A TUTORIAL INTRODUCTION TO STOCHASTIC ANALYSIS AND ITS APPLICATIONS

by

IOANNIS KARATZAS
Department of Statistics
Columbia University
New York, N.Y. 10027

September 1988

Synopsis

We present in these lectures, in an informal manner, the very basic ideas and results of stochastic calculus, including its chain rule, the fundamental theorems on the representation of martingales as stochastic integrals and on the equivalent change of probability measure, as well as elements of stochastic differential equations. These results suffice for a rigorous treatment of important applications, such as filtering theory, stochastic control, and the modern theory of financial economics. We outline recent developments in these fields, with proofs of the major results whenever possible, and send the reader to the literature for further study.

Some familiarity with probability theory and stochastic processes, including a good understanding of conditional distributions and expectations, will be assumed. Previous exposure to the fields of application will be desirable, but not necessary.

Lecture notes prepared during the period 25 July - 15 September 1988, while the author was with the Office for Research & Development of the Hellenic Navy (ΓΕΤΕΝ), at the suggestion of its former Director, Capt. I. Martinos. The author expresses his appreciation to the leadership of the Office, in particular Capts. I. Martinos and A. Nanos, Cmdr. N. Eutaxiopoulos, and Dr. B. Sylaidis, for their interest and support.

CONTENTS

Page

0. Introduction 3

1. Generalities 4

2. Brownian Motion (Wiener process) 6

3. Stochastic Integration 11

4. The Chain Rule of the new Calculus 15

5. The Fundamental Theorems 18

6. Dynamical Systems driven by White Noise Inputs 21

7. Filtering Theory 27

8. Robust Filtering 37

9. Stochastic Control 39

10. Notes 53

11. References 55

INTRODUCTION AND SUMMARY

The purpose of these notes is to introduce the reader to the fundamental ideas and results of *Stochastic Analysis* up to the point that he can acquire a working knowledge of this beautiful subject, sufficient for the understanding and appreciation of its rôle in important applications. Such applications abound, so we have confined ourselves to only two of them, namely *filtering theory* and *stochastic control*; this latter topic will also serve us as a vehicle for introducing important recent advances in the field of financial economics, which have been made possible thanks to the methodologies of stochastic analysis.

We have adopted an informal style of presentation, focusing on basic results and on the ideas that motivate them rather than on their rigorous mathematical justification, and providing proofs only when it is possible to do so with a minimum of technical machinery. For the reader who wishes to undertake an in-depth study of the subject, there are now several monographs and textbooks available, such as Liptser & Shiryaev (1977), Ikeda & Watanabe (1981), Elliott (1982) and Karatzas & Shreve (1987).

The notes begin with a review of the basic notions of Markov processes and martingales (section 1) and with an outline of the elementary properties of their most famous prototype, the Wiener-Lévy or "Brownian Motion" process (section 2). We then sketch the construction and the properties of the integral with respect to this process (section 3), and develop the chain rule of the resulting "stochastic" calculus (section 4). Section 5 presents the fundamental representation properties for continuous martingales in terms of Brownian motion (via time-change or integration), as well as the celebrated result of Girsanov on the equivalent change of probability measure. Finally, we offer in section 6 an elementary study of dynamical systems excited by white noise inputs.

Section 7 applies the results of this theory to the study of the filtering problem. The fundamental equations of Kushner and Zakai for the conditional distribution are obtained, and the celebrated Kalman-Bucy filter is derived as a special (linear) case. We also outline the derivation of the genuinely nonlinear Beneš (1981) filter, which is nevertheless explicitly implementable in terms of a finite number of sufficient statistics. A reduction of the filtering equations to a particularly simple form is presented in section 8, under the rubric of "robust filtering", and its significance is demonstrated on examples.

An introduction to stochastic control theory is offered in section 9; we present the principle of *Dynamic Programming* that characterizes the value function of this problem, and derive from it the associated Hamilton-Jacobi-Bellman equation. The notion of weak solutions (in the "viscosity" sense of P.L. Lions) of this equation is expounded upon. In addition, several examples are presented, including the so-called "linear regulator" and the portfolio/consumption problem from financial economics.

1. GENERALITIES

A stochastic process is a family of random variables $X = \{X_t\,;\, 0 \leq t < \infty\}$, i.e., of measurable functions $X_t(\omega) : \Omega \to \mathcal{R}$, defined on a probability space $(\Omega, \mathcal{F}, P)$. For every $\omega \in \Omega$, the function $t \mapsto X_t(\omega)$ is called the *sample path* (or trajectory) of the process.

1.1 Example: Let $T_1, T_2, \cdots$ be *I.I.D.* (independent, identically distributed) random variables with exponential distribution $P(T_i \in dt) = \lambda e^{-\lambda t} dt$, for $t > 0$, and define

$$S_0(\omega) = 0, \qquad S_n(\omega) = \Sigma_{j=1}^n T_j(\omega) \ \text{ for } \ n \geq 1.$$

The interpretation here is that the T_j's represent the interarrival times, and that the S_n's represent the arrival times, of customers in a certain facility. The stochastic process

$$N_t(\omega) = \#\{n \geq 1 \,:\, S_n(\omega) \leq t\}, \quad 0 \leq t < \infty$$

counts, for every $0 \leq t < \infty$, the number of arrivals up to that time and is called a *Poisson process* with intensity $\lambda > 0$. Every sample path $t \mapsto N_t(\omega)$ is a "staircase function" (piecewise constant, right-continuous, with jumps of size $+1$ at the arrival times), and we have the following properties:

(i) for every $0 = t_0 < t_1 < t_2 < \cdots < t_m < t < \theta < \infty$, the increments N_{t_1}, $N_{t_2} - N_{t_1}, \cdots, N_t - N_{t_m}, N_\theta - N_t$ are independent;

(ii) the distribution of the increment $N_\theta - N_t$ is Poisson with parameter $\lambda(\theta - t)$, i.e.,

$$P[N_\theta - N_t = k] = e^{-\lambda(\theta - t)} \frac{(\lambda(\theta - t))^k}{k!}, \quad k = 0, 1, 2, \cdots .$$

It follows from the first of these properties that

$$P[N_\theta = k | N_{t_1}, N_{t_2}, ..., N_t] = P[N_\theta = k | N_{t_1}, N_{t_2} - N_{t_1}, \cdots, N_t - N_{t_m}, N_t] = P[N_\theta = k | N_t],$$

and more generally, with $\mathcal{F}_t^N = \sigma(N_s; 0 \leq s \leq t)$:

$$P[N_\theta = k | \mathcal{F}_t^N] = P[N_\theta = k | N_s;\ 0 \leq s \leq t] \ = \ P[N_\theta = k | N_t]. \tag{1.1}$$

In other words, given the "past" $\{N_s \,:\, 0 \leq s < t\}$ and the "present" $\{N_t\}$, the "future" $\{N_\theta\}$ depends only on the present. This is the *Markov property* of the Poisson process.

