# ESTIMATION THEORY FOR CONTINUOUS TIME PROCESSES, A MARTINGALE APPROACH

by

Jan Hendrik Van Schuppen

Memorandum No. UCB/ERL M405

September 1973

# ESTIMATION THEORY FOR CONTINUOUS TIME PROCESSES, A MARTINGALE APPROACH

by

Jan Hendrik Van Schuppen

Memorandum No. ERL-M405

September 1973

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Abstract

In this thesis we present an unified approach to estimation problems, using the theory of martingales and stochastic integrals. We analyse the problem of absolute continuity of measures, and obtain the important result of translation of local martingales under a change of measure. An application of this result is the calculation of the likelihood ratio in detection problems.

In terms of martingales we define a generalized stochastic differential equation and an observation equation. This forms a stochastic system which unifies the formulation for problems of observations with Brownian motion noise and of counting process observations. The filtering, prediction and smoothing problems are considered for the two above mentioned stochastic systems. The least squares error criterion is used, and we derive stochastic differential equations for the optimal estimates. We discuss the resulting filters.

## Acknowledgement

The author expresses his sincere thanks to the professors P. P. Varaiya and E. Wong for the guidance and descussions on the research work reported on in this memorandum. This work was conducted under a research assistantship at the Electronics Research Laboratory, the funds coming from NSF Grant GK-10656X3.

Table of Contents

Table of Contents (continued)

## 1. Introduction

Over the last few years a number of new proofs and new results have been obtained for problems in estimation and stochastic control. These results concern systems with Brownian motion disturbances. There is also a recent interest in filtering problems for counting processes, sometimes, called doubly stochastic Poisson processes. When one studies the methods available to solve the above mentioned problems, then one concludes that the method that solves most of them is the theory of martingales and stochastic integrals. In this thesis we present an approach to a large class of estimation problems for continuous time processes. The method we use is the theory of martingales and stochastic integrals. We discuss both the method and the class of problems in more detail.

Martingales and stochastic integrals.

As indicated in the book by Doob [1953] on stochastic processes, there are three main classes of processes: 1. independent increment processes; 2. martingales; and 3. Markov processes. Although in this thesis martingales play a prominent role, the other classes are important but in an implicit form. The early work on martingale theory, as can be found in Doob [1953], deals mainly with martingale inequalities and the martingale convergence theorem. Further work by Meyer [1966] deals with the decomposition of martingales. The relevant martingale theory for this thesis consists of the new developments concerning the stochastic integral. The concept of stochastic integral with respect to Brownian motion was developed by Ito [1944]. In a series of articles, the main ones being [Kunita,Watanabe,1967], [Millar,1968], [Doléans-Dade,Meyer,1970], the concept of stochastic integral has been extended to a class of martingales. A related important result in applications is the differentiation formula, giving the role for martingale calculus. A complimentary result to this development is the so called martingale representation theorem, which gives a representation of certain martingales as a stochastic integral. Together these new results are the tools in our analysis of estimation theory. We will call this method the martingale approach.

Estimation problems.

The classical problems considered in estimation are detection, filtering, prediction, smoothing and identification. Except for identification we will discuss all of these. If for estimation problems we consider the least squares error criterion, then the optimal estimate is the conditional expectation of the unknown variable given the $\sigma$-field generated by the past of the observed process. This property forms the connection between estimation problems and martingales. We will show later that certain processes and their conditional expectation are related in a natural way with associated martingale processes.

The basic goal.

The basic goal of this thesis is the following:

To analyse and solve estimation problems with martingale theory. We will define a general stochastic system model in terms of martingales, that covers both the system with Brownian motion disturbances, and that with counting process observations. We will derive for this stochastic system a solution to several estimation problems, namely detection, filtering, prediction and smoothing. This way we construct a general theory that handles a large class of estimation problems. The stochastic control problem is not discussed here. For systems disturbed by Brownian motion it was considered by Davis, Varaiya [1973]. The extension to other processes is under investigation.

The results.

We have only partially succeeded in attaining the above stated goal. The restrictive factor, which is the key to the specific results, is the martingale representation theorem. This theorem has only been proven if the underlying process is Brownian motion or a Poisson process. For these two cases we derive the solution to the detection problem, and a stochastic differential equation for the optimal filtering estimate. The main contribution of this thesis is the frame work of martingale theory that is relevant in estimation problems. We comment further on the results in Chapter 7.

Outline of contents.

In chapter 2 we give the mathematical preliminaries, mainly the theory of martingales and stochastic integrals. We shortly summarize the main definitions and results, such as to introduce an unified notation and an easy reference for the reader.

In chapter 3 we discuss the problem of absolute continuity of measures, its characterization in terms of martingales, and the translation of martingales under a change of measure. These results have important applications, such as the detection problem which is discussed in section 3.4.

In chapter 4 we consider generalized stochastic differential equations, and give a new definition of a stochastic system. We also consider the projection of processes on a family of $\sigma$-fields, which includes the concept of innovation process.

In chapter 5 we approach the general estimation problem. After a review of the literature, we present the elementary concepts of estimation theory. We then give the elementary results for filtering, prediction and smoothing.

In chapter 6 we derive in detail the filtering formula's for the observation equations with Brownian motion disturbances and for counting process observations. The prediction and smoothing formula's are also derived. Finally a discussion of the martingale approach to estimation problems is given.

In chapter 7 we conclude our work with a discussion and conclusions.

Notation.

In all chapters, except chapter 2, we number all definitions, theorems and other results consecutively. The first two digits of this label indicate the section in which it is located. In chapter 2 we have labelled every statement and omitted the section labelling.
All real and vector valued variables are denoted by lower case symbols.
Matrix valued variables are denoted by capital symbols.

## 2. Stochastic processes, martingales and stochastic integrals.

### 2.1. Introduction

This chapter contains the mathematical preliminaries, necessary for our investigation of estimation theory. We will state the main definitions and results, so as to give an easy reference of notation to the reader. Slight extensions of some results and some new definitions are also given. All items are numbered for reference, a method we have adopted from Meyer. We start with some concepts from the theory of stochastic processes, and martingales. Since Brownian motion and the Poisson process play an important role in this thesis, we discuss them in Section 2.4. In the following section we define stochastic integrals and discuss its properties. In the last section we state a number of martingale representation theorems, which play a crucial role.

The topics presented in this chapter form the essential points of martingale theory. The application of these results we will call the martingale approach.

### 2.2. Stochastic processes.

In this section we will introduce the main definitions and notation concerning stochastic processes. Although the basic reference to stochastic is Doob's book [1953], we will use concepts introduced by Meyer and his co-workers in the context of martingale theory. The relevant references are Meyer's book [1966], a subsequent article [1968], and primarily the book by Dellacheric [1972]. See also the first two chapters of the new book by Meyer [1972]. The following concepts are denoted by the term general theory of stochastic processes.

We will number the following definitions and results for easy reference.

$\sigma$-fields.

1. Let $(\Omega,F,P)$ be a probability space, and let $F$ be complete with respect to $P$.

2. Let $T \subset R$, be the time interval of interest, usually we will take $T = [0,\infty)$ or $T = [0,1]$.

3. $(F_t, t \in t)$ is a family of $\sigma$-fields satisfying:

   a. Sub-$\sigma$-fields of $F$ : $F_t \subset F$, for all $t \in T$,

   b. increasing : if $s < t$ then $F_s \subset F_t$,

   c. right continuous : $\bigcap_{s>t} F_s = F_t$ for all $t \in T$

   d. $F_0$ contains all the null sets of $F$.

   Arbitrary families of $\sigma$-fields will be understood to satisfy all these conditions.

4. Define $F_\infty \triangleq \bigvee_{t \in T} F_t$, i.e. the $\sigma$-field generated by the union of $(F_t, t \in T)$.

Stochastic processes.

5. A stochastic process $x = (x_t, t \in T)$ is a collection of random variables, indexed by the parameter $t \in T$.

   Scalar or vector valued processes will be denoted by scalars, matrix valued processes by capital letters.

6. The stochastic process $(x_t, t \in T)$ is said to be adapted to the family $(F_t, t \in T)$, if for all $t \in T$, $x_t$ is $F_t$ measurable. [Meyer, 1966,IV,D31]. Notation $(x_t, F_t, t \in T)$, and this notation will always imply that $x$ is adapted to $(F_t, t \in T)$.

7. If $x$ is a stochastic process, then $(F_{xt}, t \in T)$ will always denote

the family of sub-$\sigma$-fields generated by x, i.e. $F_{xt} = \sigma(x_s, \forall s \leq t)$ We always take the right continuous family $(F_{xt+}, t \in T)$, where $F_{xt+} = \bigcap_{s > t} F_{xs}$.

8. If a stochastic process x has sample functions, which, for all most all $\omega$, are right continuous, and have left hand limits for all $t \in T$, then $\Delta x_t \triangleq x_t - x_{t-}$ is called the jump of x at time t. $x_{t-} \triangleq \lim_{s \uparrow t} x_s$.

8a. Two stochastic processes x and y, defined on the same probability space $(\Omega, F, P)$, and taking values in the same measurable space, are called modifications of each other if $x_t = y_t$ a.s. for all $t \in T$. [Meyer,1966,IVD5].

Stopping times.

9. A random variable $\tau$ taking values in T, is called a stopping time of a given family $(F_t, t \in T)$ if for all $t \in T$ : $\{\omega \in \Omega | \tau(\omega) \leq t\} \in F_t$. If $T = [0,\infty)$, $\tau$ may take the value $\infty$. [Meyer,1966,IV,D33]. There exists a classification of stopping times, for a detailed account see [Meyer,1966,IV,VII;1968; Dellacherie,1972,III].

10. If $(F_t, t \in T)$ is an increasing family, and if $\tau$ is a stopping time of it, then $F_\tau \triangleq \{A \in F_\infty | A \cap \{\omega \in \Omega | \tau(\infty) \leq t\} \in F_t, \forall t \in T\}$

$F_{\tau-} \triangleq F_0 \vee \{A \cap \{\omega \in \Omega | t < \tau(\omega)\} | A \in F_t$, for some $t \in T\}$. [Dellacherie,1972,III,D27].

11. $(F_t, t \in T)$ is called quasileft continuous if $F_\tau = F_{\tau-}$ for all predictable stopping times $\tau$. [Dellacherie,1972,III,D38].

12. If $(x_t, F_t, t \in T)$ is a process, $\tau$ a stopping time with respect to $(F_t, t \in T)$, then $x_{t\wedge\tau} \triangleq x_t\, I(\tau>t) + x_\tau\, I(\tau\leq t)$ is called the process x stopped at $\tau$. $I_A$ is the indicator function.

13. Using stopping times, many properties of processes are characterized locally. That is, there exists an increasing sequence of stopping times $\{\tau_n\}$, $\lim_n \tau_n = \infty$ a.s., such that for all n, the stopped process $x_{t \wedge \tau_n}$ has a certain property. This procedure of proving certain properties locally, will be referred to as a stopping time argument.

14. If x is a right continuous adapted process, having left hand limits then:

    1. x charges a stopping time $\tau$ if $P(x_\tau \neq x_{\tau-}, \tau < \infty) > 0$,
    2. x has a jump at $\tau$ if $x_\tau \neq x_{\tau-}$ a.s. on the set $\{\tau<\infty\}$ [Dellacherie,1972,IV,29].

15. An adapted right continuous stochastic process x, having left hand limits, is called quasi-left continuous if it satisfies one of the following equivalent conditions:

    1. the jump times of x are totally inaccessible,
    2. x does not charge any predictable stopping time,
    3. if $\{\tau_n\}$ is an increasing sequence of stopping times, then $\lim x_{\tau_n} = x_{\lim \tau_n}$ a.s. on the set $\{\lim_n \tau_n < \infty\}$. [Dellacherie,1972,IV,T32].

σ-fields on $T \times \Omega$.

16. The σ-field $\underline{P}$ on $T \times \Omega$ generated by all left continuous adapted processes on $T \times \Omega$, is called the predictable σ-field. [Meyer,1972,I,5].

17. The σ-field $\underline{W}$ on $T \times \Omega$ generated by all right continuous adapted processes that have left hand limits, is called the well measurable σ-field. [Meyer,1972,I,5;Dellacherie,1972,IV,T26]

18. A stochastic process $(x_t, F_t, t \in T)$ is called predictable (well

measurable), if it is measurable with respect to $(T \times \Omega, \underline{P})$ (respectively $(T \times \Omega, \underline{W})$). It follows that any left continuous adapted process is predictable, and any right-continuous adapted process well measurable. Under certain conditions, processes have a well measurable or predictable projection, see [Dellacharie,1972,V].

Increasing processes.

19. The real-valued process $(a_t, t \in T)$ is called an increasing process if:

 1. $a_0 = 0$, 2. for all most all $\omega$ the sample paths of a are increasing with T: if $s < t$, then $a_s \leq a_t$ a.s., 3. a is sample right continuous.

We define the following classes:

20. $BV^+ \triangleq \{a | a \text{ is an increasing process}\}$,

21. $BV = \{a | a = a_1 - a_2,\ a_1, a_2 \in BV^+\}$, $a \in BV$ is called a process of bounded variation.

22. $IV^+ = \{a \in BV^+ | \sup_{t \in T} E(a_t) < \infty\}$, IV as BV from $BV^+$. $a \in IV$ is called a process of integrable variation.

23. $LIV^+ = \{a \in BV^+ | \exists \{\tau_n\}, \lim_n \tau_n = \infty \text{ a.s.}, \forall n : a_{t \wedge \tau_n} \in IV^+\}$, LIV, $a \in LIV$ is called a process of locally integrable variation.

Integration.

24. If $a \in IV$ then $L_1(a) = \{\phi | \phi \text{ adapted, predictable, } E[\int_T |\phi_s|\ |da_s|] < \infty\}$

25. If $a \in IV$, $\phi \in L_1(a)$ then $(\int_0^t \phi_s da_s, F_t, t \in T)$ is an adapted well measurable process, the integral is well defined [Dellacharie,1972,IV,39].

26. It is known that the predictable projection of a process of bounded variation need not be of bounded variation. However the following is true.

27. If $a \in IV^+$, then there exists unique processes $a_w \in IV^+$, well measurable, $a_p \in IV^+$, predictable such that for every positive $\phi \in L_1(a)$, $E[\int_T \phi_s da_s] = E[\int_T \phi_s da_{ws}] = E[\int_T \phi_s da_{ps}]$. $a_w, a_p$ are called the dual well measurable, respectively the dual predictable projection of a. [Dellacherie,1972,V,T28].

## 2.3. Martingales.

In this section we introduce certain classes of martingales, and the main results of martingale theory. The points given here were developed by Meyer, the main references are Meyer [1966], Kunita, Watanabe [1967], Doléans-Dade,Meyer [1970]. In the following let $T = [0,\infty)$, and we suppose that some family of σ-fields $(F_t, t \in T)$ is given, satisfying the usual conditions.

1. The stochastic $(m_t, F_t, t \in T)$ is a martingale if:
   1. it is adapted, 2. $E|m_t| < \infty$ for all $t \in T$, 3. $E[m_t|F_s] = m_s$ a.s. for all $t,s \in T$, $s < t$.

We introduce the following classes of martingales:

2. $M_1 \triangleq \{m | m$ is a right continuous adapted stochastic process, having left hand limits, $m_0 = 0$, m is a martingale with respect to some specified family $(F_t, t \in T)$, and uniformly integrable.$\}$.

3. $M_2 \triangleq \{m \in M_1 | \sup_{t \in T} E(m_t^2) < \infty\}$

4. $M_2^c \triangleq \{m \in M_2 | m$ is sample continuous$\}$.

5. $M_{loc} = \{m | m$ is a right continuous adapted stochastic process, $m_0 = 0$, and there exists an increasing sequence of stopping times $\{\tau_n\}$, $\lim_n \tau_n = \infty$ a.s., such that for all n, on the set $\{\tau_n > 0\}$, $m_{t \wedge \tau_n} \in M_1$.$\}$.

6. $M^c_{1oc} \triangleq \{m \in M_{1oc} \mid m \text{ is sample continuous}\}$.

7. $M_{21oc} \triangleq \{m \in M_{1oc} \mid$ there exists $\{\tau_n\}$, $\lim_n \tau_n = \infty$ a.s., such that for all n, $m_{t\wedge\tau_n} \in M_2\}$.

8. Martingales in the above classes are called martingales, or local martingales with adjectives integrable, square integrable, sample continuous, or locally square integrable where suitable.

9. Martingale theory was developed by Doob and Meyer. Attention focused mainly on supermartingales and $M_2$ martingales. Because these classes are quite restrictive, the class of local-martingales was introduced, apparently first by Itô, Watanabe [1965]. Kunita, Watanabe [1967]also used local-martingales, which are locally square integrable martingales ($M_{21oc}$) according to our definition. Doléans-Dade, Meyer [1970] distinguished between the classes $M_{1oc}$ and $M_{21oc}$. Note that we have $M^c_{1oc} = M^c_{21oc}$ which can be proven by a stopping time argument. However $M_{1oc} \subsetneq M_{21oc}$ Doléans-Dade has given a counter-example.

Decomposition of martingales.

10. The martingales $m,n \in M_{1oc}$ are called orthogonal iff $(m_t n_t, F_t, t \in T) \in M_{1oc}$ [Doléans-Dade,Meyer,1970,Th.7.]. If $m,n \in M_2$, then they are orthogonal iff $mn \in M_1$. Similarly $m,n \in M_{21oc}$ orthogonal iff $mn \in M_{1oc}$.

11. $M^d_2 \triangleq \{m \in M_2 \mid m$ is orthogonal to all $n \in M^c_2\}$, such m is called discontinuous.

12. $M^d_{1oc} = \{m \in M_{1oc} \mid m$ is orthogonal to all $n \in M^c_{1oc}\}$.

13. $M^d_{21oc} = \{m \in M_{21oc} \mid m$ orthogonal to all $n \in M^c_{21oc}\}$.

14. If $m \in M_{1oc}$ then there exists an unique decomposition $m = m^c + m^d$, where $m^c \in M^c_{1oc}$, $m^d \in M^d_{1oc}$ [Doléan-Dade,Meyer,1970,Th.7.].

If $m \in M_2$, then $m^c \in M_2^c$, $m^d \in M_2^d$ [DD-M,1970,Th.4.].

15. If $(F_t, t \in T)$ is quasi-left continuous, than every $m \in M_{loc}$ is quasi-left continuous. Hence the jumps of m are totally inaccessible [Dellacherie, 1972,V,T42].

16. If $m \in M_{loc}^c \cap BV$, then $m = 0$ a.s. [Dellacherie,1972,V,T39].

17. If $m \in M_{loc} \cap BV$ and predictable, and if $(F_t, t \in T)$ is quasi-left continuous then $m = 0$ a.s. This follows because by 15 m charges only totally inaccessible stopping times, but since it is also predictable, it must be continuous. Then the result follows by 16.

Increasing processes and martingales.

18. Two processes $a_1, a_2 \in IV^+$ are called associated iff they have the same dual predictable projection [Dellacherie,1972,V,D35].

19. Two adapted processes $a_1, a_2 \in IV^+$ are associated iff $(a_{1t}-a_{2t}, F_t, t \in T) \in M_1$ [Dellacherie,1972,V,T36].

20. Given an adapted $a_1 \in IV^+$ and a predictable $a_2 \in IV^+$. Then $a_2$ is the unique dual predictable projection of $a_1$ iff $a_1$ and $a_2$ are associated iff $(a_{1t}-a_{2t}, F_t, t \in T) \in M_1$. [Dellacherie,1972,V,T38].

Note that this result can be extended to the classes IV, $LIV^+$, LIV.

21. If $a_2$ is the dual predictable projection of an adapted process $a_1$, then $a_2$ is sample continuous iff $a_1$ is quasi-left continuous [Dellacherie,1972,V,T40].

Martingales and associated increasing processes.

23. If $m \in M_2$ is real valued, then there exists an unique predictable increasing process $(\langle m,m\rangle_t, F_t, t \in T) \in IV^+$, such that

$(m_t^2 - \langle m,m\rangle_t,\ F_t, t \in T) \in M_1$. [Meyer,1966,VIII23]. $\langle m,m\rangle$ will be called the predictable quadratic variation of m.

24. If $m,n \in M_2$ then define $\langle m,n\rangle \triangleq \frac{1}{2}(\langle m+n,m+n\rangle - \langle m,m\rangle - \langle n,n\rangle)$.

25. If $m \in M_2$ is vector valued, then $\langle m,m\rangle$ is matrix valued and defined element wise $(\langle m^i,m^j\rangle)$.

26. If $m \in M_{2loc}$ then there exists an unique predictable process $\langle m,m\rangle \in LIV^+$ such that $(m_t^2 - \langle m,m\rangle_t, F_t, t \in T) \in M_{loc}$. The proof of this follows by a stopping time argument from 23.

27. If $m \in M_{loc}$ and if $m = m^c + m^d$ is its unique decomposition, then define $[m,m]_t = \langle m^c,m^c\rangle_t + \sum_{s\leq t} (\Delta m_s)^2$. It will be called the well measurable quadratic variation. It is well known that $[m,m]_t < \infty$ a.s. for all $t \in T$ [Doléans-Dade,Meyer,1970,Th.7.] Note that because $M_{loc}^c = M_{2loc}^c$, $\langle m^c,m^c\rangle$ is well defined.

28. If $m,n \in M_{loc}$, let $[m,n] = \frac{1}{2}([m+n,m+n] - [m,m] - [n,n])$, then also $[m,n]_t = \langle m^c,n^c\rangle_t + \sum_{s\leq t} (\Delta m_s)(\Delta n_s)$.

29. Both $[m,n]$ and $\langle m,n\rangle$, whenever they exist, have a characterization in terms of limits of sums of the quadratic variation of the process [Meyer,1967,II].

30. It will be shown later (2.5.25) that if $m,n \in M_{loc}$ then $(m_t n_t - [m,n]_t, F_t, t \in T) \in M_{loc}$.

31. Note that if $m \in M_2$, then both the well measurable $[m,m]$ and the predictable quadratic variation $\langle m,m\rangle$ are well defined and $([m,m]_t - \langle m,m\rangle_t, F_t, t \in T) \in M_1$. [Doléans-Dade,Meyer,1970,Th5.]

33. We therefore define: If $m,n \in M_{loc}$ and if $[m,n] \in LIV$ then we denote the unique dual predictable projection of $[m,n]$ by $\langle m,n\rangle$. A characterization for $\langle m,n\rangle$ is, that it is a predictable process, adapted, of bounded variation and that $([m,n]_t - \langle m,n\rangle_t, F_t, t \in T) \in$

$M_{1oc}$. The existence and uniqueness follows from (2.3.20). We will show later that there exists a case where $m \notin M_{2loc}$, but still such an $\langle m,n \rangle$ exists. (2.5.21).

33. If $(F_t, t \in T)$ is quasi-left continuous, then if $\langle m,m \rangle$ exists, it is sample continuous (by 2.3.21). $\langle m^c, m^c \rangle$ is always sample continuous.

The class $M_{2loc}$.

In subsequent chapters we concentrate our attention on martingales in the class $M_{2loc}$. It is therefore of interest to obtain a sufficient condition for a martingale to be in $M_{2loc}$.

34. The martingale $m \in M_{1oc}$ is in $M_{2loc}$ if either of the following is true:

1. $E(m_t)^2 < \infty$ for all $t \in T$.
2. $[m,m] \in LIV^+$.

Proof. 1. is obvious and 2 follows immediately because $[m,m]$ is locally integrable and $m^2 - [m,m] \in M_{1oc}$.

Semi-martingales.

35. An adapted stochastic process $(x_t, F_t, t \in T)$ is called a semi-martingale if $x_t = x_0 + a_t + m_t$, where $a \in BV$, $m \in M_{1oc}$. The class of such processes is denoted by SM. [Doléans-Dade,Meyer, 1970, § 3;Meyer,1971b,D7]. Almost all processes we will encounter in this report are semi-martingales. An important subclass are those for which the process of bounded variation is predictable.

36. The decomposition of the semi-martingale is not unique, however there are certain intrinsic properties: $x_0$ is unique, since $a_0 = 0$, $m_0 = 0$. The continuous part of the local martingale

$m^c$ is unique given the family $(F_t, t \in T)$. Hence we define $x^c \triangleq m^c$. Note however that this is not the sample continuous part of x, since a can be sample continuous too. $\Delta x_t$ is also unique. [Doléans-Dade,Meyer,1970, § 3,5].

37. Because of the above intrinsic properties we can define: if $x \in SM$ them $[x,x]_t \triangleq \langle x^c,x^c\rangle_t + \sum_{s\leq t} (\Delta x_s)^2 \in BV^+$. Both terms are well defined. Furthermore $(\langle x^c,x^c\rangle_t, F_{xt}, t \in T)$ is adapted and unique. To prove this we use the differentiation rule 2.5.23., $x_t^2 = x_0^2 + \int_0^t 2x_{s-} \, dx_s + \langle x^c,x^c\rangle_t + \sum_{s\leq t} (\Delta x_s)^2$ This implies that $(\langle x^c,x^c\rangle_t, F_{xt}, t \in T)$ is adapted, and the uniqueness follows similarly. If x has two different decompositions with respect to different families of σ-fields, $x_t = a_t + m_t = \hat{a}_t + \hat{m}_t$, where $(m_t, F_t, t \in T) \in M_{loc}$, $(\hat{m}_t, F_{xt}, t \in T)$, then the above implies that $\langle m^c,m^c\rangle = \langle x^c,x^c\rangle = \langle \hat{m}^c,\hat{m}^c\rangle$. If $x = m$ then $[x,x] = [m,m]$.

38. If $x = a + m$, $a \in BV$, $m \in M_{loc}$, and a is sample continuous, then $[x,x] = [m,m]$. This is true because the continuous part of a does not show up in the definition of $[x,x]$. If $x = a + m = \hat{a} + \hat{m}$, where a and $\hat{a}$ are sample continuous, then $[m,m] = [x,x] = [\hat{m},\hat{m}]$, irrespective of the family of σ-fields to which they are adapted.

## 2.4. Brownian motion and counting processes.

Brownian motion is a natural phenomenon, the existence of which was first published by Brown 1827. He observed 'small particles in rapid oscillatory motion.' For a detailed account on Brownian motion see Nelson [1967]. This book discusses Brown's discovery of the motion and the many mathematical theories for explaining its behavior, including Wiener integrals and certain stochastic differential equations. An interesting application is the discussion of quantum physics from a stochastic viewpoint. In this thesis the name Brownian motion denotes a stochastic process, defined below, which is a mathematical model of the natural phenomenon. For reference to the following see Wong [1971a], and Doob [1953].

1. The process $(x_t, t \in T)$ is a Gaussian process iff every finite linear combination of the form $\sum_{i=1}^{n} \alpha_i x_{t_i}$ is a Gaussian random variable.

2. A stochastic process $(w_t, t \in T)$ with values in R, is a Brownian motion if:

   1. it is a Gaussian process., 2. $E(w_t) = 0$, for all $t \in T$, 3. $E(w_t w_s) = \sigma \min(t,s)$, where $\sigma$ is a strictly positive constant.

3. Under certain conditions we can choose a separable version of the process, and this version will then be sample continuous.

4. If $\sigma = 1$, we will call w standard Brownian motion in R. Standard Brownian motion in $R^n$, will denote a vector valued process, whose components are independent standard Brownian motions in R, so $E[w_t (w_s)^T] = \min(t,s) I$ where I is the identity matrix in $R^n$.

5. Brownian motion has the following properties: 1. it is a process, with stationary independent increments, 2. it is a martingale with respect to the $\sigma$-field generated by it, 3. it is a Markov process. Brownian motion is thus a sample continuous process that is in all the three main classes of stochastic processes.

6. The martingales characterization. Note that if $T = [0,\infty)$ $(w_t, F_{wt}, t \in T) \in M^c_{2loc}$ with $\langle w,w \rangle_t = t$. If $T = [0,1]$ then

$(w_t, F_{wt}, t \in T) \in M_2^c$ since $\langle w,w \rangle$ is bounded.

7. Using the martingale characterization, there is an important equivalent condition. If $(w_t, F_t, t \in T) \in M_{loc}^c$, and if $\langle w,w \rangle_t = t$, then w is a sample continuous Brownian motion process. [Kunita,Watanabe,1967,Th.2.3.]

8. The Brownian motion process belongs to another class of processes, namely Hunt processes. This is because any process with stationary independent increments is a Hunt process. A Hunt process is essentially a strong Markov process, taking values in a specific space, that is right continuous, and quasi-left continuous, and for which the generated $\sigma$-field satisfies certain conditions. For references see [Dynkin,1965,Meyer,1967b;Blumenthal,Getoor,1968]. The only result we need is that the $\sigma$-field generated by a Hunt process is quasi-left continuous.

Counting processes.

We will now review the Poisson process, and its generalization the counting process. These topics are needed in later chapters. Let $T = [0,\infty)$.

9. A real-valued stochastic process $(n_t, t \in T)$ is a counting process if:

   1. $n_0 = 0$,
   2. n is constant, except for positive unit jumps at random times,
   3. n has right continuous sample functions almost surely.

   Note that the above definition implies that n is integer valued. Since we use only the countability of the space of values of n we could have generalized the definition, however we have not done so.

10. Define the stopping times $\tau_m = \inf\{\infty, t \in T | n_t \geq m\}$, for $m = 1,2,\ldots$ . These stopping times are and will be called the jump times of the counting process n.

11. The counting process $n \in LIV^+$, is locally of integrable variation. This follows since $n_{t\wedge\tau_m} \leq m$.

The Poisson process.

12. A stochastic process $(n_t, t \in T)$ is a Poisson process with constant rate $\lambda$ if:

    1. n is a counting process,
    2. n has independent increments,
    3. $(n_t - n_s)$ has a Poisson distribution, with parameter $\lambda(t-s)$, where $\lambda > 0$ is a real-valued constant.

13. If $\lambda = 1$ we call n a standard Poisson process. A standard Poisson process n in $R^m$, will denote a vector valued process, whose components are independent standard Poisson processes.

14. The Poisson process has the following properties: 1. n has stationary independent increments, which is obvious from the definition, 2. $(n_t - t, F_{nt}, t \in T)$ is a martingale, with respect to the $\sigma$-field generated by it, 3. n is a Markov process. The Poisson process thus is in all the main classes of stochastic processes.

15. The martingale characterization: If $T = [0,\infty)$ then $(n_t - t, F_{nt}, t \in T) \in M^d_{2loc}$ and by definition $[n_t - t, n_t - t]_t = n_t$ and hence $\langle n_t - t, n_t - t\rangle_t = t$. If $T = [0,1]$ then $(n_t - t) \in M^d_2$. This implies that n and t are associated. The martingale property follows easily from the stationary independent increment property. Because n and t are both in $LIV^+$, $(n_t - t)$ must be a discontinuous martingale.

Because $\langle n_t - t, n_t - t\rangle_t = t$, which is bounded on $T = [0,1]$ we have $(n_t - t) \in M_2^d$ on $T = [0,1]$.

16. It is interesting to compare the Brownian motion and the Poisson process. Both are in all the three main classes of stochastic processes: Stationary independent increment processes, martingales and Markov processes. These classes were discussed in detail in Doob's book [1953]. The difference between Brownian motion and the Poisson process is in the character of the sample functions, the first is sample continuous, the second is discontinuous. This is also shown in that $w \in M_{2loc}^c$ and $(n_t - t) \in M_{2loc}^d$.

17. There are other characterizations of the Poisson process. Such points can be found in Ross [1970].

18. Since the Poisson process is a process with stationary independent increments it is a strong Markov process and a Hunt process. This implies that $(F_{nt}, t \in T)$ is quasi-left continuous.

19. Since $(n_t - t, F_{nt}, t \in T) \in M_{2loc}^d$ is adapted, by 2.3.15. $(n_t - t)$ charges only totally inaccessible stopping times, and this characterizes the jump times of n.

20. Using the martingale characterization we have the following equivalent condition: if the process $(n_t, F_t, t \in T)$ satisfies: 1. it is adapted, 2. it is a counting process, 3. $(n_t - t, F_t, t \in T) \in M_{loc}$, then n is a standard Poisson process. The proof will be deferred to the end of section 2.5. Note that 2 characterizes the sample functions of n and 3 gives the martingale characterization.

Counting processes

Using our characterization of the Poisson process, in terms of martingale theory, we will now seek a similar approach for arbitrary counting processes.

21. Given a counting process $(n_t, F_t, t \in T)$ that is adapted. Then there exists an unique predictable process $(q_t, F_t, t \in T) \in LIV^+$ such that $(n_t - q_t, F_t, t \in T) \in M^d_{2loc}$. The process q is just the dual predictable projection of n, the existence and uniqueness follows from 2.2.27. and 2.3.20. [Dellacherie,1972,V,T28].
If $(F_t, t \in T)$ is quasi-left continuous, then q is sample continuous.

