $f: \mathbb{R}^m \rightarrow \mathbb{R}^m$ is Lipschitz continuous and bounded.

ii) $\sigma: \mathbb{R}^m \rightarrow \mathbb{R}^m \times \mathbb{R}^m$ is Lipschitz continuous, bounded, $\frac{\partial \sigma_{ij}}{\partial x_j}$ is

bounded for $i, j=1, \dots, m$, and $\exists \alpha > 0$ such that

$$a(x) = \sigma(x) \sigma^*(x) \geq \alpha I \quad \forall x \in \mathbb{R}^m.$$

We also assume that x_0 has a density p_0 with $p_0 \in L^2(\mathbb{R}^m)$, and let A^* denote the adjoint of the generator A operating on the Sobolev space H^1 . Then the partial differential equation

$$\frac{\partial p_t}{\partial s} = A^* p_t \text{ with initial value } p_0$$

has a unique solution in $L^2_{H^1}[0, T] \cap C_{L^2(\mathbb{R}^m)}[0, T]$. It can be shown that the solution p_t is the density of the law of x_t for the given initial value x_0 .

Pardoux takes a similar approach for obtaining a stochastic partial differential equation (SPDE) satisfied by the UCD of x_t . Assume the above conditions on $\{x_t\}$ and the same conditions on $\{y_t\}$ as in Section 2. Then the SPDE

$$dq_t = A^* q_t dt + (g_t - \delta)^* q_{t-} (dy_t - \delta dt) \text{ with } q_0 = p_0$$

has a unique solution in the appropriate space. The space depends on whether $y_t = N_t$ or $y_t = \int_0^t g_s ds + w_t$ (see [17]). It can be shown that the solution q_t is the UCD of x_t .

Finally, we mention that under more restrictive conditions on x and y , Clark in [18] obtains a robust algorithm for computing q_t . This algorithm involves solving a partial differential equation that is the diffusion process analog of equation (9) in Subsection 3.2.

CHAPTER IV - APPROXIMATE FILTERS FOR DIFFUSIONS

In many filtering applications the state process is a diffusion, but even computing the unnormalized conditional density is difficult in this situation. When the state process is a diffusion, to find the UCD we must solve either a stochastic partial differential equation or a complicated deterministic partial differential equation as discussed in Section III.3. For this reason much research has been devoted to finding filters that in some sense approximate the optimal filter of a partially observed diffusion.

In this chapter we present a general theory for developing filtering algorithms that give approximations to the optimal filter of a partially observed diffusion and that are computationally practical using a digital computer. The theory is a synthesis of the ideas of several people and these ideas are cited in particular examples throughout the chapter. To develop an algorithm we proceed in three stages. First, we must construct a sequence of finite state processes that will converge in distribution to the diffusion signal process. Then, we construct filters for the sequence of finite state processes and show that this sequence of filters will converge to the optimal diffusion filter in $L^1(P)$ uniformly on $[0, T]$. Finally, we derive numerical methods for computing the

approximate filters.

1. Finite State Diffusion Approximations

Now, we assume that the state process of our partially observed system is a diffusion

$$x_t = x_0 + \int_0^t f(x_s) ds + \int_0^t \sigma(x_s) dw_s$$

where $f: \mathbb{R}^m \rightarrow \mathbb{R}^m$ and $\sigma: \mathbb{R}^m \rightarrow \mathbb{R}^m \times \mathbb{R}^m$ are Lipschitz continuous with bounded components and $t \in [0, T]$.

1.1 Markov chain approximations

First, we define a discrete time and space grid for each $N=1, 2, \dots$. The time grid will be $\{k\Delta: k=0, 1, \dots, N\}$ where $\Delta = \frac{T}{N}$. The space grid will be $S_N = \{x \in \mathbb{R}^m: x = H_N z \text{ for some vector } z \in \mathbb{Z}^m\}$ where $H_N = \text{diag}(h_{N1}, \dots, h_{Nm})$ and $h_{Ni} \rightarrow 0$ as $N \rightarrow \infty$ for all $i=1, \dots, m$. For example, if $h_{Ni} = \frac{1}{N}$ for all $i=1, \dots, m$, then S_N consists of all m -tuples with coordinates that are integer multiples of $\frac{1}{N}$.

Suppose for each N , $\{\xi_k^N\}$ is an S_N -valued Markov chain with transition matrix (perhaps infinite dimensional) Π_N with components $\pi_N(x, y)$ for all $x, y \in S_N$. Define the process $\{x_t^N\}$ by $x_t^N = \xi_{[t/\Delta]}^N$. So, $\{x_t^N\}$ is a discrete state process on $[0, T]$. We wish to establish conditions on the transition matrices $\{\Pi_N\}$ that guarantee $x_t^N \rightarrow x_t$. Here we think of $\{x_t^N\}, x$ as $D_{\mathbb{R}^m}[0, T]$ -valued random variables and apply a corollary of Theorem

II.3.3.1 to obtain the required conditions.

Theorem 1.1.1 Suppose the processes $\{x^N\}$, x are as described above and that the distributions of the sequence $\{x_0^N\}$ converge weakly to the distribution of x_0 . Then, $x^N \rightarrow x$ when the following conditions are satisfied.

- 1) $\sum_{y \in S_N} (y-x) \pi_N(x, y) = f(x) \Delta + r_1^N(x)$ where
 $\sup_{x \in S_N} (r_1^N(x)) = o(\Delta)$.
- 2) $\sum_{y \in S_N} (y-x) (y-x)^* \pi_N(x, y) = a(x) \Delta + r_2^N(x)$ where
 $\sup_{x \in S_N} (r_2^N(x)) = o(\Delta)$ and $a(x) = \sigma(x) \sigma^*(x)$.
- 3) $\exists c > 0$ such that $\forall N, x \sum_{|y-x| > ch_N} \pi_N(x, y) = 0$ where
 $h_N = \max(h_{N1}, \dots, h_{Nm})$.

Proof See Corollary 4.2, p. 355 in [8]. ■

We should note that condition 3) is more restrictive than in the reference for the proof, but it is reasonable for computationally practical approximations.

1.2 Continuous time, discrete state Markov approximations

Notice that the approximating processes $\{x^N\}$ defined above are not Markov. However, given a Markov chain sequence

$\{\xi_k^N\}$ with transition matrices $\{\Pi_N\}$ satisfying the conditions of Theorem 1.1.1, we can also construct a sequence $\{x^N\}$ of discrete state Markov processes that will converge to x in distribution.

For each N , let $\{\Delta\tau_k^N\}$ be an iid sequence of transition times, exponentially distributed, with mean $\Delta = \frac{T}{N}$, and independent of $\{\xi_k^N\}$. Define

$$x_t^N = \xi_{N_t}^N \text{ where } N_t = \max \left\{ k : \sum_{i=0}^{k-1} \Delta\tau_i^N \leq t \right\}.$$

Then for each $t \in [0, T]$, the components of the transition matrix $\Pi_N(t)$ of x_t^N satisfy:

$$\begin{aligned} \Pi_N(t)(x, y) &= P(x_t^N = y | x_0^N = x) = \sum_{k=0}^{\infty} P(N_t = k) P(\xi_k^N = y | \xi_0^N = x) \\ &= \sum_{k=0}^{\infty} \frac{(t/\Delta)^k e^{-\frac{t}{\Delta}}}{k!} \Pi_N^k(x, y) \end{aligned}$$

where $\Pi_N^k(x, y)$ is the (x, y) component Π_N^k .

So,

$$\Pi_N(t) = \sum_{k=0}^{\infty} \frac{(t/\Delta)^k e^{-\frac{t}{\Delta}}}{k!} \Pi_N^k = e^{-\frac{t}{\Delta}} \exp\left(\frac{t}{\Delta} \Pi_N\right) = \exp(A_N t)$$

where $A_N = (\Pi_N - I) \frac{1}{\Delta}$ is the generator of $\{x_t^N\}$.

Theorem 1.2.1 Suppose the Markov processes $\{x^N\}$, x are as described above and that the distributions of the sequence $\{x_0^N\}$ converge weakly to the distribution of x_0 . Then $x^N \Rightarrow x$.

Proof We will use Theorem II.3.3.1. Define for $N_t > 0$,

$$B_t^N = \sum_{i=0}^{N_t-1} E_i(\Delta x_i^N)$$

$$A_t^N = \sum_{i=0}^{N_t-1} \left[E_i(\Delta x_i^N \Delta x_i^{N*}) - E_i(\Delta x_i^N) E_i^*(\Delta x_i^N) \right]$$

$$= \sum_{i=0}^{N_t-1} E_i \left[(\Delta x_i^N - E_i(\Delta x_i^N)) (\Delta x_i^N - E_i(\Delta x_i^N))^* \right]$$

where E_i denotes the conditional expectation given $x_{t_i}^N = x_i^N$ with

$\Delta x_i^N = x_{i+1}^N - x_i^N$. Also, assume $B_t^N = 0$ and $A_t^N = 0$ for all t with $N_t = 0$.

We must check that conditions (i)-(vii) are satisfied in

Theorem II.3.3.1. Note that $\mathcal{F}_t^N = \sigma(x_s^N, B_s^N, A_s^N; s \leq t) = \sigma(x_s^N; s \leq t)$.