1.2 Remark on Notation: For every stochastic process X, we denote by

$$\mathcal{F}_t^X = \sigma\big(X_s\,;\, 0 \leq s \leq t\big) \tag{1.2}$$

the record (history, observations, sample path) of the process up to time t. The resulting family $\{\mathcal{F}_t^X\,;\, 0 \leq t < \infty\}$ is *increasing*: $\mathcal{F}_t^X \subseteq \mathcal{F}_\theta^X$ for $t < \theta$. This corresponds to the intuitive notion that

$$\left\{ \begin{array}{l} \mathcal{F}_t^X \ \text{ represents the information about the process} \\ \qquad X \ \text{ that has been revealed up to time } t \end{array} \right\}, \tag{1.3}$$

and obviously this information cannot decrease with time.

We shall write $\{\mathcal{F}_t\,;\, 0 \le t < \infty\}$, or simply $\{\mathcal{F}_t\}$, whenever the specification of the process that generates the relevant information is not of any particular importance, and call the resulting family a **filtration**. Now if $\mathcal{F}_t^X \subseteq \mathcal{F}_t$ holds for every $t \ge 0$, we say that *the process* X *is adapted to the filtration* $\{\mathcal{F}_t\}$, and write $\{\mathcal{F}_t^X\} \subseteq \{\mathcal{F}_t\}$.

1.3 The Markov property: A stochastic process X is said to be *Markovian*, if

$$P[X_\theta \in A|\mathcal{F}_t^X] = P[X_\theta \in A|X_t]; \quad \forall\ A\ \in\ \mathcal{B}(\mathcal{R}),\quad 0 < t < \theta.$$

Just like the Poisson process, every process with independent increments has this property.

1.4 The Martingale property: A stochastic process X with $E|X_t| < \infty$ is called

$$\begin{array}{rl} \textit{martingale}, \text{ if} & E(X_t|\mathcal{F}_s) = X_s \\ \textit{submartingale}, \text{ if} & E(X_t|\mathcal{F}_s) \ge X_s \\ \textit{supermartingale}, \text{ if} & E(X_t|\mathcal{F}_s) \le X_s \end{array} \qquad \text{holds (w.p.1) for every} \qquad 0 < s < t < \infty.$$

1.5 Discussion: (i) The filtration $\{\mathcal{F}_t\}$ in 1.4 can be the same as $\{\mathcal{F}_t^X\}$, but it may also be larger. This point can be important (e.g. in the representation Theorem 5.3) or even crucial (e.g. in Filtering Theory; cf. section 7), and not just a mere technicality. We stress it, when necessary, by saying that "X is an $\{\mathcal{F}_t\}$ - martingale", or that "$X = \{X_t, \mathcal{F}_t; 0 \le t < \infty\}$ is a martingale".

(ii) In a certain sense, martingales are the "constant functions" of probability theory; submartingales are the "increasing functions", and supermartingales are the "decreasing functions". In particular, for a martingale (submartingale, supermartingale) the expectation $t \mapsto EX_t$ is a constant (resp. nondecreasing, nonincreasing) function; on the other hand, a super(sub)martingale with constant expectation is necessarily a martingale. With this interpretation, if X_t stands for the fortune of a gambler at time t, then a martinale (submartingalge, supermartingale) corresponds to the notion of a fair (respectively: favorable, unfavorable) game.

(iii) The study of processes of the martingale type is at the heart of stochastic analysis, and becomes exceedingly important in applications. We shall try in this tutorial to illustrate both these points.

1.6 The Compensated Poisson process: If N is a Poisson process with intensity $\lambda > 0$, it is checked easily that the "compensated process"

$$M_t\ =\ N_t - \lambda\, t\,,\ \mathcal{F}_t^N\,,\quad 0 \le t < \infty$$

is a martingale. $\diamond$

In order to state correctly some of our later results, we shall need to "localize" the martingale property.

1.7 Definition: A random variable $\tau : \Omega \to [0, \infty]$ is called a *stopping time* of the filtration $\{\mathcal{F}_t\}$, if the event $\{\tau \le t\}$ belongs to $\mathcal{F}_t$, for every $0 \le t < \infty$.

In other words, the determination of whether τ has occurred by time t, can be made by looking at the information $\mathcal{F}_t$ that has been made available up to time t only, without anticipation of the future.

For instance, if X has continuous paths and A is a closed set of the real line, the "hitting time" $\tau_A = \min\{t \ge 0 \, : \, X_t \; \in \; A\}$ is a stopping time.

1.8 Definition: An adapted process $X = \{X_t, \mathcal{F}_t \, ; \, 0 \le t < \infty\}$ is called a *local martingale*, if there exists an increasing sequence $\{\tau_n\}_{n=1}^{\infty}$ of stopping times with $\lim_{n\to\infty} \tau_n = \infty$ such that the "stopped process" $\{X_{t\wedge\tau_n}, \mathcal{F}_t \, ; \, 0 \le t < \infty\}$ is a martingale, for every $n \ge 1$.

It can be shown that every martingale is also a local martingale, and that there exist local martingales which are not martingales; we shall not press these points here.

1.9 Exercise: Every nonnegative local martingale is a supermartingale.

1.10 Exercise: If X is a submartingale and τ is a stopping time, then the stopped process $X_t^\tau \triangleq X_{\tau\wedge t} \, , \;\; 0 \le t < \infty$ is also a submartingale.

1.11 Exercise (optional sampling theorem): If X is a submartingale with right-continuous sample paths and σ, τ two stopping times with $\sigma \le \tau \le M$ (w.p.1) for some real constant $M > 0$, then we have $E(X_\sigma) \le E(X_\tau)$.

2. BROWNIAN MOTION

This is by far the most interesting and fundamental stochastic process. It was studied by A. Einstein (1905) in the context of a kinematic theory for the irregular movement of pollen immersed in water that was first observed by the botanist R. Brown in 1824, and by Bachelier (1900) in the context of financial economics. Its mathematical theory was initiated by N. Wiener (1923), and P. Lévy (1948) carried out a brilliant study of its sample paths that inspired practically all subsequent research on stochastic processes until today. Appropriately, the process is also known as the *Wiener-Lévy* process, and finds applications in engineering (communications, signal processing, control), economics and finance, mathematical biology, management science, etc.