22. Although q is sample continuous under certain circumstances, in general it is not absolutely continuous with respect to Lebesgue measure t. However in the case where this is true we define the following.

23. Given a counting process n, and an increasing family $(F_t, t \in T)$ satisfying the usual conditions, such that n is adapted to it.
If there exists a process $(\lambda_t, F_t, t \in T)$, satisfying $\lambda_t \geq 0$ a.s. for all $t \in T$, and $\int_T \lambda_s ds < \infty$ a.s., such that $(n_t - \int_0^t \lambda_s ds, F_t, t \in T) \in M_{loc}$, then we call $\lambda$ the rate of the counting process n with respect to $(F_t, t \in T)$.
Note that if such a process $\lambda$ exists, then it is unique almost surely by 2.2.27. It is necessary to specify the family of $\sigma$-fields with respect to the rate process, as will be seen later.
Since the process $q_t = \int_0^t \lambda_s ds$ is increasing, we see that it is necessary that $\lambda_t \geq 0$ a.s.

24. If $(F_t, t \in T)$ is quasi-left continuous, then the jump times of the counting process are totally inaccessible.

25. Let $m_t = n_t - \int_0^t \lambda_s ds$, then if $\lambda$ is the rate process,

$(m_t, F_t, t \in T) \in M^d_{2loc}$. Furthermore by definition $[m,m]_t = n_t$ and hence $\langle m,m\rangle_t = \int_0^t \lambda_s ds$.

## 2.5. Stochastic Integrals.

In this section we define stochastic integrals. We start with a short review on the development of stochastic integrals. We are interested in defining integrals of the form $\int_0^t \phi_s dx_s$. As usual we call $\phi$ the integrand process, and x the process with respect to which we integrate. If x is a process of bounded variation, and $\phi$ satisfies certain integrability conditions, then we can define the integral to be a Lebesgue-Stieltjes integral. However this does not work if x is not of bounded variation, which is the case if x is a Brownian motion process. The first to consider such integrals was Wiener, but he dealt with a limited case and did not really integrate the process. (See Wiener [1958]). Ito [1944] was the first one to define stochastic integrals. He considered the case where x is a Brownian motion process, and $\phi$ an adapted process in a suitable class. An important property of this integral is that it is martingale. Next we have to mention the decomposition theorems of Meyer for square integrable martingales. Using this result Kunita,Watanabe [1967] defined stochastic integrals with respect to square integrable martingales. Their work is based on earlier articles by Motoo, Watanabe [1965], and Watanabe [1964], where similar integrals were defined for functionals of a Markov process. Meyer [1967,I/IV] discusses the work of Kunita, Watanabe. The latest main contribution to the theory of stochastic integrals is the article by Doléans-Dade, Meyer [1970], where stochastic integrals are defined for arbitrary local-martingales and a general differentiation rule is given. See also the subsequent survey articles Meyer [1971a,b]. The main new idea is that stochastic integrals with respect to martingales should be considered as a mapping of martingales into martingales. The class of integrand processes which makes this true is precisely the class of predictable processes, satisfying certain integrability conditions. This idea of defining stochastic integrals to be martingales was published earlier by Millar [1968]. His work is an extension to the continuous-time case of some results by Burkholder [1966] for the discrete-time case. The forth coming book by Meyer on martingales and stochastic integrals of which the first two chapters have been published [1972], will undoubtedly contain these points in detail and should become a major reference.

Integrand processes.

We define several classes of integrand processes, we limit attention first to real-valued processes. The main reference for this chapter is [Doléans-Dade,Meyer,1970] which we will abbreviate by [DD-M,1970].

1. If $a \in IV$ then

   $L_1(a) = \{(\phi_t, t \in T) | \phi$ is adapted, predictable, and $E[\int_T |\phi_s| \cdot |da_s|] < \infty\}$. The family of $\sigma$-fields to which $\phi$ is adapted is specified in each individual case.

2. If $m \in M_{2loc}$ then $L_2(m) = \{\phi | \phi$ adapted, predictable, $E[\int_T |\phi_s|^2 d\langle m,m\rangle_s] < \infty\}$.

3. If $m \in M_{2loc}$ then $L_{2loc}(m) = \{\phi | \phi$ adapted, predictable, and there exists an increasing sequence of stopping times $\{\tau_n\}$, $\lim_n \tau_n = \infty$ a.s. such that for all n $E[\int_0^{\tau_n} |\phi_s|^2 d\langle m,m\rangle_s] < \infty\}$

4. If $m \in M_{1oc} \cap LIV$ then $L_{1loc}(m) \triangleq \{\phi | \phi$ adapted, predictable, and there exists an increasing sequence of stopping times

$\{\tau_n\}$, $\lim_n \tau_n = \infty$ a.s. such that for all n $E[\int_0^{\tau_n} |\phi_s|\cdot|dm_s|] < \infty\}$.

5. A process $\phi$ is called locally bounded if there exists an increasing sequence of stopping times $\{\tau_n\}$, $\lim_n \tau_n = \infty$ a.s., such that for all n, $|\phi_{t\wedge\tau_n} . I(\tau_n > 0)| \leq M_n < \infty$, where the $(M_n)$ are real positive constants.

6. $LB \triangleq \{\phi | \phi$ adapted, predictable and locally bounded$\}$.

7. A process $\phi \in LB$ is called a predictable locally bounded process. This class is somewhat restrictive, as has been indicated by Doléans-Dade,Meyer, but will undoubtedly be extended later on. This is also the reason for introducing the classes $L_{1loc}(m)$, $L_{2loc}(m)$ above.

8. The most usefull example of a locally bounded process, is the following, if x is a right continuous adapted process, having left hand limits then $(x_{t-}, t \in T) \in LB$. Note that if $a \in LIV$ or $m \in M_{loc}$, then $a_{t-}$ and $m_{t-} \in LB$.

9. For multi dimensional processes the definitions are similar using appropriate norms in $R^n$ and $(\phi_s^T \, d\langle m,m\rangle_s \, \phi_s)$ instead of $\phi_s^2 \, d\langle m,m\rangle_s$.

Stieltjes and stochastic integrals.

10. If $a \in IV$, $\phi \in L_1(a)$ then the Stieltjes integral $(\int_0^t \phi_s da_s, t \in T) \in IV$.

11. If $a \in BV$, $\phi \in LB$ then the Stieltjes integral $(\int_0^t \phi_s da_s, t \in T) \in BV$, is well defined [Meyer,1971b,D3].

12. If $m \in M_1 \cap IV$, $\phi \in L_1(m)$ then the Stieltjes integral $(\int_0^t \phi_s dm_s, F_t, t \in T) \in M_1 \cap IV$ [DD-M,1970,prop.2].

13. If $m \in M_2$, $\phi \in L_2(m)$, then there exists an unique element $(\phi.m) = (\int_0^t \phi_s dm_s, F_t, t \in T) \in M_2$, called the stochastic integral, such that for all $m_1 \in M_2$: $\langle(\phi.m), m_1\rangle_t = \int_0^t \phi_s d\langle m,m_1\rangle_s$ [DD-M,1970,th.3]. Also for all $m_1 \in M_2$: $[(\phi.m),m_1]_t = \int_0^t \phi_s d[m,m_1]_s$

[DD-M,1970,th.6].

14. If $m \in M_2 \cap IV$, $\phi \in L_2(m) \cap L_1(m)$ then the Stieltjes integral and the stochastic integral coincide [DD-M,1970,prop.3].

15. If $m \in M_{loc}$, $\phi \in LB$, then there exists an unique element $(\phi.m) = (\int_0^t \phi_s dm_s, F_t, t \in T) \in M_{loc}$, called the stochastic integral such that for all $m_1 \in M_{loc}$: $[(\phi.m),m_1]_t = \int_0^t \phi_s d[m,m_1]_s$ [DD-M,1970,prop.5].

16. If $m \in M_{2loc}$, $\phi \in L_{2loc}(m)$ then there exists an unique element $(\phi.m) \in M_{2loc}$ such that for all $m_1 \in M_{2loc}$, $\langle(\phi.m),m_1\rangle_t = \int_0^t \phi_s d\langle m,m_1\rangle_s$.

17. If $m \in M_{loc} \cap LIV$, $\phi \in L_{1loc}(m)$ then the Stieltjes integral $(\int_0^t \phi_s dm_s, F_t, t \in T) \in M_{loc} \cap LIV$.

18. The last two assertions follow by a stopping time argument from 13 respectively 12.

19. The extension of the above definitions to multi-dimensional processes is straightforward. Care should be taken in handling the quadratic variation process.

20. Examples, the definitions 16 and 17 are given to define: If w is a standard Brownian motion, $\phi \in L_{2loc}(w)$, then $(\int_0^t \phi_s dw_s, F_t, t \in T) \in M^c_{2loc} = M^c_{loc}$. If n is a standard Poisson process, then $(n_t - t, F_{nt}, t \in T) \in M^d_{2loc} \cap LIV$, $\langle n_t - t, n_t - t\rangle_t = t$. If $\phi \in L_{1loc}(n_t - t)$ then $(\int_0^t \phi_s (dn_s - ds), F_{nt}, t \in T) \in M^d_{loc} \cap LIV$. Note that $\phi \in L_{1loc}(t)$ implies that $\phi \in L_{1loc}(n_t - t)$ since $E[\int_0^{\tau_n} |\phi_s| \cdot |dn_s - ds|] \leq E[\int_0^{\tau_n} |\phi_s| \cdot (dn_s + ds)] = E[\int_0^{\tau_n} |\phi_s| \cdot 2ds] < \infty$.

21. Example: We can now give an example, that even if $m_1, m_2 \in M_{loc}$, but not in $M_{2loc}$, a predictable process $\langle m_1, m_2\rangle$ as defined in 2.3.31. exists. Let n be a standard Poisson process, then

$(n_t - t, F_{nt}, t \in T) \in M^d_{2loc}$. Let $\phi \in L_{1loc}(t)$, such that $m_t \triangleq \int_0^t \phi_s (dn_s - ds) \in M^d_{1loc}$ but not in $M^d_{2loc}$. Now $[m, n_t - t]_t = \int_0^t \phi_s dn_s$, which is in LIV by definition of $\phi$. It has the dual predictable projection $\langle m, n_t - t\rangle_t = \int_0^t \phi_s ds$.

22. Let $x \in SM$ with the decomposition $x_t = x_0 + a_t + m_t$, where $a \in BV$, $m \in M_{loc}$. Let $\phi \in LB$ then we define by 11 and 15 $\int_0^t \phi_s dx_s \triangleq \phi_0 \cdot x_0 + \int_0^t \phi_s da_s + \int_0^t \phi_s dm_s$. All terms are well defined and the integral is again a semi-martingale.

[Meyer,1971b.,D.8].

The differentiation rule.

23. If $x \in SM$ in $R^n$, and if $f : R^n \to R$ is a twice continuously differentable function, with $\frac{df}{dx} = (\frac{\partial f}{\partial x_1}, \ldots \frac{\partial f}{\partial x_n})$, $\frac{d^2 f}{dx^2} = (\frac{\partial^2 f}{\partial x_i \partial x_j})_{ij}$ then $f(x) \in SM$ of the form:

$$f(x_t) = f(x_0) + \int_0^t \frac{df}{dx}(x_{s-})\, dx_s + \int_0^t \frac{1}{2} \mathrm{Tr}[\frac{d^2 f}{dx^2}(x_{s-}) d\langle x^c, x^c\rangle_s]$$

$$+ \sum_{s \leq t} [f(x_s) - f(x_{s-}) - \frac{df}{dx}(x_{s-})\, \Delta x_s] \quad \text{[DD-M,1970,th.8].}$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix. From this equation the appropriate scalar and multi-dimensional function forms can be derived. An easy extension is to functions $f(t, x_t)$, where $f(t,x)$ is once continuously differentable in t. We give some special cases:

24. If $x, y \in SM$ in R then $x_t y_t = x_0 y_0 + \int_0^t x_{s-} dy_s + \int_0^t y_{s-} dx_s + [x,y]_t$ We will refer to this as the product rule.

25. A special case of this is if $m_1, m_2 \in M_{loc}$ then

$$m_{1t}\, m_{2t} - [m_1, m_2]_t = \int_0^t m_{1s-}\, dm_{2s} + \int_0^t m_{2s-}\, dm_{1s} \in M_{loc}.$$

26. Also if $m \in M_{loc}$ then $m_t^2 - [m,m]_t = \int_0^t 2m_{s-}\, dm_s$ which gives a characterization of $m^2 - [m,m]$ and also of $[m,m]$.

27. Example: We can now prove the sufficiency condition for a counting process n to be a Poisson process as stated in 2.4.20. Let $T = [0,1]$, the extension to $T = [0,\infty)$ follows by a stopping time argument. Since $n \in LIV^+$, it is a semi-martingale and by assumption $(n_t - t, F_t, t \in T) \in M_2^d$. We apply the differentiation rule: $e^{iun_t} = e^{iun_s} + \int_0^t iue^{iun_{\tau-}}\, dn_\tau + \sum_{s<\tau\leq t} (e^{iun_\tau} - e^{iun_{\tau-}} - iue^{iun_{\tau-}} \Delta n_\tau)$. Rearranging gives

$$E[e^{iu(n_t-n_s)}|F_s] = 1 + E[\sum_{s<\tau\leq t} e^{iu(n_{\tau-}-n_s)}(e^{iu\Delta n_\tau}-1)|F_s]$$

$$= 1 + E[\int_s^t e^{iu(n_{\tau-}-n_s)}(e^{iu}-1)dn_\tau|F_s]$$

$$= 1 + (e^{iu}-1)\int_s^t E[e^{iu(n_{\tau-}-n_s)}|F_s]d\tau$$

This is an integral equation with solution

$E[e^{iu(n_t-n_s)}|F_s] = \exp((t-s)(e^{iu}-1))$ This implies that n has stationary independent increments, and $(n_t - n_s)$ has a Poisson distribution with rate one, so n is a standard Poisson process.

## 2.6. Martingale representation.

In the previous section we have seen that stochastic integrals can be considered as a mapping of martingales into martingales. An important question now is, is this mapping onto, or in other words do certain martingales have a representation as a stochastic integral with respect to a given martingale. The first to consider this question implicitly was Ito [1951b]. He considered square integrable functionals, on the $\sigma$-field generated by a Brownian motion process, and obtained a representation as a stochastic integral with respect to the Brownian motion. Doob [1953,p.449] also considers this question, but does not specify the underlying process that generates the $\sigma$-field. Under certain conditions the existence of a Brownian motion is shown with respect to which the martingale representation exist. Wong [1971] has given an extension of this result to local martingales. The result by Ito was extended by Kunita, Watanabe [1967]. Their work is based on previous articles by Motoo, Watanabe [1965] and Watanabe [1964], who first derived a representation result for certain functionals of a Markov process. Kunita, Watanabe [1967] extended these results to square integrable martingales. The underlying process that generates the family of $\sigma$-fields to which the martingale is adapted, must be a Hunt process. The Brownian motion and the Poisson process satisfy this condition. Meyer [1967III] discusses these results. Clark [1970] independently derives a representation theorem for local martingales on the $\sigma$-field generated by a Brownian motion process. For the case where the underlying $\sigma$-field is generated by a Poisson process, the result by Kunita, Watanabe [1967] only holds for square integrable martingales. The extension to arbitrary local-martingales was done by Davis [to appear]

Let $(w_t, t \in T)$ be a standard Brownian motiom, and let $F_{wt} = \sigma(w_s, \forall s \leq t)$ be the $\sigma$-field generated by it. The family $(F_{wt}, t \in T)$ will be the underlying family of $\sigma$-fields in the following theorems.

1. Theorem: [Kunita, Watanabe, 1967; Clark, 1970]. If $(m_t, F_{wt}, t \in T) \in$ $\in M_2$, then m has the representation $m_t = \int_0^t \phi_s dw_s = (\phi \cdot w)_t$ a.s. for all $t \in T$ for an unique process $(\phi_t, F_{wt}, t \in T) \in L_2(w)$.

Then $m$ is sample continuous.

2. Theorem:

If $(m_t, F_{wt}, t \in T) \in M_{2loc}$ then $m$ has the representation $m = (\phi \cdot w)$ for an unique process $(\phi_t, F_{wt}, t \in T) \in L_{2loc}(w)$. Then $m$ is sample continuous.

3. Theorem: [Clark,1970].

If $(m_t, F_{wt}, t \in T) \in M_{loc}$ then $m$ has the representation $m = (\phi \cdot w)$ for an unique process $(\phi_t, F_{wt}, t \in T) \in L_{2loc}(w)$. Then $m$ is sample continuous.

4. Note that the given $m$ are such that $m_0 = 0$, otherwise consider $m_t - m_0$. Theorem 1 can also be deduced from Ito [1951]. Kunita, Watanabe's result is more general then 1. Theorem 2 can be deduced from 1. by a stopping time argument. Because of the difference of the classes $M_{loc}$ and $M_{2loc}$ we need result 3, however the resulting representations are the same. since $m$ will then sample continuous and $M^c_{loc} = M^c_{2loc}$. The sample continuity follows from the representation and cannot be asserted on fore hand.

5. Let $(n_t, t \in T)$ be a standard Poisson process, and let $(F_{nt}, t \in T)$ be the family generated by it, which will be the underlying family in the following theorems.

6. Theorem: [Kunita, Watanabe, 1967].

If $(m_t, F_{nt}, t \in T) \in M_2$ then $m$ has the representation $m_t = \int_0^t \psi_s(dn_s - ds) = (\psi \cdot (n_t - t))$ a.s. for all $t \in T$ for an unique process $(\psi_t, F_{nt}, t \in T) \in L_2(n_t - t)$. Then $m \in M_2^d$.

7. Theorem:

If $(m_t, F_{nt}, t \in T) \in M_{2loc}$ then $m$ has the representation $m = (\psi \cdot (n_t - t))$

for an unique process $(\psi_t, F_{nt}, t \in T) \in L_{2loc}(n_t - t)$. Then $m \in M^d_{2loc}$

8. Theorem: [Davis, to appear].

If $(m_t, F_{nt}, t \in T) \in M_{loc}$ then m has the representation $m = (\psi\cdot(n_t - t))$ for an unique process $(\psi_t, F_{nt}, t \in T) \in L_{1loc}(n_t - t)$. Then $m \in M^d_{1oc}$ and in LIV.

The martingale representation theorem plays a crucial role in derving the results of this thesis. If we could extend the representation result to a larger class of underlying processes that generate the family of σ-fields, then we could obtain many new results. A quite general representation theorem was given by Kunita, Watanabe [1967] but this works only if the martingale is square integrable, and if the underlying process is a Hunt process. The square integrability is an important restriction, and there are few examples of Hunt processes, except stationary independent increment processes. An extension of the martingale representation theorem is thus of great importance and a point of future research.

## 3. Absolute continuity of measures and related topics.

### 3.1. Introduction.

In this chapter we discuss problems of absolute continuity of measures and its relation with martingales. The main interest is in the translation of martingales by a change of measure. A partial converse problem is an abstract version of the detection problem which is well known in electrical engineering. We will analyse these problems using the martingale approach. An important concept is the exponential formula, which was introduced by Doléans-Dade [1970.a.]. Using this concept we will characterize a change of measure by a local martingale. Then we can prove the main result, theorem 3.3.5, of translation of local-martingales by a predictable process under a change of measure. This result is a generalization of a translation concept for Wiener integrals, introduced by Cameron and Martin [1944], and for Brownian motion by Girsanov [1960]. In the last section we will show how these results apply in a special case and we discuss the detection problem.

### 3.2. The exponential formula.

An important concept in the study of absolute continuity of measures is the solution to a certain stochastic differential equation, which is called the exponential formula. The sample continuous version was known earlier, but the most general case for martingales was solved by Doléans-Dade [1970 a.], which is the basic reference for this section. A special case of the exponential formula can be considered a martingale analogue of the concept of multiplicative functional of a Markov process.

3.2.1. Theorem: (Doléans-Dade, 1970 a.)

If $(x_t, F_t, t \in T)$ is a real valued semi-martingale, $x_0 = 0$,

then 1. There exists an unique semi-martingale, $(z_t, F_t, t \in T)$

satisfying $z_t = 1 + \int_0^t z_{s-} \; dx_s$

2. $z$ is given by $z_t = \exp\left(x_t - \frac{1}{2}\langle x^c, x^c\rangle_t\right) \prod_{s \le t} (1+\Delta x_s) e^{-\Delta x_s}$

Definition: We denote $z_t = \varepsilon(x_t)$ and call it the exponential formula of $x$.

Remarks:

1. The solution to the above stochastic differential equation is called the exponential formula, because it is similar to the differential equation for $f(x) = e^x$: $f(x) = 1 + \int_0^x f(y)\, dy$.

2. If $x \in M_{loc}$ then $\varepsilon(x) \in M_{loc}$, and if $x \in M^c_{loc}$ then $\varepsilon(x) \in M^c_{loc}$, a similar relation holds if $x \in M^d_{loc}$.

3. Note however that $x \in M_2$ does not imply that $z \in M_2$, it only gives that $z \in M_{loc}$.

4. Doléans-Dade also discussed when the exponential formula has a multiplicative decomposition. We will not state this result, but just note the special case: if $x \in M_{loc}$, then $z_t = \varepsilon(x_t)$ $= \varepsilon(x^c_t + x^d_t) = \varepsilon(x^c_t) \cdot \varepsilon(x^d_t)$, where $\varepsilon(x^c) \in M^c_{loc}$, $\varepsilon(x^d) \in M^d_{loc}$.

5. Some examples of exponential formula's are:

If $w$ is Brownian motion, then $\varepsilon(w_t) = \exp(w_t - \frac{1}{2}t) \in M^c_{loc}$.

If n is a standard Poisson process then

$$\varepsilon(n_t - t) = \exp(n_t \ell_n(2) - t) \in M^d_{loc}.$$

Strictly positive exponential formulas.

In the following section we need results concerning exponential formulas that are strictly positive. We will mention some conditions for this property and then state a converse result to the previous theorem. From here on we assume that $T = [0,1]$.

3.2.2. Lemma.

1. If $x \in SM$, $x_0 = 0$, $\Delta x_t > -1$ a.s. for all $t \in T$, then $z_t \triangleq \varepsilon(x_t) \in SM$ satisfies $z_t > 0$, $z_{t-} > 0$ a.s. for all $t \in T$. If $T = [0,1]$ then we require $\langle x^c, x^c \rangle_1 < \infty$ a.s.
2. If $(z_t, F_t, t \in T) \in M_1$, and if $z_1 > 0$ a.s. then $z_t$, $z_{t-} > 0$ a.s. for all $t \in T$ [Meyer, 1966, VI, T15]. If $z_t = \varepsilon(x_t)$, $x \in SM$, then $z_t$, $z_{t-} > 0$ a.s. for all $t \in T$ implies that $\Delta x_t > -1$ a.s. for all $t \in T$. If $T = [0,1]$ then $\langle x^c, x^c \rangle_1 < \infty$ a.s.

Proof. 1. From the expression for $z_t = \varepsilon(x_t)$ it follows that $\Delta x_t > -1$ a.s. implies that $z_t > 0$ a.s. Since $z_t = z_{t-} \ \Delta x_t$, we get $z_{t-} = z_t/\Delta x_t + 1 > 0$ a.s. 2. Similarly $\Delta x_t = (z_t/z_{t-}) -1 > -1$ a.s. If $T = [0,1]$ then $\varepsilon(x^c_1) > 0$ a.s. iff $\langle x^c, x^c \rangle_1 < \infty$ a.s. from the expression for $\varepsilon(x^c)$. ∎

In the case of real variables we can write $x = \exp(\ell n(x))$ if $x > 0$. We have the following analogous result in our case, which is mentioned in [Doléans-Dade, 1970a].

3.2.3. Theorem: If $(z_t, F_t, t \in T)$ is a semi-martingale, and $z_t$, $z_{t-} > 0$ a.s. for all $t \in T$ and $z_0 = 1$, then there exists a semi-martingale

$(x_t, F_t, t \in T) \in SM$, $x_0 = 0$, $\Delta x_t > -1$ a.s. for all $t \in T$, such that $z_t = \varepsilon(x_t)$.

Proof. This follows by setting $x_t \triangleq \int_0^t \frac{1}{z_{s-}} \, dz_s$, which is well defined, and applying theorem 3.2.1. to $dzt = z_{t-} \, dx_t$. Another way of getting this result is taking $\ell n(z_t)$ and applying the differentiation rule. This will yield the exponential formula directly whithout refering to 3.2.1.

Remarks:

1. In the case of a positive exponential formula we have an alternative way of writing it, which one can use. Let $x \in M_{1oc}$ then $\varepsilon(x_t) = \exp(x_t^c - \frac{1}{2} \langle x^c, x^c \rangle_t) \exp(x_t^d - \sum_{s \leq t} [\Delta x_s - \ell n(1+\Delta x_s)])$.

2. If $z \in M_{1oc}$, or $z \in M_1$ and $z = \varepsilon(x)$, then $x \in M_{1oc}$. Note however that $z \in M_1$ does not imply in general that $x \in M_{21oc}$, only if $z \in M_1^c$ then $x \in M_{21oc}^c$.

3. Some examples. If w is Standard Brownian motion and $\phi \in LB$, then $\varepsilon(\int_0^t \phi_s \, dw_s) = \exp(\int_0^t \phi_s \, dw_s - \int_0^t \frac{1}{2} \phi_s^2 \, ds)$. If n is a standard Poisson process, $\psi \in LB$ and $\psi_t > 0$ a.s. for all $t \in T$ then $\varepsilon(\int_0^t (\psi_s - 1)(dn_s - ds)) = \exp(\int_0^t \ell n(\psi_s) \, dn_s - \int_0^t (\psi_s - 1) \, ds)$. Since $\psi_t > 0$, $\Delta \int_0^t (\psi_s - 1)(dn_s - ds) = (\psi_t - 1) \, \Delta n_t > -1$ a.s., so the exponential formula is strictly positive.

Martingale exponential formula's.

A problem in some applications in the next section is, when is $\varepsilon(x)$ a martingale. We will pay some attention to this problem.

3.2.4. Lemma: If $(x_t, F_t, t \in T) \in M_{loc}$, $x_0 = 0$, and $\varepsilon(x_1) > 0$ a.s. then $\varepsilon(x) \in M_1$ iff $E[\varepsilon(x_1)] = 1$.

Proof: If $\varepsilon(x) \in M_1$, then $E[\varepsilon(x_1)] = E[\varepsilon(x_0)] = 1$

Suppose $E[\varepsilon(x_1)] = 1$. Let $\mu_t \overset{\Delta}{=} E[\varepsilon(x_1)|F_t]$, then $(\mu_t, F_t, t \in T) \in M_1$, and we show that $\mu_t = \varepsilon(x_t) \overset{\Delta}{=} z_t$ a.s. Since $z \in M_{loc}$, let $\{\tau n\} \uparrow 1$ a.s. be such that $z_{t\wedge\tau_n} \in M_1$. If $t > s$ then $E[z_{t\wedge\tau_n}|F_s] = z_{s\wedge\tau_n}$ and by Fatou's lemma $E[z_t\ |F_s] \leq z_s$ or $z$ is a super martingale. This gives $\mu_t = E[\varepsilon(x_1)|F_t] \leq z_t$ and $E(z_t) \leq E(z_s)$. Since $E[\varepsilon(x_1)] = E[\varepsilon(x_o)] = 1$ this gives $E[\varepsilon(x_t)] = 1$ for all $t \in T$. Now $E[\mu_t] = E[\varepsilon(x_1)] = 1 = E[\varepsilon(x_t)]$. So $E[\mu_t - \varepsilon(x_t)] = 0$ and $\mu_t - \varepsilon(x_t) \leq 0$ imply $\mu_t = \varepsilon(x_t)$ a.s.

3.2.5. Theorem: $T = [0,1]$

If 1. $(x_t, F_t, t \in T) \in M_{loc}$, $x_0 = 0$, $\Delta x_t > -1$ a.s. for all $t \in T$, $\langle x^c, x^c\rangle_1 < \infty$ a.s.,

2. $(\langle x,x\rangle_t, F_t, t \in T) \in LIV$ exists, the dual predictable projection of $[x,x]$, and satisfies $d\langle x,x\rangle_t = \psi_t\, dt$, where $(\psi_t, F_t, t \in T) \in L_{1loc}(t)$ satisfies $|\psi_t| \leq K(t)$ a.s. for all $t \in T$, for some positive valued function $K: T \to R$

then $E[\varepsilon(x_1)] = 1$.

Proof. Let $z_t \overset{\Delta}{=} \varepsilon(x_t)$. By definition of $z = \varepsilon(x) \in M_{loc}$, there exists an increasing sequence of stopping times $\{\tau_n\}$, $\lim_n \tau_n = 1$ a.s., such that $z_{t\wedge\tau_n} \in M_1$. Now

$E[z_{t\wedge\tau_n}] = 1 + E[\int_0^{t\wedge\tau_n} z_{s-}\ dx_s] = 1$, hence by Fatou's lemma

$E(z_t) \leq 1$ for all $t \in T$. Define the stopping times

$\tau_n = \inf\{1, t \in T \mid [z,z]_t \geq n\}$, hence $\lim_n \tau_n = 1$ a.s. We use the

differentiation rule:

$$z_t^2 = 1 + \int_0^t 2\ z_{s-}^2\ dx_s + \int_0^t z_{s-}^2\ d[x,x]_s$$

Let $y_t^n = z_t^2\ I(t<\tau_n)$, then $y_t^n \leq n$. Now

$$y_t^n \leq 1 + \int_0^t 2\ y_{s-}^n\ dx_s + \int_0^t y_{s-}^n\ d[x,x],$$

$$E(y_t^n) \leq 1 + E[\int_0^t y_{s-}^n\ d[x,x]_s] = 1 + E[\int_0^t y_{s-}^n\ d\langle x,x\rangle_s]$$

since the local martingale term vanishes, and by definition of $\langle x,x\rangle$. Using condition 2:

$$E[\int_0^t y_{s-}^n\ d\langle x,x\rangle_s] = E\int_0^t y_{s-}^n\ \psi_s\ ds] \leq \int_0^t K(s)\ E[y_{s-}^n]\ ds$$

Now $E[y_t^n] \leq 1 + \int_0^t K(s)\ E(y_{s-}^n)\ ds$ implies that by the Bellman Gronwall lemma $E(y_t^n) \leq \exp(\int_0^t K(s)\ ds) \leq \exp(\int_0^1 K(s)\ ds)$.

Since $\lim_n \tau_n = 1$ a.s. $E(z_t^2) \leq \exp(\int_0^1 K(s)\ ds)$ for all $t < 1$.

Also $E(z_1)^2 = 1 + E[\int_0^1 z_{s-}^2\ d[x,x]_s] \leq 1 + \int_0^1 E(z_{s-}^2)\ K(s)\ ds < \infty$.

Let $\{\tau_n\}$ be such that $z_{t\wedge\tau_n} \in M_1$. Thus

$E(z_t^n) = E(z_{t\wedge\tau_n}) = E(z_o^n) = 1.$ for all $t \in T$.

Since $\sup_n E(z^2_{t\wedge\tau_n}) \leq \infty$ by the above, by [Meyer, 1966, II T22], the $\{z^n\}$ are uniformly integrable. By [Meyer, 1966, II T21] and because $z^n_t \geq 0$ a.s. we get

$\lim_{n\to\infty} E(z^n_t) = E(z_t) = 1$ for all $t \in T$, hence $E[\varepsilon(x_1)] = 1$.

3.2.6. Corollary:

If $w$ is standard Brownian motion, $\phi \in LB$ and if $|\phi_t|^2 \leq K$ a.s. for all $t \in T$ where $K$ is a positive constant, then $E[\varepsilon(\int_0^1 \phi_s\, dw_s)] = 1$.

The proof is obvious by $dx_t \triangleq \phi_t\, dt$, $d\langle x,x\rangle_t = \phi^2_t\, dt$ and using 3.2.5. This result was proven earlier by [Girsanov, 1960].

Similarly we have

3.2.7. Corollary:

If $n$ is a standard Poisson process $(\psi_t, F_{nt}, t \in T) \in LB$, and if $|\psi_t|^2 \leq K$ a.s. for all $t \in T$, where $K$ is a positive constant, then

$E[\varepsilon(\int_0^1 \psi_s\, (dn_s - ds))] = 1$.

Remarks:

1. The proof of 3.2.5. is essentially an extension of the proof by Girsanov [1960] for lemma 1. The main point is to show that $\varepsilon(x) \in M_2$, which establishes the uniform integrability of $\varepsilon(x)$, which give the result.

2. An example: Let $(w_t, t\in[0,1])$ be standard Brownian motion, then $w \in M^c_{loc}$, $\langle w,w\rangle_t = t$. Hence

$E[\varepsilon(w_t)] = E[\exp(w_t - \frac{1}{2}t)] = 1$ for all $t \in [0,1]$.