To simplify notation we assume $x_0 = x_0^N = 0$.

i) We show that $m^N = x^N - B^N$ is an $\{\mathcal{F}_t^N\}$ -martingale. Obviously, m_t^N is \mathcal{F}_t^N -measurable for $t \in [0, T]$. Condition (3) on Π_N gives

$$E\{|x_t^N|\} \leq E\left\{\sum_{i=0}^{N_t-1} |\Delta x_i^N|\right\} \leq ch_N E\{N_t\} = ch_N \frac{t}{\Delta}$$

and

$$E\{|B_t^N|\} \leq E\left\{\sum_{i=0}^{N_t-1} E_i(|\Delta x_i^N|)\right\} \leq ch_N E\{N_t\} = ch_N \frac{t}{\Delta}.$$

So, m_t^N is integrable for all $t \in [0, T]$. Also,

$$\begin{aligned}
 & E \left\{ (x_t^N - B_t^N) - (x_s^N - B_s^N) \mid \mathcal{F}_s^N \right\} \\
 &= \sum_{k=1}^{\infty} E \left\{ \left[(x_t^N - B_t^N) - (x_s^N - B_s^N) \right] \chi(N_t - N_s = k) \mid \mathcal{F}_s^N \right\} \\
 &= \sum_{k=1}^{\infty} E \left\{ \chi(N_t - N_s = k) \sum_{i=N_s}^{N_s+k-1} (\Delta x_i^N - E_1(\Delta x_i^N)) \mid \mathcal{F}_s^N \right\} \\
 &= \sum_{k=1}^{\infty} P(N_t - N_s = k) E \left\{ \sum_{i=N_s}^{N_s+k-1} (\Delta x_i^N - E_1(\Delta x_i^N)) \mid \mathcal{F}_s^N \right\}
 \end{aligned}$$

($N_t - N_s$ independent of \mathcal{F}_s^N and of Δx_i^N)

$$= \sum_{k=1}^{\infty} \sum_{i=N_s}^{N_s+k-1} P(N_t - N_s = k) E \left\{ \Delta x_i^N - E_1(\Delta x_i^N) \mid \mathcal{F}_s^N \right\}$$

(N_s is \mathcal{F}_s^N measurable)

$$= \sum_{k=1}^{\infty} \sum_{i=N_s}^{N_s+k-1} P(N_t - N_s = k) \left(E \left\{ \Delta x_i^N \mid \mathcal{F}_s^N \right\} - E \left\{ \Delta x_i^N \mid \mathcal{F}_s^N \right\} \right)$$

$$= 0.$$

Thus, m_t^N is an $\{\mathcal{F}_t^N\}$ -martingale.

ii) Next we show that $m^N(m^N)^* - A^N$ is an $\{\mathcal{F}_t^N\}$ -martingale.

Again, measurability is obvious and integrability follows

from condition (3) on Π_N . First, we have

$$\begin{aligned}
& \mathbb{E} \left\{ m_t^N (m_t^N)^* - m_s^N (m_s^N)^* \mid \mathcal{F}_s^N \right\} \\
&= \mathbb{E} \left\{ \sum_{i=1}^{N_t-1} (\Delta x_i^N - E_i(\Delta x_i^N)) \sum_{j=1}^{N_t-1} (\Delta x_j^N - E_j(\Delta x_j^N))^* \right. \\
&\quad \left. - \sum_{i=1}^{N_t-1} (\Delta x_i^N - E_i(\Delta x_i^N)) \sum_{j=1}^{N_t-1} (\Delta x_j^N - E_j(\Delta x_j^N))^* \mid \mathcal{F}_s^N \right\} \\
&= \mathbb{E} \left\{ \sum_{i=N_s}^{N_t-1} \sum_{j=N_s}^{N_t-1} (\Delta x_i^N - E_i(\Delta x_i^N)) (\Delta x_j^N - E_j(\Delta x_j^N))^* \mid \mathcal{F}_s^N \right\} \\
&= \sum_{k=1}^{\infty} \sum_{i=N_s}^{N_s+k-1} \sum_{j=N_s}^{N_s+k-1} \mathbb{E} \left\{ (\Delta x_i^N - E_i(\Delta x_i^N)) (\Delta x_j^N - E_j(\Delta x_j^N))^* \mid \mathcal{F}_s^N \right\} P(N_t - N_s = k)
\end{aligned}$$

(as in i))

$$= \sum_{k=1}^{\infty} \sum_{i=N_s}^{N_s+k-1} \mathbb{E} \left\{ (\Delta x_i^N - E_i(\Delta x_i^N)) (\Delta x_i^N - E_i(\Delta x_i^N))^* \mid \mathcal{F}_s^N \right\} P(N_t - N_s = k) .$$

But, we also have

$$\begin{aligned}
& \mathbb{E} \left\{ A_t^N - A_s^N \mid \mathcal{F}_s^N \right\} = \mathbb{E} \left\{ \sum_{i=N_s}^{N_t-1} E_i \left[(\Delta x_i^N - E_i(\Delta x_i^N)) (\Delta x_i^N - E_i(\Delta x_i^N))^* \right] \mid \mathcal{F}_s^N \right\} \\
&= \sum_{k=1}^{\infty} \sum_{i=N_s}^{N_s+k-1} \mathbb{E} \left\{ E_i \left[(\Delta x_i^N - E_i(\Delta x_i^N)) (\Delta x_i^N - E_i(\Delta x_i^N))^* \right] \mid \mathcal{F}_s^N \right\} P(N_t - N_s = k) \\
&= \sum_{k=1}^{\infty} \sum_{i=N_s}^{N_s+k-1} \mathbb{E} \left\{ (\Delta x_i^N - E_i(\Delta x_i^N)) (\Delta x_i^N - E_i(\Delta x_i^N))^* \mid \mathcal{F}_s^N \right\} P(N_t - N_s = k) .
\end{aligned}$$

$$\text{So, } E\left\{ \left[m_t^N (m_t^N)^* - A_t^N \right] - \left[m_s^N (m_s^N)^* - A_s^N \right] \middle| \mathcal{F}_s^N \right\} = 0.$$

iii) Follows immediately from condition (3) on Π_N .

iv) Follows immediately from condition (1) on Π_N and the boundedness of the components of f .

v) Follows immediately from conditions (1) and (2) on Π_N and the boundedness of the components of f and σ .

vi) Now we want to show that $\sup_{t \leq T} \left| B_t^N(j) - \int_0^t f_j(x_s^N) ds \right| \rightarrow 0$ in probability for each $j=1, \dots, m$. In the following proof we drop the component index j , so when we write B_t^N we mean $B_t^N(j)$ and when we write $f(x_s^N)$ we mean $f_j(x_s^N)$. This should cause no confusion since j is arbitrary but fixed throughout. Using condition (1) on Π_N we have

$$\begin{aligned} \sup_{t \leq T} \left| B_t^N - \int_0^t f(x_s^N) ds \right| &\leq \sup_{t \leq T} \left| \sum_{i=0}^{N_t} f(x_i^N) (\Delta - \Delta \tau_i^N) \right| \\ &+ \sup_{t \leq T} \left| f(x_{N_t}^N) (\tau_{N_t+1}^N - t) \right| \\ &+ \sup_{t \leq T} \left| \left(\sum_{i=0}^{N_t-1} r_1^N(x_i^N) \right) - f(x_{N_t}^N) \Delta \right|. \end{aligned}$$

We show that each term on the right goes to zero in probability.

- a) The boundedness of f and the assumption in condition (1) on r_1^N clearly imply convergence of the last term.

$$\begin{aligned}
 \text{b)} \quad P\left(\sup_{t \leq T} \left| f(x_{N_t}^N) (\tau_{N_t+1}^N - t) \right| > \epsilon\right) &\leq P\left(\|f\|_\infty \sup_{1 \leq N_t} (\Delta \tau_1^N) > \epsilon\right) \\
 &= P\left(\sup_{1 \leq N_t} (\Delta \tau_1^N)^2 > \frac{\epsilon^2}{\|f\|_\infty^2}\right) \\
 &\leq P\left(\sum_{i=0}^{N_T} (\Delta \tau_1^N)^2 > \frac{\epsilon^2}{\|f\|_\infty^2}\right) \\
 &\leq \frac{\|f\|_\infty^2}{\epsilon^2} E\left\{\sum_{i=0}^{N_T} (\Delta \tau_1^N)^2\right\} \\
 &= \frac{\|f\|_\infty^2}{\epsilon^2} N(2\Delta^2) \\
 &= \frac{2\|f\|_\infty^2 T \Delta}{\epsilon^2}.
 \end{aligned}$$

Of course, the last two steps use the Chebychev inequality and Wald's equality respectively.

- c) To show convergence of the first term we first show that $I_t^N = \sum_{i=0}^{N_t} f(x_i^N) (\Delta - \Delta \tau_i^N)$ is an $\{\mathcal{F}_t^N\}$ -martingale. Clearly I_t^N is \mathcal{F}_t^N -measurable and the integrability of I_t^N follows easily

since f is bounded. Also,

$$\begin{aligned}
& \mathbb{E} \left\{ \sum_{i=N_s+1}^{N_t} f(x_i^N) (\Delta - \Delta \tau_i^N) \mid \mathcal{F}_s^N \right\} \\
&= \sum_{k=1}^{\infty} \mathbb{E} \left\{ \chi(N_t - N_s = k) \sum_{i=N_s+1}^{N_s+k} f(x_i^N) (\Delta - \Delta \tau_i^N) \mid \mathcal{F}_s^N \right\} \\
&= \sum_{k=1}^{\infty} \sum_{i=N_s+1}^{N_s+k} \mathbb{E} \left\{ \chi(N_t - N_s = k) (\Delta - \Delta \tau_i^N) \mid \mathcal{F}_s^N \right\} \mathbb{E} \left\{ f(x_i^N) \mid \mathcal{F}_s^N \right\} \\
&= \sum_{i=N_s+1}^{\infty} \mathbb{E} \left\{ f(x_i^N) \mid \mathcal{F}_s^N \right\} \sum_{k=i-N_s}^{\infty} \mathbb{E} \left\{ \chi(N_t - N_s = k) (\Delta - \Delta \tau_i^N) \mid \mathcal{F}_s^N \right\} \\
&= \sum_{i=N_s+1}^{\infty} \mathbb{E} \left\{ f(x_i^N) \mid \mathcal{F}_s^N \right\} \left[\mathbb{E} \left\{ \Delta - \Delta \tau_i^N \mid \mathcal{F}_s^N \right\} - \mathbb{E} \left\{ (\Delta - \Delta \tau_i^N) \chi(N_t < i) \mid \mathcal{F}_s^N \right\} \right] \\
&= \sum_{i=N_s+1}^{\infty} \mathbb{E} \left\{ f(x_i^N) \mid \mathcal{F}_s^N \right\} \mathbb{E} \left\{ \Delta - \Delta \tau_i^N \mid \mathcal{F}_s^N \right\} P(N_t \geq i) = 0.
\end{aligned}$$