2.1 Motivational considerations (in one dimension): Consider a particle that is subjected to a sequence of *I.I.D.* (independent, identically distributed) Bernoulli "kicks" $\xi_1, \xi_2, \cdots$ with $P[\xi_1 = \pm 1] = 1/2$, of size $h > 0$, at the end of regular time-intervals of constant length $\delta > 0$. Thus, the location of the particle after n kicks is given as $h \cdot \Sigma_{j=1}^n \xi_j(\omega)$; more generally, the location of the particle at time t is

$$S_t(\omega) = h. \sum_{j=1}^{[t/\delta]} \xi_j(\omega) \, , \quad 0 \le t < \infty .$$

The resulting process S has right-continuous and piecewise constant sample paths, as well as stationary and independent increments (because of the independence of the ξ_j's). Obviously, $ES_t = 0$ and

$$Var(S_t) = h^2 \left[\frac{t}{\delta}\right] \cong \frac{h^2}{\delta} t .$$

We would like to get a continuous picture, in the limit, by letting $h \downarrow 0$ and $\delta \downarrow 0$, but at the same time we need a *positive* and *finite* variance for the limiting random variable S_t. This can be accomplished by maintaining $h^2 = \sigma^2 \delta$ for a finite constant $\sigma > 0$; in particular, by taking $\delta_n = 1/n$, $h_n = \sigma/\sqrt{n}$, and thus setting

$$S_t^{(n)}(\omega) \triangleq \frac{\sigma}{\sqrt{n}} \sum_{j=1}^{[nt]} \xi_j(\omega); \quad 0 \le t < \infty, \quad n \ge 1. \tag{2.1}$$

Now a direct application of the Central Limit Theorem shows that

(i) for fixed t, the sequence $\{S_t^{(n)}\}_{n=1}^{\infty}$ converges in distribution to a random variable $W_t \sim \mathcal{N}(0, \sigma^2 t)$.

(ii) for fixed $m \ge 1$ and $0 = t_0 < t_1 < ... < t_{m-1} < t_m < \infty$, the sequence of random vectors

$$\{(S_{t_1}^{(n)}, S_{t_2}^{(n)} - S_{t_1}^{(n)}, \dots, S_{t_m}^{(n)} - S_{t_{m-1}}^{(n)})\}_{n=1}^{\infty}$$

converges in distribution to a vector

$$(W_{t_1}, W_{t_2} - W_{t_1}, \dots, W_{t_m} - W_{t_{m-1}})$$

of independent random variables, with

$$W_{t_j} - W_{t_{j-1}} \sim \mathcal{N}(0, \sigma^2(t_j - t_{j-1})), \quad 1 \le j \le m.$$

You can easily imagine now that the entire *process* $S^{(n)} = \{S_t^{(n)}; 0 \le t < \infty\}$ converges in distribution (in a suitable sense) as $n \to \infty$, to a process $W = \{W_t; 0 \le t < \infty\}$ with the following properties:

(i) $W_0 = 0$;

(ii) $W_{t_1}, W_{t_2} - W_{t_1}, \cdots, W_{t_m} - W_{t_{m-1}}$ are independent,

(2.2) for every $m \ge 1$ and $0 = t_0 < t_1 < \dots t_m < \infty$;

(iii) $W_t - W_s \sim \mathcal{N}(0, \sigma^2(t - s))$, for every $0 < s < t < \infty$;

(iv) the sample path $t \mapsto W_t(\omega)$ is continuous, $\forall\, \omega \in \Omega$.

2.2 Definition: A process W with the properties of (2.2) is called a (one-dimensional) *Brownian motion* with variance σ^2; if $\sigma = 1$, the motion is called *standard.*

If $W^{(1)}, \ldots, W^{(d)}$ are d independent, standard Brownian motions, the vector-valued process $W = (W^{(1)}, \ldots, W^{(d)})$ is called a *standard Brownian motion in* $\mathcal{R}^d$. We shall take routinely $\sigma = 1$ from now on.

One cannot overstate the significance of this process. It stands out as the prototypical

(a) process with *stationary, independent increments;*

(b) *Markov* process;

(c) *Martingale* with continuous sample paths; and

(d) *Gaussian* process (with covariance function $R(t, s) = t \wedge s$).

2.3 Exercise: (i) Show that $W_t,\ W_t^2 - t$ are martingales.

(ii) Show that for every $\theta \in \mathcal{R}$, the processes below are martingales:

$$Z_t = \exp\left(\theta W_t - \frac{1}{2}\theta^2 t\right), \quad Y_t = \exp\left(i\theta W_t + \frac{1}{2}\theta^2 t\right) .$$

2.4 White Noise: For every integer $n \geq 1$, consider the Gaussian process

$$\xi_t^{(n)} \triangleq n\left[W_t - W_{t-1/n}\right]; \qquad 0 \leq t < \infty$$

with $E\left(\xi_t^{(n)}\right) = 0$ and covariance function $R_n(t, s) \triangleq E\left(\xi_t^{(n)}\xi_s^{(n)}\right) = Q_n(t-s)$, where

$$Q_n(\tau) = \left\{ \begin{array}{ccc} n^2(\frac{1}{n} - \tau) & ; & |\tau| \leq 1/n \\ 0 & ; & |\tau| \geq 1/n \end{array} \right\} .$$

As $n \to \infty$, the sequence of functions $\{Q_n\}_{n=1}^{\infty}$ approaches the Dirac delta. The "limit"

$$\xi_t = \text{``} \lim_{n \to \infty} \xi_t^{(n)} \text{''}$$

(in a generalized, distributional sense) is then a "zero mean Gaussian process with covariance function $R(t, s) = E(\xi_t \xi_s) = \delta(t - s)$". It is called *White Noise* and is of tremendous importance in communications and system theory.

Nota Bene: Despite its continuity, the sample path $t \mapsto W_t(\omega)$ *is not differentiable anywhere on* $[0, \infty)$.

Now fix a $t > 0$ and consider a sequence of partitions $0 = t_0^{(n)} < t_1^{(n)} < \ldots < t_k^{(n)} < \ldots < t_{2^n}^{(n)} = t$ of the interval $[0, t]$, say with $t_k^{(n)} = kt2^{-n}$, as well as the quantity

$$V_p^{(n)}(\omega) \triangleq \sum_{k=1}^{2^n} \left| W_{t_k^{(n)}}(\omega) - W_{t_{k-1}^{(n)}}(\omega) \right|^p, \quad p > 0, \tag{2.3}$$

the "variation of order p of the sample path $t \mapsto W_t(\omega)$ along the n^{th} partition". For $p=1$, $V_1^{(n)}(\omega)$ is simply the length of the polygonal approximation to the Brownian path; for $p=2$, $V_2^{(n)}(\omega)$ is the "quadratic variation" of the path along the approximation.