3.3. Absolute continuity and translation of martingales.

In this section we discuss some results on absolute continuity of measures and its relation to martingales. We start by characterizing a transformation of measure by a local martingale. Next we present our main result, the translation of a local martingale by a predictable process into a new local martingale under a transformation of measure. Some historical comments will be provided later. The result we give here is widely used in problems of absolute continuity, especially showing existence of solutions to stochastic differential equations and in detection problems.

Absolute continuity.

3.3.1. Definition. Given the measurable space $(\Omega,F)$ and two probability measures $P$ and $P_0$ defined on it. The measure $P$ is said to be absolutely continuous with respect to $P_0$, if for all $A \in F$ such that $P_0(A) = 0$ we have that $P(A) = 0$. We denote this by $P << P_0$. $P$ and $P_0$ are said to be mutually absolutely continuous, or equivalent, if $P << P_0$ and $P_0 << P$ , which we denote by $P \sim P_0$.

If $P << P_0$ then the Radon-Nikodym theorem says that there exists a measurable integrable function such that $P(A) = \int_A \mu(w)\, P_0(dw)$.

A major point of interest is the characterization of the Radon-Nikodym derivative $\mu$, which we also denote by

$\mu = \frac{dP}{dP_0}$. If $P \sim P_0$ then $\mu > 0$ a.s. In the following we denote expectation with respect to the measures $P_0$, $P$ by $E_0(\cdot)$, $E(\cdot)$ respectively. Using the martingale approach and the results concerning the exponential formula we have the following characterization of $\frac{dP}{dP_0}$.

3.3.2. Theorem:

1. Given a probability space $(\Omega,F,P_0)$. Let $(x_t,F_t,t\in T)\in M_{loc}$, be such that $x_0=0$, $\langle x^c,x^c\rangle_1<\infty$ a.s., $\Delta x_t>-1$ a.s. for all $t\in T$, and $E_0[\varepsilon(x_1)]=1$, then $(\varepsilon(x_t),F_t,t\in T)\in M_1$. The formula $\frac{dP}{dP_0}=\varepsilon(x_1)$ introduces a new probability measure P on $(\Omega,F)$ and P is equivalent to $P_0$.

2. Given a measurable space $(\Omega,F)$ and two probability measures P and $P_0$ defined on it, and assume that P and $P_0$ are equivalent. Then $\frac{dP}{dP_0}>0$ a.s. Let $(F_t,t\in T)$ be any family of sub-$\sigma$-fields with the usual conditions. Let $\mu_t \triangleq E_0[\frac{dP}{dP_0}|F_t]$. Then there exists a process $(\hat{x}_t,F_t,t\in T)\in M_{loc}$, $\hat{x}_0=0$, $\Delta\hat{x}_t>-1$ a.s. for all $t\in T$, $\langle\hat{x}^c,\hat{x}^c\rangle_1<\infty$ a.s., such that $\mu_t=\varepsilon(\hat{x}_t)$ a.s. for all $t\in T$.

Remark: The above theorem shows that a local martingale x, satisfying certain conditions introduces a new probability measure. Conversely the estimate of the Radon-Nikodym derivative of two measures given some family of $\sigma$-fields is characterized by a local martingale $\hat{x}$.

Proof. 1. Given $(x_t,F_t,t\in T)\in M_{loc}$, we have that $\varepsilon(x)\in M_{loc}$ and $\varepsilon(x_1)>0$ by the conditions assumed. The condition $E_0[\varepsilon(x_1)]=1$ guarantees that in fact $\varepsilon(x)\in M_1$ as was shown in 3.2.4. The set function $P(A)=\int_A \varepsilon(x_1)(w)\,P_0(dw)$ now defines a probability.

2. Note that $\frac{dP}{dP_0} > 0$ a.s. by assumption. Let

$\mu_t \overset{\Delta}{=} E_0[\frac{dP}{dP_0} |F_t]$, then $(\mu_t, F_t, t \in T) \in M_1$, and $\mu_1 > 0$ a.s. implies

that $\mu_t$, $\mu_{t-} > 0$ a.s. for all $t \in T$. By 3.2.3. there exists

$(\hat{x}_t, F_t, t \in T) \in M_{loc}$ such that $\mu_t = \varepsilon(\hat{x}_t)$ and $\hat{x}$ has the given

properties.

Remark: The main reason for using absolute continuity of measures is in calculating expectations or in doing derivations based on it. Suppose that z is an integrable random variable, let $P \sim P_0$ then $E[z] = E_0[z\varepsilon(x_1)]$. The last integral might in some cases be easier to integrate than the first. This point was made earlier by Cameron and Martin [1944] and by Benes [1971], and is the main argument for the results of this section.

Translation of martingales.

We now are going to look at how a transformation of probability measure influences martingales. Before reaching our main result we have to do some preliminary work.

3.3.3. Definition. The notation $(x_t, F_t, t \in T) \in M_{loc}\ (P_0)$ denotes that $x$ is a local martingale under the measure $P_0$.

3.3.4. Lemma. If P, $P_0$ are equivalent probability measures on $(\Omega, F)$ and

if $\frac{dP}{dP_0} = \varepsilon(x_1)$ where $(x_t, F_t, t \in T) \in M_{loc}(P_0)$

then $(m_t, F_t, t \in T) \in M_1(P)$ iff $(m_t\ \varepsilon(x_t), F_t, t \in T) \in M_1\ (P_0)$

Similarly $m \in M_{loc}\ (P)$ iff $m\ \varepsilon(x) \in M_{loc}\ (P_0)$.

Proof. We do the proof in one direction, assume that $(m_t\ \varepsilon(x_t),$ $F_t, t \in T) \in M\ (P_0)$. Since P and $P_0$ are equivalent

$$\frac{dP}{dP_0} = \varepsilon(x_1) > 0 \text{ a.s.}$$

Note that $E|mt| = E_0[\varepsilon(x_1)\,|m_t|] = E_0[|m_t\ \varepsilon(x_1)|] < \infty$ for all $t \in T$.
Now for any s, $t \in T$, by [Loève, 1963, p. 344]

$$E[m_t|F_s] = \frac{E_0[m_t\ \varepsilon(x_1)|F_s]}{E_0[\varepsilon(x_1)|F_s]}.$$

Let $s < t$ then

$$E[m_t|F_s] = E_0[m_t\varepsilon(x_t)|F_s]/\varepsilon(x_s) = m_s$$

hence $(m_t, F_t, t \in T) \in M_1(P)$

The converse direction is similar and the extension to local martingales follows from a stopping time argument.

Remark: We discuss a condition assumed in the following theorem. If x, $y \in M_{loc}$ then $[x,y] \in BV$. In 2.3.31. we have defined $\langle x,y\rangle \in BV$ to be the dual predictable projection of $[x,y]$, whose existence follows if $[x,y] \in LIV$.

3.3.5. Theorem.

If 1. $(\Omega, F, P_0)$ is a probability space.

2. We do a transformation of measure $\frac{dP}{dP_0} = \varepsilon(x_1)$, characterized by the local martingale $(x_t, F_t, t \in T) \in M_{loc}(P_0)$, real-valued, $x_0 = 0, \langle x^c, x^c\rangle_1 < \infty$ a.s. and $\Delta x_t > -1$ a.s. for all $t \in T$ and satisfying $E_0[\varepsilon(x_1)] = 1$,

3. $(y_t, F_t, t \in T) \in M_{loc}(P_0)$ in $R^n$,

4. there exists a process, denoted $(\langle y,x\rangle_t, F_t, t \in T) \in BV$, predictable, such that $([y,x]_t - \langle y,x\rangle_t, F_t, t \in T) \in M_{loc}(P_0)$

then P is a probability measure on $(\Omega.F)$, and the process m defined by $m_t \triangleq y_t - \langle y,x\rangle_t$ satisfies $(m_t, F_t, t \in T) \in M_{loc}(P)$, i.e. m is a local martingale under the measure P. If in addition $\langle y,x\rangle$ is sample continuous, then $[m,m] = [y,y]$.

Proof. By condition 2, and theorem 3.3.2, P is a probability measure on $(\Omega, F)$. To prove that $m \in M_{loc}(P)$ by lemma 3.3.4. it suffices to show that $m\varepsilon(x) \in M_{loc}(P_0)$. We apply the differentiation rule to $m\varepsilon(x)$:

$$m_t\varepsilon(x_t) = \int_0^t m_{s-}\, d\varepsilon(x_s) + \int_0^t \varepsilon(x_s)_- dm_s + [m, \varepsilon(x)]_t$$

$$= \int_0^t m_{s-}\, \varepsilon(x_s)_-\, dx_s + \int_0^t \varepsilon(x_s)_-\, dy_s - \int_0^t \varepsilon(x_{s-})\, d\langle y,x\rangle_s$$

$$+ \int_0^t \varepsilon(x_s)_-\, d[m,x]_s$$

Note that under $P_0$ $m_t = y_t - \langle y,x\rangle_t$ is a semi-martingale, where $y \in M_{loc}$, $\langle y,x\rangle \in BV$, so $m_t^c = y_t^c$ and $\Delta m_t = \Delta y_t - \Delta\langle y,x\rangle_t$

Now $[m,x]_t = \langle m^c, x^c\rangle_t + \sum_{s\leq t} \Delta m_s \Delta x_s = [y,x]_t - \sum_{s\leq t} \Delta\langle y,x\rangle_s \Delta x_s$

$$= [y,x]_t - \int_0^t \Delta\langle y,x\rangle_s\, dx_s =$$

$$= [y,x]_t - \int_0^t (\langle y,x\rangle_s - \langle y,x\rangle_{s-})\, dx_s.$$

$$m_t\varepsilon(x_t) = \int_0^t m_{s-}\, \varepsilon(x_s)_-\, dx_s + \int_0^t \varepsilon(x_s)_-\, dy_s +$$

$$+ \int_0^t \varepsilon(x_s)_-\, d([y,x]_s - \langle y,x\rangle_s) - \int_0^t \varepsilon(x_s)_-\, (\langle y,x\rangle_s - \langle y,x\rangle_{s-})\, dx_s$$

Since this is a sum of stochastic integrals of predictable processes with respect to local martingales, $m\,\varepsilon(x) \in M_{loc}(P_0)$ hence $m \in M_{loc}(P)$. If in addition $\langle y,x \rangle$ is sample continuous, then by 2.3.38. $[m,m] = [y,y]$.

Remarks:

1. Condition 4 in the theorem is satisfied in the following cases:

   1. If $x,y \in M_{2loc}$ then $\langle y,x \rangle$ exists, see 2.3.25.
   2. If $x,y \in M^c_{loc}$ then $\langle y,x \rangle$ exists and $[y,x] - \langle y,x \rangle = 0$ a.s.
   3. If $[y^d,X^d] = 0$ a.s. then $\langle y^c,x^c \rangle$ will do, since $[y,x] - \langle y^c,x^c \rangle = 0$ a.s.
   4. In general the existence of the process $\langle y,x \rangle$ must be shown, this is most easily done by first calculating $[y,x]$ and then guessing the form of the process $\langle y,x \rangle$.

2. The foregoing result does not hold if it is changed to the form $m_t \triangleq y_t - [y,x]_t$.

   Example: let $x_t = y_t = n_t - t$, where $(n_t, F_{nt}, t \in T)$ is a standard Poisson process. Then after a transformation of probability $m_t = y_t - [y,x]_t = (n_t - t) - n_t = -t$ which is not a local martingale.

3. The theorem as given here seems the most general result possible, and it includes many earlier versions as special cases. It seems by the remark above that the predictability of $\langle y,x \rangle$ is required. Another way of looking at the foregoing result is considering it as a transformation of a martingale y, into a semi-martingale $y = m + \langle y,x \rangle$, where the associated

process of bounded variation is predictable.

4. Some historic comments. The concept of translation was first introduced by Cameron and Martin [1944] in the context of Wiener integrals. The translation was that of the Wiener process y into the process $y - \langle y,x \rangle$. Using the definition of a stochastic Ito integral for a Brownian motion process, Girsanov [1960] gave a similar approach for Brownian motion processes (Corollary 3.4.3), which he called the transformation of a stochastic process. This result is widely used in problems of absolute continuity, for the Brownian motion process. The extension of Girsanov's result, first to Poisson processes [Brémaud, 1972] and then to the martingale case [Van Schuppen, Wong, to appear], was done in cooperation with Wong.

5. We give some general examples.

Let $x \in M_{loc}(P_0)$ be given, and characterize the transformation of measure. We consider different forms for y.

1. $y = x \in M_{2loc}(P_0)$, then $m_{1t} = x_t - \langle x,x \rangle_t \in M_{loc}(P)$

2. If $y_t = \int_0^t \phi_s \, dx_s \in M_{loc}(P_0)$, $x \in M_{2loc}(P_0)$ then

$$m_{2t} = \int_0^t \phi_s \, dx_s - \int_0^t \phi_s \, d\langle x,x \rangle = \int_0^t \phi_s \, dm_{1s}$$

3. If $x_t = \int_0^t \psi_s \, dz_s$ where $(z_t, F_t, t \in T) \in M_{2loc}(P_0)$

then $m_{3t} = z_t - \int_0^t \psi_s \, d\langle z,z \rangle_s \in M_{loc}(P)$

4. If $x_t = \int_0^t \psi_s \, dz_s$, $y_t = \int_0^t \phi_s \, dz_s$, where $(z_t, F_t, t \in T) \in M_{2loc}(P_0)$

Then $m_{4t} = \int_0^t \phi_s \, dz_s - \int_0^t \phi_s \psi_s \, d\langle z,z\rangle_s = \int_0^t \phi_s (dz_s - \psi_s \, d\langle z,z\rangle_s)$

$= \int_0^t \phi_s \, dm_{3s}$

6. Now we have stated the translation theorem one could consider a converse problem. Given two measures $P_0$, and P under one y is martingale, under the other y is a semi-martingale, under which conditions are P and $P_0$ equivalent and what is the characterization of the Radon-Nikodym derivative. This problem cannot be solved whithout additional assumptions, we will discuss this problem in the next section.

Using the translation theorem we can give a more general condition on a local martingale x such that $E[\varepsilon(x_1)] = 1$. The following proof is inspired by a similar proof for the Brownian motion case by Clark.

3.3.6. Theorem: $T = [0,1]$

If 1. $(x_t, F_t, t \in T) \in M^c_{loc}$, $x_0 = 0$,

2. $\langle x,x\rangle_1 \leq K$ a.s. for some positive constant K,

then $E[\varepsilon(x_1)] = 1$.

Proof. From 3.2.5. $E[\varepsilon(x_t)] \leq 1$ for all $t \in T$. Note that $z_t \triangleq \varepsilon(x_t) = \exp(x_t - \frac{1}{2}\langle x,x\rangle_t) \in M^c_{loc} = M^c_{2loc}$. Let $(\tau_n)$ be an increasing sequence of stopping times, $\lim_n \tau_n = 1$ a.s., such that for all n $z^n_t = z_{t\wedge\tau_n} \in M_2$. Since $\sup_{t\in T} E(z^n_t)^2 < \infty$, for fixed n, we conclude by uniform integrability that for all n $E[z^n_t] = 1$ for all $t \in T$.

Next we establish the uniform integrability of $(z_1^n, n=1,2,\ldots)$.

First $\sup_n E(z_1^n) \leq 1$. Fix n, then $E(z_1^n) = 1$, so we can define

the probability measure $\frac{dP_n}{dP} = z_1^n$ on $(\Omega, F)$. Let $x_t^n = x_{t\wedge\tau_n}$.

Now $\int_{(z_1^n \geq c)} z_1^n \, dP(w) = P_n(z_1^n \geq c) = P_n(x_1^n - \frac{1}{2}\langle x^n, x^n\rangle_1 \geq \log c)$

$= P_n(x_1^n - \langle x^n, x^n\rangle_1 + \frac{1}{2}\langle x^n, x^n\rangle_1 \geq \log c) \quad \leq$

$P_n(x_1^n - \langle x^n, x^n\rangle_1 \geq \frac{1}{2}\log c) + P(\langle x^n, x^n\rangle \geq \log c)$

By the translation theorems $y_t^n \triangleq x_t^n - \langle x^n, x^n\rangle_t$,

$(y_t^n, F_t, t \in T) \in M_{loc}^c(P_n)$. Also $\langle y^n, y^n\rangle = \langle x^n, x^n\rangle$ since $x^n \in M_{loc}^c(P)$.

Now $\sup_{t\in T} E_n(y_t^n)^2 \leq \sup_{t\in T} E_n(\langle y^n, y^n\rangle_t) \leq K$, so $y^n \in M_2^c(P_n)$.

We now use the martingale inequality by Doob:

$P_n(x_1^n - \langle x^n, x^n\rangle_1 \geq \frac{1}{2}\log c) = P_n(y_1^n \geq \frac{1}{2}\log c) \leq$

$$P_n(\sup_{t\in T}|y_t^n| \geq \frac{1}{2}\log c) \leq \frac{4\, E_n|y_1^n|^2}{\log^2 c} \leq \frac{4\, E_n(\langle y^n, y^n\rangle_1)}{\log^2 c} \leq \frac{4\, K}{\log^2 c}$$

Note that if $\log c > K$ or $c > e^K$ then

$P_n(\langle x^n, x^n\rangle_1 \geq \log c) = 0$ for all n. Now

$$\sup_n \int_{(z_1^n \geq c)} z_1^n \, dP(w) \leq \frac{4K}{\log^2 c} \xrightarrow{c\to\infty} 0 \text{ independently of } n.$$

This establishes the uniform integrability of $(z_1^n, n=1,2,\ldots)$.

Now by [Meyer, 1966, II T21]

$\lim_{n\to\infty} E(z_1^n) = 1 = E(z_1) = E[\varepsilon(x_1)]$.

## 3.4. Applications of the translation theorem.

In this section we discuss an important special case of the translation theorem, this result will be used in the next section for showing existence of solution to a certain stochastic differential equation. Its two corollaries, one for the Brownian motion process, which is known as Girsanov's theorem, and one for counting processes, first given by Brémaud, are stated. Next we discuss an abstract version of the detection problem.

3.4.1. Theorem:

If 1. $(\Omega, F, P_0)$ is a probability space, $T = [0,1]$,

2. $(v_t, F_t, t \in T) \in M_{2loc}(P_0)$ in $R^n$, and $(F_t, t \in T)$ is quasi left continuous,

3. $(\phi_t, F_t, t \in T) \in LB$ in $R^n$, $\int_0^1 \phi_s^T \, d\langle v^c, v^c\rangle_s \, \phi_s < \infty$ a.s.,

$(\psi_t, F_t, t \in T) \in LB$ in $R^n$, and $\psi_t^T \, \Delta v_t + 1 > 0$ a.s. for all $t \in T$.

4. Define $z_t \triangleq \int_0^t \phi_s^T \; dv_s^c + \int_0^t \psi_s^T \; dv_s^d$, then $(z_t, F_t, t \in T) \in M_{loc}(P_0)$

5. Let $\phi$ and $\psi$ be such that $E_0[\varepsilon(z_1)] = 1$,

6. We introduce a new measure P on $(\Omega, F)$ by $\dfrac{dP}{dP_0} = \varepsilon(z_1)$,

then 1. P is a probability on $(\Omega, F)$,

2. $m_t \triangleq v_t - \langle v, z\rangle_t = v_t - \int_0^t d\,\langle v^c, v^c\rangle_s \, \phi_s - \int_0^t d\langle v^d, v^d\rangle_s \, \psi_s$,

$(m_t, F_t, t \in T) \in M_{loc}(P)$, $[m,m] = [v,v]$, $m = m^c + m^d$,

3. $m^c_t \triangleq v^c_t - \langle v^c, z^c \rangle_t = v^c_t - \int_0^t d\langle v^c, v^c \rangle_s \, \phi_s,$

$(m^c_t, F_t, t \in T) \in M^c_{loc}(P), \; \langle m^c, m^c \rangle = \langle v^c, v^c \rangle.$

4. $m^d_t = v^d_t - \langle v^d, z^d \rangle_t = v^d_t - \int_0^t d\langle v^d, v^d \rangle_s \, \psi_s,$

$(m^d_t, F_t, t \in T) \in M^d_{loc}(P), \; [m^d, m^d] = [v^d, v^d].$

Proof. We apply 3.3.5. and check its conditions:

$$[v,z]_t = \int_0^t d\langle v^c, v^c \rangle_s \, \phi_s + \int_0^t d[v^d, v^d]_s \, \psi_s,$$

where $\phi \in R^n$ and $\langle v^c, v^c \rangle$ takes values in $R^{n \times n}$.

Now $\langle v, z \rangle_t = \int_0^t d\langle v^c, v^c \rangle_s \, \phi_s + \int_0^t d\langle v^d, v^d \rangle_s \, \psi_s$ since it is the dual predictable projection of $[v,z]$. Assertion 1 and 2 now follow from 3.3.5. Because $(F_t, t \in T)$ is quasi left continuous by assumption, $\langle v,z \rangle$ is sample continuous, and hence by 2.3.38 $[m,m] = [v,v]$. If $m^c$ is defined as in 3. then by applying 3.3.5 again we see that $m^c \in M^c_{loc}(P)$. Now $m^d \triangleq m - m^c \in M_{loc}(P)$. Let $k \in M^c_{loc}(P)$ be arbitrary. Consider the semi-martingale $m^d$ under $P_0$, where $\langle v^d, z^d \rangle$ is sample continuous, and let us apply 2.3.38.

$$[m^d, k]_t = [v^d, k]_t = \sum_{s \leq t} (\Delta v^d_s)(\Delta k_s)^T = 0 \text{ a.s., hence}$$

$m^d (k)^T \in M_{loc}$ and thus $m^d \in M^d_{loc}(P)$.

Remarks:

There are several points to note in the previous theorem.

1. First we wanted to make explicit the decomposition of the

martingales into their continuous and discontinuous part, and the characterization of the translation. Note that $m^c$ is characterized by $v^c$ and $z^c$ and a similar characterization holds for $m^d$. Another point we want to mention is the special form of the martingale z as a stochastic integral with respect to v. Note that is is necessary that $v \in M_{2loc}$.

2. We discuss the condition $\phi_t^T \Delta v_t + 1 > 0$ a.s. for all $t \in T$. If v has only jumps of a fixed height $a \in R$, then we can easily find a condition on $\psi$. If the jumps are bounded we can find a condition too, but if v has arbitrary positive jumps, then it is better to take $\psi_t > 0$ a.s. for all $t \in T$. Remark that it is necessary to distinguish between the positive and negative jumps of v.
3. The above theorem leaves an important question open: when is $m^d \in M^d_{2loc}(P)$ and if so, what is the process $\langle m^d, m^d \rangle$ in terms of processes defined previously. This question can easily be answered if v has only jumps of a fixed height. The proof is straightforward, one considers the translation of the local martingale $[v,v] - \langle v,v \rangle$. In general, when v has arbitrary jumps we cannot solve this question. However this problem can be approached better if we analyse discontinuous martingales using the Levy measure [Watanabe, 1964], but this theory has some difficult and non-rigorous points, so that we have decided to omit this.

We give two main applications, first the Brownian motion case.

3.4.2. Corollary: [Girsanov, 1960]

If 1. $(\Omega, F, P_0)$ is a probability space.

2. $(w_t, F_t, t \in T)$ is a sample continuous Brownian motion, in $R^n$,

3. $(\phi_t, F_t, t \in T) \in L_{2loc}(w)$ in $R^n$, $\int_0^1 |\phi_s|^2 \, ds < \infty$ a.s. $P_0$,

and satisfies $E_0[\varepsilon(\int_0^1 \phi_s^T \, dw_s)] = 1$

4. we introduce a new measure $\frac{dP}{dP_0} = \varepsilon(\int_0^1 \phi_s^T \, dw_s)$

then P is a probability measure,

$$m_t \triangleq w_t - \int_0^t \phi_s \, ds$$

$$(m_t, F_t, t \in T) \in M^c_{loc}, \ \langle m,m \rangle_t = \langle w,w \rangle_t = t,$$

hence m is a sample continuous Brownian motion under P. (by 2.4.7.)

Comment: This result was first published by Girsanov [1960], whose proof is rather complicated. Several other proofs exist in the literature [Benes, 1971; Kailath, Zakai, 1971]. In condition 3 we have imposed the condition that $\phi \in L_{2loc}(w)$, while in 3.4.1. $\phi \in LB$. The result of 3.4.1 with $\phi \in L_{2loc}(w)$ also holds, the proof is similar. The same remark applies to the process $\lambda$ in 3.4.3. We now discuss the case for the Poisson process.

3.4.3. Corollary:

If 1. $(\Omega, F, P_0)$ is a probability space,

2. $(n_t, F_t, t \in T)$ is a real-valued standard Poisson process,

$(n_t - t, F_t, t \in T) \in M_{loc}(P_0)$,

3. $(\lambda_t, F_t, t \in T) \in L_{1loc}(t), \lambda_t > 0$ a.s. for all $t \in T$, and

satisfies $E_0[\varepsilon(\int_0^1 (\lambda_s - 1)\ (dn_s - ds))] = 1$

4. we introduce a new measure

$$\frac{dP}{dP_0} = \varepsilon(\int_0^1 (\lambda_s - 1)\ (dn_s - ds))$$

then P is a probability measure.

$$m_t \triangleq (n_t - t) - \int_0^t (\lambda_s - 1)\ ds = n_t - \int_0^t \lambda_s\ ds$$

$(m_t, F_t, t \in T) \in M_{2loc}(P)$ with

$$[m,m]_t = n_t,\ \langle m,m \rangle_t = \int_0^t \lambda_s\ ds$$

Proof. The proof follows from theorem 3.4.1., and since $\int_0^t (\lambda_s - 1)\ ds$ is sample continuous $[m,m]_t = [n_t - t, n_t - t] = n_t$.

By definition of m we conclude that $\langle m,m \rangle_t = \int_0^t \lambda_s\ ds$ since it is predictable and well defined by assumption. Brémaud [1972] has given a first version of this result.

The detection problem

We will now look in more detail to the detection problem, of which a version was formulated at the end of section 3.3. In this section we will formulate an abstract version of this problem, and then discuss the likelihood ratio method for solving it. Next we will try to solve the problem in terms of martingale theory.

3.4.4. The detection problem.

Given a measurable space $(\Omega, F)$ and two probability measures $P$, $P_0$, defined on it.

Under $P_0$: $(y_t, F_t, t \in T) \in M_{loc}$

Under $P$: $y_t = a_t + m_t$, where $(a_t, F_t, t \in T) \in BV$ and predictable and $(m_t, F_t, t \in T) \in M_{loc}$

The measures are associated with two hypotheses $H_0$, $H_1$ concerning the process y. The process y is being observed. The problem now is, given the observed process y, to decide which hypothesis to accept.

The likelihood ratio method.

Recent interest in solving the detection problem centers on the likelihood ratio method, using the martingale approach. In a particular problem it can be shown that the measures P and $P_0$ are mutually aboslutely continuous, hence $\frac{dP}{dP_0}$ exists. Using our previous results concerning the translation theorem, we will show how to estimate $\frac{dP}{dP_0}$ given the observations. Having done this, the likelihood ratio method then prescribes a statistical test on the estimate of $\frac{dP_0}{dP_0}$, to decide which hypothesis to accept.

We first discuss the problem of absolute continuity of the measures P and $P_0$.

In the case the detection problem is formulated for the Brownian motion process this problem has been investigated in several articles. They all vary in the conditions assumed and in the method of the proofs. The most detailed are [Kadota, Shepp, 1970;

Kailath, Zakai, 1971].

For the general problem that we have formulated here, we have not found any useful conditions to guarantee absolute continuity or equivalence of P and $P_0$. One method would be to use the translation result of the previous section, but this leads to strong conditions even in the Brownian motion case.

Calculating likelihood ratios.

We now assume that the measures are absolutely continuous and discuss the problem of calculating the likelihood ratio. We first give a lemma that combines the general martingale theory arguments in the derivation.

3.4.5. Lemma: $T = [0,1]$.

Given the detection problem 3.4.4., where

1. under $P_0$ $y \in M_{loc}$
2. under P $y_t = a_t + m_t$, $(a_t, F_t, t \in T) \in LIV$ and predictable, $(m_t, F_t, t \in T) \in M_{loc}(P)$,
3. We assume that P and $P_0$ are equivalent,

then $\mu_t \triangleq E_0[\frac{dP}{dP_0} | F_{yt}]$ the likelihood ratio given the observations, is characterized by a local martingale $(\hat{z}_t, F_{yt}, t \in T) \in M_{loc}(P_0)$, satisfying $\Delta\hat{z}_t > -1$ a.s. for all $t \in T$, $\hat{z}_0 = 0$, $\langle \hat{z}^c, \hat{z}^c \rangle_1 < \infty$ a.s.,

$$\mu_t = \varepsilon(\hat{z}_t) = 1 + \int_0^t \mu_{s-} \; d\hat{z}_s$$

Proof. By condition 3 and P and $P_0$ are equivalent, so $\frac{dP}{dP_0} > 0$ a.s.

Let $\mu_t \triangleq E_0[\frac{dP}{dP_0} | F_{yt}]$, then $\mu_1 > 0$ a.s., and $(\mu_t, F_{yt}, t \in T) \in M_1(P_0)$.

By 3.3.2. there exists a process $(z_t, F_{yt}, t \in T) \in M_{loc}(P_0)$

having the above described properties and such that

$$\mu_t = \varepsilon(\hat{z}_t) = 1 + \int_0^t \mu_{s-} \, d\hat{z}_s \qquad \blacksquare$$

Note that in general we cannot assert more than that $\hat{z} \in M_{loc}$

With this lemma we can derive the solution to the detection problems for the Brownian motion case and the counting process case. The reason why we cannot solve the general detection problem as formulated in 3.4.4. is that a general martingale representation theorem does not yet exist.

3.4.6. Theorem: [Duncan, 1970; Kailath, 1970 c.] $T = [0,1]$.

Given the detection problem:

1. under $P_0$ $y \in M_2^c$ is standard Brownian motion in $R^n$
2. under $P$ $y_t = \int_0^t h_s \, ds + m_t$ where $(h_t, t \in T)$ is an adapted measurable process, $E[\int_0^1 |h_s|^2 \, ds] < \infty$, and $(m_t, F_t, t \in T) \in M_2^c$ is standard Brownian motion in $R^n$.
3. Assume that $P$ and $P_0$ are equivalent,

then the likelihood ratio given the observations y can be calculated by

$$\mu_t = 1 + \int_0^t \mu_s \, \hat{h}_s^T \, dy_s \quad \text{where,}$$

$$\hat{h}_t = E[h_t | F_{yt}],$$

$$\mu_t = \exp\left(\int_0^t \hat{h}_s^T \, dy_s - \int_0^t \tfrac{1}{2} |\hat{h}_s|^2 \, ds\right).$$

Proof. We apply 3.4.5., for which the conditions are satisfied. Then by 2.6.3. $(z_t, F_{yt}, t \in T) \in M_{loc}(P_0)$ has the representation

$d\hat{z}_t = \phi_t \, dy_t$, where $(\phi_t, F_{yt}, t \in T) \in L_{2loc}(y)(P_0)$. We apply the translation theorem 3.3.5.: $m_{1t} \triangleq y_t - \langle y, \hat{z} \rangle_t = y_t - \int_0^t \phi_s \, ds$,

$(m_{1t}, F_{yt}, t \in T) \in M_{loc}(P)$. Since $E[\int_0^1 |h_s|^2 \, ds] < \infty$, we can define $E[h_t | F_{yt}]$ a.e. on T and put it zero otherwise. We get a process $(E(h_t | F_{yt}), F_{yt}, t \in T)$. Then it is easy to prove (see also 4.4.3.) that

$\hat{m}_t \triangleq y_t - \int_0^t E(h_s | F_{ys}) \, ds$, $(\hat{m}_t, F_{yt}, t \in T) \in M_{loc}(P)$. Now

$m_{1t} - \hat{m}_t = \int_0^t [E(h_s | F_{ys}) - \phi_s] \, ds \in M^c_{loc} \cap BV$, hence by 2.3.16.

it vanishes. Define $\hat{h}_t = \phi_t$, then $\hat{h}_t = E(h_t | F_{yt})$ a.s. a.e. on T. The result follows.