Now, since I_t^N is a martingale, we can show $\mathbb{E}\{(I_T^N)^2\}$ converges to zero to get convergence in probability for the first term. (We can assume $N_T \geq 1$ a.s., otherwise redefine x_T on $\{N_T = 0\}$ to get a jump at time T .) We have

$$\begin{aligned}
& \mathbb{E} \left\{ \left(\sum_{i=0}^{N_T} f(x_i^N) (\Delta - \Delta \tau_i^N) \right)^2 \right\} \\
&= \sum_{k=1}^{\infty} \sum_{j=0}^k \sum_{i=0}^k \mathbb{E} \left\{ f(x_i^N) f(x_j^N) (\Delta - \Delta \tau_i^N) (\Delta - \Delta \tau_j^N) \chi(N_T = k) \right\}
\end{aligned}$$

$$\begin{aligned}
&= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} E\{f(x_i^N) f(x_j^N)\} \sum_{k=\max(i,j)}^{\infty} E\{(\Delta - \Delta\tau_i^N)(\Delta - \Delta\tau_j^N) \chi(N_T=k)\} \\
&= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} E\{f(x_i^N) f(x_j^N)\} \left[E\{(\Delta - \Delta\tau_i^N)(\Delta - \Delta\tau_j^N)\} \right. \\
&\quad \left. - E\{(\Delta - \Delta\tau_i^N)(\Delta - \Delta\tau_j^N) \chi(N_T < \max(i,j))\} \right] \\
&= \sum_{i=1}^{\infty} E\{f^2(x_i^N)\} E\{(\Delta - \Delta\tau_i^N)^2\} P(N_T \geq i) \\
&\leq \|f\|_{\infty}^2 \Delta^2 E\{N_T\} \\
&= \|f\|_{\infty}^2 \Delta T.
\end{aligned}$$

vii) The computations to show $\sup_{t \leq T} \left| A_t^N(i,j) - \int_0^t a_{i,j}(x_s^N) ds \right| \rightarrow 0$ in probability for each $i, j=1, \dots, m$ are identical to those in part (vi). ■

1.3 Example

We have seen that for a given sequence of transition matrices satisfying conditions (1)-(3) of Theorem 1.1.1, we can construct two sequences of discrete state processes that will converge in distribution to the diffusion x . For one of the sequences the processes have fixed deterministic transition times and for the other sequence the processes are Markov with random transition times. So, now we must derive

an appropriate sequence of transition matrices. An extensive exposition on how to derive transition matrices for diffusion approximations is given in [12]. Here we give one example of a procedure for deriving these transition matrices. To simplify notation we will assume that the diffusion x is one dimensional, i.e., $m=1$. The procedure for $m>1$ is given in [12].

First, we note that $w: \mathbb{R} \times [0, T] \rightarrow \mathbb{R}$ defined $w(x, t) = E\{\phi(x_T) | x_t = x\}$ for given $\phi \in D_A$ satisfies the backward Kolmogorov equation

$$\frac{\partial w(x, t)}{\partial t} + Aw(x, t) = 0, \quad w(x, T) = \phi(x)$$

where A is the diffusion generator $f \frac{\partial}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2}{\partial x^2}$. Then, we construct a finite difference approximation to the above parabolic pde. Use

$$\frac{1}{\Delta} [u(x, t+\Delta) - u(x, t)] \text{ to approximate } \frac{\partial u(x, t)}{\partial t},$$

$$\begin{cases} \frac{1}{h_N} [u(x+h_N, t+\Delta) - u(x, t+\Delta)] & \text{if } f(x) \geq 0 \\ \frac{1}{h_N} [u(x, t+\Delta) - u(x-h_N, t+\Delta)] & \text{if } f(x) < 0 \end{cases} \text{ to approximate } \frac{\partial u(x, t)}{\partial x},$$

$$\frac{1}{h_N^2} [u(x+h_N, t+\Delta) + u(x-h_N, t+\Delta) - 2u(x, t+\Delta)] \text{ to approximate } \frac{\partial^2 u(x, t)}{\partial x^2}.$$

If $w^N(x, t)$ is a solution to the resulting finite difference equation, then

$$\begin{aligned} & \frac{f^+(x)}{h_N} [w^N(x+h_N, t+\Delta) - w^N(x, t+\Delta)] - \frac{f^-(x)}{h_N} [w^N(x, t+\Delta) - w^N(x-h_N, t+\Delta)] \\ & + \frac{\sigma^2(x)}{2h_N^2} [w^N(x+h_N, t+\Delta) + w^N(x-h_N, t+\Delta) - 2w^N(x, t+\Delta)] \\ & = \frac{1}{\Delta} [w^N(x, t) - w^N(x, t+\Delta)] \end{aligned}$$

or

$$\begin{aligned} (1) \quad w^N(x, t) &= w^N(x, t+\Delta) \left[1 - \frac{|f(x)|\Delta}{h_N} - \frac{\sigma^2(x)\Delta}{h_N^2} \right] \\ &+ w^N(x+h_N, t+\Delta) \left[\frac{f^+(x)\Delta}{h_N} + \frac{\sigma^2(x)\Delta}{2h_N^2} \right] \\ &+ w^N(x-h_N, t+\Delta) \left[\frac{f^-(x)\Delta}{h_N} + \frac{\sigma^2(x)\Delta}{2h_N^2} \right] \end{aligned}$$

where $w^N(x, T) = \phi(x)$. Take $\{\xi_k^N\}$ to be the S_N -valued Markov chain with transition matrix Π_N having components

$$\begin{aligned} \Pi_N(x, x) &= 1 - \frac{|f(x)|\Delta}{h_N} - \frac{\sigma^2(x)\Delta}{h_N^2} \\ \Pi_N(x, x+h_N) &= \frac{f^+(x)\Delta}{h_N} + \frac{\sigma^2(x)\Delta}{2h_N^2} \\ \Pi_N(x, x-h_N) &= \frac{f^-(x)\Delta}{h_N} + \frac{\sigma^2(x)\Delta}{2h_N^2} \end{aligned}$$

$$\pi_N(x, y) = 0 \text{ for } y \neq x - h_N, x, x + h_N.$$

Then, it is easy to check that Π_N satisfies conditions (1)-(3) of Theorem 1.1.1.

Kushner in [13] worked with the fixed transition time process $x_t^N = \xi_{[t/\Delta]}^N$ corresponding to the above transition matrix Π_N . He noted that $w^N(x, t)$, the solution to the finite difference equation, satisfies $w^N(x, 0) = E\{\phi(x_T^N) | x_0^N = x\}$. To see this, use equation (1) to get $w^N(x, t) = E\{w^N(x_{t+\Delta}^N, t+\Delta) | x_t^N = x\}$ and recall that $w^N(x, T) = \phi(x)$.

Runggaldier and Di Masi in [14] worked with the Markov, random transition time process x^N corresponding to the above transition matrix Π_N . Using Subsection 1.2 we can see that the generator $A_N = (\Pi_N - I) \frac{1}{\Delta}$ of x^N has components

$$a_N(x, x) = -\left(\frac{|f(x)|}{h_N} + \frac{\sigma^2(x)}{h_N^2} \right)$$

$$a_N(x, x+h_N) = \frac{f^+(x)}{h_N} + \frac{\sigma^2(x)}{2h_N^2}$$

$$a_N(x, x-h_N) = \frac{f^-(x)}{h_N} + \frac{\sigma^2(x)}{2h_N^2}$$

$$a_N(x, y) = 0 \text{ for } y \neq x - h_N, x, x + h_N.$$

1.4 Truncation of processes

The diffusion approximations we have constructed so far are discrete state processes. But, for obvious computational reasons we want finite state approximations. This is achieved by truncating the diffusion and the approximating processes. Specifically, let $G \subset \mathbb{R}^m$ be open and bounded and let $G_N = G \cap S_N$.

Define the stopping times

$$\begin{aligned}\tau &= \begin{cases} T & \text{if } x_t \in G \text{ for all } t \in [0, T] \\ \inf\{t: x_t \notin G\} & \text{otherwise} \end{cases} \\ \tau^N &= \begin{cases} T & \text{if } x_t^N \in G_N \text{ for all } t \in [0, T] \\ \inf\{t: x_t^N \notin G_N\} & \text{otherwise} \end{cases},\end{aligned}$$

and let $\bar{x}_t = x_{t \wedge \tau}$, $\bar{x}_t^N = x_{t \wedge \tau^N}^N$ represent the corresponding stopped processes.

Of course, we want $\bar{x}^N \Rightarrow \bar{x}$ whenever $x^N \Rightarrow x$. Let \bar{G} be the closure of G and define

$$\tau' = \begin{cases} T & \text{if } x_t \in \bar{G} \text{ for all } t \in [0, T] \\ \inf\{t: x_t \notin \bar{G}\} & \text{otherwise} \end{cases}.$$

Proposition 1.4.1 If $P(\tau = \tau') = 1$ and $x^N \Rightarrow x$, then $\bar{x}^N \Rightarrow \bar{x}$.

Proof (See Theorem 2 in [14].) ■

For the remainder of the paper when we talk about the

diffusion x or an approximating sequence $\{x^N\}$, we will assume that these are the truncated processes described above.

Provided G is properly chosen, this assumption is not very restrictive.

2. Filter Approximations

Consider the following situation. We are given processes $\{x^N\}, x, y$ defined on a probability space (Ω, \mathcal{F}, P) that are measurable w.r.t. the complete right continuous filtration $\{\mathcal{F}_t\}$ for $t \in [0, T]$. Assume $\{x^N\}$ is a sequence of finite state processes converging in distribution to the (truncated) diffusion x . Also, assume there exists a probability measure Q on \mathcal{F} with $P \ll Q$ and define $R_t = E^Q \left\{ \frac{dP}{dQ} \middle| \mathcal{F}_t \right\}$ for $t \in [0, T]$. Assume we are given a sequence of positive square integrable random variables $\{R_T^N(x^N, y)\}$ with $E^Q\{R_T^N\} = 1$, $R_t^N = E^Q\{R_T^N | \mathcal{F}_t\}$, and $\frac{dP^N}{dQ} = R_T^N$. Finally, for a time discretization process assume $\{\alpha^N\}$ is a sequence of $Y_t = \sigma\{y_s : s \leq t\}$ -measurable nondecreasing real valued processes with $\alpha_0^N = 0$ for all N .