2.5 Theorem: *With probability one, we have*

$$\lim_{n\to\infty} V_p^{(n)} = \left\{ \begin{array}{ccll} \infty & ; & \text{for} & 0 < p < 2 \\ t & ; & \text{for} & p = 2 \\ 0 & ; & \text{for} & p > 2 \end{array} \right\} . \tag{2.4}$$

In particular:

$$\sum_{k=1}^{2^n} \left| W_{t_k^{(n)}} - W_{t_{k-1}^{(n)}} \right| \underset{n\to\infty}{\longrightarrow} \infty , \tag{2.5}$$

$$\sum_{k=1}^{2^n} \Big(W_{t_k^{(n)}} - W_{t_{k-1}^{(n)}} \Big)^2 \underset{n\to\infty}{\longrightarrow} t . \tag{2.6}$$

Remark: The relations (2.5), (2.6) become easily believable, if one considers them in L^1 rather than with probability one. Indeed, since

$$\begin{aligned} E\left| W_{t_k^{(n)}} - W_{t_{k-1}^{(n)}} \right| &= c \cdot \sqrt{t_k^{(n)} - t_{k-1}^{(n)}} = c \cdot 2^{-n/2} t^{\frac{1}{2}} \\ E\Big(W_{t_k^{(n)}} - W_{t_{k-1}^{(n)}} \Big)^2 &= t_k^{(n)} - t_{k-1}^{(n)} = 2^{-n}\, t , \end{aligned}$$

with $c = \sqrt{2\pi}$, we have as $n \to \infty$:

$$E \sum_{k=1}^{2^n} \left| W_{t_k^{(n)}} - W_{t_{k-1}^{(n)}} \right| = c.2^{n/2} \to \infty , \quad E \sum_{k=1}^{2^n} \Big(W_{t_k^{(n)}} - W_{t_{k-1}^{(n)}} \Big)^2 = t . \qquad \diamond$$

Arbitrary (local) martingales with continuous sample paths do not behave much differently. In fact, we have the following result.

2.6 Theorem: *For every nonconstant (local) martingale M with continuous sample paths, we have the analogues of (2.5), (2.6):*

$$\sum_{k=1}^{2^n} \left| M_{t_k^{(n)}} - M_{t_{k-1}^{(n)}} \right| \overset{\mathrm{P}}{\underset{n\to\infty}{\longrightarrow}} \infty \tag{2.7}$$

(2.8) $$\sum_{k=1}^{2^n}\Big(M_{t_k^{(n)}} - M_{t_{k-1}^{(n)}}\Big)^2 \xrightarrow[n\to\infty]{\text{P}} \langle M\rangle_t \ ,$$

where $\langle M\rangle$ is a process with continuous, nondecreasing sample paths. Furthermore, the analogue of (2.4) holds, if one replaces t by $\langle M\rangle_t$ on the right-hand side, and convergence with probability one by convergence in probability.

2.7 Remark on Notation: The process $\langle M\rangle$ of (2.8) is called the *quadratic variation* process of M; it is the unique process with continuous and nondecreasing paths, for which

(2.9) $$M_t^2 \ - \ \langle M\rangle_t \ = \ \text{local martingale.}$$

In particular, if M is a square-integrable martingale, i.e., if $E(M_t^2) < \infty$ holds for every $t \geq 0$, then

(2.9)′ $$M_t^2 \ - \langle M\rangle_t \ = \ \text{martingale.}$$

2.8 Corollary: *Every (local) martingale M, with sample paths which are continuous and of finite first variation, is necessarily constant.*

2.9 Exercise: For the compensated Poisson process M of 1.6, show that $M_t^2 - \lambda t$ is a martingale, and thus $\langle M\rangle_t = \lambda t$ in (2.9)′.

2.11 Exercise: For any two (local) martingales M and N with continuous sample paths, we have

(2.10) $$\sum_{k=1}^{2^k}[M_{t_k^{(n)}} - M_{t_{k-1}^{(n)}}][N_{t_k^{(n)}} - N_{t_{k-1}^{(n)}}] \xrightarrow[n\to\infty]{\text{P}} \langle M,N\rangle_t \ \triangleq \ \frac{1}{4}\Big[\,\langle M+N\rangle_t - \langle M-N\rangle_t\,\Big].$$

2.12 Remark: The process $\langle M,N\rangle$ of (2.10) is continuous and of bounded variation (difference of two nondecreasing processes); it is the unique process with these properties, for which

(2.11) $$M_t\,N_t \ - \ \langle M,N\rangle_t \ = \text{local martingale}$$

and is called the *cross-variation* of M and N. If M, N are independent, then $\langle M,N\rangle \equiv 0$.

For square-integrable martingales M, N the pairing $\langle\cdot,\cdot\rangle$ plays the rôle of an inner product: the process of (2.11) is then a martingale, and we say that M, N are *orthogonal* if $\langle M,N\rangle \equiv 0$ (which amounts to saying that MN is a martingale).

2.13 Burkholder-Gundy Inequalities: Let M be a local martingale with continuous sample paths, $\langle M\rangle$ the associated process of (2.9), and $M_t^* = \max_{0\leq s\leq t}|M_s|$, for $0 \leq t < \infty$. Then for any $p > 0$ and any stopping time τ we have:

$$k_p \cdot E\langle M\rangle_\tau^p \ \leq E(M_\tau^*)^{2p} \ \leq \ K_p \cdot E\langle M\rangle_\tau^p$$

where k_p, K_p are universal constants (depending only on p).

2.14 Doob's Inequality: If M is a nonnegative submartingale with right-continuous sample paths, then

$$E\left(\sup_{0\le t\le T} M_t\right)^p \le \left(\frac{p}{p-1}\right)^p \cdot E\big(X_T\big)^p, \quad \forall\ p>1\,.$$

3. STOCHASTIC INTEGRATION

Consider a Brownian motion W adapted to a given filtration $\{\mathcal{F}_t\}$; for a suitable *adapted* process X, we would like to define the stochastic integral

$$I_t(X) = \int_0^t X_s dW_s \tag{3.1}$$

and to study its properties as a process indexed by t. We see immediately, however, that $\int_0^t X_s(\omega)dW_s(\omega)$ cannot possibly be defined for any $\omega\in\Omega$ as a Lebesgue-Stieltjes integral, because the path $s\mapsto W_s(\omega)$ is of infinite first variation on any interval $[0,t]$; recall (2.5).

Thus, we need a new approach, one that can exploit the fact that *the path has finite and positive second* (quadratic) *variation*; cf. (2.6). We shall try to sketch the main lines of this approach, leaving aside all the technicalities (which are rather demanding!).

Just as with the Lebesgue integral, it is pretty obvious what everybody's choice should be for the stochastic integral, in the case of particularly simple processes X. Let us place ourselves, from now on, on a finite interval $[0,T]$.

3.1 Definition: A process X is called *simple*, if there exists a partition $0=t_0<t_1\ldots<t_r<t_{r+1}=T$ such that $X_s(\omega)=\theta_j(\omega);\ t_j<s\le t_{j+1}$ where θ_j is a bounded, $\mathcal{F}_{t_j}-$measurable random variable.