3.4.7. Theorem: [Brémaud, 1972; Davis, to appear]. $T = [0,1]$.

Given the detection problem:

1. under $P_0$ $y_t = n_t - t$, where n is a standard Poisson process,
2. under P $y_t = n_t - t = \int_0^t (\lambda_s - 1) \, ds + m_t$ or $n_t = \int_0^t \lambda_s \, ds + mt$

   where $(\lambda_t, t \in T)$ is an adapted measurable process, $\lambda_t > 0$ a.s. for all $t \in T$, $E[\int_0^1 \lambda_s \, ds] < \infty$

3. assume that P and $P_0$ are equivalent.

then $\mu_t = E_0[\frac{dP}{dP_0} | F_{yt}]$,

the likelihood ratio given the observation y, can be calculated by

$$\mu_t = 1 + \int_0^t \mu_{s-} \; (\hat{\lambda}_s - 1) \; (dn_s - ds)$$

where $(\hat{\lambda}_t, F_{nt}, t \in T)$ is a predictable modification of $E(\lambda_t | F_{nt})$,

and $\mu_t = \exp\left(\int_0^t \ell n\,(\hat{\lambda}_s)\; dn_s - \int_0^t (\hat{\lambda}_s - 1)\; ds\right)$

Proof. We apply 3.4.5. for which the conditions are satisfied. By the martingale representation theorem 2.6.8. $(\hat{z}_t, F_{nt}, t \in T) \in M_{loc}$ has the representation $d\hat{z}_t = \psi_t(dn_t - dt)$, where $(\psi_t, F_{nt}, t \in T) \in L_{1loc}\,(n_t - t)$. We apply the translation theorem 3.3.5. so $m_{1t} \triangleq (n_t - t) - \int_0^t \psi_s\, ds$, $(m_{1t}, F_{nt}, t \in T) \in M_{loc}(P)$.

Since $\lambda_t > 0$ a.s. and $E[\int_0^1 \lambda_s\, ds]$ we can define $E(\lambda_t | F_{nt})$.

Note that $E(\lambda_t | F_{nt}) > 0$ a.s. Define

$\hat{m}_t \triangleq (n_t - t) - \int_0^t [E(\lambda_s | F_{ns}) - 1]\, ds$ then it is easy to prove

(see 4.4.4) that $(\hat{m}_t, F_{nt}, t \in T) \in M_{loc}(P)$. Now

$m_{1t} - \ddot{m}_t = \int_0^t (E(\lambda_s | F_{ns}) - 1 - \psi_s)\, ds \in M^c_{loc} \cap BV$ hence it

vanishes by 2.3.16. Define $\hat{\lambda}_t = \psi_t + 1$ then $(\hat{\lambda}_t, F_{nt}, t \in T)$ is an adapted predictable process and $\hat{\lambda}_t = E(\lambda_t | F_{nt})$ a.s., a.e. on T. Also $\hat{\lambda}_t > 0$ a.s. for all $t \in T$. The result follows.

Remarks:

1. Note that we do not calculate $\frac{dP}{dP_0}$, but $\mu_1 = E[\frac{dP}{dP_0} | F_{y1}]$ which is the estimate of $\frac{dP}{dP_0}$ given the observed process y in T.

The calculation of the likelihood ratio is done recursively while the process y is being observed.

2. Note that $\hat{h}_t$ or $\hat{\lambda}_t$ in 3.4.6 respectively 3.4.7. are estimates of quantities that are not directly observed. They must therefore be estimated given the observations. This indicates that before we can calculate $\mu_t$ we must first solve an estimation problem and this requires further assumptions on the processes h and $\lambda$. This point is the fundamental connection between filtering theory and detection theory.

3. The results mentioned before are important by itself, but also because they can be used in applications where one needs an expression for the likelihood ratio, as for example in certain methods in filtering theory.

4. The result 3.4.6. for the Brownian motion case is well known and several different proofs exist. Duncan [1970] was the first to approach this problem in terms of martingale theory. Several articles by Kailath give the solution and alternative proofs [Kailath 1970 a; 1970 b; 1970 c; Kailath, Zakai, 1971; Kailath 1971 b].

5. Recently several articles have been published discussing the problem for the Poisson process case. Snyder [1972 a.] discusses the problem. Brémaud [1972] gave a result similar to 3.4.7. but the proof contained an error. The proof could only be given correctly when the martingale representation theorem on a Poisson process $\sigma$-field was found and proven,

as was done in [Davis, to appear]. This article also derives the solution to the detection problem but in a less direct way.

6. Skorokhod [1957] discusses the problem of absolute continuity and characterization of the likelihood ratio, for the case of independent increment and Markov processes. We could rederive these results and obtain other new results, by following the martingale approach as outlined in this section, if only we had martingale representation theorems covering these cases.

## 4. Stochastic system equations

### 4.1. Introduction.

The purpose of this chapter is to define a stochastic system, using concepts from martingale theory. In section 4.2. we discuss stochastic differential equations, where we use several results from the previous chapter. In section 4.3. we discuss stochastic system equations that have been used in the past, and we will define our semi-martingale model. In section 4.4. we discuss the concept of the innovation process, and a generalization of it, both topics are needed in the next chapter on estimation theory.

### 4.2. Stochastic differential equations.

In system theory one usually considers dynamical systems for which the state equation is described by a differential equation of the form:

$$\frac{dx(t)}{dt} = f(t,x(t)).$$

When one started investigating dynamical systems disturbed by noise processes, one considered the equation

$$\frac{dx(t)}{dt} = f(t,x(t)) + v(t)$$

where v is usually taken to be white Gaussian noise. This is sometimes known as the Langevin equation, after Langevin who considered this equation in connection with investigations of the phenomenon of Brownian motion. See Nelson [1967] and Wonham [1970]. It soon became clear that a more rigorous approach should be taken, using stochastic differential equations as introduced by Ito [1951 a]. He proved the existence and uniqueness of the solution to a stochastic differential equation of the form:

$$dx_t = f(t,x_t)\, dt + g(t,x_t)\, dv_t$$

where f and g satisfy certain conditions. The disturbance process v is a process with stationary independent increments, for which the well known Levy representation is used. Wonham [1970], in notes written earlier, introduces the work by Ito for Brownian motion processes to workers in system and control theory. A more extensive recent reference for this is Wong [1971 a]. There is a recent interest in extending the concept of stochastic differential equations, and the natural class of processes to consider is martingales. In this section we will deal with this problem in detail.

To be specific we consider the following stochastic differential equation, first published by Kazamaki [1972]:

$$dx_t = f(x_t)\, d\langle m,m\rangle_t + g(x_t)\, dm_t,\ x_0\ ,$$

where $(m_t, F_t, t \in T) \in M_{2loc}$. If one lets $m = w$ a standard Brownian motion, then the above equation reduces to the familiar form. The main question is of course the existence and uniqueness of the solution to the above stochastic differential equation.

4.2.1. Theorem: [Kazamaki, 1972].

Given the stochastic differential equation

$$dx_t = f(x_t)\, d\langle m,m\rangle_t + g(x_t)\, dm_t,\ x_0\ ,$$

where $(m_t, F_t, t \in T) \in M_{2loc}$, $(F_t, t \in T)$ is quasi-left continuous, and all processes are scalar. $f,g \in C'(R)$ and f and g satisfy a uniform Lipshitz condition i.e. $|f(x) - f(y)| \leq K |x-y|$, $\forall x, y \in R$. Then the stochastic differential equation has an unique solution x. ■

The proof is the classical method of Picard iteration, as used by Ito [1951 a] and Wong [1971 a].

Remarks:

1. The above result can be extended to the multi-dimensional case, and f and g can be allowed to be time varying. No new concepts are needed.

2. Note that when the stolution exists, then x is a semi-martingale.

3. The above theorem covers several special cases which were known before.

   If m = w, a standard Brownian motion, we get the usual form.

   If $m = n_t - t$, where n is a standard Poisson process, then we get $dx_t = f(x_t)\,dt + g(x_t)\,(dn_t - dt)$.

   Both these cases are covered by the form considered by Ito [1951 a], where m = v, v a process with stationary independent increments.

4. In the Brownian motion case, where one is concerned with stochastic control theory, it was felt that the uniform Lipschitz condition on f and g in 4.2.1. was an important restriction. In the control problem formulation f depends on the control u, so that only a limited class of control laws can be considered. Because of this problem Benes [1971], has introduced an alternative way of defining a

solution to the stochastic differential equation for the Brownian motion case.

The translation method for proving existence.

We will first state a theorem with two corollary's, afterwards we will comment on these results and interpret them.

4.2.2. Theorem:

If 1. $(\Omega,F,P_0)$ is a probability space,

2. $(r_t,F_t,t \in T) \in M_{2loc}(P_0)$ in $R^n$ with the decomposition $r = r^c + r^d$, and where $(F_t,t \in T)$ is quasi left continuous,

3. $(\phi_t,F_t,t \in T) \in LB$, $(\psi_t,F_t,t \in T) \in LB$ satisfying

$$\int_0^1 \phi_s^T \, d\langle r^c,r^c\rangle_s \, \phi_s < \infty \text{ a.s.}, \quad \psi_t^T \Delta r_t + 1 > 0 \text{ a.s. for all}$$

$t \in T$,

4. Let $z_t = \int_0^t \phi_s^T \, dr_s^c + \int_0^t \psi_s^T \, dr_s^d$

5. Let $\phi$ and $\psi$ be such that $E_0[\varepsilon(z_1)] = 1$,

then there exists a measure P, equivalent to $P_0$, such that $(r_t,F_t,t \in T)$ is a solution to,

$$dr_t = d\langle m^c,m^c\rangle_t \, \phi_t + d\langle r^d,r^d\rangle_t \, \psi_t + dm_t$$

where $(m_t,F_t,t \in T) \in M_{loc}(P)$, $m = m^c + m^d$, $\langle m^c,m^c\rangle = \langle r^c,r^c\rangle$ and $[m^d,m^d] = [r^d,r^d]$.

Alternatively written

$$dr_t^c = d\langle m^c,m^c\rangle_t \quad \phi_t + dm_t^c, \qquad m^c \in M_{loc}^c,$$

$$dr_t^d = d\langle r^d,r^d\rangle_t \quad \psi_t + dm_t^d, \qquad m^d \in M_{loc}^d.$$

Proof. By 3. both terms of z are well defined, we then apply

3.4.1. to get the result. Before making some comments let us state the two most important corollary's.

4.2.3. Corollary: [Benes', 1971].

If 1. $(\Omega, F, P_0)$ is a probability space,

2. $(r_t, F_t, t \in T)$ is a standard Brownian motion in $R^n$,

3. $(f_t, F_t, t \in T) \in LB$, $\int_0^1 |f_s|^2 \, ds < \infty$ a.s. and

$$E_0[\varepsilon(\int_0^1 f_s \, dr_s)] = 1,$$

then there exists a probability measure P equivalent to $P_0$, such that r is a solution to

$dr_t = f_t \, dt + dm_t$

where $(m_t, F_t, t \in T) \in M^c_{loc}(P)$, $\langle m,m \rangle_t = I.t$ hence m is a standard Brownian motion.

4.2.4. Corollary:

If 1. $(\Omega, F, P_0)$ is a probability space,

2. $(n_t, F_t, t \in T)$ is a standard Poisson process in $R^n$,

3. $(\lambda_t, F_t, t \in T) \in LB$, $\lambda_t > 0$ a.s. for all $t \in T$, such that

$$E_0[\varepsilon(\int_0^1 (\lambda_s - 1)\ (dn_s - ds))] = 1,$$

then there exists a probability measure P equivalent to $P_0$, such that $(n_t, F_t, t \in T)$ is a counting process satisfying

$dn_t = \lambda_t \, dt + dm_t$, where $(m_t, F_t, t \in T) \in M_{2loc}(P)$,

$\langle m,m \rangle_t = \int_0^t \Lambda(\lambda_s) \, ds.$

Notation $\Lambda(\lambda_s) = \mathrm{diag}(\lambda_s^1, \lambda_s^2, \ldots \lambda_s^n)$. ■

The corollary's follow from 4.2.2. or from 3.4.2. and 3.4.3.

Remarks:

1. In 4.2.3. notice that the essential point we have proven, is that the process m, defined by $dm_t \triangleq dr_t - f_t\, dt$ is a local martingale under the constructed measure P. Suppose that f is a function of $r_t$, or of the past of r. Then we can interpret this result as a solution to the stochastic differential equation

   $$dr_t = f_t(r)\, dt + dm_t$$

   where m is a standard Brownian motion. What we have proven this way is the existence of a solution to this stochastic differential equation, under the conditions given. This method of showing existence of solutions was first given by Benes [1971]. In connection with the solution to the stochastic differential equation we need to pose the question of the uniqueness of the solution. This problem can be considered as part of a measurability problem, which we discuss in section 4.4. However, this problem has not yet been solved satisfactorily.

2. We now consider the general case of 4.2.2., which shows the existence of the martingale m under P. This implies, if $\phi$ and $\psi$ depend on r, the existence of a solution to the stochastic differential equation

   $$dr_t = d\langle m^c, m^c\rangle_t \quad \phi_t + d\langle r^d, r^d\rangle_t \quad \psi_t + dm_t.$$

   Note that $\phi$ and $\psi$ can depend on r in any way, satisfying the conditions of the theorem. The method we have used is constructing a new measure P under which a given process satisfies the above equation. Again we have to ask the question about the uniqueness of the constructed solution but this question has not been resolved in general.

3. Consider the result of 4.2.2. as the solution to the equation
$dr_t^d = d\langle r^d,r^d\rangle_t \cdot \psi_t + dm_t^d$
with $m^d \in M_{1loc}^d$, but $m^d \notin M_{2loc}^d$. However $\langle r^d,r^d\rangle$ is defined under $P_0$ but has no interpretation under P. In special cases $\langle r^d,r^d\rangle$ is absolutely continuous with respect to $\langle m^d,m^d\rangle$, or an explicit form for it can be found in which cases we can simplify the equation. In general no satisfactory explanation has been found to interpret $\langle r^d,r^d\rangle$ under P. The result of 4.2.2. is thus in an important way different from 4.2.1.

4. We discuss some of the conditions that are assumed in 4.2.2. Note that the conditions of 4.2.2. do not require that $\phi$ and $\psi$ satisfy an uniform Lipschitz condition, as is necessary in 4.2.1. However 4.2.2. has the condition $E_0[\varepsilon(z_1)] = 1$, which is an implicit condition on $\phi$ and $\psi$. We have sufficient conditions for this to hold 3.2.5., and 3.3.6. but those seem rather strong. No other sufficient condition for this is known at this moment. In the case of 4.2.3. it is known [Benes, 1971], that if $f_t$ depends on $r_t$ only, a Lipschitz condition on f guarantees that $E_0[\varepsilon(z_1)] = 1$.

5. Corollary 4.2.4. as given here seems new. It deals with a counting process n, satisfying the stochastic differential equation

$dn_t = \lambda_t(n)\ d_t + dm_t$, with $m \in M^d_{2loc}$.

In 2.4.23. we have defined $\lambda$ to be the rate process of the family $(F_t, t \in T)$ associated with n. Note that $\lambda_t$ can depend on the complete past of the counting process. The result is that we have a stochastic differential equation with as solution a counting process that influences its own rate.

## 4.3. Stochastic Systems

In this section we consider the problem of modelling continuous-time processes by stochastic systems. We first review some of the models that are used in stochastic problems. We then define a semi-martingale model and show that it covers most known problem formulations.

Stochastic processes.

Up to recently, most engineers used the white Gaussian noise 'process' in stochastic problems. This approach however is non-rigorous, and was mainly used because it leads to simple analytical calculations. It was Wonham [1970] who introduced the Brownian motion process for stochastic problems in estimation and control. Brownian motion is a well defined stochastic process, and it relates to white Gaussian noise, because this process can be considered as a generalized 'derivative' of Brownian motion.

The reason one takes Brownian motion as a noise process is because it is completely determined by the properties: 1. it has independent increments,, 2. it is sample continuous. Both assumptions seem reasonable justified from a physical point of view, the independent increment assumption because a noise process is generated at microscopic levels.

Stochastic system models.

We discuss several stochastic system models for observation processes, which are used in filtering and estimation problems. In the filtering problem phrased by Wiener [1949], the observed process y is assumed to be the sum of a signal process h and a noise process v: $y_t = h_t + v_t$. The processes h and v are assumed to be stationary and of second order. The model is further specified by giving the covariance matrices or their Fourrier transforms, of the processes h and v and their dependence relation. In the article by Kalman, Bucy [1961] on linear filtering, it is assumed that the above noise process v is a white Gaussian noise process, and further the signal process h is modelled as the output of a linear system disturbed by another white Gaussian noise process:

$\dot{x}(t) = A(t)\ x(t) + w_t,\ h_t = c(t)\ x(t).$

Observe that the form of the above linear model for the signal process is an essential assumption.

This way of modelling, which allows non stationary processes, was inspired by concepts from linear system theory as developed by Kalman and others during the 1950's. This model makes more explicit the dynamical structure of the signal process. The above model can be recast in terms of Brownian motion processes, using stochastic differential equations, we get

$$dx_t = A(t)\, x_t\, dt + dw_{1t}$$

$$dy_t = C(t)\, x_t\, dt + dw_{2t}$$

where $w_1$, $w_2$ are Brownian motion processes [Wonham, 1970]. This stochastic system model can be further extended to account for control operations and nonlinear systems.

The counting process observation equation.

Because later on we consider certain applications of counting processes we will define a stochastic system equation for such processes. Snyder [1972 a] considered the same problem but used the concept of doubly stochastic Poisson processes. Since that method has certain drawbacks, we will follow here the martingale approach first applied by Brémaud [1972]. Recall from section 2.4. that if n is a counting process, then $\lambda$ is it's associated rate process if $(n_t - \int_0^t \lambda_s\, ds,\ F_t,\ t \in T) \in M_{loc}$, which we can rewrite as $dn_t = \lambda_t\, dt + dm_t$, $n_0 = 0$. Note the analogy with the output equation for the stochastic system with Brownian motion disturbances. We can complete the stochastic system model by specifying the dynamical equation for the rate process $\lambda$, say $d\lambda_t = f(t,\lambda_t)\, dt + dm_{1t}$. We call this stochastic system model the counting process observation system.

The stochastic system model.

We now introduce a stochastic system, using concepts from our earlier investigations concerning martingales. Our goal is a stochastic system that covers both the equation with Brownian motion disturbances, and the counting process equation as special cases.

4.3.1. Definition: The stochastic system.

We assume given the following equations forming the stochastic system:

1. the observation equation,

   $dy_t = d\langle m_2,m_2\rangle_t \; h_t + dm_{2t},\ y_0 = 0,\ y \in R^k,$

2. the state equation,

   $dx_t = d\langle m_1,m_1\rangle_t \; f_t + dm_{1t},\ x_0,\ x \in R^n$

3. where $(F_t, t \in T)$ is some increasing family satisfying the usual conditions, and quasi left continuous,

   $(m_{1t},F_t,t \in T) \in M_{2loc}$ in $R^n$, $(m_{2t},F_t,t \in T) \in M_{2loc}$, in $R^k$,

4. $(f_t,F_t,t \in T)$ is adapted, predictable, and

   $$\int_T |f_s| \; |d\langle m_2,m_2\rangle_s| < \infty \text{ a.s.},$$

   $(h_t,F_t,t \in T)$ is adapted, predictable, and

   $$\int_T |h_s| \; |d\langle m_2,m_2\rangle_s| < \infty \text{ a.s., and } f_t \; \Delta m_{1t} + 1 > 0 \text{ a.s.,}$$

   $h_t \; \Delta m_{2t} + 1 > 0$ a.s. for all $t \in T$.

   f, h can depend on x in any way, such that $dx_t$ is a stochastic differential equation and y a semi-martingale.

Remarks:

1. We consider the problem of the existence of a solution to the stochastic differential equation, $dx_t = d\langle m_1,m_1\rangle_t \; f_t + dm_{1t},\ x_0$. If $f_t$ is a function of $x_t$, $f\colon T \times R^n \to R^n$, satisfying a Lipschitz condition in x, then by 4.2.1. the stochastic differential equation

has an unique solution. If $m_1 \in M^c_{loc}$ then we can show existence of solution for arbitrary$(f_t, t \in T)$, depending on the past of x by 4.2.2. The general case of showing existence of solution has not been solved yet.

2. We call the equation for x the state equation, and x the state process, although it does not satisfy the usual definition of a state in deterministic system theory. We assume that the process h depends on x in some way. Since y is observed it will be called the observed process, it provides information about the state process x.

3. Note that x and y are semi-martingales. We can write the above equation in different forms. One is to write it in one equation, if

$$r_t = \binom{x_t}{y_t},\ m_t = \binom{m_{1t}}{m_{2t}},\ \phi_t = \binom{f_t}{h_t},\ \langle m_1, m_2 \rangle = 0 \text{ a.s.,}$$

$$dr_t = d\langle m,m\rangle_t \quad \phi_t + dm_t,\ r_0.$$

A different way of considering the above equation, in analogy with 4.2.2., is

$$dx_t = d\langle m_1^c, m_1^c\rangle_t \quad f_{1t} + d\langle m_1^d, m_1^d\rangle_t \quad f_{2t} + dm_{1t},$$
$$dy_t = d\langle m_2^c, m_2^c\rangle_t \quad h_{1t} + d\langle m_2^d, m_2^d\rangle_t \quad h_{2t} + dm_{2t}.$$

However there is no general existence proof for the stochastic differential equation for x.

By specializing the definition, we get the equation with Brownian motion disturbances, and the counting process equation.

4.3.2. Definition: the equation with Brownian motion disturbances.

We define the equation with Brownian motion, by taking in 4.3.1.

$m \in m^c_{2loc}$, $\langle m,m \rangle_t = I \quad t$, hence m is standard Brownian motion. We write the stochastic system as:

$dr_t = f_t\, dt + dm_t,\ r_0.$

4.3.3. Definition: The counting process equation.

We define the counting process equation, by taking in 4.3.1. $(m_t, F_t, t \in T) \in M^d_{2loc}$ in $R^n$, $d\langle m,m \rangle_t = \Lambda(\lambda_t)\, dt$, where $(\lambda_t, F_t, t \in T)$ is a predictable process in $R^n$, $\Lambda(\lambda_t) = \mathrm{diag}(\lambda^1_t, \ldots \lambda^n_t)$, $\lambda_t > 0$ a.s. for all $t \in T$ and

$\int_0^1 \lambda_s\, ds < \infty$ a.s. Furthermore $\phi_t = 1$ and r is a counting process denoted by n. We write the stochastic system as

$dn_t = \lambda_t\, dt + dm_t,\ n_0 = 0.$

Note that m is the martingale associated with the counting process n with rate process $\lambda$, as defined in 2.4.23.

4.4. Representation of the stochastic system, the innovation process.

In this section we consider the concept of projection of the semi-martingale model on an increasing family of $\sigma$-fields, and the related concept of an innovation process. Both of these topics are needed in the next chapter on estimation theory. We first give an example of what an innovation process is, and then discuss it for the more general case of our stochastic system model.

An example.

Let $T = [1,2,..N]$ be discrete. Let $y = (y_t, t \in T)$ be a discrete time observed process satisfying $E|y_t| < \infty$ for all

$t \in T$. Let $F_{yn} = \sigma(y_t, \forall t \leq n)$. We define the process:

$\hat{m}_1 \triangleq y_1 - E(y_1)$, $\hat{m}_t - \hat{m}_{t-1} \triangleq y_t - E[y_t | F_{yt-1}]$, $t = 2, \ldots N$.

Suppose now we observe the process y, say we received the observations up to (t-1), so all our information is contained in $F_{yt-1}$. Our estimate of the new observation $y_t$ given $F_{yt-1}$, in the least squares sense is $E[y_t | F_{yt-1}]$. Now we observe $y_t$. The new information we have received by observing $y_t$ is $y_t - E[y_t | F_{yt-1}] = \hat{m}_t - \hat{m}_{t-1}$. This is why this random variable is called the innovation, or new information, and the process $\hat{m}$ is called the innovation process. The name innovation process was introduced by Masani and Wiener [1950, p. 136] in the above context.

Note that the innovation process is defined in terms of the observed process only, and that $\hat{m}$ is actually a martingale with respect to $F_{yt}$. We can rewrite the above process $\hat{m}$ as

$\hat{m}_t - \hat{m}_{t-1} = (y_t - y_{t-1}) - E[y_t - y_{t-1} | F_{yt-1}]$.

We now consider the continuous-time case heuristically. In analogy with the above equation one can define the innovation process by

$d\hat{m}_t \triangleq dy_t - E[dy_t | F_{yt-1}]$,

which again can be interpreted as the new information gained by observing $dy_t$. One sees intuitively that $(\hat{m}_t, F_{yt}, t \in T)$ is a martingale.

A short review

As mentioned earlier the innovation process was introduced by Masani and Wiener [1958]. In a series of articles Kailath

[1968], Kailath, Frost [1968], Frost, Kailath [1971] used the innovation process to analyse and derive results concerning filtering. The use of the innovation process has been much advocated by Kailath, for example in detection problems [1970 a]. There is a more recent article [Kailath, 1971 a] where the innovation process is considered in terms of martingale theory, an approach also followed by others. Meyer [1973] discusses and slightly extends the innovation process result.

The innovation process for the equation with Brownian motion disturbances is well known to be a Brownian motion process, hence a martingale. We now consider the innovation process in a martingale context and derive several results.

4.4.1. Theorem:

If 1. given the semi-martingales

$$x_t = x_0 + a_{1t} + m_{1t}$$

$$y_t = a_{2t} + m_{2t}$$

2. $E|x_0| < \infty$,

3. $(m_{1t}, F_t, t \in T)$, $(m_{2t}, F_t, t \in T)$ are martingales, $m_{10} = 0$, $m_{20} = 0$.

4. $(a_{1t}, F_t, t \in T) \in IV$, $(a_{2t}, F_t, t \in T) \in IV$,

then $\hat{x}_t = \hat{x}_0 + \hat{a}_{1t} + \hat{m}_{1t}$

$$y_t = \hat{a}_{2t} + \hat{m}_{2t}$$

where $(\hat{a}_{1t}, F_{yt}, t \in T)$ is $\hat{a}_{1t} = E(a_{1t}|F_{yt})$, similarly

$(\hat{a}_{2t}, F_{yt}, t \in T)$ is $\hat{a}_{2t} = E(a_{2t}|F_{yt})$,

$(\hat{x}_t, F_{yt}, t \in T)$ is $\hat{x}_t \triangleq E(x_t|F_{yt})$,

$(\hat{m}_{1t}, F_{yt}, t \in T)$ and $(\hat{m}_{2t}, F_{yt}, t \in T)$ are martingales.

Proof. $a_1 \in IV$ implies that $\sup_{t \in t} E|a_{1t}| < \infty$, hence we can define $\hat{a}_{1t} = E(a_{1t}|F_{yt})$, and the process $(\hat{a}_{1t}, F_{yt}, t \in T)$. Similarly we define $\hat{a}_2$. Since $m_1$ is a martingale by the conditions 2, 3 and 4

$E|x_t| \le E|x_0| + E|a_{1t}| + E|m_{1t}| < \infty$ for all $t \in T$.

We let $\hat{x}_t = E(x_t|F_{yt})$ and get the process $(\hat{x}_t, F_{yt}, t \in T)$. Let $\hat{m}_{1t} \triangleq \hat{x}_t - \hat{x}_0 - \hat{a}_{1t}$, then it is adapted $(\hat{m}_{1t}, F_{yt}, t \in T)$. By the above $E|\hat{m}_{1t}| < \infty$ for all $t \in T$, and $E[\hat{m}_{1t} - \hat{m}_{1s}|F_{ys}] = E[\hat{x}_t - \hat{x}_s - \hat{a}_{1t} + \hat{a}_{1s}|F_{ys}]$ $= E[E[m_{1t} - m_{1s}|F_s]|F_{ys}] = 0$ so $(\hat{m}_{1t}, F_{yt}, t \in T)$ is a martingale. The proof for $\hat{m}_2$ is similar.

In subsequent chapters we need a special version of this result which cannot be deduced from it directly. We state it in 4.4.2.

4.4.2. Theorem: $T = [0,1]$.

If 1. given the semi-martingales

$$x_t = x_0 + \int_0^t f_s \, ds + m_{1t}$$

$$y_t = \int_0^t h_s \, ds + m_{2t}$$

2. $(m_{1t}, F_t, t \in T)$ is a martingale

$(m_{2t}, F_t, t \in T) \in M_{loc}$

3. $(f_t, F_t, t \in T)$ is an adapted measurable process,

$\sup_{t \in T} E|f_t| < \infty$,

$(h_t, F_t, t \in T)$ is an adapted measurable process,

$\sup_{t \in T} E|h_t| < \infty$,

4. $E|x_0| < \infty$,

then 1. $\hat{x}_t = \hat{x}_0 + \int_0^t \hat{f}_s \, ds + \hat{m}_{1t}$

$$y_t = \int_0^t \hat{h}_s \, ds + \hat{m}_{2t}$$

2. where $(\hat{f}_t, F_{yt}, t \in T)$ is $\hat{f}_t = E(f_t | F_{yt})$, similarly $(\hat{h}_t, F_{yt}, t \in T)$ is $\hat{h}_t = E(h_t | F_{yt})$.

3. $(x_t, F_{yt}, t \in T)$ is a modification of $E(x_t | F_{yt})$, which is right continuous, having left hand limits,

4. $(\hat{m}_{1t}, F_{yt}, t \in T)$ is a martingale,

5. $(\hat{m}_{2t}, F_{yt}, t \in T) \in M_{loc}$, $\langle \hat{m}_2^c, \hat{m}_2^c \rangle = \langle m_2^c, m_2^c \rangle$ a.s., $[\hat{m}_2^d, \hat{m}_2^d] = [m_2^d, m_2^d]$ a.s.

Proof. Since $\sup_{t \in T} E|f_t| < \infty$, we can define $\hat{f}_t = E[f_t | F_{yt}]$.

$$E[\int_0^1 \hat{f}_s \, ds] \leq E[\int_0^1 E[|f_s| | F_{ys}] \, ds] = E[\int_0^1 |f_s| ds]$$

$\leq 1 \cdot \sup_{t \in T} E|f_t| < \infty$. Similarly we can define $(\hat{h}_t, F_{yt}, t \in T)$ and $E[\int_0^1 |\hat{h}_s| ds] < \infty$. By the conditions 2, 3, 4, $E|x_t| < \infty$ for all $t \in T$, so we can define $E(x_t | F_{yt})$. Since $F_{y0}$ is the trivial $\sigma$-field $E(x_0 | F_{y0}) = E(x_0)$. Let $m_{3t} \triangleq E(x_t | F_{yt}) - E(x_0) - \int_0^t \hat{f}_s \, ds$, then $E|m_{3t}| < \infty$ for all $t \in T$. Now

$$E[m_{3t} - m_{3s} | F_{ys}] = E[E(x_t | F_{yt}) - E(x_s | F_{ys}) - \int_s^t E(f_\tau | F_{y\tau}) \, d\tau | F_{ys}] = E[E[m_{1t} - m_{1s} | F_s] | F_{ys}] = 0.$$

So $(m_{3t}, F_{yt}, t \in T)$ is a martingale, and has thus a right continuous modification $(\hat{m}_{1t}, F_{yt}, t \in T)$, having left hand limits. Define $\hat{x}_t = E(x_0) + \int_0^t \hat{f}_s \, ds + \hat{m}_{1t}$, thus $\hat{x}$ is a modification of $E(x_t | F_{yt})$ and has the properties given above. Similarly we can show that

$$\hat{m}_{2t} \triangleq y_t - \int_0^t \hat{h}_s \, ds, \; (\hat{m}_{2t}, F_{yt}, t \in T) \in M_{loc} \text{ by a}$$

stopping time argument. By 2.3.38., $[m_2, m_2] = [y,y] = [\hat{m}_2, \hat{m}_2]$ which characterizes $\hat{m}_2$.

Note the difference between the two previous theorems. In 4.4.2. we have given a more explicit form to the processes of bounded variation. This allows us to obtain a more explicit result. Applying just 4.4.1. to the semi-martingales x and y of 4.4.2., does not give a valuable result, since we would get $E[\int_0^t f_s \, ds | Fy_t]$, which is not of bounded variation in general. However the form of 4.4.2. gives us a process of bounded variation.

We state two corollary's to 4.4.2.

4.4.3. Corollary: $T = [0,1]$

If 1. $dy_t = h_t \, dt + dm_t$

2. $(m_t, F_t, t \in T) \in M_2^c$ in $R^n$, $\langle m,m \rangle_t = I.t$, so m is standard Brownian motion,

3. $(h_t, F_t, t \in T)$ is an adapted measurable process, $\sup_{t \in T} E|h_t| < \infty$, let $\hat{h}_t \triangleq E(h_t | F_{yt})$

then $dy_t = \hat{h}_t \, dt + d\hat{m}_t$

$(\hat{m}_t, F_{yt}, t \in T) \in M_2^c$, $\langle \hat{m},\hat{m} \rangle_t = I.t$, hence $\hat{m}$ is standard Brownian motion. We call $\hat{m}$ the innovation process.

4.4.4. Corollary: $T = [0,1]$,

If 1. given the scalar counting process observation equation:

$$dn_t = \lambda_t \, dt + dm_t$$

2. $(m_t, F_t, t \in T) \in M_2^d$, $d\langle m,m \rangle_t = \lambda_t \, dt$,

3. $(\lambda_t, F_t, t \in T)$ is an adapted measurable process, $\lambda_t > 0$ a.s. for all $t \in T$, $\sup_{t \in T} E(\lambda_t) < \infty$, let $\hat{\lambda}_t \triangleq E(\lambda_t | F_{nt})$

then $dn_t = \hat{\lambda}_t \, dt + d\hat{m}_t$

$(\hat{m}_t, F_{nt}, t \in T) \in M_2^d$, $d\langle m,m \rangle_t = \hat{\lambda}_t \, dt$, we call $\hat{m}$ the innovation process.