With this setup the Kallianpur-Striebel formula gives us for $\phi \in C_b(\mathbb{R}^m)$:

$$E^P\{\phi(x_t) | Y_t\} = \frac{E^Q\{\phi(x_t) R_t | Y_t\}}{E^Q\{R_t | Y_t\}},$$

$$E^{P^N}\{\phi(x^N(\alpha_t^N)) | Y_t\} = \frac{E^Q\{\phi(x^N(\alpha_t^N)) R_t^N | Y_t\}}{E^Q\{R_t^N | Y_t\}}.$$

In Subsection 2.2 we will show that the sequence of filters, $\{E^{P^N}\{\phi(x^N(\alpha_t^N))|Y_t\}\}$ converges to the filter $E^P\{\phi(x_t)|Y_t\}$ in $L^1(P)$ uniformly on $[0, T]$ under appropriate conditions on $\{R^N\}$ and $\{\alpha^N\}$. But, first we provide some motivational examples.

2.1 Examples

In all of the examples, the process y is an observation process of the form:

$$y_t = \int_0^t g(x_s) ds + n_t$$

where g and n are described in Chapter III. So, as in Section III.2, under Q the process y is a standard process (Brownian motion or Poisson martingale) independent of x and $\{x^N\}$, $\{R_t\}$ is a (Q, \mathcal{F}_t) martingale given by one of the Girsanov theorems (Theorem II.2.2.5 or Theorem II.2.2.6), and the $D_{R^N}[0, T]$ distributions induced by x and $\{x^N\}$ under Q are the same as the $D_{R^N}[0, T]$ distributions induced by x and $\{x^N\}$ under P . Thus, $E^P\{\phi(x_t)|Y_t\}$ is the optimal mean square filter of $\phi(x_t)$ given the observations in Y_t and $E^{P^N}\{\phi(x^N(\alpha_t^N))|Y_t\}$ is an approximating filter. We give five examples of how to construct the approximating filters, i.e., how to choose $\{R^N\}$ and $\{\alpha^N\}$.

- i) Kushner in [13] considers the case where n is SBM. He uses fixed transition-time processes like those described

in Section 1.1 for the diffusion approximations $\{x^N\}$.

He takes $\alpha_t^N = [t/\Delta]\Delta$ and

$$R_t^N(x^N, y) = \prod_{k=0}^{[t/\Delta]-1} \exp \left[g^*(x_{k\Delta}^N) \Delta y_{k\Delta} - \frac{1}{2} g^*(x_{k\Delta}^N) g(x_{k\Delta}^N) \Delta \right]$$

where $\Delta y_{k\Delta} = y_{(k+1)\Delta} - y_{k\Delta}$.

- ii) Di Masi and Runggaldier in [14] also consider the case where n is SBM, but they use random transition-time Markov processes like those described in Section 1.2 for the diffusion approximations $\{x^N\}$. They take $\alpha_t^N = t$ and

$$R_t^N(x^N, y) = \exp \left(\int_0^t g^*(x_s^N) dy_s - \frac{1}{2} \int_0^t g^*(x_s^N) g(x_s^N) ds \right).$$

- iii) Goggin in [15] considers the same observation process and diffusion approximations as Kushner, but she works with different processes $\{R^N\}$ and $\{\alpha^N\}$. When the observation process $\{y_t\}$ is one dimensional, she takes $\alpha_t^N = N_t \Delta$ where $\Delta = \frac{T}{N}$, $N_t = \max\{k: \tau_k^N \leq t\}$, $\tau_0^N = 0$, and $\tau_k^N = \inf\{t > \tau_{k-1}^N: |y_t - y_{\tau_{k-1}^N}| > \sqrt{\Delta}\}$. For R_t^N she uses

$$R_t^N(x^N, y) = \prod_{k=0}^{N_t-1} (1 + g(x_{k\Delta}^N) \Delta y(\tau_k^N))$$

where $\Delta y(\tau_k^N) = y(\tau_{k+1}^N) - y(\tau_k^N)$. (In [15] she also discusses

the case where $\{y_t\}$ is r -dimensional, $r > 1$.)

- iv) When the observation is a Poisson process and the diffusion approximations have fixed transition-times, we can use $\alpha_t^N = [t/\Delta]\Delta$ and

$$R_t^N(x^N, y) = \prod_{k=0}^{[t/\Delta]-1} \prod_{i=1}^r [g_i(x_{k\Delta}^N)]^{\Delta y_{k\Delta}(i)} \exp[(1 - g_i(x_{k\Delta}^N))\Delta].$$

- v) When the observation is a Poisson process and the diffusion approximations are random transition-time Markov processes, we can use $\alpha_t^N = t$ and

$$R_t^N(x^N, y) = \exp\left(\int_0^t \sum_{i=1}^r \ln(g_i(x_s^N)) dy_s(i) + \int_0^t (1 - g_i(x_s^N)) ds\right).$$

2.2 Convergence

We show that $E^P\{\phi(x^N(\alpha_t^N)) | Y_t\} \rightarrow E^P\{\phi(x_t) | Y_t\}$ in $L^1(P)$ uniformly on $[0, T]$ for any $\phi \in C_b(\mathbb{R}^m)$ provided the following conditions are satisfied.

- 1) $R_t^N(x^N, y) \rightarrow R_t(x, y)$ in (Q) probability uniformly on $[0, T]$.
- 2) Under Q , $\{\sup_{t \leq T} R_t^N(x^N, y)\}$ is a uniformly integrable sequence.
- 3) $\alpha_t^N - t \rightarrow 0$ in (Q) probability uniformly on $[0, T]$.

Theorem 2.2.1 Assuming conditions 1)-3),

$$E^Q\{\phi(x^N(\alpha_t^N))R_t^N|Y_t\} \rightarrow E^Q\{\phi(x_t)R_t|Y_t\}$$

in $L^1(Q)$ uniformly on $[0, T]$ for any $\phi \in C_b(\mathbb{R}^m)$.

Proof This proof is a generalization of the discussion in [15] on p. 17.

$$\begin{aligned} & E^Q\left\{\left|E^Q\{\phi(x^N(\alpha_t^N))R_t^N|Y_t\} - E^Q\{\phi(x_t)R_t|Y_t\}\right|\right\} \\ & \leq E^Q\left\{\left|\phi(x^N(\alpha_t^N))R_t^N - \phi(x_t)R_t\right|\right\} \\ & \leq E^Q\left\{\left|\phi(x^N(\alpha_t^N)) - \phi(x(\alpha_t^N))\right|R_t^N\right\} \\ & \quad + E^Q\left\{\left|\phi(x(\alpha_t^N)) - \phi(x_t)\right|R_t^N\right\} + E^Q\left\{|\phi(x_t)||R_t^N - R_t|\right\} \\ & \leq E^Q\left\{\left[\phi(x^N(\alpha_t^N)) - \phi(x(\alpha_t^N))\right]^2\right\}^{\frac{1}{2}} E^Q\left\{(R_t^N)^2\right\}^{\frac{1}{2}} + \\ & \quad E^Q\left\{\left[\phi(x(\alpha_t^N)) - \phi(x_t)\right]^2\right\}^{\frac{1}{2}} E^Q\left\{(R_t^N)^2\right\}^{\frac{1}{2}} + \|\phi\|_{\infty} E^Q\left\{|R_t^N - R_t|\right\}. \end{aligned}$$

Conditions 1) and 2) guarantee that $E^Q\{|R_t^N - R_t|\} \rightarrow 0$ uniformly on $[0, T]$. So, the third term in the last inequality goes to zero uniformly on $[0, T]$.

Since $\{R_t^N\}$ is a (Q, \mathcal{F}_t) -martingale, $E^Q\{(R_t^N)^2\} \leq E^Q\{(R_T^N)^2\}$.

But, $E^Q\{(R_T^N)^2\} \rightarrow E^Q\{(R_T)^2\}$. Thus, $E^Q\{(R_t^N)^2\}$ is uniformly bounded for all N and all $t \in [0, T]$. Wlog we can assume $x^N \rightarrow x$ a.s. in the Skorohod topology. Then, since x is a continuous process, we have $|x^N(\alpha_t^N) - x(\alpha_t^N)| \rightarrow 0$ a.s. uniformly on $[0, T]$ (see [16] p. 112). So, the first term in the last inequality goes to zero uniformly on $[0, T]$, since ϕ is continuous and bounded.

To show that the second term in the last inequality goes to zero uniformly on $[0, T]$, we must show that

$E^Q\left\{\left[\phi(x(\alpha_t^N)) - \phi(x_t)\right]^2\right\} \rightarrow 0$ uniformly on $[0, T]$. For this suppose $\epsilon > 0$ is given. Choose $\gamma > 0$ such that $(\phi(x) - \phi(y))^2 < \epsilon$ whenever $|x - y| \leq \gamma$ and $x, y \in \bar{G}$ (recall \bar{G} is the closure of the truncation set G in Subsection 1.4). Then, choose $\delta > 0$ such that $Q\left(\sup_{|s-t| \leq \delta} |x_s - x_t| > \gamma\right) < \epsilon$. (This can be done since x is a truncated diffusion with bounded coefficients.) Finally, choose N_0 such that $Q\left(\sup_{t \leq T} |\alpha_t^N - t| > \delta\right) < \epsilon$ for all $N \geq N_0$. We have for $N \geq N_0$,

$$\begin{aligned} E^Q\left\{\left[\phi(x(\alpha_t^N)) - \phi(x_t)\right]^2\right\} &= E^Q\left\{\left[\phi(x(\alpha_t^N)) - \phi(x_t)\right]^2 \chi\left(\sup_{t \leq T} |\alpha_t^N - t| > \delta\right)\right\} \\ &\quad + E^Q\left\{\left[\phi(x(\alpha_t^N)) - \phi(x_t)\right]^2 \chi\left(\sup_{t \leq T} |\alpha_t^N - t| \leq \delta\right) \chi\left(\sup_{|s-t| \leq \delta} |x_s - x_t| > \gamma\right)\right\} \\ &\quad + E^Q\left\{\left[\phi(x(\alpha_t^N)) - \phi(x_t)\right]^2 \chi\left(\sup_{t \leq T} |\alpha_t^N - t| \leq \delta\right) \chi\left(\sup_{|s-t| \leq \delta} |x_s - x_t| \leq \gamma\right)\right\} \\ &\leq 4\|\phi\|_\infty^2 \epsilon + 4\|\phi\|_\infty^2 \epsilon + \epsilon \end{aligned}$$

$$= (8 \|\phi\|_{\infty}^2 + 1) \epsilon.$$

Theorem 2.2.2 Assuming conditions 1)-3),

$$E^P \left\{ \phi(x^N(\alpha_t^N)) | Y_t \right\} \rightarrow E^P \left\{ \phi(x_t) | Y_t \right\}$$

in $L^1(P)$ uniformly on $[0, T]$ for any $\phi \in C_b(\mathbb{R}^m)$.