For such a process, we define in a natural way:

$$\begin{aligned} I_t(X) \ = \int_0^t X_s dW_s &\triangleq \sum_{j=0}^{m-1}\theta_j(W_{t_{j+1}}-W_{t_j})+\theta_m(W_t-W_{t_m});\quad t_m<t\le t_{m+1} \\ &= \sum_{j=0}^{r}\theta_j(W_{t\wedge t_{j+1}}-W_{t\wedge t_j}). \end{aligned} \tag{3.2}$$

There are several properties of the integral that follow easily from this definition; pretending that $t=t_{m+1}$ to simplify notation, we obtain

$$EI_t(X) = E\sum_{j=0}^{m}\theta_j(W_{t_j+1}-W_{t_j}) \ = \sum_{j=0}^{m}E\big[\theta_j\cdot E(W_{t_{j+1}}-W_{t_j}|\mathcal{F}_{t_j})\big]=0\,,$$

and more generally, for $s < t$:

(3.3) $$E[I_t(X)|\mathcal{F}_s] \;=\; I_s(X)\,.$$

In other words, the integral is a *martingale with continuous sample paths.* What is the quadratic variation of this martingale? We can get a clue, if we compute the second moment

(3.4)
$$\begin{aligned}
E(I_t(X))^2 \;&=\; E\left(\sum_{j=0}^{m}\theta_j(W_{t_{j+1}}-W_{t_j})\right)^2 \\
&= E\sum_{j=0}^{m}\theta_j^2(W_{t_{j+1}}-W_{t_j})^2 + 2\cdot E\sum_{j=0}^{m}\sum_{i=j+1}^{m}\theta_i\theta_j(W_{t_{i+1}}-W_{t_i})(W_{t_{j+1}}-W_{t_j}) \\
&= E\sum_{j=0}^{m}\theta_j^2 E[(W_{t_{j+1}}-W_{t_j})^2|\mathcal{F}_{t_j}]+ \\
&\qquad + 2\cdot E\sum_{j=0}^{m}\sum_{i=j+1}^{m}\theta_i\theta_j(W_{t_{j+1}}-W_{t_j}).E[W_{t_{i+1}}-W_{t_i}|\mathcal{F}_{t_i}] \\
&= E\sum_{j=0}^{m}\theta_j^2(t_{j+1}-t_j) \;=\; E\int_0^t X_u^2\,du.
\end{aligned}$$

A similar computation leads to

(3.5) $$E[(I_t(X)-I_s(X))^2\,|\,\mathcal{F}_s] \;=\; E\left[\int_s^t X_u^2\,du\,\middle|\,\mathcal{F}_s\right],$$

which shows that the quadratic variation of $I(X)$ is $\langle I(X)\rangle_t \;=\; \int_0^t X_u^2\,du\,.$

On the other hand, if Y is another simple process, a computation similar to (3.4), (3.5) gives $E[I_t(X)I_t(Y)] = E\int_0^t X_uY_u du$ and, more generally,

$$E[(I_t(X)-I_s(X))(I_t(Y)-I_s(Y)\,|\,\mathcal{F}_s] \;=\; E\left[\int_s^t X_uY_u\,du\,|\,\mathcal{F}_s\right].$$

We are led to the following.

3.2 Proposition: *For simple processes X and Y, the integral of (3.1) is defined as in (3.2), and is a square-integrable martingale with continuous paths and quadratic* (respectively, cross-) *variation process given by*

(3.6) $$\langle I(X)\rangle_t \;=\; \int_0^t X_u^2\,du\;, \quad \langle I(X),I(Y)\rangle_t \;=\; \int_0^t X_uY_u\,du\,.$$

In particular, we have

(3.7)

$$E[I_t(X)] = 0\,, \quad E(I_t(X))^2 = E\int_0^t X_u^2 du\,, \quad E[I_t(X)I_t(Y)] = E\int_0^t X_u Y_u du. \qquad \diamond$$

The idea now is that an arbitrary measurable, adapted process X with

$$E\int_0^T X_u^2 du < \infty$$

can be approximated by a sequence of simple processes $\{X^{(n)}\}_{n=1}^{\infty}$, in the sense

$$E\int_0^T |X_u^{(n)} - X_u|^2 du \underset{n\to\infty}{\longrightarrow} 0\,.$$

Then the corresponding sequence of stochastic integrals $\{I(X^{(n)})\}_{n=1}^{\infty}$ *converges* in the sense of $L^2(dt\otimes dP)$, and the limit $I(X)$ is called the *stochastic integral of* X with respect to W. It also turns out that most of the properties of Proposition 3.2 are maintained.

3.3 Theorem: *For every measurable, adapted process X with the property $\int_0^T X_u^2\,du < \infty$ (w.p.1), one can define the stochastic integral $I(X)$ of X with respect to W. This process is a local martingale with continuous sample paths, and quadratic (and cross –) variation processes given by (3.6).*

Furthermore, if we have $E\int_0^T X_u^2 du < \infty$, then the local martingale $I(X)$ is actually a square-integrable martingale and the properties of (3.7) hold. $\diamond$

Predictably, nothing in all this development is terribly special about Brownian motion. Indeed, if we let M, N be arbitrary (local) martingales with continuous sample paths, we have the following analogue of Theorem 3.3:

3.4 Theorem: *For any measurable, adapted process X with*

$$\int_0^T X_u^2\, d\langle M\rangle_u < \infty \qquad \text{(w.p.1)},$$

one can define the stochastic integral $I^M(X)$ of X with repect to M; the resulting process is a local martingale with continuous paths and quadratic (and cross–) variations

(3.8)
$$\langle I^M(X)\rangle_t = \int_0^t X_u^2 d\langle M\rangle_u, \quad \langle I^M(X), I^M(Y)\rangle_t = \int_0^t X_u Y_u\, d\langle M\rangle_u\,.$$

(Here, Y is a process with the same properties as X.)

Furthermore, if Z is a measurable, adapted process with $\int_0^T Z_u^2 d\langle N\rangle_u < \infty$ *(w.p.1), we have*

$$\langle I^M(X), I^N(Z)\rangle_t = \int_0^t X_u Z_u d\langle M, N\rangle_u . \tag{3.9}$$

If now $E\int_0^T (X_u^2 + Y_u^2) d\langle M\rangle_u < \infty$, $E\int_0^T Z_u^2 \, d\langle N\rangle_u < \infty$, *then the processes* $I^M(X)$, $I^M(Y)$ *and* $I^N(Z)$ *are actually square-integrable martingales, with*

$$E(I_t^M(X)) = 0, \qquad E\big(I_t^M(X)\big)^2 = E\int_0^t X_u^2 \, d\langle M\rangle_u \ , \tag{3.10}$$

$$E[I_t^M(X) I_t^M(Y)] = E\int_0^t X_u Y_u \, d\langle M\rangle_u \ , \tag{3.11}$$

$$E[I_t^M(X) I_t^N(Z)] = E\int_0^t X_u Z_u \, d\langle M, N\rangle_u . \tag{3.12}$$

3.5 Remark: If we take $Z \equiv 1$, then obviously $I^N(Z) \equiv N$, and (3.9) becomes

$$\langle I^M(X), N\rangle_t = \int_0^t X_u \, d\langle M, N\rangle_u .$$

It turns out that this property characterizes the stochastic integral, in the following sense: suppose that for some continuous local martingale Λ we have

$$\langle \Lambda, N\rangle_t = \int_0^t X_u \, d\langle M, N\rangle_u \ , \qquad 0 \le t \le T$$

for every continuous local martingale N; then $\Lambda \equiv I^M(X)$.