Proof. By 4.4.2. $[\hat{m},\hat{m}]_t = [n,n]_t = n_t$ hence $\langle \hat{m},\hat{m} \rangle$ has the given form.

Remarks:

1. Consider 4.4.1. and the semi-martingale y. What we have achieved is the following $y_t = a_{2t} + m_{2t} = \hat{a}_{2t} + \hat{m}_{2t}$. Both are representations of the semi-martingale y, but with respect to different families of σ-fields, the first $(F_t, t \in T)$, the second $(F_{yt}, t \in T)$. We will refer to this property as the representation of a semi-martingale with respect to a family of σ-fields. Now consider the semi-martingales x and $\hat{x}$ of

4.4.1. and their representation. This representation is a straightforward extension of the different representations for y, the only difference being the conditional expectation of x with respect to $(F_{yt}, t \in T)$.

2. If the semi-martingale y is an observed process, then we will call $\hat{m}_2$ the innovation process associated with it. Comparing 4.4.2. with our earlier discussion on the innovation process we note that $d\hat{m}_t = dy_t - \hat{h}_t\, dt = dy_t - \hat{h}_t\, dt - E[E(dm_t | F_t)\ F_{yt}) =$ $= dy_t - E[dy_t | F_{yt}]$ which has the intuitive interpretation given earlier.

3. In 4.4.1. and 4.4.2. we have considered special cases of the semi-martingale system model as defined in 4.3.1. At this moment it is not quite clear how 4.4.2. should be generalized. The corollaries 4.4.3. and 4.4.4. have been given for later reference.

Equality of $\sigma$-fields.

A related problem of considerable interest is the following, we state an abstract version of it.

4.4.5. Definition.

Given a semi-martingale y and its representation with respect to $(F_{yt}, t \in T)$, $y_t = \hat{a}_t + \hat{m}_t$, where $(\hat{a}_t, F_{yt}, t \in T) \in LIV$ and predictable, $(\hat{m}_t, F_{yt}, t \in T) \in M_{loc}$. Let $F_{\hat{m}t} = \sigma(\hat{m}_s, \forall s \leq t)$, then $F_{\hat{m}t} \subset F_{yt}$ for all $t \in T$. The problem is under what conditions do we have that $F_{yt} = F_{\hat{m}t}$, for all $t \in T$.

Remarks:

1. The version of this problem which has been considered, is

that of 4.4.3. The equality of $\sigma$-fields was proven under general conditions for the case of linear filters for linear systems [Kailath, 1968a; Kailath, 1972]. A more general proof under the condition that h and m are independent and h is uniformly bounded is due to Clark. Several attempts to prove the above equality are known but no rigorous proof has been published yet. This is a point of active research. The equality of the $\sigma$-fields has the interpretation that all the information contained in the observed process, is also contained in the innovation process.

2. If under certain conditions a result as defined in 4.4.5. exists then it has two main applications, the first one being estimation theory. If $F_{yt} = F_{\hat{m}t}$ and if we consider the output equation with the Brownian motion disturbances, then by 4.4.3. $\hat{m}$ is standard Brownian motion. By 2.6.3. any $(m_t, F_{yt}, t \in T) = (m_t, F_{\hat{m}t}, t \in T)$ thus has the representation $dm_t = \phi_t \, d\hat{m}_t$. Using this point Kailath [1968] and Frost, Kailath [1971] derived results for the filtering problem. However since the equality of the $\sigma$-fields can only be proven under strict conditions, this approach is not useful. We will see in section 5.3. that it is not necessary to prove the equality of $\sigma$-fields to obtain the above representation result. A second application of the possible equality of $\sigma$-fields would be the uniqueness of solution to stochastic differential equations, using the method given in the second part of section 4.2.

# 5. Estimation theory.

## 5.1. Introduction.

In this chapter we start our investigation of estimation theory. The purpose is to solve the estimation problem using martingale theory. In section 5.2. we give a review of earlier work on estimation, primarily filtering theory for the equation with Brownian motion disturbances and for counting process observations. In section 5.3. we derive two crucially important martingale representation theorems. In section 5.4. we pose the filtering, prediction and smoothing problem, discuss the least squares estimation method, and outline the solution to the filtering problem. In section 5.5. we derive general results concerning prediction and smoothing.

## 5.2. A short review of estimation theory.

The estimation problem.

We start by defining the problems that are usually considered under the heading of estimation problems. Let $(\Omega,F,P)$ be a probability space, and T be the time interval of interest. It is called either discrete-time or continuous-time depending on the character of T. All estimation problems deal with an observed process, denoted by y, defined on $(\Omega,F,P)$ and T. There are three main estimation problems:

1. The detection problem: where one wants to choose between two hypotheses concerning the distribution of the process y. This problem was considered in section 3.4. and some references were

given there. We will not comment on this problem further here.

2. The filtering, smoothing and prediction problem: here another process x, also defined on $(\Omega,F,P)$ and T, is present. The filtering problem is to estimate the process x at time $t \in T$, given the observations y up to time t, and this for all $t \in T$. The smoothing and prediction problems are related, see the definitions, 5.5.1, 5.5.3.
3. The identification problem: here the distribution of the process y, or alternatively the dynamical equation for y and related processes, depend on an unknown parameter. The identification problem is to identify this parameter. We will not discuss this problem in this report.

The time interval.

There are several important distinctions between the discrete or continuous-time case. In the discrete-time case with finite observations the problem is relatively simple. After some calculations one can then generate the estimate recursively. This approach however does not work in the continuous-time case. Another point is that in the continuous-time case it is difficult to store the past of the observed process, without special devices. This point leads to the concept of recursive estimation, where we store only the current estimate, and this estimate and the observed process, generate the new estimate on a continuous-time basis. We will now restrict our discussion to continuous-time processes and concentrate mainly on the filtering problem.

General solution methods.

The estimation problem as given before, should be further specified as to how to choose the estimate from the many possible ones. Depending on the performance criterion for the estimate, the solution method will vary.

The least-squares method.

Estimation theory started with Gauss, who developed the least squares method in 1795. The method was used for estimation problems that appear in astronomy, (for comments see Sorenson [1970]). The least-squares estimation error method, as we will see later, leads to the optimal estimate $E(x_t | F_{yt})$, i.e. the conditional expectation of $x_t$ given $F_{yt}$. Using this an estimator has to be derived which accounts for the dynamical evolution of the x process. This can be done in a relatively straightforward manner, using ideas of martingale theory as developed in this report. An alternative way of deriving an estimator for $\hat{x}$ is finding an expression in terms of the unnormalized or normalized conditional density for $x_t$ given $F_{yt}$. Then however an equation should be derived for the dynamical evolution of this density.

An alternative method for solving estimation problems is by first deriving the conditional density of $x_t$ given $F_{yt}$, which then is maximized with respect to $x_t$. This is called the maximum likelihood method, which we will not consider here.

Filtering theory for the Brownian motion model.

The following methods all deal with the estimation problem for versions of systems with Brownian motion disturbances. We restrict the attention to the main methods and papers. Although Kolmogorov [1941] first published a discussion on the estimation problem for discrete-time processes, it was Wiener [1949] who first worked on the continuous-time problem. The problem formulation was: given an observed process, which is the sum of a signal and a noise process, both, stationary; the covariance matrix of both processes and their correlation is given. The least-squares estimation error method was discussed and applied to this problem. The optimal linear filter specified by its impulse response is being sought. A minimization approach then leads to the so-called Wiener-Hopf equation, which is in terms of the covariance matrices only. The problem now is entirely non-probabilistic, and the Wiener-Hopf equation is solved by a frequency factorization method, which was discovered earlier by Wiener. The design procedure was rather cumbersome and was difficult to apply to multi-dimensional observations. In 1961 Kalman and Bucy published their paper with a new approach to linear filtering and prediction. Their problem statement is more general; they assume the observation and state process to be modelled by a linear, possibly, time-varying system representation; the noise process is assumed to be white, but not necessarily stationary; the system model is a multivariable system, allowing multi-output observations; and crucially the filter is restricted to be linear. The method they use is applying the orthogonal projection theorem. From this they derive the Wiener-Hopf equation, which then gives both the form of the optimal filter equation and the non linear Riccati equation for the error covariance. For a discussion on the history and a perspective of the Kalman-Bucy filter see Sorenson [1970]. A large number of investigators have considered the linear filtering problem, and special versions of it, and have derived the same results by different methods.

The innovation process in estimation theory.

As described earlier the innovation process was defined by Masani and Wiener [1958], but its importance in estimation was first stressed in a series of articles by Kailath [1968] and others. The main point is that if the system and the filter are restricted to be linear then the observation and the innovation process generate the same σ-field. It is this property that allows one to derive the linear filter equations much easier, avoiding the Wiener-Hopf equation. Kailath [1968], Kailath, Frost [1968] show how this method can be applied to the linear filtering and smoothing problem. In Frost, Kailath [1971], the innovation process approach is used for nonlinear filtering problems.

A stochastic differential equation for the optimal estimate. In the following articles the same stochastic system model is discussed. The observation equation has the usual Brownian motion model, but x is assumed to be a Markov process. There are no constraints on the form of the filter. The first to consider this problem was Stratonovich, in the context of conditional Markov processes. His result however must be interpreted in a special stochastic calculus, different from the martingale or Ito calculus. Kushner [1967a] first derived a stochastic differential equation for the optimal estimate. The conditions stated are rather strict, but they are necessary to derive a dynamical dquation for the conditional density of $x_t$ given $F_{yt}$. From this density the stochastic differential equation for the optimal estimate follows. Bucy [1965] in a short note gives a similar approach but with a different proof. Next follow the articles by Kallianpur and Striebel [1968, 1969]. They rederive Kushner's result but under more relaxed conditions, and using martingale theory arguments. Fujisaki, Kallianpur, Kunita [1972] finally clarified the proof and extended the result. Their method is the martingale approach which we are following in this thesis. The basic points in the derivation are the martingale representation theorem on the σ-field of the observations, given in section 5.3., and the analysis using the innovation process as given in section 4.4. Using this method under quite general conditions the stochastic differential equation for the optimal estimate can be derived.

The unnormalized conditional density method.

Zakai [1969] gives an alternative method to the approach by Kushner. See also the similar approach in the book by Wong [1971, Ch. 6]. Instead of the conditional density, the unnormalized conditional density of $x_t$ given $F_{yt}$ is considered. Using results concerning the likelihood ratio as developed in Chapter 3, one can derive an equation for the optimal estimate in terms of a conditional likelihood ratio. This last quantity in turn satisfies a dynamical equation. However the application of these formula's seems to be complicated and few examples of filters are known. It gives however additional insight in the problem.

Approximations to nonlinear filters.

Kushner [1967b] first considered applications of nonlinear filters. It was his analysis that no finite dimensional filters exist, that one gets a sequence of conditional moments of the optimal estimate. Then some approximation procedures were given. Several other approximation procedures have been published, usually they involve an expansion of the nonlinear functions around the optimal estimate, and retaining only first or first and second order terms. Such filters are mentioned in Sage, Melsa [1971].

Estimation problems for counting processes.

Given one cbserves a counting process. As defined in section 2.4. under certain conditions, we can associate with it a rate process. The problem is given the observed process, to detect or to estimate the rate process, or some other unobserved process, that is of interest. Specialized problems of the above are of course the detection problem and the filtering, prediction and smoothing problem. The above problem is an abstract version of problems encountered in nuclear medicine, communication theory and operations research. See Snyder [1972a] and Brémaud [1972] for detailed problem areas. The first to consider the filtering and detection problems in detail was Snyder [1972a]. Instead of using the martingale approach as we do here, counting processes are modelled as doubly stochastic Poisson processes. The method given by Snyder is to derive the evolution of the conditional density of the unknown random process with respect to the $\sigma$-field of the observations. This is the analogue of Kushner's method for the Brownian motion model. Using this equation a general dynamic equation can be derived for the optimal estimate. Then the usual first order approximation procedure is used to derive finite dimensional filters. In a subsequent article, Snyder [1972b], discusses the smoothing problem, the same approach is used. Brémaud [1972] in his thesis, deals also with counting processes or point processes, but uses the martingale approach. The existence of a point process with a stochastic rate is shown. A formula for the likelihood ratio is given, however the proof contains an error. The general form of the result however is correct. Next the filtering problem is considered, the unnormalized conditional density is derived. This approach is analogous to the method by Zakai [1969] for the Brownian motion model.

5.3. Martingale representation theorems.

In section 2.6. we have seen that certain martingales on the $\sigma$-field generated by a Brownian motion or a Poisson process, have a representation as a stochastic integral with respect to the underlying process. In this section we will derive a representation theorem for certain martingales on the $\sigma$-field of the observation process y. Since y is not a martingale, the representation theorem for martingales as stochastic integrals will be with respect to the martingale associated with y, the innovation process $\hat{m}$.

5.3.1. Theorem: $T = [0,1]$.

If 1. $dy_t = \hat{h}_t\, dt + d\hat{m}_t$, $y \in R^k$

2. $(\hat{m}_t, F_{yt}, t \in T) \in M_2^c$, $\langle \hat{m}, \hat{m} \rangle_t = I.t$, so $\hat{m}$ is Brownian motion,

3. $(\hat{h}_t, F_{yt}, t \in T)$ is an adapted measurable process, $\int_T |\hat{h}_s|^2\, ds < \infty$ a.s.

4. let $(m_t, F_{yt}, t \in T) \in M_{loc}$ in $R^n$

then m has the representation

$$m_t = \int_0^t \Sigma_s\, d\hat{m}_s \quad \text{a.s. for all } t \in T,$$

for an unique process $(\Sigma_t, F_{yt}, t \in T) \in L_{2loc}(\hat{m})$ in $R^{n\times k}$

Then also m is sample continuous.

Proof. We give the proof only for the scalar case, without loss of generality.

Step 1. Define $\hat{z}_t \triangleq \int_0^t - \hat{h}_s\, d\hat{m}_s$, then $(\hat{z}_t, F_{yt}, t \in T) \in M_{2loc}^c$.

Define the stopping times $\tau_n = \inf\{1, t \in T | \langle \hat{z}, \hat{z} \rangle_t \geq n\}$

and let them be also such that $m_t^n \overset{\Delta}{=} m_{t\wedge\tau_n} \in M_1$ for all n.

Since $\langle \hat{z},\hat{z}\rangle_1 = \int_0^1 |\hat{h}_s|^2 \, ds < \infty$ a.s. by condition 3.,

$\lim_n \tau_n = 1$ a.s. Define $\hat{z}_t^n \overset{\Delta}{=} \hat{z}_{t\wedge\tau_n}$, $m_t^n \overset{\Delta}{=} m_{t\wedge\tau_n}$, $\hat{m}_t^n = \hat{m}_{t\wedge\tau_n}$.

Step 2. Since $\langle \hat{z}^n,\hat{z}^n\rangle_1 \leq n$, by 3.3.6. $E_0[\varepsilon(\hat{z}_1^n)] = 1$,

so $\frac{dP^n}{dP} = \varepsilon(\hat{z}_1^n)$ introduces a new probability on $(\Omega,F)$.

We now apply the translation theorem 3.3.5.:

$$\hat{m}_t^n - \langle \hat{m}^n,\hat{z}^n\rangle_t = \hat{m}_t^n + \int_0^{t\wedge\tau_n} \hat{h}_s \, ds = y_t^n = y_{t\wedge\tau_n} \text{ and}$$

$(y_t^n, F_{yt\wedge\tau_n}, t \in T) \in M_{2loc}^c\ (P^n)$. Since $\langle \hat{m}^n,\hat{z}^n\rangle$ is sample

continuous, $\langle y^n,y^n\rangle_t = [y^n,y^n]_t = [\hat{m}^n,\hat{m}^n]_t = t\wedge\tau_n$.

Now $(y_t^n, F_{yt\wedge\tau_n}, t \in T)$ is standard Brownian motion under $P^n$.

Step 3. Note that because $\hat{z}^n \in M_2^c$, $[m^n,\hat{z}^n] = \langle m^{nc},\hat{z}^n\rangle = \langle m^n,\hat{z}^n\rangle$.

We now again apply the translation theorem: let

$\hat{q}_t^n \overset{\Delta}{=} m_t^n - \langle m^n,\hat{z}^n\rangle_t, (\hat{q}_t^n, Fy_t, t \in T) \in M_{loc}(P^n)$. We now

apply the representation theorem 2.6.3. Since under $P^n$ on

$[0,\tau_n], y^n$ is standard Brownian motion, $d\hat{q}_t^n = \sigma_s^n \, dy_s^n$, on

$[0,\tau_n]$ where $(\sigma_t^n, F_{yt}, t \in T) \in L_{2loc}(\hat{m})$ under $P^n$.

We calculate

$$\hat{q}_t^n = \int_0^t \sigma_s^n \, dy_s^n = \int_0^{t\wedge\tau_n} \sigma_s^n \, d\hat{m}_s^n + \int_0^{t\wedge\tau_n} \sigma_s^n \, d\langle \hat{m}^n,\hat{z}^n\rangle_s$$

Because $\sigma^n \in L_{2loc}\ (y^n)$, there exists a sequence of stopping

times $\{s_m\}$ such that for all m $E_n[\int_0^{s_m} |\sigma_s^n|^2 \, ds] < \infty$ hence

$\int_0^{s_m} |\sigma_s^n|^2 \, ds < \infty$ a.s. $P^n$ and by equivalence also a.s. P.

Now under P measure $(\int_0^{t\wedge\tau_n} \sigma_s^n \, d\hat{m}_s, F_{yt}, t \in T)$ is a local martingale. From above we rewrite

$$m_t^n - \int_0^{t\wedge\tau_n} \sigma_s^n \, d\hat{m}_s^n = \langle m^n, \hat{z}^n \rangle_t + \int_0^{t\wedge\tau_n} \sigma_s^n \, d\langle \hat{m}^n, \hat{z}^n \rangle_s$$

$\in M_{1oc}^c(P) \cap BV$ and hence zero by 2.3.16.

So $m_t^n = \int_0^{t\wedge\tau_n} \sigma_s^n \, d\hat{m}_s$ for all n.

Step 4. Define $\sigma_t = \sigma_t^n$ on the set $\{(t,\omega) \in T\times\Omega | \; t \leq \tau_n(\omega)\}$ then since $\lim_n \tau_n = 1$ a.s.

$$m_t = \int_0^t \sigma_s \, d\hat{m}_s \text{ a.s. for all } t \in T.$$

This implies that $m \in M_{1oc}^c = M_{2oc}^c$, so there exists a sequence of stopping times $\{s_n\}$, $\lim_n s_n = 1$ a.s., such that for all n $m_t^n \in M_2$. Now

$$E[\langle m^n, m^n \rangle_1] = E[\int_0^{s_n} |\sigma_s|^2 \, ds] < \infty. \text{ hence}$$

$\sigma \in L_{21oc}(\hat{m})$ under P.

5.3.2. Theorem: $T = [0,1]$.

If 1. $dy_t = dn_t - dt = (\hat{\lambda}_t - 1) \, dt + d\hat{m}_t$,

2. where $(n_t, t \in T)$ is a counting process in $R^k$,

3. $(\hat{\lambda}_t, F_{nt}, t \in T)$ an adapted measurable process, right continuous, having left hand limits, and $\hat{\lambda}_t$, $\hat{\lambda}_{t-} > 0$ a.s. for all $t \in T$,

4. $(\hat{m}_t, F_{nt}, t \in T) \in M^d_{2loc}$,

5. given any $(m_t, F_{nt}, t \in T) \in M_{loc}$ in $R^n$,

then m has the representation,

$$m_t = \int_0^t \Sigma_s \, d\hat{m}_s \text{ a.s. for all } t \in T,$$

for an unique predictable process $(\Sigma_t, F_{nt}, t \in T) \in L_{1loc}(\hat{m}_2)$ in $R^{n\times k}$.

Proof. We give the proof only for the scalar real valued case.

Step 1. Since $\hat{\lambda}$ has the properties given above, we can define

$d\hat{z}_t \triangleq (-1 + 1/\hat{\lambda}_{t-}) \, d\hat{m}_t$, $(\hat{z}_t, F_{nt}, t \in T) \in M_{loc}$

Define the stopping times $\tau_n = \inf\{1, t \in T | \hat{\lambda}_t \geq n, 1/\hat{\lambda}_t \geq n\}$, and such that $m_{t\wedge\tau_n} \in M_1$. By the properties of $\hat{\lambda}$, $\lim_n \tau_n = 1$ a.s. Note that $d[\hat{z},\hat{z}]_t = (-1 + 1/\hat{\lambda}_{t-})^2 \, dn_t$, so using 1 we conclude that $d\langle \hat{z},\hat{z}\rangle_t = (-1 + 1/\hat{\lambda}_{t-})^2 \, \hat{\lambda}_{t-} \, dt =$

$= (\hat{\lambda}_{t-} - 2 + 1/\hat{\lambda}_{t-}) \, dt$

Define $\hat{\lambda}^n_t \triangleq \hat{\lambda}_{t\wedge\tau_n}$, $\hat{z}^n_t = \hat{z}_{t\wedge\tau_n}$, $m^n_t = m_{t\wedge\tau_n}$, $\hat{m}^n_t = \hat{m}_{t\wedge\tau_n}$, $y^n_t = y_{t\wedge\tau_n}$. Now $|\hat{\lambda}^n_{t-} - 2 + 1/\hat{\lambda}^n_{t-}| \leq 2n + 2$, so by 3.2.5. $E[\varepsilon(\hat{z}^n_1)] = 1$.

Step 2. Let $\frac{dP^n}{dP} = \varepsilon(\hat{z}^n_1)$, then $P^n$ is a new probability measure on $(\Omega, F)$. We have $d[\hat{m}^n, \hat{z}^n]_t = (-1 + 1/\hat{\lambda}^n_{t-}) \, dn_t$, which has the dual predictable projection $d\langle \hat{m}^n, \hat{z}^n\rangle_t = (-1 + 1/\hat{\lambda}_t) \, \hat{\lambda}_{t-} \, dt = -(\hat{\lambda}_{t-} - 1) \, dt$. We now apply the translation theorem 3.3.5.,

$$\hat{m}^n_t - \langle \hat{m}^n, \hat{z}^n\rangle_t = \hat{m}^n_t + \int_0^{t\wedge\tau_n} (\hat{\lambda}_{t-} - 1) \, dt = y_{t\wedge\tau_n} = y^n_t =$$

$= (n_t - t)_{t\wedge\tau_n}$, $(y^n_t, F_{nt}, t \in T) \in M_{loc}(P^n)$ and n is a counting process. Now by 2.4.20. n is a standard Poisson process on $[0,\tau_n]$, under the measure $P^n$.

Step 3. Note that $(m_t^n, F_{nt}, t \in T) \in M_1(P)$ so by 3.3.4. $m_t^n[\varepsilon(\hat{z}_t^n)]^{-1} \varepsilon M_1(P^n) \subset M_{loc}(P^n)$. By 2.6.8. under $P^n$ this martingale has the representation $\int_0^{t\wedge\tau_n} \phi_s^n \, dy_s^n$ for $(\phi_t^n, F_{nt}, t \in T) \in L_{1loc}(y^n)$ under $P^n$.

We now apply the differentiation rule under $P^n$ to $m_t^n = \varepsilon(\hat{z}_t^n)(\int_0^{t\wedge\tau_n} \phi_s^n \, dy_s^n)$. Let $\mu_t^n \triangleq \varepsilon(\hat{z}_t^n)$, and note that this is a semi-martingale under $P^n$, which can be proven by 3.3.5.

We calculate:

$$d\mu_t^n = \mu_{t-}^n \, d\hat{z}_t^n = \mu_{t-}^n \, (-1 + 1/\hat{\lambda}_{t-}^n) \, d\hat{m}_t^n,$$

$$d[\mu^n, \int_0^t \phi_s^n \, dy_s^n]_t = \mu_{t-}^n \, \phi_t^n \, (-1 + 1/\hat{\lambda}_{t-}^n) \, dn_t,$$

$$dm_t^n = (\int_0^{t\wedge\tau_n -} \phi_s^n \, dy_s^n) \, \mu_{t-}^n \, (-1 + 1/\hat{\lambda}_{t-}^n) \, d\hat{m}_t + \mu_{t-}^n \, \phi_t^n \, d\hat{m}_t$$

$$+ \mu_{t-}^n \, \phi_t^n \, (\hat{\lambda}_{t-}^n - 1) \, dt + \mu_{t-}^n \, \phi_t^n \, (-1 + 1/\hat{\lambda}_{t-}^n) \, dn_t$$

$$= [(\mu_{t-}^n \phi_t^n / \hat{\lambda}_{t-}^n) + \mu_{t-}^n \, (-1 + 1/\hat{\lambda}_{t-}^n) \, (\int_0^{t\wedge\tau_n -} \phi_s^n \, dy_s^n)] \, d\hat{m}_s$$

$$= \sigma_s^n \, d\hat{m}_s$$

which also holds under P.

Now $m_t^n = \int_0^{t\wedge\tau_n} \sigma_s^n \, d\hat{m}_s$

Step 4. Define $\sigma_t = \sigma_t^n$ on $[0,\tau_n)$ then $(\sigma_t, F_{nt}, t \in T)$ is adapted, predictable. Since $\lim_n \tau_n = 1$ a.s.,

$m_t = \int_0^t \sigma_s \, d\hat{m}_s$ a.s. for all $t \in T$. Since $m, \hat{m}_2 \in M_{loc}$,

and the integral is defined, it follows that

$\sigma \in L_{1loc}(\hat{m}_2)$.

Remarks:

1. Theorem 5.3.1. for the case where m is square integrable, was first proven by [Fujisaki, Kallianpur, Kunita, 1972]. Davis [1971] has given a slight extension of their result to local martingales. The last part of the proof of 5.3.1. differs from [Fujisaki et al., 1972], where we used techniques developed in Chapter 3 of this thesis.

2. The main points in the proof are the following: we do a transformation of measure, such that under the new measure the observed process y becomes a martingale, actually a Brownian motion. The transformation martingale $\hat{z}$ is chosen such that this is the case. Next the given martingale m is also translated into a local martingale on $(F_{yt}, t \in T)$. Then the martingale representation theorem on the $\sigma$-field generated by a Brownian motion is used. The result then follows by returning to P measure.

3. Theorem 5.3.2. as given here is new, the method of the proof is similar to that of 5.3.1., only some minor details are different. Any martingale representation theorem can by the above outlined procedure be converted in a martingale representation theorem on the $\sigma$-field generated by the associated semi-martingale.

4. The two previous theorems are of great importance in any problem where we deal with martingales on the σ-field generated by the observed process. The two main applications are estimation theory and stochastic control (see Davis [1971]).

5. Under the condition that the σ-fields generated by the observation and innovation processes are equal, we can derive 5.3.1. in a different way. Consider the assumptions of 5.3.1. then $\hat{m}$,the innovation process, is a Brownian motion. Now if $F_{yt} = F_{\hat{m}t}$ then by the representation theorem 2.6.3. $(m_t, F_{yt}, t \in T) \in M_{loc}$ has the representation $dm_t = \phi_t \, d\hat{m}_t$. The real problem however is that the equality of the σ-fields $F_{yt} = F_{\hat{m}t}$ has only been proven in some special cases, see the discussion at the end of section 4.4.

## 5.4. Estimation theory for continuous-time stochastic processes.

In this section we will review some of the basic aspects of estimation theory. We give the least squares estimation method for obtaining the optimal estimate. Then we discuss the general filtering problem and give a solution in the form of a dynamical equation. We start our discussion by defining the problem.

5.4.1. Definition: The estimation problem.

Given 1. $T \subset R$, the time interval of interest, usually $T = [0,1]$, or $[0,\infty)$,

2. $(x_t, t \in T)$ a stochastic process of interest that is not observed,

3. $(y_t, t \in T)$ a stochastic process that is observed, and that provides information concerning the process x,

4. let $F_{yt} \triangleq \sigma(y_s, \forall s \leq t)$, $(F_{yt}, t \in T)$.

The filtering problem: to obtain for all $t \in T$ an estimate of $x_t$ given the observations y on $[0,t]$, or equivalently $F_{yt}$.

The prediction problem: to obtain for $t, s \in T$, with $s < t$, the predicted estimate of $x_t$ given $F_{ys}$.

The smoothing problem: to obtain for $t, s \in T$, with $t < s$, the smoothed estimate of $x_t$ given $F_{ys}$.

Remarks:

1. We have not yet specified the dynamical equations for x and y nor the way y depends on x. This will be done later.
2. There are obviously many ways of obtaining estimates of $x_t$ given $F_{yt}$. We will now specify a cost function that associates a cost with each estimate. We will take the least squares error criterion, because it requires the least number of assumptions and does not require specification of a priori statistics or distributions. As pointed out by Wiener [1949, section 0.7.] this criterion has a physical interpretation, since the square of the estimation error is proportional to power.

5.4.2. Definition: The least squares error estimation problem.

Given 1. $(y_t, F_{yt}, t \in T)$ the observed stochastic process,

2. $(x_t, t \in T)$ the unknown stochastic process, with $E|x_t|^2 < \infty$ for all $t \in T$.

3. let $t, s \in T$ be given but fixed.

Problem: to find a random variable $z(t|s)$ such that

1. $z(t|s)$ is measurable with respect to $F_{ys}$

2. $E\,|z(t|s)|^2 < \infty$

3. $z(t|s)$ minimizes the cost $E|x_t - z(t|s)|^2 =$

$= E([x_t - z(t|s)]^T [x_t - z(t|s)])$

Definition: 1. such $z(t|s)$ is called the least squares error estimate of $x_t$ given by $F_{ys}$.

2. $e^x(t|s) \triangleq x_t - z(t|s)$ denotes the estimation error.

5.4.3. Lemma:

Given the above problem, then

1. the minimal least squares error estimate of $x_t$ given $F_{ys}$ is $\hat{x}(t|s) \triangleq E[x_t|F_{ys}]$ a.s.
2. if $z_s$ is any random variable measurable with respect to $F_{ys}$, and $E|z_s|^2 < \infty$ then

$$E[z_s^T\, e^x(t|s)|F_{ys}] = 0 = E[z_s^T\, e^x(t|s)]$$

Proof: Since $E|x_t|^2 < \infty$, $E[x_t|F_{ys}]$ is well defined.

Assertion two follows by the property of conditional expectations

$$E[z_s^T\, e^x(t|s)|\, F_{ys}] = z_s^T\, E[x_t - E[x_t|F_{ys}]|F_{ys}] = 0$$

Now take $x_t - z(t|s) = x_t - \hat{x}(t|s) + \hat{x}(t|s) - z(t|s)$

hence we get

$$E|x_t - z(t|s)|^2 = E|x_t - \hat{x}(t|s)|^2 + E\,|\hat{x}(t|s) - z(t|s)|^2$$
$$+ 2E[(\hat{x}(t|s) - z(t|s))^T (x_t - \hat{x}(t|s))] \geq$$
$$E|x_t - \hat{x}(t|s)|^2$$

where we used assertion 2. Since this holds for all $z(t|s)$ with equality only for $\hat{x}(t|s)$, the result follows.

Remarks:

1. The above proof is essentially the projection theorem in Hilbert space. This is related to the fact that conditional expectation is projection of a random variable on a $\sigma$-field Hence the above proof uses only properties of conditional expectation.
2. We remark that here we have taken the projection on $F_{ys}$. In Kalman and Bucy [1961], the orthogonal projection was also considered, but then on the space of all linear functionals of the past of the observed process.
3. The above two equations are the essential points for deriving the optimal filter. The optimality of the filter is determined by using these results.
4. We note that the least squares error estimate $\hat{x}(t|s)$ has the property that it is unbiased:

   $$E[e^{x}(t|s)] = E[x_t - \hat{x}(t|s)] = 0.$$
5. We now limit our attention to the case where $s = t$, the filtering problem. The prediction and smoothing problem are considered in section 5.5.

Filtering theory.

Remarks:

1. We can define as before $\hat{x}_t \triangleq E(x_t|F_{yt})$ for all $t \in T$ and hence obtain a stochastic process $(\hat{x}_t, F_{yt}, t \in T)$. Note that if $y_0 = 0$, then $F_{y0}$ contains only null sets, so we get $\hat{x}_0 = E(x_0|F_{y0}) = E(x_0)$.
2. We now have to specify the model or the dynamical equations

for the processes x and y. This has to be done to get some structure in the problem. We will assume that x and y are semi-martingales of a form as assumed in the previous chapter.