Proof (See Theorem 3.5 in [15].)

Now we show that the examples of the last subsection satisfy conditions 1)-3). The conditions are verified for example iii) in [15]. To simplify notation we verify the conditions for the other examples when the observation process y is one dimensional. For all the other examples, $\alpha_t^N = [t/\Delta]\Delta$ or t and R_t^N, R_t can be written

$$R_t^N = \exp \left(\int_0^{\alpha_t^N} F(x_s^N) dz_s + \int_0^{\alpha_t^N} H(x_s^N) ds \right),$$

$$R_t = \exp \left(\int_0^t F(x_s) dz_s + \int_0^t H(x_s) ds \right)$$

where F, H are bounded continuous real-valued functions and $z_t = y_t - \int_0^t \delta ds$ as in Section III.2.

For R_t^N, R_t described above condition (iii) is immediate. Condition ii) is easy. Just note $E^Q \left\{ \sup_{t \leq T} (R_t^N)^2 \right\} \leq 4 E^Q \left\{ (R_T^N)^2 \right\}$ and $E^Q \left\{ (R_T^N)^2 \right\}$ is uniformly bounded because F, H are bounded. So, we

will consider condition i). Using the Skorohod representation and the continuity of $\{x_t\}$ we have $x_t^N \rightarrow x_t$ a.s. uniformly on $[0, T]$. Therefore, since H is continuous,

$$\int_0^t H(x_s^N) ds \rightarrow \int_0^t H(x_s) ds \text{ a.s. uniformly on } [0, T].$$

Also, $[F(x_t^N) - F(x_t)]^2 \rightarrow 0$ a.s. uniformly on $[0, T]$ which implies $E^Q\{[F(x_t^N) - F(x_t)]^2\} \rightarrow 0$ uniformly on $[0, T]$ since F is bounded. Thus,

$$E^Q\left\{\left[\int_0^t F(x_s^N) - F(x_s) dz_s\right]^2\right\} = \int_0^t E^Q\{[F(x_s^N) - F(x_s)]^2\} ds \rightarrow 0$$

uniformly on $[0, T]$. So, the exponents of $\{R_t^N\}$ converge to the exponent of R_t in (Q) probability uniformly on $[0, T]$. Condition 1) then follows easily.

3. Computational Considerations

For a given diffusion x and observation process y we have seen how to construct filter approximations in a variety of ways. The particular filter approximation chosen will usually depend on the computability of the filter approximation and on the efficiency of the computational algorithm. In this section we first describe algorithms for computing the approximate filters in the examples of Section 2. Then, we

discuss the relative efficiency of some of these algorithms using the computer simulations in the appendix.

3.1 Computational algorithms

In examples ii) and v) where each diffusion approximation x^N is a finite state Markov process, R^N satisfies the conditions of Section III.2. So, the UCE's $E^0\{\phi(x_t^N)R_t^N|Y_t\}$ satisfy the Zakai equation, and since x^N is a finite state Markov process, we can compute the UCD process using the differential equations in Subsection III.3.2.

In the other examples we can compute the UCD's recursively via a matrix multiplication algorithm. The key to this algorithm is the multiplicative nature of R^N . That is,

$$R_t^N(x^N, y) = \prod_{k=0}^{N_t-1} r_k^N(x_{k\Delta}^N, \Delta y_k^N).$$

Specifically, in example i)

$$N_t = [t/\Delta], \quad \Delta y_k^N = y_{(k+1)\Delta} - y_{k\Delta}, \quad \text{and}$$

$$r_k^N(x_{k\Delta}^N, \Delta y_k^N) = \exp[g^*(x_{k\Delta}^N)\Delta y_k^N - \frac{1}{2}g^*(x_{k\Delta}^N)g(x_{k\Delta}^N)\Delta];$$

in example iii)

$$N_t = \max\{k: \tau_k^N \leq t\}, \quad \tau_0^N = 0, \quad \tau_k^N = \inf\{t > \tau_{k-1}^N: |y_t - y_{\tau_{k-1}^N}| > \sqrt{\Delta}\},$$

$$\Delta y_k^N = y(\tau_{k+1}^N) - y(\tau_k^N), \text{ and } r_k^N(x_{k\Delta}^N, \Delta y_k^N) = (1 + g(x_{k\Delta}^N) \Delta y_k^N);$$

in example iv)

$$N_t = [t/\Delta], \quad \Delta y_k^N = y_{(k+1)\Delta} - y_{k\Delta}, \text{ and}$$

$$r_k^N(x_{k\Delta}^N, \Delta y_k^N) = \prod_{i=1}^r [g_i(x_{k\Delta}^N)]^{\Delta y_k^{(i)}} \exp[(1 - g_i(x_{k\Delta}^N))\Delta].$$

The algorithm is derived as follows.

Recall that in these examples $x_t^N = \xi_{[t/\Delta]}^N$ for an appropriate Markov chain ξ^N with state space S_N and transition matrix Π_N . Suppose $t \in [0, T]$ and $\phi \in B(S_N)$ are fixed. Also, denote $x_{k\Delta}^N$ by x_k^N and $r_k^N(x_k^N, \Delta y_k^N)$ by r_k^N . Then for $k = N_t, \dots, 1$ define a sequence $\{V_k(x_k^N)\}$ of random variables recursively by

$$\begin{aligned} V_{N_t}(x_{N_t}^N) &= \phi(x_{N_t}^N) \\ V_{k-1}(x_{k-1}^N) &= E^Q \{ r_k^N V_k | Y_t \vee x_{k-1}^N \} \end{aligned}$$

$h = \frac{1}{4}$ where $Y_t \vee x_{k-1}^N$ is the σ -algebra generated by $Y_t \cup \sigma\{x_{k-1}^N\}$.

Clearly, each $V_k(x_k^N)$ depends on the observations $\{y_s: 0 \leq s \leq t\}$ also.

Using the independence of x^N, y and the Markov property of ξ^N , we have

$$\begin{aligned} \text{a) } V_{N_t-1}(x_{N_t-1}^N) &= E^Q \{ r_{N_t}^N V_{N_t}(x_{N_t}^N) | Y_t \vee x_{N_t-1}^N \} \\ &= E^Q \{ r_{N_t}^N \phi(x_{N_t}^N) | Y_t \vee x_{N_t-1}^N \}; \end{aligned}$$

b) If $V_k(x_k^N) = E^Q \left\{ \left(\prod_{i=k}^{N_t-1} r_i^N \right) \phi(x_{N_t}^N) | Y_t V x_t^N \right\}$, then

$$\begin{aligned} V_{k-1}(x_{k-1}^N) &= E^Q \left\{ r_{k-1}^N E^Q \left\{ \left(\prod_{i=k}^{N_t-1} r_i^N \right) \phi(x_{N_t}^N) | Y_t V x_k^N \right\} | Y_t V x_{k-1}^N \right\} \\ &= E^Q \left\{ \left(\prod_{i=k-1}^{N_t-1} r_i^N \right) \phi(x_{N_t}^N) | Y_t V x_{k-1}^N \right\}. \end{aligned}$$

So, by induction $V_0(x_0^N) = E^Q \{ R_t^N \phi(x_{N_t}^N) | Y_t V x_0^N \}$.

Now, suppose S_N has d_N elements $\{x_1, \dots, x_{d_N}\}$. Define $G_N(k) = \text{diag}[r_k^N(x_1, \Delta y_k), \dots, r_k^N(x_{d_N}, \Delta y_k)]$, i.e., $G_N(k)$ is a $d_N \times d_N$ diagonal random matrix dependent on Δy_k . Also, denote by V_k the random vector $(V_k(x_1), \dots, V_k(x_{d_N}))$. Finally, denote the UCD of x_k^N by q_k^N for $k=0, 1, \dots, N_t$. Then, using the definition of $V_k(x_k^N)$ and the result in the last paragraph, we have (as always $*$ denotes the transpose operation):

$$\begin{aligned} \phi^* q_{N_t}^N &= E^Q \{ \phi(x_{N_t}^N) R_t^N | Y_t \} \\ &= V_0^* q_0^N \\ &= (G_N(0) \prod_N V_1)^* q_0^N \\ &= (G_N(0) \prod_N G_N(1) \prod_N V_2)^* q_0^N \\ &\vdots \\ &= (G_N(0) \prod_N G_N(1) \prod_N \dots G_N(N_t-1) \prod_N V_{N_t})^* q_0^N \\ &= \phi^* \prod_N^* G_N(N_t-1) \dots \prod_N^* G_N(1) \prod_N^* G_N(0) q_0^N. \end{aligned}$$

Since ϕ is arbitrary in $B(S)$ it follows that

$q_{N_t}^N = \prod_{N_t}^* G_N(N_t-1) \dots \prod_{N_t}^* G_N(1) \prod_{N_t}^* G_N(0) q_0^N$. Furthermore, since $t \in [0, T]$ is arbitrary we can compute the sequence $\{q_k^N\}$ using the difference equation $q_{k+1}^N = \prod_{N_t}^* G_N(k) q_k^N$.

3.2 Algorithm comparison

To compare some of the algorithms discussed above, consider the following partially observed real-valued system:

$$dx_t = f(x_t)dt + \sigma(x_t)dv_t$$

(2)

$$dy_t = g(x_t)dt + rdw_t$$

where $\{v_t\}, \{w_t\}$ are SBM processes. We have run computer simulations of the evolution of this system. Also, we have filtered the observation process $\{y_t\}$ in these simulations using the Di Masi-Runggaldier algorithm, the Kushner algorithm, and the Goggin algorithm to obtain estimations $\{\bar{x}_t\}$ of the signal process $\{x_t\}$.