3.6 Exercise: In the context of Theorem 3.4, suppose that $N = I^M(X)$ and $Q = I^N(Z)$; show that we have then $Q = I^M(XZ)$.

4. THE CHAIN RULE OF THE NEW CALCULUS

The definition of the integral carries with it a certain "calculus", i.e., a set of rules that can make the integral amenable to more-or-less mechanical calculation. This is true for the Riemann and Lebesgue integrals, and is just as true for the stochastic integral as well; it turns out that, in this case, there is a simple chain rule that complements and extends that of the ordinary calculus.

Suppose that $f : \mathcal{R} \to \mathcal{R}$ and $\psi : [0, \infty) \to \mathcal{R}$ are C^1 functions; then the ordinary chain rule gives

$$(f(\psi(t))' = f'(\psi(t)) \cdot \psi'(t),$$

or equivalently:

$$f(\psi(t)) = f(\psi(0)) + \int_0^t f'(\psi(s)) d\psi(s). \tag{4.1}$$

Actually (4.1) holds even when ψ is simply of bounded variation (not necessarily continuously differentiable), provided that the integral on the right-hand side is interpreted in the Stieltjes sense. We cannot expect (4.1) to hold, however, if $\psi(\cdot)$ is replaced by $W_{.}(\omega)$, because the path $t \mapsto W_t(\omega)$ is of infinite variation for almost every $\omega \in \Omega$. It turns out that we need a second-order correction term.

4.1 Theorem: *Let $f : \mathcal{R} \to \mathcal{R}$ be of class C^2, and W be a Brownian motion. Then*

$$f(W_t) \ = \ f(W_0) + \int_0^t f'(W_s) dW_s + \frac{1}{2} \int_0^t f''(W_s) ds \ , \quad 0 \le t < \infty \, . \tag{4.2}$$

More generally, if M is a (local) martingale with continuous sample paths:

$$f(M_t) \ = \ f(M_0) + \int_0^t f'(M_s) dM_s + \frac{1}{2} \int_0^t f''(M_s) \, d\langle M \rangle_s \ , \quad 0 \le t < \infty \, . \tag{4.3}$$

Idea of proof: Consider a partition $0 = t_0 < t_1 < \ldots < t_m < t_{m+1} = t$ of the interval $[0, t]$, and do a Taylor expansion:

$$f(W_t) - f(W_0) \ = \ \sum_{j=1}^m \{ f(W_{t_{j+1}}) - f(W_{t_j}) \} \ =$$

$$= \sum_{j=1}^m f'(W_{t_j})(W_{t_{j+1}} - W_{t_j}) + \frac{1}{2} \sum_{j=1}^m f''(W_{t_j} + \theta_j (W_{t_{j+1}} - W_{t_j})) \cdot (W_{t_{j+1}} - W_{t_j})^2,$$

where θ_j is an $\mathcal{F}_{t_{j+1}}-$measurable random variable with values in the interal $[-1, 1]$. In this last expression, as the partition becomes finer and finer, the first sum approximates the stochastic integral $\int_0^t f'(W_s)dW_s$, whereas the second sum approximates $\int_0^t f''(W_s)ds$; cf. (2.6).

4.2 Example: In order to compute $\int_0^t W_s dW_s$, all you have to do is take $f(x) = x^2$ in (4.2), and obtain

$$\int_0^t W_s dW_s = \frac{1}{2}(W_t^2 - t)\,.$$

(Try to arrive at the same conclusion, by evaluating the approximating sums of the form $\sum_{j=1}^m W_{t_j}(W_{t_{j+1}} - W_{t_j})$ along a partition, and then letting the partition become dense in $[0,t]$. Notice how harder you have to work this way!).

4.3 Theorem: *Let M be a (local) martingale with continuous sample paths and $\langle M \rangle_t = t$. Then M is a Brownian motion.*

Proof: We have to show that M has independent increments, and that the increment $M_t - M_s$ has a normal distribution with mean zero and variance $t-s$, for $0 < s < t < \infty$. Both these claims will follow, as soon as it is shown that

$$E\left[e^{i\theta(M_t - M_s)} \,\middle|\, \mathcal{F}_s\right] = e^{-\frac{1}{2}\theta^2(t-s)}; \quad \forall\, \theta \in \mathcal{R}\,. \tag{4.4}$$

With $f(x) = e^{i\theta x}$, we have from (4.3):

$$e^{i\theta M_t} = e^{i\theta M_s} + \int_s^t i\theta e^{i\theta M_u} dM_u - \frac{\theta^2}{2}\int_s^t e^{i\theta M_u} du\,,$$

and this leads to:

$$E\left[e^{i\theta(M_t - M_s)} \,\middle|\, \mathcal{F}_s\right] = 1 + i\theta \cdot e^{-i\theta M_s} E\left[\int_s^t e^{i\theta M_u} dM_u \,\middle|\, \mathcal{F}_s\right] - \frac{\theta^2}{2}\int_s^t E\left[e^{i\theta(M_u - M_s)} \,\middle|\, \mathcal{F}_s\right] du\,.$$

Because the conditional expectation of the stochastic integral is zero (the martingale property!), we are led to the conclusion that the function

$$g(t) \triangleq E\left[e^{i\theta(M_t - M_s)} \,\middle|\, \mathcal{F}_s\right] \;;\;\; t \geq s$$

satisfies the integral equation $g(t) = 1 - \frac{1}{2}\theta^2 \int_s^t g(u)du$. But there is only one solution to this equation, namely $g(t) = e^{-\frac{1}{2}\theta^2(t-s)}$, proving (4.4). ◇

Theorem 4.1 can be generalized in several ways; here is a version that we shall find most useful, and which is established in more or less the same way.