5.4.4. Definition: The filtering problems

Given: 1. $(F_t, t \in T)$ some family of sub-$\sigma$-fields,

2. $(x_t, t \in T)$, $x_t = x_0 + a_{1t} + m_{1t}$, satisfying $E|x_t| < \infty$, for all $t \in T$, $(a_{1t}, F_t, t \in T) \in IV$, $(m_{1t}, F_t, t \in T) \in M_{loc}$,

3. $(y_t, F_{yt}, t \in T)$ the observed process,

$y_t = \int_0^t d\langle m_2, m_2\rangle_s \, h_s + m_{2t}$, $(h_t, F_t, t \in T)$ is an adapted measurable process, $(m_{2t}, F_t, t \in T) \in M_{2loc}$.

The problem is to derive a generalized stochastic differential equation for the optimal estimate $\hat{x}$, given the observed process. This equation will then be called the filter.

Remarks: 1. Note that the processes x and y are specified in the form of semi-martingale equations, where $a_1$ and h are unspecified. No distributions for x and y are given. Note also that the equations for x and y are non-anticipative, by the adaptivity of $a_1, h, m_1, m_2$. The relation between x and y is implicitly specified by the adaptedness to $(F_t, t \in T)$.

2. In the above definition we do not specify the structure of the filter or of the equation that $\hat{x}$ must satisfy. This is done in the case of linear filtering by Wiener [1949], and by Kalman, Bucy [1961].

3. If we derive a generalized stochastic differential equation for $\hat{x}$, then by adaptivity it is necessarily non-anticipative, in

the sense that it does not need the future observations. However we cannot say whether the equation depends only on the current estimate, or on the complete past of the estimate.

5.4.5. Assertion:

Given 1. the filtering problem of 5.4.4., where the observations contain Brownian motion noise $(m_{2t}, F_t, t \in T)$;

$$dy_t = h_t \, dt + dm_{2t}$$

2. the innovation process as defined in 4.4.3.,

$$d\hat{m}_{2t} \triangleq dy_t - \hat{h}_t \, dt$$

then the optimal least square error estimate satisfies

$$d\hat{x}_t = d\hat{a}_{1t} + \Sigma_t \, d\hat{m}_{2t}, \; \hat{x}_0 = E(x_0)$$

where $(\Sigma_t, F_{yt}, t \in T)$ is a predictable process, and

$\hat{a}_{1t} = E[a_{1t} | F_{yt}]$.

5.4.6. Assertion:

Given 1. the following problem of 5.4.4., with counting process observations

$$dn_t = \lambda_t \, dt + dm_{2t}$$

2. the innovation process as defined in 4.4.4.

$$d\hat{m}_{2t} = dn_t - \hat{\lambda}_t \, dt$$

then the optimal least squares error estimate satisfies

$$d\hat{x}_t = d\hat{a}_{1t} + \Sigma_t \, d\hat{m}_{2t}, \; \hat{x}_0 = E(x_0),$$

where $(\Sigma_t, F_{yt}, t \in T)$ is a predictable process, and

$\hat{a}_{1t} = E[a_{1t} | F_{yt}]$.

The precise results are deferred to the sections 6.2. and 6.4. Here we only point out the main points. Similar to 4.4.1. we get that $\hat{x}$ satisfies $\hat{x}_t = \hat{x}_0 + \hat{a}_{1t} + \hat{m}_{1t}$, where $(\hat{m}_{1t}, F_{yt}, t \in T) \in M_{loc}$.

Now using the martingale representation theorem 5.3.1. or 5.3.2., we get $d\hat{m}_{1t} = \Sigma_t \, d\hat{m}_{2t}$, where $\hat{m}_2$ is the innovation process, and this gives the result. These assertions give the basic result, note the importance of the martingale representation theorem. One problem that still rests is, what is the process $\Sigma$. In the next chapter we will see that with a further specification of the process $a_1$ we can find an explicit expression for $\Sigma$. Note also that by the martingale approach we can give a similar derivation to the counting process observation problem, and for the equation with Brownian motion disturbances.

5.5. General prediction and smoothing problems.

In this section we will discuss in a general format the prediction and smoothing estimation problems.

5.5.1. Definition: The prediction problem.

Given the stochastic system model, satisfying the assumptions of theorem 4.4.1.,

$$x_t = x_0 + a_{1t} + m_{1t}$$

and $(y_t, t \in T)$ the observed process.

The prediction problem consists of finding a generalized stochastic differential equation for $\hat{x}(t|s)$, the least squares error estimate of $x_t$ given $F_{ys}$ where $s < t$. The following prediction problems can be distinguished: $\hat{x}(t|s)$

1. if t is fixed, we call it fixed point prediction,
2. if s is fixed, we call it fixed interval prediction,
3. if $t - s = \theta > 0$ is fixed, we call it fixed increment prediction.

5.5.2. Theorem:

Given the prediction problem of 5.5.1., and let by theorem 4.4.1

$$\hat{x}_t = \hat{x}_0 + \hat{a}_{1t} + \hat{m}_{1t}$$

then the predicted variable satisfies, $s < t$

$$\hat{x}(t|s) = \hat{x}(s|s) + E[\hat{a}_{1t}|F_{ys}] - \hat{a}_{1s}$$

where $\hat{x}(s|s) = \hat{x}_s$ is the filtered estimate.

Proof. By conditional expectation $\hat{x}(t|s) = E[x_t|F_{ys}] = E[\hat{x}_t|F_{ys}]$.

Furthermore: $\hat{x}(t|s) - \hat{x}(s|s) = E[\hat{x}_t - \hat{x}_s|F_{ys}] = E[\hat{a}_{1t} - \hat{a}_{1s} +$

$+ \hat{m}_{1t} - \hat{m}_{1s}|F_{ys}] = E[\hat{a}_{1t}|F_{ys}] - \hat{a}_{1s}$.

5.5.3. Definition: the smoothing problem.

Given the unknown process $(x_t, t \in T)$ satisfying $E|x_t| < \infty$ for all $t \in T$, and the observation process $y$ satisfying

$$y_t = \int_0^t d\langle m_2, m_2\rangle_s \, h_s + m_{2t}$$

with the usual assumptions.

The smoothing problem consists of finding a generalized stochastic differential equation for the estimate $\hat{x}(t|s)$, of $x_t$ given $F_{ys}$ where $t < s$. The following smoothing problems can be distinguished $\hat{x}(t|s)$

1. if $t$ is fixed we call it fixed point smoothing,
2. if $s$ is fixed, we call it fixed interval smoothing,
3. if $s - t = \theta > 0$ is fixed, we call it fixed lag smoothing.

Remark: The distinction between these special problems was given in [Kailath, Frost, 1968].

5.5.4. Theorem:

Given 1. the smoothing problem of definition 5.5.3.,

2. let $\hat{m}_2$ be the innovation process,

3. we restrict us to the observation equation with Brownian motion disturbances or to counting process observations

then the smoothed estimate satisfies

$$\hat{x}(t|s) = \hat{x}(t|t) + \int_t^s \Sigma(t,\tau)\, d\hat{m}_{2\tau} \;, \; t < s$$

for a predictable process $(\Sigma(t,\tau),\ F_{y\tau}, \tau \in [t,1], t \in T)$

Proof. Let t be fixed, and define

$\hat{m}_s \triangleq \hat{x}(t|s) - \hat{x}(t|t)$ then $(\hat{m}_s, F_{ys}, s \in [t,1]) \in M_1$

because $t < \tau < s$: $E[\hat{m}_s - \hat{m}_\tau | F_{y\tau}] = E[\hat{x}(t|s) - \hat{x}(t|\tau) | F_{y\tau}] = 0$.

Now for the Brownian motion or counting process case we have the martingale representation theorem of section 5.3., which gives the result.

Remark: The foregoing basic results will allow us to derive the detailed prediction and smoothing equations in the next chapter, for more detailed stochastic system models.

6. Stochastic differential equations for the optimal estimate.

6.1. Introduction.

In this chapter we derive in detail the stochastic differential equations which the optimal estimates satisfy. In Section 6.2. we derive the stochastic differential equation for the filtering problem for observations with Brownian motion noise. In Section 6.3. we discuss the question when the derived filters will be finite dimensional. In Section 6.4. we derive the stochastic differential equations for the filtering problem for counting process observations. In Section 6.5. we derive the stochastic differential equations for some special estimation problems, such as prediction, smoothing and systems with delays. In Section 6.6. we discuss in general the martingale approach to estimation problems, primarily to filtering.

6.2. Filtering from observations with Brownian motion noise.

In this section we derive the general filtering equations for the case where the observations are disturbed by Brownian motion.

6.2.1. Definition: the observation equation.

In this section we assume that:

1. $T = [0,1]$,
2. the observation equation is given by

   $dy_t = h_t dt + dm_{2t}, \; y_0 = 0$
3. where $(m_{2t}, F_t, t \in T) \in M_2^c$ is standard Brownian motion,
4. $(h_t, F_t, t \in T)$ is an adapted measurable process, $\sup_{t \in T} E|h_t|^2 < \infty$.

The conditions of 6.2.1. are necessary to derive the following lemma, which recalls some points proven earlier.

6.2.2. Lemma: Under the conditions of 6.2.1.

1. We can define $\hat{h}_t = E(h_t|F_{yt})$, $(\hat{h}_t, F_{yt}, t \in T)$ adapted, measurable,

2. let $e_t^h \triangleq h_t - \hat{h}_t$, then $\sup_{t \in T} E|e_t^h|^2 < \infty$,

3. $E[\int_T |h_s|^2 ds] < \infty$, $\qquad E[\int_T |\hat{h}_s|^2 ds] < \infty$,

4. $d\hat{m}_{2t} \triangleq dy_t - \hat{h}_t dt = e_t^h\, dt + dm_{2t}$, $(\hat{m}_{2t}, F_{yt}, t \in T) \in M_2^c$, is the innovation process as defined in 4.4.3., it is Brownian motion,

5. the conditions for 5.3.1 are satisfied, so if $(m_t, F_{yt}, t \in T) \in M_{loc}$, then $dm_t = \Sigma_t\, d\hat{m}_{2t}$, for an unique process $\Sigma \in L_{2loc}(\hat{m}_2)$.

Proof: 1 and 2 follow from 6.2.1.4. By the same condition $E[\int_T |h_s|^2 ds] = \int_T E|h_s|^2 ds \leq 1.\ \sup_{t \in T} E|h_t|^2 < \infty$. Since $E|\hat{h}_s|^2 = E|E(h_s|F_{ys})|^2 \leq E(E[|h_s|^2|F_{ys}]) = E|h_s|^2$ we have $E[\int_T |\hat{h}_s|^2 ds] = \int_T E|\hat{h}_s|^2 ds \leq \int_T E|h_s|^2 ds < \infty$. 4 follows from 4.4.3, and because $T = [0,1]$, $\hat{m}_2 \in M_2^c$. Because $E[\int_T |\hat{h}_s|^2 ds] < \infty$, the conditions for 5.3.1 are satisfied.

The stochastic differential equation for the optimal estimate.

6.2.3. Theorem:

Given the observation equation with the assumptions of 6.2.1. Given the semi-martingale $dx_t = f_t dt + dm_{1t}$, $x_0$,

1. where $(m_{1t}, F_t, t \in T) \in M_1$, hence $\langle m_1^c, m_2 \rangle$ exists, assume that $d\langle m_1^c, m_2 \rangle_t = \phi_{12t} dt$, where $(\phi_{12t}, F_t, t \in T) \in L_1(t)$,

2. $(f_t, F_t, t \in T)$ is an adapted measurable process, $\sup_{t \in T} E|f_t| < \infty$,

3. $\sup_{t \in T} E|x_t|^2 < \infty$ ,

then $\qquad d\hat{x}_t = \hat{f}_t dt + (\Sigma_t(x,h) + E[\phi_{12t}|F_{yt}])\, d\hat{m}_{2t}$, $\hat{x}_0 = E(x_0)$,

$\qquad d\hat{m}_{2t} = dy_t - \hat{h}_t dt$, the innovation process,

1. where $(\hat{x}_t, F_{yt}, t \in T)$ is a right continuous modification of $E[x_t|F_{yt}]$, having left hand limits,

2. $\hat{f}_t = E[f_t|F_{yt}]$,

3. $\Sigma_t(x,h) \triangleq E[e_t^x (e_t^h)^T|F_{yt}]$, will be called the conditional covariance of x and h.

Proof. By condition 2 we can define the process $(\hat{f}_t, F_{yt}, t \in T)$, $\hat{f}_t = E[f_t | F_{yt}]$. Note that $E[\int_T |f_s| ds] = \int_T E|f_s| ds \leq 1.\ \sup_{t\in T} E|f_t| < \infty$. We now apply 4.4.2 and get $d\hat{x}_t = \hat{f}_t dt + d\hat{m}_{1t}$, where $(\hat{x}_t, F_{yt}, t \in T)$ is a modification of $E[x_t | F_{yt}]$ having the above described properties and $(\hat{m}_{1t}, F_{yt}, t \in T) \in M_{loc}$. By 6.2.2.5 we can apply 5.3.1 to get the representation $d\hat{m}_{1t} = \Sigma_t dm_{2t}$, when $(\Sigma_t, F_{yt}, t \in T) \in L_{2loc}(\hat{m}_2)$. The problem now is to determine an expression for $\Sigma$. Let $e_t^f \triangleq f_t - \hat{f}_t$ then we calculate:

$$de_t^x \triangleq dx_t - d\hat{x}_t = e_t^f dt + dm_{1t} - \Sigma_t\, d\hat{m}_{2t}$$

$$d\hat{m}_{2t} = e_t^h dt + dm_{2t}$$

$$d[e^x, \hat{m}_2]_t = d[x, \hat{m}_2]_t - d[\hat{x}, \hat{m}_2]_t = d\langle m_1^c, m_2\rangle_t - \Sigma_t dt = (\phi_{12t} - \Sigma_t) dt$$

by condition 1. We use the differentiation rule,

$$e_t^x (\hat{m}_{2t})^T = e_s^x (\hat{m}_{2s})^T + \int_s^t e_\tau^x (d\hat{m}_{2\tau})^T + \int_s^t de_\tau^x (\hat{m}_{2\tau})^T + \int_s^t d[e^x, \hat{m}_2]_\tau$$

$$= e_s^x (\hat{m}_{2s})^T + \int_s^t e_\tau^x (e_\tau^h)^T d\tau + \int_s^t e_\tau^x (dm_{2\tau})^T + \int_s^t e_\tau^f (\hat{m}_{2\tau})^T d\tau$$

$$+ \int_s^t dm_{1\tau} (\hat{m}_{2\tau})^T - \int_s^t \Sigma_\tau d\hat{m}_{2\tau} (\hat{m}_{2\tau})^T + \int_s^t (\phi_{12\tau} - \Sigma_\tau) d\tau$$

By condition 3 and 6.2.2.4, $e^x$ and $\hat{m}_2$ are square integrable so by 5.4.3 we have, if $s \leq t$, $E[e_t^x (\hat{m}_{2t})^T | F_{ys}] = 0$. Now eliminating the $F_{yt}$ and $F_t$ local martingales and using that

$$E[\int_s^t e_\tau^f (\hat{m}_{2\tau}) d\tau | F_{ys}] = E[\int_s^t E[e_\tau^f (\hat{m}_{2\tau})^T | F_{y\tau}] d\tau | F_{ys}] = 0,$$

$$E[e_t^x (\hat{m}_{2t})^T | F_{ys}] = 0 = E[\int_s^t (e_\tau^x (e_\tau^h)^T + \phi_{12} - \Sigma_\tau) d\tau | F_{ys}]$$

Because of the integrability of the first two terms by the conditions 1, 3 and 6.2.1.4 we get $\Sigma_t = E[e_t^x (e_t^h)^T | F_{yt}] + E[\phi_{12t} | F_{yt}]$.

Remarks

1. The semi-martingale x represents any unobserved process, which can

be modelled this way. The processes f and $m_1$ are unspecified. Note that f can be any adapted process, which can depend on the past of x, and not necessarily just on the current state $x_t$. This formulation is an extension of earlier stochastic system equations.

2. The conditions in 6.2.3 on x,f, and $m_1$ are of course related by the expression for x. The way the conditions are stated, is the form in which they are needed in the proof. In a subsequent proof we will use this result again. Of course equivalent or stricter conditions could be stated, however it was left that these conditions were most easily to apply.

3. The result stated is similar to that of [Fujisaki, Kallianpur, Kunita, 1972], except that we have considered a more general case. The proof is new, and was partially inspired by a simple example given by Wong [1972], also Wong [1973]. For a further discussion on the method used see Section 6.6.

4. Note that the derived result has a similar interpretation as the Kalman-Bucy filter. If the conditional covariance matrix $\Sigma_t(x,h)$ is large, then the optimal estimate $\hat{x}$ relies more on the innovation process. If this covariance is small, then the innovation process plays a lesser role, and the optimal estimate relies more on the estimate of $(f_t, t \in T)$.

The stochastic differential equation for the conditional covariance. We now define the stochastic system, for which we will derive a stochastic differential equation for the optimal estimate and the conditional covariance.

6.2.4. Definition: the stochastic system equations.

Given the observation equation as defined in 6.2.1 with the conditions assumed there. Given an unobserved process x, a semi-martingale, and suppose that the process h, as defined in 6.2.1, is also a semi-martingale:

$$dx_t = f_t dt + dm_{1t},\ x_0,$$

$$dh_t = r_t dt + dm_{3t},\ h_0,$$

where we impose the following conditions:

1. $(m_{1t}, F_t, t \in T) \in M_2$, $(m_{3t}, F_t, t \in T) \in M_2$, $\langle m_1, m_2 \rangle = 0$, $\langle m_3, m_2 \rangle = 0$, $d\langle m_1, m_3 \rangle_t = \phi_{13t} dt$, $d\langle m_3, m_3 \rangle_t = \phi_{33t} dt$, $(\phi_{13t}, F_t, t \in T) \in L_1(t)$, $(\phi_{33t}, F_t, t \in T) \in L_1(t)$,
2. $(f_t, F_t, t \in T)$, $(r_t, F_t, t \in T)$ are adapted measurable processes,
3. $\sup_{t \in T} E|x_t|^4 < \infty$, $\sup_{t \in T} E|h_t|^4 < \infty$,
4. $\sup_{t \in T} E|f_t|^2 < \infty$, $\sup_{t \in T} E|r_t|^2 < \infty$,
5. $\sup_{t \in T} E|\phi_{13t}| < \infty$, $\sup_{t \in T} E|\phi_{33t}| < \infty$.

Remark: consider the conditional covariance $\Sigma_t(x,h) = E[e_t^x (e_t^h)^T | F_{yt}] = E[x_t (h_t)^T | F_{yt}] - \hat{x}_t (\hat{h}_t)^T$. One usually derived a stochastic differential equation for the estimate of $x_t (h_t)^T$. Here we take a different approach and obtain directly a stochastic differential equation for $\Sigma_t(x,h)$.

6.2.5. Theorem: the general filtering equations.

Given the stochastic system as defined in 6.2.4. The optimal filter has the form:

$$d\hat{m}_{2t} = dy_t - \hat{h}_t dt, \text{ the innovation process,}$$

$$d\hat{x}_t = \hat{f}_t dt + \Sigma_t(x,h) d\hat{m}_{2t},\ \hat{x}_0 = E(x_0),$$

$$d\hat{h}_t = \hat{r}_t dt + \Sigma_t(h,h) d\hat{m}_{2t},\ \hat{h}_0 = E(h_0),$$

$$d\Sigma_t(x,h) = [\Sigma_t(x,r) + \Sigma_t(f,h) - \Sigma_t(f,h)\ \Sigma_t(h,h) + E[\phi_{13t}|F_{yt}]]dt +$$
$$+ \Sigma_t(h,h,h)d\hat{m}_{2t},\ \Sigma_0(x,h) = E[e_0^x(e_0^h)^T],$$
$$d\Sigma_t(h,h) = [\Sigma_t(h,r) + \Sigma_t(r,h) - \Sigma_t(h,h)\ \Sigma_t(h,h) + E[\phi_{33t}|F_{yt}]]dt +$$
$$+ \Sigma_t(h,h,h)d\hat{m}_{2t},\ \Sigma_0(h,h) = E[e_0^h(e_0^h)^T].$$

Remark: Still unknown, $\hat{f}_t$, $\hat{r}_t$ and several terms in the equations for $\Sigma_t(x,h)$, $\Sigma_t(h,h)$.

Notation: $\Sigma_t(x,r) \triangleq E[(x_t-\hat{x}_t)(r_t-\hat{r}_t)|F_{yt}]$ and similar expressions for the other conditional covariance matrices.

$\Sigma_t(x,h,h)d\hat{m}_{2t} \triangleq (\Sigma_t(x,h^1,h)d\hat{m}_{2t},\ldots,\Sigma_t(x,h^k,h)d\hat{m}_{2t})$, in the scalar case $\sigma_t(x,h,h) = E[e_t^x(e_t^h)^2|F_{yt}]$.

Proof: Because of notational problems we give the proof for the scalar case only. From 6.2.2.4. $d\hat{m}_{2t} = e_t^h dt + dm_{2t}$. The conditions 1, 3, 4 of 6.2.4 imply that we can apply 6.2.3 to the semi-martingales x and h,

$$d\hat{x}_t = \hat{f}_t dt + \sigma_t(x,h)d\hat{m}_{2t},$$
$$d\hat{h}_t = \hat{r}_t dt + \sigma_t(h,h)d\hat{m}_{2t},$$

The problem now is to derive a stochastic differential equation for $\sigma_t(x,h)$. We calculate, let

$e_t^x \triangleq x_t - \hat{x}_t$, $e_t^f = f_t - \hat{f}_t$, $e_t^r = r_t - \hat{r}_t$, then

$$de_t^x = e_t^f dt + dm_{1t} - \sigma_t(x,h)d\hat{m}_{2t},$$
$$= [e_t^f - \sigma_t(x,h)e_t^h]dt + dm_{1t} - \sigma_t(x,h)dm_{2t},$$
$$d[e^x,e^h]_t = d[m_1,m_3]_t + \sigma_t(x,h)\sigma_t(h,h)dt,$$
$$d(e_t^x e_t^h) = a_t dt + dm_{4t},$$
$$a_t \triangleq e_t^x(e_t^r - \sigma_t(h,h)e_t^h) + (e_t^f - \sigma_t(x,h)e_t^h)e_t^h + \phi_{13t} + \sigma_t(x,h)\ \sigma_t(h,h)$$
$$dm_{4t} \triangleq e_t^h dm_{1t} + e_t^x dm_{3t} - [\sigma_t(x,h)e_t^h + e_t^x\sigma_t(h,h)]dm_{2t}$$
$$+ d([m_1,m_3]_t - \langle m_1,m_3\rangle_t),$$

$(m_{4t},F_t,t \in T) \in M_{loc}$. Note that $(e_t^x e_t^h)$ is a real valued semi-martingale, of the form assumed in 6.2.3, which result we now want to apply. We check the conditions. We have to check that $\sup_{t\in T} E|a_t| < \infty$. This follows from the conditions 3, 4 and 5 of 6.2.4, for example:

$$E|e_t^x e_t^h \sigma_t(h,h)| \leq E|e_t^x e_t^h|^2 \quad E|\sigma_t(h,h)|^2 \leq E|e_t^x|^4 \quad (E|e_t^h|^4)^3 \leq$$

$$K \quad E|x_t|^4 \quad (E|h_t|^4)^3 < \infty, \text{ where K is a positive constant.}$$

Similarly by condition 3 of 6.2.4 we have

$$E|e_t^x e_t^h|^2 \leq K \quad E|x_t|^4 \quad E|h_t|^4 < \infty$$

Now by the equation for $(e_t^x e_t^h)$, $E|m_{4t}| < \infty$, hence $m_4$ is a martingale and $d[m_4,m_2]_t = d\langle m_4^c,m_2\rangle_t = -[\sigma_t(x,h)e_t^h + e_t^x \sigma_t(h,h)]dt$. We now can apply 6.2.3 to $(e_t^x e_t^h)$ and using that

$E[\dfrac{d\langle m_4^c,m_2\rangle_t}{dt} | F_{yt}] = 0$, and the notation introduced before, we get

$$d\sigma_t(x,h) = [\sigma_t(x,r) - \sigma_t(x,h)\sigma_t(h,h) + \sigma_t(f,h) - \sigma_t(x,h)\sigma_t(h,h) +$$
$$+ E[\phi_{13t}|F_{yt}] + \sigma_t(x,h)\sigma_t(h,h)]dt + \sigma_t(x,h,h)d\hat{m}_{2t}, \text{ which}$$

gives the result. The equation for $\sigma_t(h,h)$ follows from this.

Note that

$$\sigma_t(x,h,h) = E[(e_t^x e_t^h - \sigma_t(x,h))(h_t-\hat{h}_t)|F_{yt}] = E[e_t^x(e_t^h)^2|F_{yt}].$$

Remark:

Theorem 6.2.5 is the main result of this section. It gives the stochastic differential equation for the optimal estimates $\hat{x},\hat{h}$ and for the conditional covariance $\Sigma_t(x,h)$. The filtering problem is not solved with these equations, as indicated by the unknown variables mentioned. In the next section we discuss in more detail the implementation of this filter, which in general will be infinite dimensional.

Applications.

At this point we want to point out a modelling guideline and give a special case of the above formula's. We assumed the unknown process x to be a semi-martingale, some special forms of which are: $x_t = x_0$, $x_t = x_0 + a_t$, $x_t = x_0 + m_t$, where $a \in BV$, $m \in M_{loc}$. Now $x_t = x_0$ for all $t \in T$, leads to $dx_t = 0$, hence $d\hat{x}_t = 0$, $\hat{x}_0 = E(x_0)$, so this is not usefull. Similarly $x_t = x_0 + a_t$ leads to $d\hat{x}_t = d\hat{a}_t$. Note that in both cases there is no 'feedback' correction using the innovation process. This follows because the equation for x does not include any martingale term, which represents the disturbance. So as a modelling guideline processes to be estimated should be modelled by an equation which includes a martingale term. In the case where we suspect x to be approximately constant it can be modelled as $x_t = x_0 + m_t$, where $m \in M_{loc}$. We state the result.

6.2.6. Corollary:

Given the system of 6.2.4, with scalar quantities

$dy_t = h_t dt + dm_{2t}$, $y_0 = 0$,

$h_t = h_0 + m_{3t}$,

$h_0$ an unknown random variable, $E(h_0)$, $E(h_0)^2$ given,

$(m_{3t}, F_t, t \in T) \in M_2$, $d\langle m_3, m_3\rangle_t = \phi_{3t}dt$, $\langle m_3, m_2\rangle = 0$. Then the optimal filter for h is:

$d\hat{m}_{2t} = dy_t - \hat{h}_t dt$, the innovation process,

$d\hat{h}_t = \sigma_t(h,h)d\hat{m}_{2t}$, $\hat{h}_0 = E(h_0)$,

$d\sigma_t(h,h) = [E(\phi_{3t}|F_{yt}) - \sigma_t^2(h,h)]dt + \sigma_t(h,h,h)d\hat{m}_{2t}$,

$\sigma_0(h,h) = (h_0 - E(h_0))^2$.

This follows from 6.2.5.

The stochastic system model with Brownian motion disturbances.

We now specialize the discussion to what is generally regarded as the nonlinear system, with Brownian motion disturbances.

6.2.7. Corollary:

Given the stochastic system

$dx_t = f(t,x_t)dt + G(t,x_t)dm_{1t}, \; x_0,$

$dy_t = h(t,x_t)dt + dm_{2t}, \; y_0 = 0,$

that satisfies the assumptions in 6.2.1 and 6.2.4

Furthermore $(m_{1t},F_t,t \in T)$ is a standard Brownian motion in $R^{n_1}$, independent of $m_2$.

$f : T \times R^{n_x} \to R^{n_x}$, $h : T \times R^{n_x} \to R^{n_y}$, $G : T \times R^{n_x} \to R^{n_x \times n_1}$ are jointly measurable functions, $f(t,x)$, $h(t,x)$ are assumed to be twice continuously differentiable in x and once in t. Furthermore f and G are such that the above stochastic differential equation has an unique solution.

Then the optimal filter is

$d\hat{x}_t = \hat{f}_t dt + \Sigma_t(x,h)d\hat{m}_{2t}, \; \hat{x}_0 = E(x_0),$

$d\hat{h}_t = \hat{r}^h_t dt + \Sigma_t(h,h)d\hat{m}_{2t}, \; \hat{h}_0 = E(h(0,x_0)),$

and the equations of the conditional covariances as in 6.4.5.

$\hat{f}_t \triangleq E[f(t,x_t)|F_{yt}], \; \hat{h}_t = E[h(t,x_t)|F_{yt}].$

Proof. The theorem follows from 6.4.5, since by the assumptions, $f(t,x_t)$ and $h(t,x_t)$ are semi-martingales, for which we can derive a detailed semi-martingale expression by the differentiation rule.

6.2.8. Corollary: The Kalman-Bucy filter.

Given the stochastic system, satisfying the assumptions of 6.2.7,

$dx_t = A(t)x_t dt + B(t)dm_{1t}, \; x_0,$

$dy_t = C(t)x_t dt + dm_{2t}, \; y_0 = 0,$

where $F_t = \sigma(m_{1s},m_{2s},x_0,\forall s \le t)$, $m_1,m_2$ are standard Brownian motions, $\langle m_1,m_2\rangle = 0$, $E|x_0|^2 < \infty$, $x_0$ is a Gaussian random variable.

The optimal filter is

$$d\hat{m}_{2t} = dy_t - C(t)\hat{x}_t dt,$$

$$dx_t = A(t)\hat{x}_t dt + \Sigma_t(x,x)C^T(t)d\hat{m}_{2t},\ \hat{x}_0 = E(x_0),$$

$$d\Sigma_t(x,x) = [\Sigma_t(x,x)A(t) + A^T\Sigma_t(x,x) + B(t)B^T(t) - \Sigma_t(x,x)C^T(t)C(t)\Sigma_t(x,x)]dt,$$

$$\Sigma_0(x,x) = E[e_0^x(e_0^x)^T].$$

The proof follows from 6.2.5 with the following observations. Since $x_0$ is Gaussian, and $m_1$ is Brownian motion, x is Gaussian and by linearity so is y. Actually $e^x = x - \hat{x}$ and y are jointly Gaussian, and since they are uncorrelated, they are independent. Now

$$\sigma_t(x,x,h) = E[(e_t^x)^2(e_t^h)|F_{yt}] = C\ E[(e_t^x)^3|F_{yt}] = C\ E(e_t^x)^3 = 0.$$

The general formulation of the first part of this section allows us to obtain the equations for another class of problems. We discuss here the scalar case only.

6.2.9. Definition: Given the system equations.

$$dn_t = \lambda_t dt + dm_{1t},\ n_0 = 0,$$

$$dy_t = n_t dt + dm_{2t},\ y_0 = 0,$$

satisfying the assumptions of 6.2.4, where $(n_t,F_t,t \in T)$ is a counting process, $(\lambda_t,F_t,t \in T)$ its rate process, where $\lambda_t > 0$ a.s. for all $t \in T$, $\lambda \in L_1(t)$, $(m_{1t},F_t,t \in T) \in M_2$, $d\langle m_1,m_1\rangle_t = \lambda_t dt$. The problem is to obtain an estimate of the counting process n, given the observations y.

6.2.10. Corollary:

Given the filtering problem of 6.2.9 the optimal filter is:

$d\hat{m}_{2t} = dy_t - \hat{n}_t dt$, the innovation process,

$$d\hat{n}_t = \hat{\lambda}_t dt + \sigma_t(n,n)d\hat{m}_{2t},\ \hat{n}_0 = 0,$$

$$d\sigma_t(n,n) = [2\sigma_t(n,\lambda) + \hat{\lambda}_t - \sigma_t^2(n,n)]dt + \sigma_t(n,n,n)d\hat{m}_{2t},$$

$$\sigma_0(n,n) = 0.$$

Still unknown $\hat{\lambda}_t, \sigma_t(n,\lambda)$, $\sigma_t(n,n,n)$.

This result follows from 6.2.5. Note that the estimate $\hat{n}$ of the counting process does not necessarily assume integer values. The above theorem is a general result, more specific formula's can be obtained by assuming certain models for the rate process $\lambda$.

## 6.3. Nonlinear filters.

In this section we discuss in more detail the nonlinear filters derived in the previous section. Little research has been done on this subject, but without these points this thesis would have been incomplete. What follows is a nonrigorous discussion of the problems and an outline of possible solutions to implement the derived filters.

Problem statement.

Given the stochastic system as defined in 6.2.1 and 6.2.4, where all variables are real valued:

$dy_t = h_t dt + dm_{2t},\ y_0 = 0,$

$dx_t = f_t dt + dm_{1t},\ x_0,$

$dh_t = r_t dt + dm_{3t},\ h_0,$

where $(f_t, t \in T)$, $(r_t, t \in T)$ are still unspecified.