We begin by describing the details of the system evolution simulations. First, we generate the trajectories of the time discretized signal process. For this we assume the signal is a linear diffusion, i.e., $f(x_t) = -ax_t$ and $\sigma(x_t) = b$ where a, b are constants. There are two reasons for this linearity assumption. We want to compare the performance of the three filter algorithms mentioned above to the Kalman

filter which is optimal for a linear system. Also, the discrete time analog of a linear SDE is easy to determine (see [19] p. 189). Then, we generate the discrete time observation trajectories for various functions $g(x)$. The noise coefficient r in these observations must be fairly small (say $r \leq 2^{-3}$) for reasonable filtering results. So, to simulate the evolution of the partially observed system, the computer evaluates the discrete time analog of system (2):

$$x_k = e^{-a\Delta} x_{k-1} + \frac{b^2}{2a} (1 - e^{-2a\Delta}) v_k \quad (3)$$

$$\Delta y_k = g(x_k) \Delta + r \sqrt{\Delta} w_k$$

where Δ is the time increment, $\{v_k\}$ and $\{w_k\}$ are iid $N(0,1)$ sequences of random numbers. The initial value x_0 is a given number or random number, and the initial value y_0 is always zero.

Next we apply the three filtering algorithms to the observation process $\{y_k\}$ with $k=0,1,\dots,t/\Delta$ to obtain estimation processes. We use the following discrete space-time grid for the filters. The finite state space $S=\{x:x=-m+kh \text{ and } k=0,1,\dots,2m/h\}$ where m is the bound for the space and h is the space increment. The time grid is $\{t:t=kd \text{ and } k=0,1,\dots,T/d\}$ where d is the time increment. The relation between the time and space increments is discussed below. Recall that all of the filtering algorithms use the

transition matrix Π of a Markov chain that takes values on the state space S . We choose Π to be the $\left(\frac{2m}{h}+1\right) \times \left(\frac{2m}{h}+1\right)$ dimensional tridiagonal matrix with entries described on page 71 of this paper. But, the entries described on page 71 must be between zero and one in order for Π to be a transition matrix. This requires that the time increment d be sufficiently small compared to the space increment h . For our filters we take $d = \frac{h^2}{4}$.

Within the framework described above, we summarize each of the three filtering algorithms.

- i) The Di Masi-Runggaldier algorithm solves Clark's differential equation (see p. 57):

$$r'_t = R_t^{-1} A^* R_t r_t, \quad r_0 = p_0, \quad t \in [0, T]$$

where $A^* = (\Pi^* - I) \frac{1}{d}$, $R_t = \exp(Gy_t - \frac{1}{2}G^2t)$, $G = \text{diag}(g(-m), g(-m+h), \dots, g(m-h), g(m))$, and p_0 is the initial distribution of the Markov chain signal approximation. Then $\bar{x}_t = E\{x_t | Y_t\}$ is computed using the conditional distribution ρ_t where

$$\rho_t = \frac{R_t r_t}{\sum_{x \in S} R_t(x, x) r_t(x)}.$$

To numerically solve Clark's differential equation we use an Euler approximation.

- ii) The Kushner algorithm solves the difference equation:

$$q_{k+1} = \Pi^* G(k) q_k, \quad q_0 = p_0, \quad k=0, 1, \dots, T/d$$

where $G(k) = \text{diag}(r_k(-m, \Delta y_k), \dots, r_k(m, \Delta y_k))$,

$\Delta y_k = y_{(k+1)d} - y_{kd}$, and $r_k(x, \Delta y_k) = \exp(g(x) \Delta y_k - \frac{1}{2} g^2(x) d)$.

Then the conditional distribution ρ_k is given by

$$\rho_k = q_k \left(\sum_{x \in S} q_k(x) \right)^{-1}.$$

- iii) The Goggin algorithm solves a difference equation of the same form as the Kushner algorithm except that

$$r_k(x, \Delta y_k) = (1 + g(x) \Delta y_k),$$

where $\Delta y_k = \pm \frac{\sqrt{d}}{r}$ (see p. 77 for a description of Δy_k in the Goggin algorithm).

Because it is designed around matrix computations the MATLAB software environment is ideal for running the above system and filter simulations. The MATLAB code for a typical run is given in the Appendix. Also, in the Appendix we display the results of several simulation runs. For each simulation we vary the observation function $g(x)$ and/or the space increment h . We display the signal trajectory $\{x_t\}$ and the filtered estimation $\{\bar{x}_t\}$ for each of the three filters described above and the linearized Kalman filter. However, for some of the simulations we display the unfiltered

observation process instead of the Goggin filter. We do this because $r_k(x, \Delta y_k)$ must be positive and this requires that $\|g\|_\infty \frac{\sqrt{d}}{r}$ be less than one in the Goggin algorithm. So, when this condition is not satisfied we do not run the Goggin filter. Also, note that in each simulation we record the CPU time required to run each filter algorithm.

Comparing the filtered estimates in the simulations in the Appendix, we observe the following. First, it appears that the accuracy of all three filters compare well with the linearized Kalman filter when $g(x)$ is approximately linear near zero (i.e., $g'(0) \neq 0$). Of course, they outperform the linearized Kalman filter when $g(x)$ is "very" nonlinear near zero (e.g., $g(x) = x^3$). Next, there is a significant difference in the computation times required for the three filters. Since the Di Masi-Runggaldier algorithm requires the solution of an ODE it is much slower than the others. The reason the Goggin algorithm is faster than the Kushner algorithm is that the $G(k)$ matrix in the Goggin algorithm is not reevaluated for each k since $G(k)$ has only two possible forms depending on whether $\Delta y_k = \frac{\sqrt{d}}{r}$ or $\Delta y_k = -\frac{\sqrt{d}}{r}$. Note that even though the Goggin algorithm cannot be run for specific $g(x)$ functions when $h = \frac{1}{4}$, the CPU time for the Goggin filter when $h = \frac{1}{8}$ is comparable to the CPU time for the Kushner filter when $h = \frac{1}{4}$.

Of course, much work still needs to be done to get an

accurate assessment of the relative merits of these filter algorithms and other algorithms that have been developed. For example, we should compare the algorithms when the signal process is nonlinear. Also, we need to develop a quantitative measure for comparing the accuracy of the filters. Finally, these filters should be compared to other filtering schemes such as numerical methods for solving the Zakai equation. We hope to investigate these topics in the future.

REFERENCES

1. Elliott, R.J.; Stochastic Calculus and Applications; Springer-Verlag, New York, 1982.
2. Kopp, P.E.; Martingales and Stochastic Integrals; Cambridge University Press, New York, 1984.
3. Brémaud, P.; Point Processes and Queues; Springer-Verlag, New York, 1981.
4. Elliott, R.J.; "Filtering and Control for Point Process Observations"; Recent Advances in Stochastic Calculus, ed. Baras and Mirelli; Springer-Verlag, New York, 1990.
5. Wong, E.; "Recent Progress in Stochastic Processes: A Survey"; IEEE Transactions on Information Theory, (1973), v.19.
6. Allinger, D. and Mitter, S.K.; "New Results on the Innovations Problem for Nonlinear Filtering"; Stochastics, (1981), v.4.
7. Fujisaki, M. and Kallianpur, G. and Kunita, H.; "Stochastic Differential Equations for the Nonlinear Filtering Problem"; Osaka Journal of Mathematics, (1972), v.1.
8. Ethier, S.M. and Kurtz, T.G.; Markov Processes: Characterization and Convergence; Wiley-Interscience, New York, 1990.
9. Bhattacharya, R.N. and Waymire, E.C.; Stochastic Processes with Applications; Wiley, New York, 1990.
10. Davis, M.H.A.; Linear Estimation and Stochastic Control; Chapman and Hall, London, 1977.
11. Davis, M.H.A. and Marcus, S.I.; "An Introduction to Nonlinear Filtering"; Stochastic Systems: The Mathematics of Filtering and Identification and Applications, ed. Hazewinkel and Willems; D.Reidel, Boston, 1981.
12. Kushner, H.J. and Dupuis, P.; Numerical Methods for Stochastic Control Problems in Continuous Time; Springer-Verlag, New York, 1992.

13. Kushner, H.J.; Probability Methods for Approximations in Stochastic Control and for Elliptic Equations; Academic Press, New York, 1977.
14. Di Masi, G.B. and Runggaldier, W.J.; "Continuous Time Approximations for the Nonlinear Filtering Problem"; Applied Mathematics and Optimization, (1981), v.7.
15. Goggin, E.M.; "An L^1 Approximation for Conditional Expectations"; 1994; to appear.
16. Billingsley, P.; Convergence of Probability Measures; Wiley, New York, 1968.
17. Pardoux, E.; "Nonlinear Filtering, Prediction, and Smoothing"; Stochastic Systems: The Mathematics of Filtering and Identification and Applications, ed. Hazewinkel and Willems; D. Reidel, Boston, 1981.
18. Clark, J.M.C.; "The Design of Robust Approximations to the Stochastic Differential Equations of Nonlinear Filtering"; Communication Systems and Random Process Theory, ed. Skwizynski; Alphen aan den Rijn, Sijlhoff and Noordhoff, 1978.
19. Brown, R.G.; Introduction to Random Signal Analysis and Kalman Filtering; Wiley, New York, 1983.
20. Chen, C.; Linear System Theory and Design; Holt, Rinehart, and Winston, New York, 1984.