4.4 Proposition: *Let X be a* semimartingale, *i.e., a process of the form*

$$X_t = X_0 + M_t + V_t\,, \quad 0 \leq t < \infty$$

where M is a local martingale with continuous sample paths, and V a process with continuous sample paths of finite first variation. Then, for every function $f : \mathcal{R} \rightarrow \mathcal{R}$ of class C^2, we have

(4.5)

$$f(X_t) \;=\; f(X_0) + \int_0^t f'(X_s)\, dM_s + \int_0^t f'(X_s)\, dV_s + \frac{1}{2}\int_0^t f''(X_s)\, d\langle M\rangle_s\,, \quad 0 \le t < \infty\,.$$

More generally, let $X = (X^{(1)}, \cdots, X^{(d)})$ be an $\mathcal{R}^d$ - valued process with components $X_t^{(i)} = X_0^{(i)} + M_t^{(i)} + V_t^{(i)}$ of the above type, and $f : \mathcal{R}^d \rightarrow \mathcal{R}$ a function of class C^2. We have then

(4.6)

$$\begin{aligned} f(X_t) \;&=\; f(X_0) + \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x_i}(X_s) dM_s^{(i)} + \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x_i}(X_s) dV_s^{(i)} + \\ &+ \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s)\; d\langle M^{(i)}, M^{(j)}\rangle_s\,, \quad 0 \le t < \infty. \end{aligned}$$

4.5 Example: If M is a local martingale with continuous sample paths, then so is

(4.7)
$$Z_t \;=\; \exp\left[M_t - \frac{1}{2}\langle M\rangle_t \right]\,, \quad 0 \le t < \infty$$

and satisfies the elementary *stochastic integral equation*

(4.8)
$$Z_t \;=\; 1 + \int_0^t Z_s dM_s\,, \quad 0 \le t < \infty\,.$$

Indeed, apply (4.5) to the semimartingale $X = M - \frac{1}{2}\langle M\rangle$ and the function $f(x) = e^x$. The local martingale property of Z follows from the fact that it is a stochastic integral; on the other hand, Exercise 1.9 shows that Z is also a supermartingale.

When is this supermartingale actually a *martingale*? It turns out that

(4.9)
$$E\Big[\exp\Big\{\frac{1}{2}\langle M\rangle_T\Big\}\Big] \;<\; \infty$$

is a sufficient condition. For instance, if

(4.10)
$$M_t \;=\; \int_0^t X_s dW_s \;\triangleq\; \sum_{i=1}^d \int_0^t X_s^{(i)} dW_s^{(i)}$$

with $\int_0^T ||X_t||^2 dt < \infty$ (w.p.1), then the exponential supermartingale

(4.11)
$$Z_t \;=\; \exp\left(\int_0^t X_s dW_s - \frac{1}{2}\int_0^t ||X_s||^2 ds \right)$$

satisfies the equation

$$Z_t = 1 + \int_0^t Z_s X_s dW_s\,, \quad 0 \le t < \infty \tag{4.12}$$

and is a martingale if

$$E\Big[\exp\Big\{\frac{1}{2}\int_0^T ||X_s||^2 ds\Big\}\Big] < \infty. \tag{4.13}$$

4.6 Example: Integration-by-parts. With $d = 2$ and $f(x_1, x_2) = x_1 x_2$ in (4.6), we obtain

$$X_t^{(1)} X_t^{(2)} = X_0^{(1)} X_0^{(2)} + \int_0^t X_s^{(1)} dX_s^{(2)} + \int_0^t X_s^{(2)} dX_s^{(1)} + \langle M^{(1)}, M^{(2)} \rangle_t\,. \tag{4.14}$$

4.7 Exercise: Using the formula (4.6), establish the following multi-dimensional analogue of Theorem 4.3: "*If $M = (M^{(1)}, \ldots, M^{(d)})$ is a vector of (local) martingales with continuous paths and $\langle M^{(i)}, M^{(j)} \rangle_t = t\delta_{ij}$, then M is an $\mathcal{R}^d$ - valued Brownian motion.*"

Here, $\delta_{ij} = \left\{ \begin{matrix} 0 & , & i \neq j \\ 1 & , & i = j \end{matrix} \right\}$ is the Kronecker delta.

5. THE FUNDAMENTAL THEOREMS

In this section we expound on the theme that *Brownian motion is the fundamental martingale with continuous sample paths.* We illustrate this point by establishing "representation results" for such martingales in terms of Brownian motion. We conclude with the celebrated result of Girsanov, according to which "Brownian motion is invariant under the combined effect of a particular translation, and of a change of probability measure".

Our first result states that "every local martingale with continuous sample paths, is nothing but a Brownian motion, run under a different clock".

5.1 Theorem: *Let M be a continuous local martingale with There exists then a Brownian motion W, such that:*

$$M_t = W_{\langle M \rangle_t}\,; \quad 0 \le t < \infty.$$

Sketch of proof in the case $\langle M \rangle$ is strictly increasing *: In this case $\langle M \rangle$ has an inverse, say T, which is continuous (as well as strictly increasing).

* E.g., if $M_t = \int_0^t X_s\, dB_s$, where B is Brownian motion and X takes values in $\mathcal{R}\backslash\{0\}$.

Then it is not hard to see that the process

$$W_s = M_{T(s)}\,, \quad 0 \le s < \infty \tag{5.1}$$

is a martingale (with respect to the filtration $\mathcal{G}_s = \mathcal{F}_{T(s)}, s \ge 0$) with continuous sample paths, as being the composition of the two continuous mappings $T : \mathcal{R}^+ \to \mathcal{R}^+$ and $M : \mathcal{R}^+ \to \mathcal{R}$. On the other hand, it is "intuitively clear" that $\langle W\rangle_s = \langle M\rangle_{T(s)} = s$, so W is Brownian motion by Theorem 4.3. Furthermore, replacing s by $\langle M\rangle_t$ in (5.1), we obtain $W_{\langle M\rangle_t} = M_{T(\langle M\rangle_t)} = M_t$, which is the desired representation. ◇

A second representation result, similar in spirit to Theorem 5.1, is left as an Exercise.

5.2 Exercise: Let M be a local martingale with continuous sample paths, and quadratic variation of the form $\langle M\rangle_t = \int_0^t X_s^2\, ds$ for some adapted process X. Then M admits the representation

$$M_t = M_0 + \int_0^t X_s\, dW_s\,, \quad 0 \le t < \infty \tag{5.2}$$

as a stochastic integral of X with respect to a suitable Brownian motion W.

(*Hint:* If X takes values in $\mathcal{R}\backslash\{0\}$, show that one can take

$$W_t = \int_0^t \frac{1}{X_s}\, dM_s \tag{5.3}$$

in (5.2). For the general case assume, as you may, that the probability space is rich enough to support a Brownian motion B independent of M, and use B to modify accordingly the definition of W in (5.3)). ◇

Here is now our final, and most important, representation result for martingales in terms of Brownian motion. In constrast to both Theorem 5.1 and Exercise 5.2, the Brownian motion W in Thorem 5.3 is *given* and fixed.