The stochastic differential equations for the optimal estimates are

$d\hat{m}_{2t} = dy_t - \hat{h}_t dt,$

$d\hat{x}_t = \hat{f}_t dt + \sigma_t(x,h) d\hat{m}_{2t},\ \hat{x}_0 = E(x_0),$

$d\hat{h}_t = \hat{r}_t dt + \sigma_t(h,h) d\hat{m}_{2t},\ \hat{h}_0 = E(h_0),$

still unknown $\hat{f}_t$, $\hat{r}_t$, $\sigma_t(x,h)$, $\sigma_t(h,h)$.

The problem is this: given further specification of the stochastic system, how many variables and which ones do we have to estimate to determine $(\hat{x}_t, t \in T)$, in other words what is the order of the filter for $\hat{x}$.

Consider which variables need to be estimated, first we need $\hat{x}$ and $\hat{h}$.

The equations for these quantities contain $\hat{f}_t$, $\hat{r}_t$, $\sigma_t(x,h)$, $\sigma_t(h,h)$ as unknowns. To further specify the above stochastic system, we assume that f and r are also semi-martingales of the form assumed for x. By this assumption we get two more unknown processes, and two unknown conditional covariances. Depending on further specification of the system we continue this way indefinitely. Even if $\hat{f}_t$, $\hat{r}_t$ are known, then the stochastic differential equations for $\sigma_t(x,h)$,$\sigma_t(h,h)$ (see 6.2.5) still contain the unknown variables $\sigma_t(x,h,h)$, $\sigma_t(h,h,h)$. For these we can again derive stochastic differential equations, which however contain fourth order conditional moments and so etc.

The conclusion from this argument is that in general the filter is infinite dimensional. This fact is well known from the literature, for the case of nonlinear stochastic systems with Brownian motion disturbances, see for example Kushner [1967b]. The question now is for which stochastic systems do we get finite dimensional filters. From our preceding discussion we see that to get finite dimensional filters two conditions need to be satisfied:

1. the processes $\hat{f}$ and $\hat{r}$ need to be known,
2. the sequence of conditional higher moments must stop after a finite number.

We discuss both these conditions in more detail.

Filtering equations and Hermite polynomials.

Rather than talking in general terms we will give an example that demonstrates some essential ideas. It satisfies one condition for obtaining finite dimensional filters.

Example:

Given the stochastic system model of 6.2.7 of the form:

$dx_t = -x_t dt + dm_{1t}$, $x_0$,

$dy_t = h(x_t)dt + dm_{2t}$, $y_0 = 0$,

where $m_1$, $m_2$ are independent Brownian motions, and $h : R \mapsto R$ is a twice continuously differentiable function. By the differentiation rule 2.5.23 we get that

$dh(x_t) = [-x_t h'(x_t)] + \frac{1}{2} h''(x_t)]dt + h'(x_t)dm_{1t}$.

To prevent the filter from growing in order, we now restrict the class of functions, h, and demand that the equation for h is linear i.e.

$[-x_t h^1(x_t) - \frac{1}{2} h''(x_t)] = c\, h(x_t)$ for some constant c.

However this equation is a differential equation:

$h''(x) - 2xh'(x) - 2ch(x) = 0$.

For $c = -n$, $n = 0,1,2,\ldots.$ it has as solution the Hermite polynomials

$h_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} (e^{-x^2})$.

The first few polynomials are: $h_0(x) = 1$, $h_1(x) = 2x$, $h_2(x) = 4x^2 - 2$, $h_3(x) = 8x^3 - 12x$, $h_4(x) = 16x^4 - 48x^2 + 12$.

We now give the filtering equations for our example.

Let $h(x_t) = h_3(x_t) = 8x_t^3 - 12x_t$, and let $h_2(x_t) = 4x_t^2 - 2$,

$\hat{h}_{2t} = E[h_2(x_t)|F_{yt}]$, $\hat{h}_{3t} = E[h_3(x_t)|F_{yt}]$. Then

$d\hat{m}_{2t} \triangleq dy_t - \hat{h}_{3t}dt$

$d\hat{x}_t = -\hat{x}_t dt + \sigma_t(x,h_3)d\hat{m}_{2t}$, $\hat{x}_0$,

$d\hat{h}_{2t} = -2\hat{h}_{2t}dt + \sigma_t(h_2,h_3)d\hat{m}_{2t}$, $\hat{h}_{20}$,

$d\hat{h}_{3t} = -\hat{h}_{3t}dt + \sigma_t(h_3,h_3)d\hat{m}_{2t}$, $\hat{h}_{30}$,

We can also state the stochastic differential equations for $\sigma_t(x,h_3)$, $\sigma_t(h_2,h_3)$, $\sigma_t(h_3,h_3)$, which after some simplifications only

contain the unknown $\sigma_t(x,h_3,h_3)$, $\sigma_t(h_2,h_3,h_3)$, $\sigma_t(h_3,h_3,h_3)$.
It turns out to be necessary also to estimate the lower order polynomial $h_2(x_t)$.

Let us review what has been done and draw some conclusions. By taking a special stochastic system, where we take in the observation equation a Hermite polynomial, we can derive the filtering equations, in which as only unknowns the higher conditional moments are left. We will discuss this problem in the sequel.

The way we obtain the Hermite polynomials, is by enforcing linearity in the stochastic differential equation for h. That we get the Hermite polynomials is not so surprising, since they are closely related to the Gaussian distribution, see the way they are defined. More research could be done on the use of Hermite polynomials in this context. Several extensions of the above ideas have been considered but no usefull results have been obtained. Hermite polynomials have been used before in statistics and estimation. One application is the approximation of arbitrary functions by a finite number of Hermite polynomials. This method is well known in mathematical statistics, it is called the Gram-Charlier series representation. For this see Cramer [1946], which has references to earlier work, and Deutsch [1969 section 8.4.3]. A recent reference on the use of Hermite polynomials in estimation is Srinivasan [1970].

Conditional covariances.

The next problem in obtaining finite dimensional filters is the sequence of conditional moments. Not much research has been done on this problem. The problem is to find conditions such that the sequence of conditional moments stop, by having one of them vanish.

In the scalar case this leads to the problem, when is

$E[(e_t^h)^3|F_{yt}] = 0$, or $E[e_t^x(e_t^h)^2|F_{yt}] = 0$.

For the case of a linear system, with Brownian motion disturbances we have that $e_t^h = c\ e_t^x$. Then $e_t^x$ and $y_t$ are jointly Gaussian, uncorrelated, hence independent, so

$E[(e_t^x)^3|F_{yt}] = E[(e_t^x)^3] = 0$.

Actually to obtain this result, only the symmetry of the conditional distribution of $(e_t^h)$ given $F_{yt}$ is needed. No further points have been found concerning this problem, specifically for the example derived earlier we have not been able to show that the third order conditional moment vanishes.

The foregoing gives a short summary of the little work we have done on nonlinear filters. We have concentrated attention on the problem, when is the optimal filter finite dimensional. We have not considered approximations to the optimal filter, since this approach is well known and established in the literature. For references see Schwartz, Stear [1968], which compares several nonlinear filters and gives further references.

## 6.4. Filtering for counting processes.

In this section we derive the general filtering equations for the case where the observation process is a counting process.

### 6.4.1. Definition. The observation equation.

In this section we assume that $T = [0,1]$,

1. the observation equation is given by

   $dn_t = \lambda_t dt + dm_{2t},\ n_0 = 0,$

2. the rate process $\lambda$ is given by

   $d\lambda_t = r_t dt + dm_{1t},\ \lambda_0,$

3. where $n$ is a counting process in $R^k$,

4. $(m_{2t}, F_t, t \in T) \in M_2^d$, $d\langle m_2, m_2\rangle_t = \Lambda(\lambda_t)dt = \text{diag}(\lambda_t^1, \ldots \lambda_t^k)dt$,

5. $(\lambda_t, F_t, t \in T)$ is a supermartingale, $\lambda_1 > 0$ a.s., $\sup_{t \in T} E(\lambda_t) < \infty$

6. $(m_{1t}, F_t, t \in T)$ a right continuous martingale, having left hand limits,

7. $(r_t, F_t, t \in T)$ an adapted measurable process, $\sup_{t \in T} E|r_t| < \infty$.

We recall some points derived earlier.

6.4.2. Lemma. Under the conditions of 6.4.1:

1. $d\hat{\lambda}_t = \hat{r}_t dt + d\hat{m}_{1t}$, $\hat{\lambda}_0 = E(\lambda_0)$,

2. where $(\hat{r}_t, F_{nt}, t \in T)$ is an adapted measurable process, $\hat{r}_t = E(r_t | F_{nt})$,

3. $(\hat{m}_{1t}, F_{nt}, t \in T)$ is a right continuous martingale, having left hand limits,

4. $(\hat{\lambda}_t, F_{nt}, t \in T)$ a right continuous modification of $E(\lambda_t | F_{nt})$, having left hand limits, a supermartingale,

5. $\hat{\lambda}_t > 0$, $\hat{\lambda}_{t-} > 0$ a.s. for all $t \in T$,

6. let $e_t^\lambda \triangleq \lambda_t - \hat{\lambda}_t$, then $\sup_{t \in T} E|e_t^\lambda| < \infty$,

7. $E[\int_T \lambda_s ds] < \infty$, $E[\int_T \hat{\lambda}_s ds] < \infty$,

8. $d\hat{m}_{2t} \triangleq dn_t - \hat{\lambda}_t dt = e_t^\lambda dt + dm_{2t}$ is the innovation process as defined in 4.4.4, $(\hat{m}_{2t}, F_{nt}, t \in T) \in M_2^d$, $d\langle \hat{m}_2, \hat{m}_2\rangle_t = \Lambda(\hat{\lambda}_t)dt$,

9. the conditions for the martingale representation theorem 5.3.2 are satisfied.

Proof. 1, 2, 3, 4 follow from 4.4.2. Similarly 8 follows from 4.4.4, and 6, 7 and 9 are obvious. By definition of $\hat{\lambda}_1 = E(\lambda_1 | F_{n1})$ $\hat{\lambda}_1 > 0$ a.s., Since $E[\hat{\lambda}_t | F_{ns}] = E[E[\lambda_t | F_s] | F_{ns}] \leq \hat{\lambda}_s$, $\hat{\lambda}$ is a supermartingale, so by [Meyer,1966,VI,T15] it follows that $\hat{\lambda}_t > 0$, $\hat{\lambda}_{t-} > 0$ a.s. for all $t \in T$.

Remark: The supermartingale condition is necessary to prove that $\hat{\lambda}_{t-} > 0$ a.s. for all $t \in T$. This last point is absolutely necessary in the subsequent formula's, since we will use $\hat{\lambda}^{-1}_{t-}$. We suspect that we can drop the supermartingale condition on $\lambda$, and replace it by the condition $\lambda$ is a semimartingale of the form given in 6.4.1.2 satisfying $\lambda_t > 0$, $\lambda_{t-} > 0$ a.s. for all $t \in T$. However we have as yet no proof that this implies that $\hat{\lambda}^{-1}_{t-} > 0$ a.s.

The stochastic differential equation for the optimal estimate.

6.4.3. Theorem:

Given the observation equation 6.4.1 and the conditions assumed there.

Given the semi-martingale:

$$dx_t = f_t dt + dm_t, \quad x_0,$$

1. Where $(m_t, F_t, t \in T)$ is a martingale, and there exists $\langle m, m_2 \rangle$, satisfying $d\langle m, m_2\rangle_t = \phi_t dt$, $(\phi_t, F_t, t \in T) \in L_1(t)$,
2. $(f_t, F_t . t \in T)$ is an adapted measurable process, $\sup_{t \in T} E|f_t| < \infty$,
3. $\sup_{t \in T} E|x_t|^2 < \infty$,
4. in addition to 6.4.1 $\sup_{t \in T} E|\lambda_t|^2 < \infty$,

then $d\hat{m}_{2t} = dn_t - \hat{\lambda}_t dt$, the innovation process,

$$d\hat{x}_t = \hat{f}_t dt + (\Sigma_t(x,\lambda) + E[\phi_t | F_{nt}]) \wedge(\hat{\lambda}^{-1}_{t-}) d\hat{m}_{2t}, \quad \hat{x}_0 = E(x_0),$$

1. where $(\hat{x}_t, F_{yt}, t \in T)$ is a right continuous modification of $E(x_t | F_{yt})$, having left hand limits,
2. $\hat{f}_t = E[f_t | F_{nt}]$,
3. $\Sigma_t(x,\lambda) = E[(x_t - \hat{x}_t)(\lambda_t - \hat{\lambda}_t) | F_{nt}]$ a.s., called the conditional covariance of $x$ and $\lambda$,
4. $\wedge(\hat{\lambda}^{-1}_t) \triangleq \text{diagonal } ((\hat{\lambda}^1_t)^{-1}, \ldots . (\hat{\lambda}^k_t)^{-1})$

Proof. By condition 2 we can define the adapted process $(\hat{f}_t, F_{yt}, t \in T)$, $\hat{f}_t = E[f_t | F_{yt}]$. Now $E[\int_T |f_s| ds] \leq 1. \sup_{t \in T} E|f_t| < \infty$.

By 4.4.2 we have that $d\hat{x}_t = \hat{f}_t dt + d\hat{m}_t$, where $\hat{x}$ has the properties stated above, and $(\hat{m}_t, F_{yt}, t \in T) \in M_{loc}$. By 6.4.2.5 we can apply 5.3.2 to get the representation $d\hat{m}_t = \Sigma_t d\hat{m}_{2t}$, where $\Sigma$ is an unique predictable process. The problem now is to determine an expression for $\Sigma$. Let $e_t^f \triangleq f_t - \hat{f}_t$, then we calculate

$$de_t^x \triangleq dx_t - d\hat{x}_t = e_t^f dt + dm_t - \Sigma_t d\hat{m}_{2t},$$

$$d\hat{m}_{2t} = e_t^\lambda dt + dm_{2t}$$

$$d[e^x, \hat{m}_2]_t = d[x, \hat{m}_2]_t - d[\hat{x}, \hat{m}_2]_t = d[m, m_2]_t - \Sigma_t d[\hat{m}_2, \hat{m}_2]_t$$

$$= (\phi_t - \Sigma_t \Lambda(\hat{\lambda}_{t-})dt + d([m,m_2]_t - \langle m, m_2\rangle_t) - \Sigma_t d([\hat{m}_2, \hat{m}_2]_t - \langle \hat{m}_2, \hat{m}_2\rangle_t)$$

$$e_t^x(\hat{m}_{2t})^T = e_s^x(\hat{m}_{2s})^T + \int_s^t de_\tau^x(\hat{m}_{2\tau-})^T + \int_s^t e_{\tau-}^x(d\hat{m}_{2\tau})^T + \int_s^t d[e^x, \hat{m}_2]_\tau$$

$$= e_s^x(\hat{m}_{2s})^T + \int_s^t e_\tau^x(e_\tau^\lambda)^T d\tau + \int_s^t e_{\tau-}^x(dm_{2\tau})^T + \int_s^t e_\tau^f(\hat{m}_{2\tau})^T d\tau$$

$$+ \int_s^t dm_\tau(\hat{m}_{2\tau-})^T - \int_s^t \Sigma_\tau d\hat{m}_2\ (\hat{m}_{2\tau-})^T + \int_s^t [\phi_\tau - \Sigma_\tau \Lambda(\hat{\lambda}_{\tau-})]d\tau$$

$$+ \int_s^t (dm_{4\tau} + dm_{5\tau})$$

By condition 3 and 6.4.2.8, $e^x$ and $\hat{m}_2$ are square integrable, so by 5.4.3 we have, if $s \le t$ $E[e_t^x(\hat{m}_{2t})^T | F_{ns}] = 0$. Now eliminating the $F_{nt}$ and $F_t$ martingales and using that

$$E[\int_s^t e_\tau^f(\hat{m}_{2\tau})^T d\tau | F_{ns}] = E[\int_s^t E[e_\tau^f(\hat{m}_{2\tau})^T | F_{n\tau}]d\tau | F_{ns}] = 0,$$

we get

$$E[e_t^x(\hat{m}_{2t})^T | F_{ns}] = 0 = 0 + E[\int_s^t [e_\tau^x(e_\tau^\lambda)^T + \phi_\tau - \Sigma_\tau \Lambda(\hat{\lambda}_{\tau-})]d\tau | F_{ns}]$$

By the conditions 3 and 4 the first term on the right hand side is integrable, by condition 1 the second, hence we get

$$\Sigma_t \Lambda(\hat{\lambda}_{t-}) = E[e_t^x(e_t^\lambda)^T | F_{nt}] + E[\phi_t | F_{nt}]$$

$$\triangleq \Sigma_t(x, \lambda) + \hat{\phi}_t$$

which we define predictable.

Since $\hat{\lambda}_{t-} > 0$ a.s. for all $t \in T$, we can define $\Lambda(\hat{\lambda}_{t-}^{-1}) \triangleq \mathrm{diag}((\hat{\lambda}_{t-}^i)^{-1})$, so $\Sigma_t = [\Sigma_t(x,\lambda) + \hat{\phi}_t]\Lambda(\hat{\lambda}_{t-}^{-1})$.

The stochastic differential equation for the conditional covariance.

6.4.4. Theorem.

Given the observation equation 6.4.1 and the conditions assumed there.

In addition assume that

1. $m_1 \in M_2$, and $d\langle m_1, m_1\rangle_t = \phi_t dt$, $(\phi_t, F_t, t \in T) \in L_1(t)$, $[m_1, m_2] = 0$ a.s.,
2. $\sup_{t\in T} E|\lambda_t|^4 < \infty$,
3. let $Q_t \triangleq e_t^\lambda[e_t^r - \Sigma_t(\lambda,\lambda)\wedge(\hat{\lambda}_t^{-1})e_t^\lambda]^T + [e_t^r - \Sigma_t(\lambda,\lambda)\wedge(\hat{\lambda}_t^{-1})e_t^\lambda](e_t^\lambda)^T + \phi_t$
$$+ \Sigma_t(\lambda,\lambda)\wedge(\hat{\lambda}_t^{-2}\lambda_t)\,\Sigma_t(\lambda,\lambda)$$
4. $dm_{3t} \triangleq e_{t-}^\lambda(dm_{1t})^T + dm_{1t}(e_{t-}^\lambda)^T - [\Sigma_t(\lambda,\lambda)\wedge(\hat{\lambda}_{t-}^{-1})(e_{t-}^\lambda)^T +$
$$+ (e_{t-}^\lambda)\wedge(\hat{\lambda}_{t-}^{-1})\Sigma_t(\lambda,\lambda)]dm_{2t} + d([m_1,m_1]_t - \langle m_1,m_1\rangle_t) +$$
$$+ \Sigma_t(\lambda,\lambda)\wedge(\hat{\lambda}_{t-}^{-1})d([m_2,m_2]_t - \langle m_2,m_2\rangle_t)\wedge(\hat{\lambda}_{t-}^{-1})\Sigma_t(\lambda,\lambda).$$
then $d\langle m_3, m_2\rangle_t = \phi_{32t}dt$, where
$$\phi_{32t} = \Sigma_t(\lambda,\lambda)\wedge(\hat{\lambda}_t^{-2}\lambda_t)\Sigma_t(\lambda,\lambda) - \Sigma_t(\lambda,\lambda)\wedge(\hat{\lambda}_t^{-1}\lambda_t)(e_t^\lambda)^T - e_t^\lambda\wedge(\hat{\lambda}_t^{-1}\lambda_t)\Sigma_t(\lambda,\lambda),$$
5. Assume that $\sup_{t\in T} E|Q_t| < \infty$, and $E[\int_0^1 \phi_{32s}ds] < \infty$.

then $d\Sigma_t(\lambda,\lambda) = [\Sigma_t(\lambda,r) + \Sigma_t(r,\lambda) - \Sigma_t(\lambda,\lambda)\wedge(\hat{\lambda}_t^{-1})\Sigma_t(\lambda,\lambda) + E[\phi_t|F_{nt}]]dt$
$$+ [\Sigma_t(\lambda,\lambda,\lambda) - \Sigma_t(\lambda,\lambda)\wedge(\hat{\lambda}_t^{-1})\Sigma_t(\lambda,\lambda)]\wedge(\hat{\lambda}_{t-}^{-1})d\hat{m}_{2t}$$

Notation: $\Sigma_t(\lambda,\lambda) = E[e_t^\lambda(e_t^\lambda)^T|F_{nt}]$ is a right continuous modification having left hand limits. $\Sigma_t(r,\lambda) = E[e_t^r(e_t^\lambda)^T|F_{nt}]$ is a predictable modification, $\Sigma_t(\lambda,\lambda,\lambda)d\hat{m}_{2t} \triangleq (\Sigma_t(\lambda,\lambda^1,\lambda)d\hat{m}_{2t},\ldots.\Sigma_t(\lambda,\lambda^k,\lambda)d\hat{m}_{2t})$.

Proof. We give the proof for the scalar case only.

The conditions above imply that we can apply 6.4.3 to the semi-martingale $\lambda$,

$d\hat{\lambda}_t = \hat{r}_t dt + \sigma_t(\lambda,\lambda)\lambda_{t-}^{-1}\,d\hat{m}_{2t}$,

where $\sigma_t(\lambda,\lambda) = E[(e_t^\lambda)^2|F_{nt}]$. Let $e_t^r \triangleq r_t - \hat{r}_t$, then

$$de_t^\lambda = e_t^r dt + dm_{1t} - \sigma_t(\lambda,\lambda)\hat{\lambda}_{t-}^{-1} d\hat{m}_{2t}$$
$$= [e_t^r - \sigma_t(\lambda,\lambda)\hat{\lambda}_t^{-1}(e_t^\lambda)]dt + dm_{1t} - \sigma_t(\lambda,\lambda)\hat{\lambda}_{t-}^{-1} dm_{2t}.$$

Now $d[e^\lambda,e^\lambda]_t = d[m_1,m_1]_t + \sigma_t^2(\lambda,\lambda)\hat{\lambda}_{t-}^{-2} d[m_2,m_2]_t$
$$= d([m_1,m_1]_t - \langle m_1,m_1\rangle_t) + \phi_t dt + \sigma_t^2(\lambda,\lambda)\hat{\lambda}_t^{-2}\lambda_t dt$$
$$+ \sigma_t^2(\lambda,\lambda)\hat{\lambda}_{t-}^{-2} d([m_2,m_2]_t - \langle m_2,m_2\rangle_t)$$

$$d(e_t^\lambda)^2 = 2e_{t-}^\lambda de_t^\lambda + d[e^\lambda,e^\lambda]_t$$
$$= [2e_t^\lambda(e_t^r - \sigma_t(\lambda,\lambda)\hat{\lambda}_t^{-1}e_t^\lambda) + \phi_t + \sigma_t^2(\lambda,\lambda)\hat{\lambda}_t^{-2}\lambda_t]dt$$
$$+[2e_{t-}^\lambda dm_{1t} - 2e_{t-}^\lambda\sigma_t(\lambda,\lambda)\hat{\lambda}_{t-}^{-1}dm_{2t} + d([m_1,m_1]_t - \langle m_1,m_1\rangle_t)$$
$$+ \sigma_t^2(\lambda,\lambda)\hat{\lambda}_{t-}^{-2} d([m_2,m_2]_t - \langle m_2,m_2\rangle_t)]$$
$$\triangleq q_t dt + dm_{3t} \quad \text{as defined above.}$$

Note that $[m_2,m_2]_t - \langle m_2,m_2\rangle_t = m_{2t}$, and

$[[m_1,m_1] - \langle m_1,m_1\rangle, m_2] = 0$ a.s. Using that $[m_1,m_2] = 0$ a.s.

$d[m_3,m_2]_t = [\sigma_t^2(\lambda,\lambda)\hat{\lambda}_{t-}^{-2} - 2e_{t-}^\lambda\sigma_t(\lambda,\lambda)\hat{\lambda}_{t-}^{-1}]d[m_2,m_2]_t$

such that $d\langle m_3,m_2\rangle_t = [\sigma_t^2(\lambda,\lambda)\hat{\lambda}_t^{-2} - 2e_t^\lambda\sigma_t(\lambda,\lambda)\hat{\lambda}_{t-}^{-1}]\lambda_t dt = \phi_{32t}dt$

which gives $\phi_{32t}$ as above.

Note that by condition 2. $\sup_{t\in T} E|e_t|^4 \le 2 \sup_{t\in T} E|\lambda_t|^4 < \infty$.

Now the conditions of 6.4.3 are satisfied for the semi-martingale

$d(e_t^\lambda)^2 = q_t dt + dm_{3t}$. We get that

$$d\sigma_t(\lambda,\lambda) = [2\sigma_t(r,\lambda) - 2\sigma_t^2(\lambda,\lambda)\hat{\lambda}_t^{-1} + E(\phi_t|F_{nt}) + \sigma_t^2(\lambda,\lambda)\hat{\lambda}_t^{-1}]dt$$
$$+ [\sigma_t(\lambda,\lambda,\lambda) + \sigma_{t-}^2(\lambda,\lambda)\hat{\lambda}_{t-}^{-1} - 2\sigma_{t-}^2(\lambda,\lambda)\hat{\lambda}_{t-}^{-1}]\ (\hat{\lambda}_{t-}^{-1})d\hat{m}_{2t}$$

where $\sigma_t(\lambda,\lambda,\lambda) = E[(e_t^\lambda)^3|F_{nt}]$ the predictable modification,

which gives the result.

Remarks:

1. Theorem 6.4.3 is the first result, it gives the stochastic differential equation for the optimal estimate. The semi-martingale x represents any unobserved process, and by this formulation we obtain a quite general result. The conditions

in 6.4.3 on the variables x, f and m are of course related by the equation for x. The way the conditions are stated, is how they are needed in the proof.

2. Theorem 6.4.5 is the second main result of this section, it gives the stochastic differential equation that the conditional covariance $\Sigma_t(h,h)$ satisfies. To be able to use 6.4.3 in the proof, it is necessary to impose quite complicated conditions. Condition 5 of 6.4.5 could be replaced by proper conditions on the processes $\lambda$ and r, but this has not been done to keep the proof simple.

3. The derivation of 6.4.3 is similar to the proof of 6.2.3 for the filtering problem with observations with Brownian motion noise. A result similar to that of 6.4.3 was given by Snyder [1972a], but a different method was used, and a more restricted problem was considered. Little research has been done on the implementation of the filter derived in this section. We have not found any finite dimensional filters. Snyder [1972a] outlines a method to obtain approximations to the optimal filter and gives several examples.

Applications

We now limit our attention to real valued scalar processes, and concentrate on filtering for the rate process only. We first make an important observation. In 6.4.1 we assumed that the super-martingale $\lambda$ is such that $\lambda_1 > 0$ a.s. which implies that $\lambda_t > 0$, $\lambda_{t-} > 0$ a.s. for all $t \in T$. Since $\lambda$ is a super-martingale it is also a semi-martingale. By 3.2.3 there now exists a semi-martingale $(x_t, F_t, t \in T)$, such that $\lambda_t = \lambda_0 \ \varepsilon(x_t)$, or $d\lambda_t = \lambda_{t-} \ dx_t$ and x has certain properties. This implies that every rate process can be modelled by an exponential formula of another semi-martingale x. We rephrase the previous result using this point.

6.4.5. Corollary:

Given the stochastic system, satisfying the assumptions of 6.4.4, when all variables are real valued:

$dn_t = \lambda_t dt + dm_{2t},\ n_0 = 0,$

$dx_t = f_t dt + dm_{1t},\ x_0 = 0,$

$d\lambda_t = \lambda_{t-}dx_t = \lambda_t f_t dt + \lambda_{t-}dm_{1t},\ \lambda_0,$

where $\lambda = \varepsilon(x)$ is an exponential formula.

Assume that $(m_{1t}, F_t, t \in T) \in M_2$, $d\langle m_1, m_1\rangle_t = \phi_t dt$, $(\phi_t, F_t, t \in T) \in L_1(t)$, $\langle m_1, m_2\rangle = 0$.

The optimal filter for the rate process is:

$d\hat{m}_{2t} = dn_t - \hat{\lambda}_t dt$, the innovation process,

$d\hat{\lambda}_t = E[f_t\lambda_t | F_{nt}]dt + \sigma_{t-}(\lambda,\lambda)\hat{\lambda}_{t-}^{-1}d\hat{m}_{2t},\ \hat{\lambda}_0 = E(\lambda_0),$

$$d\sigma_t(\lambda,\lambda) = [2\sigma_t(\lambda f,\lambda) - \hat{\lambda}_t^{-1}\sigma_t^2(\lambda,\lambda) + E[\lambda_t^2\phi_t | F_{nt}]]dt +$$
$$+ [\sigma_t(\lambda,\lambda,\lambda) - \hat{\lambda}_{t-}^{-1}\sigma_{t-}^2(\lambda,\lambda)]\hat{\lambda}_{t-}^{-1}\, d\hat{m}_{2t},$$

$\sigma_0(\lambda,\lambda) = E[(\lambda_0 - \hat{\lambda}_0)^2]$

Still unknown $E[f_t\lambda_t | F_{nt}]$, $\sigma_t(\lambda f,\lambda) = E[(f_t\lambda_t - E[f_t\lambda_t | F_{nt}])(\lambda_t - \hat{\lambda}_t) | F_{nt}]$, $\sigma_t(\lambda,\lambda,\lambda)$ and $E[\lambda_t^2\phi_t | F_{nt}]$.

We have done little research on the question whether there exists finite dimensional filters, for this type of problem. However one special case is given.

6.4.6. Corollary: $T = [0,1]$.

Given the stochastic system, satisfying the assumptions of 6.4.4,

$dn_t = \lambda_t dt + dm_{2t},\ n_0 = 0,$

$d\lambda_t = a(t)\lambda_t dt + \lambda_t dm_{1t},\ \lambda_0,$

where $(m_{1t}, F_t, t \in T)$ is standard Brownian motion,

and $a : T \to R$ is a measurable deterministic function, $a(t) \leq 0$ for all $t \in T$

then $\lambda_t = \lambda_0 \exp(m_{1t} - \frac{1}{2} t) \exp(\int_0^t a(s)ds)$.

The optimal filter is:

$dm_{2t} = dn_t - \hat{\lambda}_t dt$, the innovation process,

$$d\hat{\lambda}_t = a(t)\hat{\lambda}_t dt + \sigma_{t-}(\lambda,\lambda)\hat{\lambda}_{t-}^{-1} d\hat{m}_{2t},\ \hat{\lambda}_0 = E(\ _0),$$

$$d\sigma_t(\lambda,\lambda) = [(2a(t)+1)\ \sigma_t(\lambda,\lambda) + \hat{\lambda}_t^2 - \hat{\lambda}_t^{-1}\sigma_t^2(\lambda,\lambda)]dt +$$
$$+ [\sigma_t(\lambda,\lambda,\lambda) - \hat{\lambda}_{t-}^{-1}\sigma_{t-}^2(\lambda,\lambda)]\hat{\lambda}_{t-}^{-1} d\hat{m}_{2t},$$

$$\sigma_0(\lambda,\lambda) = E[(\lambda_0 - \hat{\lambda}_0)^2].$$

Still unknown $\sigma_t(\lambda,\lambda,\lambda)$.

The proof of 6.4.6 follows from 6.4.5, and using that $\sigma_t(\lambda,\lambda) = E[\lambda_t^2 | F_{nt}] - \hat{\lambda}_t^2$. From an example following 3.2.6, we see that on $T = [0,1]$, $E[\varepsilon(m_{1t})] = E[\exp(m_{1t} - \frac{1}{2} t)] = 1$, hence, $E[\lambda_t/\lambda_0] = \exp(\int_0^t a(s)ds)$ which gives the interpretation for $\lambda$.

If $a(t) = a < 0$, the rate process is a decaying exponential.

The above result has a certain analogy with the Kalman-Bucy filter.

6.5. Some special estimation problems.

In this section we discuss some special estimation problems. We start with the case where the observation equation is disturbed by Brownian motion.

6.5.1. Theorem:

Given the prediction problem as defined in 5.5.1 for the following stochastic system. Given the observation equation with Brownian motion disturbances as defined and with the assumptions of 6.2.1.

Consider the semi-martingale x:

$$dx_t = f_t dt + dm_{1t},\ x_0,$$

1. where $(f_t, F_t, t \in T)$ is an adapted measurable process, $\sup_{t \in T} E|f_t| < \infty$,

2. $(m_{1t}, F_t, t \in T)$ is a right continuous martingale, having left hand limits,

3. $E|x_0| < \infty$,

then the optimal prediction estimator is given by:

$d\hat{x}(t|s) = E[f_t|F_{ys}]dt$, $\hat{x}(s|s) = \hat{x}_s$, where $s < t$ or

$$\hat{x}(t|s) = \hat{x}_s + \int_s^t E[f_\tau|F_{ys}]d\tau$$

The proof is an easy application of 5.5.2, using 4.4.2 and that

$$E[\int_s^t E(f_\tau|F_{y\tau})d\tau|F_{ys}] = \int_s^t E(f_\tau|F_{ys})d\tau.$$

6.5.2. Theorem:

Given the smoothing problem as defined in 5.5.3 for the following processes. Given the observation equation with Brownian motion disturbances as defined and with the assumptions of 6.2.1. Given any $(x_t, t \in T)$ measurable process, $\sup_{t \in T} E|x_t|^2 < \infty$.