APPENDIX

```

% INITIALIZE PARAMETERS
%
dt=1/2^10;
m=1/dt;
t=dt*[1:m]';
a=1;
b=sqrt(2);
s2=b^2/(2*a);
ex=0;
ep=0;
%ep=s2;
xa=ex*exp(-a*dt*[1:m]');
f=exp(-a*dt);
q2=s2*(1-exp(-2*a*dt));
r2=1/2^6;
%
h=1/2^3;
d=(h^2)/4;
bd=2;
c=bd/h;
s=h*[-c:c]';
%q0=exp(-(s.^2)/(2*s2))*h/sqrt(2*pi*s2);
q0=[zeros(c,1);1;zeros(c,1)];
pu=[0;0;(a*s2*d/(h^2))*ones(2*c-1,1)+[(-a*d)*[-c+1:-1]';zeros(c,1)]];
pl=flipud(pu);
pd=[1;ones(2*c-1,1)-pl(1:2*c-1,1)-pu(3:2*c+1,1);1];
p=spdiags([pl pd pu],[-1:1,2*c+1,2*c+1]);
clear pu pl pd;

% DEFINE G(X)
%
function y=g(x)
y=x;

% SYSTEM EVOLUTION
%
%x0=sqrt(s2)*randn(1);
x0=0;
v=randn(m,1);
x=ltitr(f,sqrt(q2),v,x0);
w=randn(m,1);
z=g(x)*dt+sqrt(dt*r2)*w;
y0=0;
y=ltitr(1,1,z,y0);
clear v w;

% LINEARIZED KALMAN FILTER
%

```

```

%gl=zeros(m,1);
gl=dt*ones(m,1);
xb=f*ex;
pb=f*ep*f+q2;
z=z-g(xa)*dt;
xkal=zeros(m,1);
for k=1:m
    kg=pb*gl(k)/(gl(k)*pb*gl(k)+r2*dt);
    ph=(1-kg*gl(k))*pb;
    xkal(k)=xb+kg*(z(k)-gl(k)*xb);
    xb=f*xkal(k);
    pb=f*ph*f+q2;
end
xkal=xkal+xa;
clear gl z;

% DI MASI FILTER
%
q=q0;
yc=y0;
xrdc=zeros(m,1);
bv=exp(g(s)./r2);
bg=diag(exp(-(dt/(2*r2))*(g(s).^2)))*((1-dt/d)*eye(2*c+1)+(dt/d)*p');
for k=1:m
    dy=y(k)-yc;
    yc=y(k);
    q=diag(bv.^dy)*bg*q;
    xrdc(k)=(s'*q)/sum(q);
end
clear bv bg q;

% KUSHNER FILTER
%
q=q0;
yc=y0;
xkusc=ex;
xkus=zeros(m,1);
bv=exp(g(s)./r2);
bm=p'*diag(exp(-(d/(2*r2))*(g(s).^2)));
for k=1:m
    if rem(k,m*d)==0
        dy=y(k)-yc;
        yc=y(k);
        q=bm*diag(bv.^dy)*q;
        xkus(k)=(s'*q)/sum(q);
        xkusc=xkus(k);
    else
        xkus(k)=xkusc;
    end
end

```

```

    end
end
clear bv bm q;

% GOGGIN FILTER
%
q=q0;
yc=y0;
xgogc=ex;
xgog=zeros(m,1);
bn=p'*diag(ones(2*c+1,1)-sqrt(d/r2)*g(s));
bp=p'*diag(ones(2*c+1,1)+sqrt(d/r2)*g(s));
for k=1:m
    if abs(y(k)-yc)>=sqrt(r2*d)
        dy=sign(y(k)-yc);
        yc=y(k);
        if dy==1
            q=bp*q;
        else
            q=bn*q;
        end
        xgog(k)=(s'*q)/sum(q);
        xgogc=xgog(k);
    else
        xgog(k)=xgogc;
    end
end
clear bn bp q;

% DRIVER
%
simsys
time=cputime;
kalfilt
tkal=cputime-time
xrdc=y;
if h>=1/2^3;
time=cputime;
rdcfilt
trdc=cputime-time
end
xkus=y;
if h>=1/2^3
time=cputime;
kusfilt
tkus=cputime-time
end
xgog=y;

```

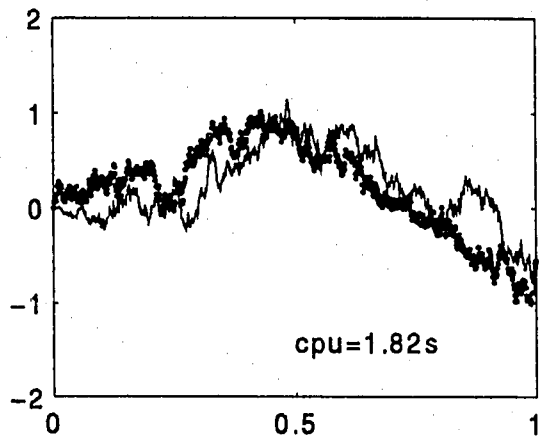
```

if max(abs(g(s)))*sqrt(d/r2)<=1
time=cputime;
gogfilt
tgog=cputime-time
end
%
subplot(2,2,1),plot(t,x,'y.',t,xkal,'r')
title('KALMAN'),axis([0 1 -2 2])
subplot(2,2,2),plot(t,x,'y.',t,xrdc,'r')
title('DI MASI'),axis([0 1 -2 2])
subplot(2,2,3),plot(t,x,'y.',t,xkus,'r')
title('KUSHNER'),axis([0 1 -2 2])
subplot(2,2,4),plot(t,x,'y.',t,xgog,'r')
title('GOGGIN'),axis([0 1 -2 2])

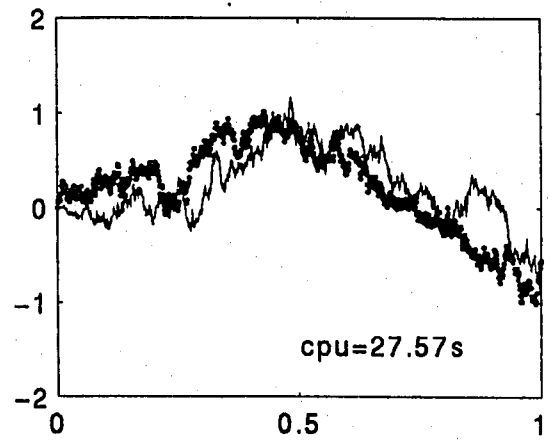
%ANNOTATION
%
gtext('signal .... filter ____')
gtext('g(x)=x    h=1/8')
gtext('cpu=1.82s')
gtext('cpu=27.57s')
gtext('cpu=8.13s')
gtext('cpu=1.48s')