Then the smoothing estimator is

$$\hat{x}(a|t) = \hat{x}(a|a) + \int_a^t \Sigma(s,a,x,h)d\hat{m}_{2s}, \quad a < t, \ a,t \in T, \text{ where}$$

$$\hat{x}(a|t) \triangleq E(x_a|F_{yt}),$$

$$\Sigma(t,a,x,h) = E[(x_a - \hat{x}(a|t))(h_t - \hat{h}_t)|F_{yt}].$$

Proof. From 5.5.4 we have that if $a < t$,

$$\hat{x}(a|t) = x(a|a) + \int_a^t \Sigma_s d\hat{m}_{2s}.$$

The problem is to determine $\Sigma_s$. Let

$$r_t \triangleq x_a - \hat{x}(a|t) = x_a - \hat{x}_a + \hat{x}(a|a) - \hat{x}(a|t) = e_a^x - \int_a^t \Sigma_s d\hat{m}_{2s}.$$

Note that by 5.4.3 if $a < s < t$

$$E[r_t(y_t)^T|F_{ys}] = E[E[r_t(y_t)^T|F_{yt}]|F_{ys}] = 0.$$

$y_t = y_a + \int_a^t \hat{h}_s ds + \hat{m}_{2t} - \hat{m}_{2a}$ so applying the differentiation rule

$$r_t(y_t)^T = r_s(y_s)^T + \int_s^t dr_\tau(y_\tau)^T + \int_s^t r_\tau(dy_\tau)^T + \int_s^t d[r,y]_\tau$$

and $d[r,y]_t = -\Sigma_t dt$ so

$$E[r_t(y_t)^T|F_{ys}] = 0 = 0 + E[-\int_s^t \Sigma_\tau d\hat{m}_{2\tau}(y_\tau)^T + \int_s^t r_\tau(h_\tau)^T d\tau + \int_s^t r_\tau(dm_{2\tau})^T - \int_s^t \Sigma_\tau d\tau|F_{ys}]$$

now the local-martingale terms drop out, and we get

$$\int_s^t E[r_\tau(h_\tau)^T - \Sigma_\tau|F_{ys}]d\tau = 0 \quad \text{or}$$

$$\Sigma_t = E[r_t(h_t)^T|F_{yt}] = E[x_a-\hat{x}(a|t))(h_t-\hat{h}_t)^T|F_{yt}] \triangleq \Sigma(t,a,x,h)$$

6.5.3. Theorem:

Given the smoothing problem as defined in 5.5.3.

If in addition to 6.5.2 we assume that

1. $dh_t = r_t dt + dm_{3t}$,
2. $(r_t, F_t, t \in T)$ is an adapted measurable process, $\sup_{t\in T} E|r_t|^2 < \infty$.
3. $(m_{3t}, F_t, t \in T) \in M_1$, $\langle m_3, m_2\rangle = 0$ a.s.,
4. $(h_t, F_t, t \in T)$ is an adapted measurable process, $\sup_{t\in T} E|h_t|^4 < \infty$
5. $(x_t, t \in T)$ is any measurable process, $\sup_{t\in T} E|x_t|4 < \infty$,

then the optimal smoothing estimator is

$$\hat{x}(a|t) = \hat{x}(a|a) + \int_a^t \Sigma(s,a,x,h)d\hat{m}_{2s},$$

$$d\Sigma(t,a,x,h) = [\Sigma(t,a,x,r) - \Sigma(t,a,x,h)\Sigma_t(h,h)]dt + H_t d\hat{m}_{2t}$$

$$\Sigma(a,a,x,h) = \Sigma_a(x,h) = E[e_a^x(e_a^h)^T|F_{ya}],$$

where

$$\Sigma(t,a,x,h) = E[(x_a-\hat{x}(a|t))(h_t-\hat{h}_t)^T|F_{yt}],$$

$$\Sigma(t,a,x,r) = E[(x_a-\hat{x}(a|t))(r_t-\hat{r}_t)^T|F_{yt}],$$

and

$H_t d\hat{m}_{2t} = (H_t^1 d\hat{m}_{2t},..H^k d\hat{m}_{2t})$ , where $(H_t^n)$ an matrix valued stochastic processes, still unknown.

Proof. We give the proof for the scalar case only.

Note that

$$e^x(a|t) = x_a - \hat{x}(a|t) = e_a^x - \int_a^t \sigma(s,a,x,h)d\hat{m}_{2s} = e_a^x - \int_a^t \sigma(s,a,x,h)(e_s^h ds+dm_{2s})$$

$$e_t^h = e_a^h + \int_a^t e_s^r ds + \int_a^t dm_{3s} - \int_a^t \sigma_s(h,h)d\hat{m}_{2s}$$

$$= e_a^h + \int_a^t [e_s^r - \sigma_s(h,h)e_s^h]ds + \int_a^t dm_{3s} - \int_a^t \sigma_s(h,h)dm_{2s}$$

$$[e^x(a|t),e^h]_t = \int_a^t \sigma(s,a,x,h)\sigma_s(h,h)ds$$

We apply the product rule again:

$$e^x(a|t)e_t^h = e_a^x e_a^h + \int_a^t e^x(a|s)de_s^h + \int_a^t e_s^h de^x(a|s) + \int_a^t d[e^x(a|s),e^h]_s$$

$$= e_a^x e_a^h + \int_a^t [e^x(a|s)e_s^r - e^x(a|s)\sigma_s(h,h)e_s^h - \sigma(s,a,x,h)(e_s^h)^2$$

$$+ \sigma(s,a,x,h)\sigma_s(h,h)]ds + \int_a^t e^x(a|s)dm_{3s} - \int_a^t e^x(a|s)\sigma_s(h,h)dm_{2s}$$

$$- \int_a^t e_s^h \sigma(s,a,x,h)dm_{2s}$$

$$= q_t dt + dm_{4t}.$$

We now want to apply 6.2.3 to this semi-martingale. The conditions 2,4 and 5 imply that $\sup_{t\in T} E|q_t| < \infty$. Also

$\sup_{t\in T} E|e^x(a|t)e_t^h|^2 \leq k \sup_{t\in T} E|x_t|^4 \cdot \sup_{t\in T} E|h_t|^4 < \infty.$

Furthermore $(m_{4t},F_t,t \in T) \in M_{loc}$ and a process $\langle m_4,m_2\rangle$ eixsts, since

$d\langle m_4,m_2\rangle_t = [-e^x(a|t)\sigma_t(h,h) - e_t^h\sigma(t,a,x,h)]dt.$

We can now apply 6.2.3 and using the definitions given earlier

$d\sigma(t,a,x,h) = [\sigma(t,a,x,r) - \sigma(t,a,x,h)\sigma_t(h,h)]dt + H_t dm_{2t}.$

Note that $\sigma(a,a,x,h) = E[(x_a-\hat{x}(a|a))e_a^h|F_{ya}] = \sigma_a(x,h).$

Theorem 6.5.3 works for any pair $a,t \in T$, so it covers all the three smoothing problems as defined in 5.5.3. Kailath, Frost [1968], discuss the smoothing problem for linear stochastic systems with Brownian motion. Because of the linearity of the system, they are able to derive a more detailed recursive solution. Such a solution does not seem to exist in the nonlinear case.

Systems with delays.

The foregoing theory, with the general results derived earlier, allows us to derive the filtering equations for systems with delays. The general nonlinear case with Brownian motion disturbances follows easily using 6.2.4, and the previous results of this section. Here we only give an example, of a linear system.

6.5.4. **Theorem:**

Given the stochastic system model, which contains a delay, scalar equations:

$$dx_t = a\, x(t-h)dt + dm_{1t},\quad x(s): -h \leq s \leq 0 \text{ given.}$$

$$dy_t = x_t dt + dm_{2t}$$

satisfying the assumptions of 6.2.1, where $a,h \in R$ are constants, $h > 0$, and $m_1$ is standard Brownian motion.

The optimal estimator is

$$d\hat{m}_{2t} = dy_t - \hat{x}_t dt$$

$$d\hat{x}_t = a\, \hat{x}(t-h|t)dt + \sigma_t(x,x)d\hat{m}_{2t},\quad \hat{x}_0 = E(x_0)$$

$$\hat{x}(t-h|t) = \hat{x}(t-h|t-h) + \int_{t-h}^{t} \sigma(s,s-h,x,x)d\hat{m}_{2s}$$

$$d\sigma_t(x,x) = [2a\sigma(t,t-h,x,x) - a\sigma(t,t-h,x,x)\sigma_t(x,x) + 1]dt$$
$$+ \sigma_t(x,x,x)d\hat{m}_{2t},\quad \sigma_0(x,x) = E[(x_0-\hat{x}_0)^2]$$

$$d\sigma(t,t-h,x,x) = [a\sigma(t,t-h,x,x) - \sigma(t,t-h,x,x)\sigma_t(x,x)]dt + k_t d\hat{m}_{2t}$$

$$\sigma(t-h,t-h,x,x) = \sigma_{t-h}(x,x)$$

Still unknown: $\sigma_t(x,x,x), k_t$.

The proof follows easily from 6.2.5 and 6.5.3.

The above result was derived earlier by Kwakernaak [1967].

It is difficult to say something about the third order conditional moment.

Prediction and smoothing for counting process observations.

6.5.5. Theorem:

Given the prediction problem as defined in 5.5.1 for the following stochastic system. Given the equations with counting process observations, as defined and with the assumptions of 6.4.1.,

$dn_t = \lambda_t dt + dm_{2t}, \; n_0 = 0$

$d\lambda_t = r_t dt + dm_{1t}, \; \lambda_0$

Then the optimal prediction estimator of the rate process $\lambda$ is given by

$d\hat{\lambda}(t|s) = E[r_t|F_{ns}]dt, \quad \hat{\lambda}(t|s) = E(\lambda_t|F_{ns}), \; s < t$ or

$\hat{\lambda}(t|s) = \hat{\lambda}_s + \int_s^t E(r_\tau|F_{ns})d\tau.$

The proof is immediate from 5.5.2, similar to the proof of 6.5.1

6.5.6. Example:

Consider the prediction problem 6.5.5 for the real valued semi-martingale $\lambda$ given by

$d\lambda_t = a(t)\lambda_t dt + \lambda_t dm_{1t}, \; \lambda_0,$

where $a : T \to R$, is measurable.

Then the optimal prediction estimator is

$\hat{\lambda}(t|s) = \hat{\lambda}(s|s) + \int_s^t a(\tau)\, \hat{\lambda}(\tau|s)d\tau \;, \; s < t$ or

$\hat{\lambda}(t|s) = \phi(t,s)\hat{\lambda}(s|s)$

where $\phi(t,s)$ is the transition function associated with $a(t)$.

6.5.7. Theorem:

Given the smoothing problem as defined in 5.5.3. Consider the equations with counting process observations, as defined and with the assumptions of 6.4.1,

$dn_t = \lambda_t dt + dm_{2t}, \; n_0 = 0$

and $(\lambda_t, F_t, t \in T)$ an adapted measurable process real valued,

$\sup_{t \in T} E|\lambda_t|^2 < \infty$ .

Then the optimal smoothing estimator for the rate process $\lambda$ is:

$$\hat{\lambda}(a|t) = \hat{\lambda}(a|a) + \int_a^t \sigma(s,a,\lambda,\lambda)\hat{\lambda}_{s-}^{-1}d\hat{m}_{2s},$$

where $\hat{\lambda}(a|t) = E(\lambda_a|F_{nt})$, $a < t$,

$\sigma(t,a,\lambda,\lambda) = E[(\lambda_a-\hat{\lambda}(a|t))(\lambda_t-\hat{\lambda}_t)|F_{nt}]$ a.s.

Proof. From 5.5.4 we have that if $a < t$ then

$$\hat{\lambda}(a|t) = \hat{\lambda}(a|a) + \int_a^t k_s d\hat{m}_{2s}.$$

The problem is to find $(k_t, F_{nt}, t \in T)$. This follows easily from the by now well established procedure.

$$e^{\lambda}(a|t) = \lambda_a-\hat{\lambda}(a|t) = e_a^{\lambda} + \hat{\lambda}_a - \hat{\lambda}(a|t) = e_a^{\lambda} - \int_a^t k_s d\hat{m}_{2s} =$$

$$= e_a^{\lambda} - \int_a^t k_s e_s^{\lambda} ds - \int_a^t k_s dm_{2s}.$$

By 5.4.3 $E[e^{\lambda}(a|t)\hat{m}_{2t}|F_{nt}] = 0$, so

$$E[e^{\lambda}(a|t)\hat{m}_{2t}|F_{ns}] = 0 = 0 + E[\int_a^t e^{\lambda}(a|t)e_{\tau}^{\lambda}d\tau + \int_a^t e^{\lambda}(a|\tau-)dm_{2\tau}$$

$$- \int_a^t \hat{m}_{2\tau}k_{\tau}e_{\tau}^{\lambda}d\tau - \int_a^t \hat{m}_{2\tau-}\, k_{\tau}dm_{2\tau} - \int_a^t k_{\tau}\hat{\lambda}_{\tau}d\tau - \int_a^t k_{\tau}d\hat{m}_{2\tau}|F_{ns}].$$

Now the martingale terms drop out, and we get

$$\int_a^t E[e^{\lambda}(a|\tau)e^{\lambda} - k_{\tau}\hat{\lambda}_{\tau}|F_{ns}]d\tau = 0$$

$k_t\hat{\lambda}_{t-} = E[e^{\lambda}(a|t)e_t^{\lambda}|F_{nt}] = \sigma(t,a,\lambda,\lambda)$ a predictable modification.

This gives the result.

Just as in 6.5.3 under suitable assumptions one can derive a stochastic differential equation for $\sigma(t,a,\lambda,\lambda)$.

## 6.6. Comments on the martingale approach to filtering problems.

In this section we want to discuss the martingale approach to the filtering problem, which we have given in this chapter. We compare our method with previous methods.

### The stochastic system model.

The filtering problem that was considered in recent years, was for a nonlinear dynamical system, that was disturbed by a Brownian motion process. In this thesis we have considered a more general model, with semi-martingale equations, as defined in Section 4.3:

$$dx_t = f_t dt + dm_{1t}$$

$$dy_t = h_t dt + dm_{2t}.$$

This is a generalization in two directions: first we allow $m_1, m_2$ to be certain martingales, and secondly the processes f and h need only be adapted. Most of the recent liberature deals with the case where x or h are Markov processes, and where thus $f_t$ depends only on $x_t$. The generalization allows us to solve at the same time, problems where we have systems with delays, or similar type of problems. Snyder [1972a] discusses the filtering problem for what he calls the doubly stochastic Poisson process. He also noted the similarity in the filtering equations for problems with Brownian motion noise and for counting process filtering. This similarity was also noted by Brémaud [1972]. We now of course know that the similarity arises because both are derived from similar semi-martingale equations. This similarity allows us to solve both problems with analogeous methods. We will discuss the Brownian motion and the counting process filtering problem together.

The derivation of the filtering equations.

We emphasize the main points in the derivation of the filter equations. The first one is the definition of the innovation process, and the associated projection of semi-martingale equations on the σ-field of the observations. We have

$$d\hat{x}_t = \hat{f}_t dt + d\hat{m}_{1t}$$

$$dy_t = \hat{h}_t dt + d\hat{m}_{2t}$$

where $\hat{m}_1, \hat{m}_2 \in M_{loc}$. The innovation process was emphasized by Kailath [1968] for the Brownian motion model. The important point however is, that it is a martingale. The projection of the semimartingale x is seen to be an extension of the innovation process property. The second important point, which is the crucial one, is the martingale representation theorem on the σ-field of the observations, i.e. if $(\hat{m}_{1t}, F_{yt}, t \in T) \in M_{loc}$, then $\hat{m}_{1t} = \int_0^t \phi_s dm_{2s}$, for some process $\phi$. This representation theorem has been proven for the case of Brownian motion and for counting processes. In Frost, Kailath [1971] a similar result was obtained, but through the equality of $F_{yt} = F\hat{m}_{2t}$, which has only been proven under rather strict conditions. Although this equality may hold in more general cases, there is no real proof. Here we have followed the approach by Fujisaki, Kallianpur, Kunita [1972], who prove the martingale representation theorem by a translation argument. Using the above two points the calculation of the optimal filter equation is a straightforward operation. One only uses the stochastic calculus, and the optimality of the estimate, i.e. the orthogonality of the estimation error and the observations. A similar equation can be derived for the conditional covariance.

The derivation of the filter equation for the Brownian motion case is the one given by Fujisaki et al. [1972], except for a generalization of the model and for slightly different proofs. Here we have given a stochastic differential equation for the conditional covariance. The approach given also uses ideas from Wong [1972]. The extension of this approach to the filtering problem for counting processes is relatively obvious. One only needs to use the respective martingale properties.

Extension of the filtering problem to other processes.

We are interested in extending the results of this chapter to more general stochastic system equations. It turns out that the only limiting point is the absence of a martingale representation theorem. It is thus of prime importance to prove for a large class of underlying martingale processes the martingale representation theorem. This was done in a somewhat limited case by Kunita-Watanabe [1967], but their result needs careful interpretation. Once such a martingale representation theorem has been proven, we can consider stochastic system models disturbed by such processes, as defined in Section 4.3. The innovation process property and the martingale representation on the $\sigma$-field of the observations, then easily follow from the approach given here. The equation for the optimal estimate then results.

## 7. Discussion and conclusions.

The goal for this thesis was to analyse and solve estimation problems using the theory of martingales and stochastic integrals. Here we discuss the results obtained and state some of the problems that are still open. The main tools in the martingale approach are the stochastic integral, the differentiation formula, the martingale representation theorem, and martingale theory proper. The main concept used from estimation theory is the least squares error criterion, which leads to the optimal estimate of conditional expectation. All our results are derived from this basic principle using stochastic claculus.

### Discussion of results.

It should be kept in mind that all practical results depend on the martingale representation theorem, which has only been proven if the underlying process is Brownian motion or a Poisson process. In Chapter 3, the main result is section 3.3., where we characterize a change of measure by a local martingale, and where we obtain the translation of local martingales under a change of measure. The converse of the translation theorem is the abstract version of the detection problem, which we can solve only for the two cases for which we have a martingale representation theorem. In Chapter 4 the main result is the generalized definition of a stochastic system and of stochastic differential equations. A complementary result is the projection of semi-martingale processes on a family of σ-fields, which includes the concept of innovation process. In Chapter 5 the main result is the general formulation of and the elementary solution to the least squares error filtering, prediction and smoothing problem. Of equal importance are the two martingale representation theorems on the $\sigma$-field of the observations. In Chapter 6 the main result is the derivation of the stochastic differential equations for the optimal filter, prediction and smoothing estimates and for the conditional covariances.

The novel points in the thesis.

The results we have obtained in this thesis are a generalization in the direction of martingale theory. Instead of considering separately, systems with Brownian motion disturbances, as was done usually before, or the counting observations problem, martingale theory gives a unified approach to both of these problems, and to possible new problems which can be formulated in this framework. The analysis of absolute continuity with martingale theory has opened a new way of thinking, and led to a more general formulation of the detection problem. The stochastic system consisting of the two semi-martingales x and y, is also an extension of previous system equations. It covers systems with Brownian motion disturbances, including systems with delays, as well as counting processes with a rate process. Regarding the filtering, prediction and smoothing problem, martingale theory has many advantages. The derivations, apart from difficult details, are straightforward and clearly show the basic principles involved. It avoids assumptions on the existence of conditional densities as was necessary in the work by Kushner and others.

Open questions and future research.

There are essentially three problem areas for future research.

1. Martingales and stochastic integrals.

The first problem area is the development of the stochastic integral and the associated martingale theory. The theory as known now is still far from complete, there are still open questions concerning the class of integrand processes and discontinuous martingales. The main problem however is the extension of the martingale representation theorem, to a larger class of underlying processes.

2. Absolute continuity.

The second problem area is that of absolute continuity and related topics. In section 3.4. we formulated an abstract version of the detection problem, which can be solved for any case where we have a martingale representation theorem. Apart from this there are several applications of the likelihood ratio and other properties of absolute continuity, where the results of chapter 3 are of interest.

3. Filtering problems.

It is believed that more filtering problems, some not yet formulated, might be put in the framework of martingale thoery, specifically in the semi-martingale equations presented in chapter 4. It is further believed that for the defined stochastic system, the stochastic differential equation for the optimal estimate can be obtained. However the limiting factor is the martingale representation theorem. Such a generalization was attempted by the author but did not seem useful. The results derived in Chapter 6 have several limitations. The first is that we only considered the time interval $T = [0,1]$, which is equivalent to any finite time interval. The extension of this result to $T = [0,\infty)$ is unsolved. Related to this, as experience with linear systems shows, is the concept of observability, which has not been defined for stochastic systems. The major point is that the derived filters are in general infinite dimensional. The question now is, which systems give finite dimensional filters. This question and related topics on the implementation of these filters remain unsolved. Whether martingale theory can provide any help in obtaining finite dimensional filters and provide further insight in estimation problems is open to research.

## REFERENCES


V. E. Benes, 'Existence of optimal stochastic control laws', SIAM J. Control 9, 3, (Aug., 1971), pp. 446-472.

R. M. Blumenthal, R. K. Getoor, 'Markov processes and potential theory', Academic Press, N. Y., 1968, 313 pp.

P. M. Brémaud, 'A martingale approach to point processes', Ph.D. Thesis, Univ. of Cal., Dep. EECS, Berkeley, Cal., Aug., 1972. Also Memo, ERL-M-345; Aug. 4, 1972. Elec. Res. Lab., Berkeley, Cal.

R. S. Bucy, 'Non-linear filtering theory', IEEE Tr. AC 10, 2, (Apr., 1965) p. 198.

D. L. Burkholder, 'Martingale transforms', Ann. Math. Stat., 37, 6, (Dec., 1966), pp. 1494-1504.

R. H. Cameron, W. T. Martin, 'Transformations of Wiener integrals under translations', Ann. Math., 45, 2, (April, 1944), pp. 386-396.

J. M. C. Clark, 'The representation of functionals of Brownian motion by stochastic integrals', Ann. Math. Stat., 41, 4, (Aug., 1970), pp. 1282-1295.

J. M. C. Clark, 'A measurability property of the solutions of stochastic differential equations', preliminary draft.

H. Cramér, 'Mathematical methods of statistics', Princeton Univ. Press, Princeton, N. J., 1946.

M. H. A. Davis, 'Dynamic programming conditions for partially observable stochastic systems', Ph.D. Thesis, Univ. of Cal., Berkeley, Cal., 1971. Also Memo ERL-M-304, Electronics Research Laboratory, Univ. of Cal., Berkeley.

M. H. A. Davis, 'Detection theory of Poisson processes', to be published.

M. H. A. Davis, P. P. Varaiya, 'Dynamic programming conditions for partially observable stochastic systems', SIAM J. Control 11, 2, (May, 1973), pp. 226-261.

C. Dellacherie, 'Capacites et processus stochastiques', Springer, Berlin, (N. Y.), 1972.

R. Deutsch, 'System analysis techniques', Prentice Hall, Englewood Cliffs, 1969.

C. Doléans-Dade, 'Quelques applications de la formule de changement de variables pour les semi martingales', Z. für Wahrsch. Th. verw. Gb. 16, 3, (1970), p. 181-194.

C. Doléans-Dade, P. A. Meyer, 'Intégrales stochastiques par rapport aux martingales locales', Séminaire de probabilités IV, Université de Strasbourg, Lecture notes in mathematics, 124, Springer-Verlag, Berlin. p. 77-107, 1970.

Doob, 'Stochastic Processes', Wiley, New York, 1953.

T. E. Duncan, 'On the absolute continuity of measures', Annals Math. Stat., 41, 1, (Feb., 1970), p. 30-38.

E. B. Dynkin, 'Markov Processes', Vol. I., Academic Press, N. Y., 1965.

P. A. Frost, T. Kailath, 'An innovations approach to least squares estimation - Part III: Nonlinear estimation in White Gaussian Noise', IEEE-T-AC 16, 3, (June, 1971) p. 217-226.

M. Fujisaki, G. Kallianpur, H. Kunita, 'Stochastic differential equations for the non-linear filtering problem', Osaka J. Math 9, 1, (Apr., 1972), p. 19-40.

I. V. Girsanov, 'On transforming a certain class of stochastic processes by absolutely continuous substitution of measures', Theory of Prob. and Applic. 5, 3, (1960), pp. 285-301.

K. Ito, 'Stochastic Integral', Proc. Imp. Acad. Tokyo 20, 8, (Oct., 1944), p. 519-524.

K. Ito, 'On stochastic differential equations', Mem. Am. Math. Soc. no. 4, (1951a).

K. Ito, 'Multiple Wiener Integral', J. of the Math Soc. Japan, 3, 1, (May, 1951b), pp. 157-169.

K. Ito, S. Watanabe, 'Transformation of Markov processes by multiplicative functionals', Ann. Inst. Fourier, Grenoble, 15, 1, (1965), pp. 13-30.

T. T. Kadota, L. A. Shepp, 'Conditions for absolute continuity between a certain pair of probability measures', Z. Wahrsch. th. verw. Gb. 16, 3, (1970), pp. 250-260.

T. Kailath, 'An innovations approach to least-squares estimation part I: linear filtering in additive white noise', IEEE-T-AC 13, 6, (Dec., 1968), p. 646-655.

T. Kailath, 'The innovations approach to detection and estimation theory', Proc. IEEE 58, 5, (May, 1970a), p. 680-695.

T. Kailath, 'Likelihood Ratios for Gaussian processes', IEEE-T-IT 16, 3, (May, 1970b), p. 276-287.

T. Kailath, 'A further note on a general likelihood formula for random signals in Gaussian noise', IEEE-T-IT 16, 4, (July, 1970c), p. 393-396.

T. Kailath, 'Some extensions of the innovations theorem', Bell Syst. Techn. J. 50, 4, (Apr., 1971a) p. 1487-94.

T. Kailath, 'The structure of Radon-Nikodym derivatives with respect to Wiener and related measures', Ann. Math. Stat. 42, 3, (June, 1971b), pp. 1054-1067.

T. Kailath, 'A note on least squares estimation by the innovations method', SIAM J. on Control 10, 3, (Aug., 1972), pp. 477-486.

T. Kailath, P. Frost, 'An innovations approach to least squares estimation part II: linear smoothing in additive white noise', IEEE-T-AC 13, 6, (Dec., 1968), p. 655-660.

T. Kailath, M. Zakai, 'Absolute continuity and Radon-Nikodym derivatives for certain measures relative to Wiener measure', Ann. Math. Stat. 42, 1, (Feb., 1971), pp. 130-140.

T. Kailath, R. A. Geesey, 'An innovations approach to least squares estimation - part IV: Recursive estimation given lumped covariance functions', IEEE T-AC 16, 6, (Dec., 1971), pp. 720-727.

G. Kallianpur, C. Striebel, 'Estimation of stochastic systems arbitrary system process with additive white noise observation errors', Ann. Math. Stat. 39, 3, (June, 1968), pp. 785-801.

G. Kallianpur, C. Striebel, 'Stochastic differential equations occurring in the estimation of continuous parameter stochastic processes', Th. of Prob. & Appl. 14, 4, (1969), pp. 567-594.

R. E. Kalman, R. S. Bucy, 'New results in linear filtering and prediction theory', Trans. ASME, J. Basic. Engrg., Ser. D, 83, 1, (Mar., 1961), pp. 95-108.

N. Kazamaki, 'Notes on a stochastic integral equation', Sem. de Prob. VI, lecture notes in math vol. 258, pp. 105-108, Springer, Berlin, 1972.

H. Kunita, S. Watanabe, 'On square integrable martingales', Nagoya Math. J. 30, (1967), pp. 209-245.

H. J. Kushner, 'Dynamical equations for optimal nonlinear filtering', J. Diff. Eq. 3, (1967a), pp. 179-190.

H. J. Kushner, 'Approximations to optimal nonlinear filters', IEEE-T-AC, 12, 5, (Oct., 1967b), pp. 546-556.

H. Kwakernaak, 'Optimal filtering in linear systems with time delays', IEEE-T-AC 12, 2, (Apr, 1967), pp. 169-173.

P. Masani, N. Wiener, 'The prediction theory of multivariate stochastic processes, part I: the regularity condition', Acta Math. 98, (1958), pp. 111-150.

P. A. Meyer, 'Probabilités et Potentiel', Hermann, Paris, 1966. English translation: 'Probability and potentials', Blaisdell, Waltham, Mass., 1966.

P. A. Meyer, 'Intégrales stochastique I, II, III, IV', Séminaire de probabilités I, Univ. de Strasbourg, Lecture notes in math. vol. 39, Springer, Heidelberg, 1967a, pp. 72-162.

P. A. Meyer, 'Processus de Markov', Lecture notes in math. vol. 26, Springer, Berlin, 1967b.

P. A. Meyer, 'Guide détaillé de la théorie générale des processus', Séminaire de probabilités II, Lecture notes in math, vol. 51, pp. 140-165. Springer, Berlin, 1968.

P. A. Meyer, 'Square integrable martingales, a survey', In 'Lecture notes in mathematics', vol. 190, 'Martingales', Springer Verlag, Berlin, 1971a, p. 32-37.

P. A. Meyer, 'Non-square integrable martingales, etc.', In 'Lecture notes in mathematics', vol. 190, 'Martingales', Springer-Verlag, Berlin, 1971b, pp. 38-43.

P. A. Meyer, 'Martingales and stochastic integrals I', Lecture notes in mathematics, vol. 284, Springer, Berlin, 1972.

P. A. Meyer, 'Sur un probleme de filtration', Sem. de probabilités, Univ. de Strasbourg, 1971/72, Lecture notes in mathematics, vol. 321 , Springer, Berlin.

P. W. Millar, 'Martingale integrals', Trans. Amer. Math. Soc. 133, 1, (Aug., 1968), pp. 145-166.

M. Motoo, S. Watanabe, 'On a class of additive functionals of Markov processes', J. of Math. Kyoto Univ., 4, 3, (July, 1965), pp. 429-469.

E. Nelson, 'Dynamical theories of Brownian motion', Princeton Univ. Press, Princeton, N. J., 1967

S. M. Ross, 'Applied probability models with optimization applications', Holden-Day, San Francisco, Cal., 1970.

A. P. Sage, J. L. Melsa, 'Estimation theory with applications to communication and control', McGraw Hill, N. Y., 1971, 529 pp.

L. Schwartz, E. B. Stear, 'A computational comparison of several nonlinear filters', IEEE-T-AC 13, 1, (Feb., 1968), pp. 83-86.

A. V. Skorokhod, 'Differentiability of measures corresponding to random processes: I. Processes with independent increments', Th. of Prob. & Applic. 2, 4, (1957), pp. 407-432.

D. L. Snyder, 'Filtering and detection for doubly stochastic Poisson processes', IEEE-T-IT 18, 1, (Jan., 1972a), pp. 91-102.

D. L. Snyder, 'Smoothing for doubly stochastic Poisson processes', IEEE-T-IT 18, 5, (Sept., 1972b), pp. 558-562.

H. W. Sorenson, 'Least squares estimation: From Gauss to Kalman', Spectrum 7, 7, (July, 1970), pp. 63-68.

K. Srinivasan, 'State estimation by orthogonal expansion of probability distributions', IEEE-T-AC 15, 1, (Feb., 1970), pp. 3-10.

J. H. Van Schuppen, E. Wong, 'Transformation of local martingales under a change of law', Electronics Research Laboratory Memo, ERL-M385, University of Calif., Berkeley.

S. Watanabe, 'On discontinuous additive functionals and Lévy measures of a Markov process', Jap. J. Math. 34, (1964), pp. 53-70.

N. Wiener, 'The extrapolation, interpolation, and smoothing of stationary time series with engineering applications', Wiley, N. Y. 1949.

N. Wiener, 'Nonlinear problems in random theory', MIT Press, Cambridge, Mass., 1958.

E. Wong, 'Stochastic processes in information and dynamical systems', McGraw-Hill, N. Y., 1971a.

E. Wong, 'Representation of martingales, quadratic variation and applications', SIAM J. Control 9, 4, (Nov., 1971b), pp. 621-632.

E. Wong, 'Martingale theory and applications to stochastic problems in dynamical systems', Research report, Imperial College, Dep. of Computing and Control, London, England, June 1972.

E. Wong, 'Recent progress in stochastic processes - a survey', IEEE-T-IT 19, 3, (May, 1973), pp. 262-275.

W. M. Wonham, 'Random differential equations in control theory', in 'Probabilistic methods in applied mathematics', vol. 2, Bharucha-Reid ed., Academic Press, N. Y., 1970.

M. Zakai, 'On the optimal filtering of diffusion processes', Z. Wahrsch. th. verw. Geb. 11, 3, (1969), p. 230-243.