```

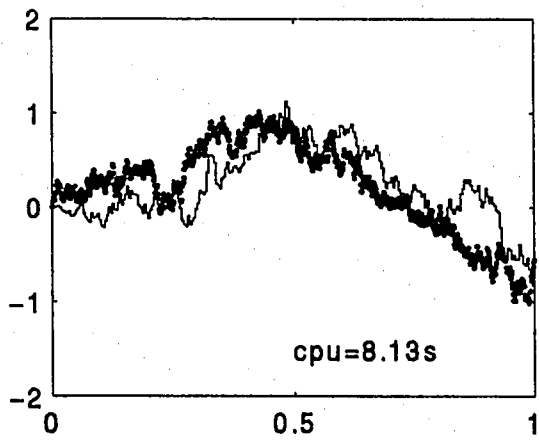
KALMAN



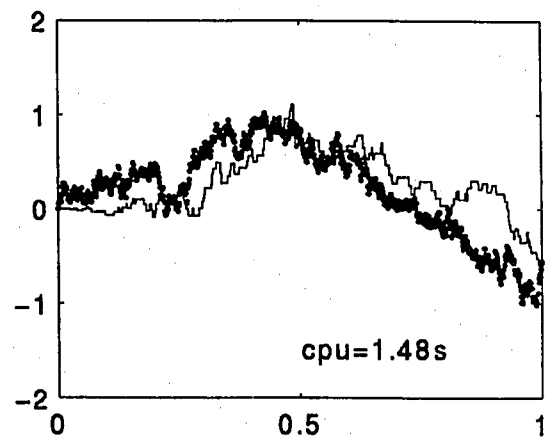
DI MASI



KUSHNER



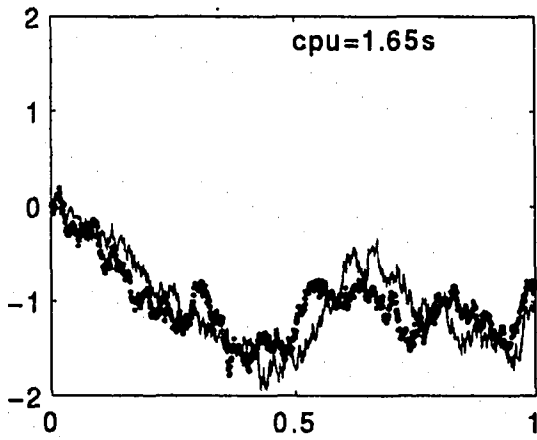
GOGGIN



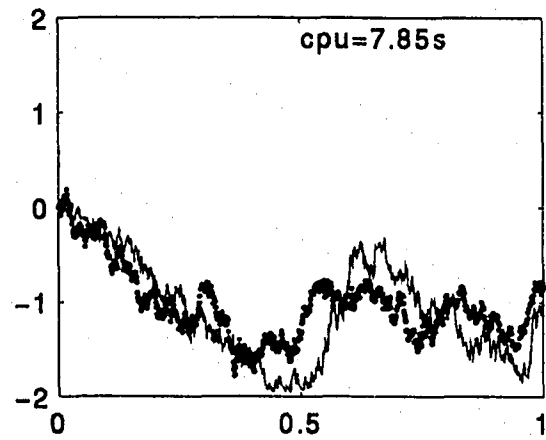
signal filter _____

$g(x)=x$ $h=1/8$

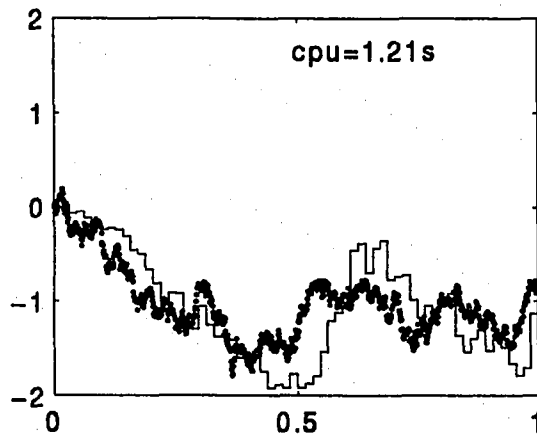
KALMAN



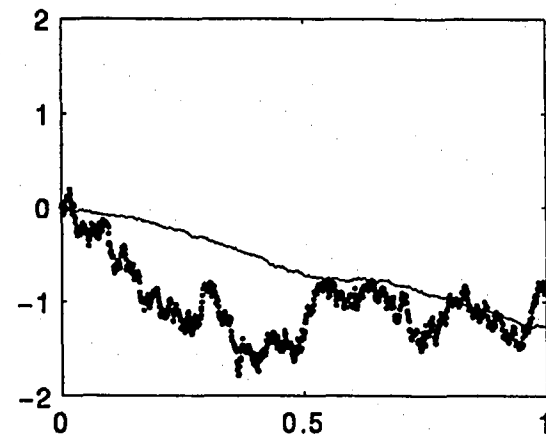
DI MASI



KUSHNER



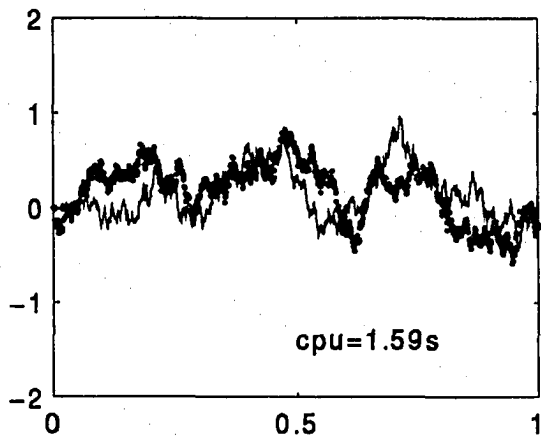
OBSERVATION



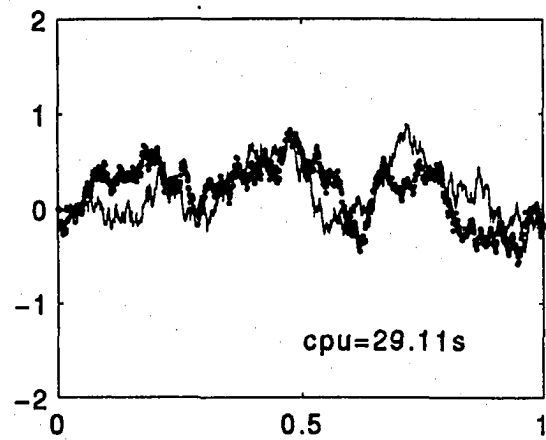
signal filter _____

$g(x)=x$ $h=1/4$

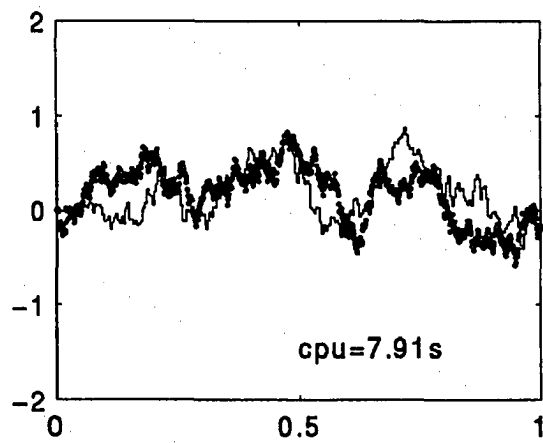
KALMAN



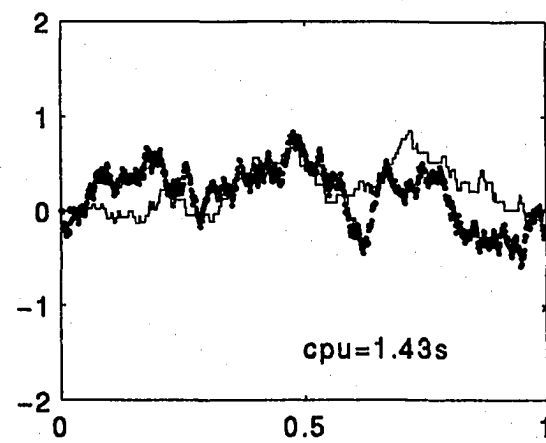
DI MASI



KUSHNER



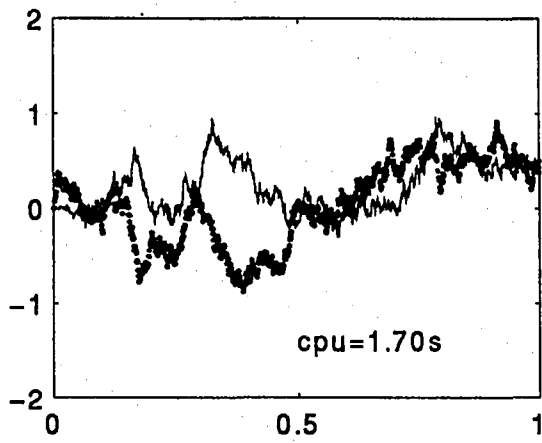
GOGGIN



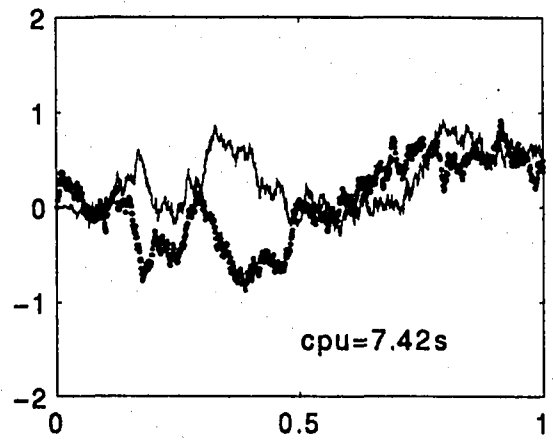
signal filter _____

$g(x) = \sin x$ $h = 1/8$

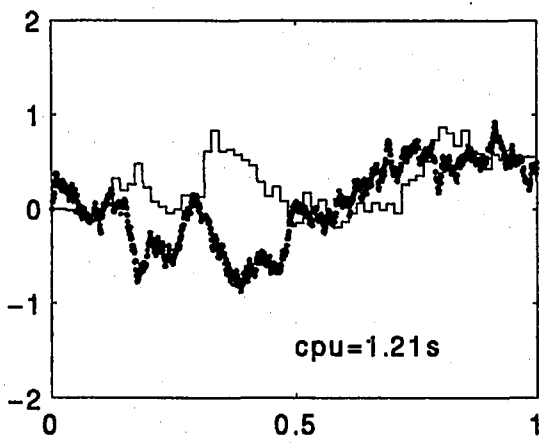
KALMAN



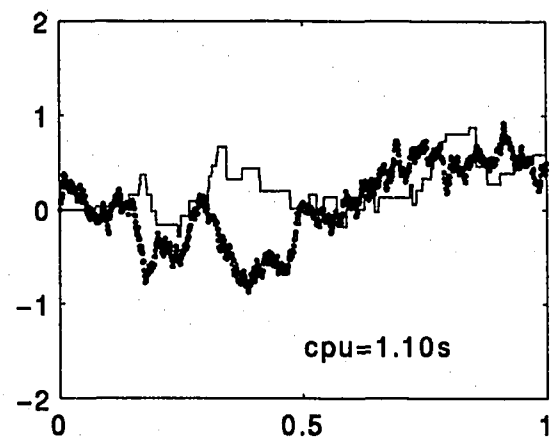
DI MASI



KUSHNER



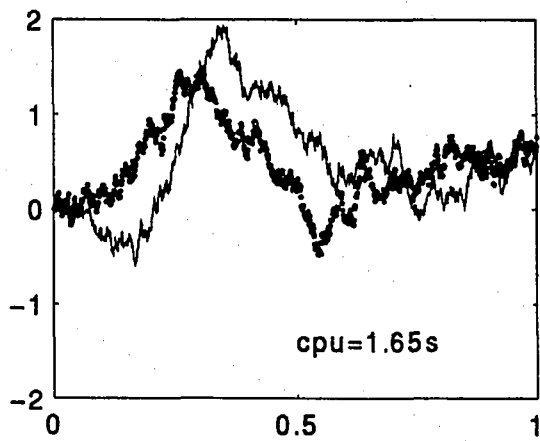
GOGGIN



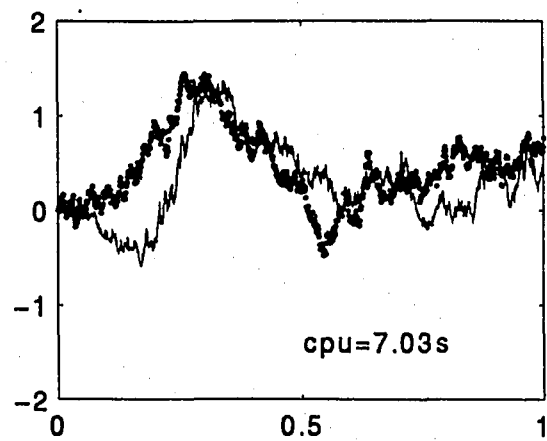
signal filter _____

$g(x) = \sin x$ $h = 1/4$

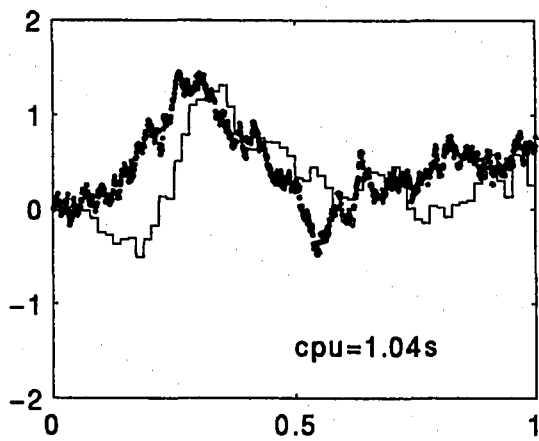
KALMAN



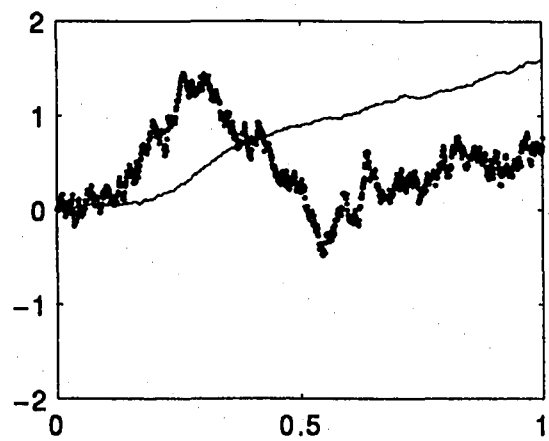
DI MASI



KUSHNER



OBSERVATION



signal filter _____

$g(x) = \exp x$ $h = 1/4$