5.3 Theorem: *Let W be a Brownian motion, and recall the notation $\mathcal{F}_t^W = \sigma(W_s; 0 \le s \le t)$ of (1.2) for the history of the process up to time t. Every local martingale M with respect to the filtration $\{\mathcal{F}_t^W\}$ admits a representation of the form*

$$M_t = M_0 + \int_0^t X_s\, dW_s\,, \quad 0 \le t < \infty\,, \tag{5.4}$$

for a measurable process X which is adapted to $\{\mathcal{F}_t^W\}$ and satisfies $\int_0^T X_s^2\, ds < \infty$ (w.p.1) for every $0 < T < \infty$. In particular, every such M has continuous sample paths. ◇

5.4 Remark: If M in Theorem 5.3 happens to be a square-integrable martingale, then X in (5.4) can be chosen to satisfy $E\int_0^T X_s^2 ds < \infty$ for every $0 < T < \infty$.

Finally, we present the fundamental "change of probability measure" result (Theorem 5.5). By way of motivation, let us consider a probability space $(\Omega, \mathcal{F}, P)$ and independent, standard normal random variables $\xi_1, \xi_2, \ldots, \xi_n$ on it. For an arbitrary vector $\mu \in \mathcal{R}^n$, introduce a new measure $\widetilde{P}$ on $(\Omega, \mathcal{F})$ by

$$\widetilde{P}(d\omega) = \exp\left\{\sum_{i=1}^n \mu_i \xi_i(\omega) - \frac{1}{2}\sum_{i=1}^n \mu_i^2\right\} \cdot P(d\omega),$$

which is actually a probability, since

$$\widetilde{P}(\Omega) = e^{-\frac{1}{2}\sum_1^n \mu_i^2} \cdot \prod_{i=1}^n E\left(e^{\mu_i \xi_i}\right) = 1.$$

What is the distribution of $(\xi_1, \ldots, \xi_n)$ under $\tilde{P}$? We have

$$\begin{aligned}\widetilde{P}[\xi \in dz_1, \ldots, \xi_n \in dz_n] &= \exp\left(\sum_1^n \mu_i z_i - \frac{1}{2}\sum_1^n \mu_i^2\right) \cdot P[\xi_1 \in dz_1, \ldots, \xi_n \in dz_n] \\ &= (2\pi)^{-n/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^n (z_i - \mu_i)^2\right\} dz_1 \ldots dz_n.\end{aligned}$$

In other words, under $\widetilde{P}$ the random variables $(\xi_1, \ldots, \xi_n)$ are independent, and $\xi_i \sim \mathcal{N}(\mu_i, 1)$. Equivalently, with $\tilde{\xi}_i = \xi_i - \mu_i$, $1 \le i \le n$: *the random variables* $(\tilde{\xi}_1, \ldots, \tilde{\xi}_n)$ *have the same law under* $\tilde{P}$, *as the random variables* $(\xi_1, \ldots, \xi_n)$ under P (namely, independent and standard normal).

The following result, which extends this idea to processes, is of paramount importance in stochastic analysis. We formulate it directly in its multidimensional form.

5.5 Theorem (Girsanov (1960)): *Let* $W = \{W_t, \mathcal{F}_t\,;\, 0 \le t \le T\}$ *be* $d-$*dimensional Brownian motion,* $X = \{X_t, \mathcal{F}_t\,;\, 0 \le t \le T\}$ *a measurable, adapted,* $\mathcal{R}^d$ *– valued process with* $\int_0^T ||X_t||^2 dt < \infty$ *(w.p.1), and suppose that the exponential supermartinale* Z *of (4.11) is actually a martingale:*

(5.5) $$E(Z_T) = 1.$$

Then under the measure

(5.6) $$\widetilde{P}(d\omega) = Z_T(\omega) P(d\omega),$$

which is actually a probability by virtue of (5.5), the process

$$(5.7) \qquad \widetilde{W}_t = W_t - \int_0^t X_s\, ds, \quad \mathcal{F}_t\ , \quad 0 \le t \le T$$

is Brownian motion.

5.6 Remark: Recall the sufficient condition (4.13) for the validity of (5.6); in particular, Z is a martingale if X is bounded.

5.7 Remark: We have the following generalization of Theorem 5.5. Suppose that M is a local martingale with continuous sample paths and $M_0 = 0$, for which the exponential local martingale $Z = \exp[M - \frac{1}{2}\langle M\rangle]$ of (4.7) is actually a *martingale*, and define a new probability measure $\tilde{P}$ as in (5.6). Then for any continuous, P-local martingale N, the process

$$\tilde{N}_t \triangleq N_t - \langle M, N\rangle_t = N_t - \int_0^t \frac{1}{Z_s}\, d\langle Z, N\rangle_s, \qquad 0 \le t \le T$$

is a (continuous) local martingale under $\widetilde{P}$.

6. DYNAMICAL SYSTEMS DRIVEN BY WHITE NOISE INPUTS

Consider a dynamical system described by an ordinary differential equation of the form

$$(6.1) \qquad \dot{X}_t = b(t, X_t) + \sigma(t, X_t)\xi_t\ ,$$

where X_t is the ($\mathcal{R}^d-$valued) state of the system at time t, and ξ_t is the ($\mathcal{R}^n-$valued) input at t.

The study of dynamical systems of this form, for deterministic inputs ξ, is well-established. We should like to develop a similar theory for stochastic inputs ξ as well; in particular, *we shall develop a theory for equations of the type (6.1) when ξ is a white noise process* as in 2.4.

Since, formally at least, we have $\xi_t = dW_t/dt$, we shall prefer to look at the *integral version*

$$(6.2) \qquad X_t = \eta + \int_0^t b(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dW_s\ , \quad 0 \le t \le T$$

of (6.1), where W is an $\mathcal{R}^n$ - valued Brownian motion independent of the initial condition $X_0 = \eta$, and the convention of (4.10) is being used. For suitable *drift* b and *volatility* σ coëfficients, we should like to answer questions concerning the existence, the uniqueness, and the various properties of the solution X to the resulting *stochastic integral equation*

(6.2). Note that, before we can even pose such questions, we need to have developed a theory of stochastic integration with respect to Brownian motion W (in order to make sense of the last integral in (6.2)). This was part of the motivation behind the theory of section 3.

6.1 Example: Linear Systems. In the case $b(t,x) = A(t)x + a(t), \sigma(t,x) = B(t)x + b(t)$, the resulting equation

$$dX_t = [A(t)X_t + a(t)]dt + [B(t)X_t + b(t)]dW_t \tag{6.3}$$

(in differential form) is *linear* in X_t, and can be solved in principle explicitly; the solution X is then a Gaussian process, if the initial random vector X_0 has a normal distribution and is independent of the driving Brownian motion W.

More precisely, one considers the deterministic linear system

$$\dot{\xi}(t) = A(t)\xi(t) + a(t) \tag{6.3$'$}$$

associated with (6.3), and its homogeneous version $\dot{\xi}(t) = A(t)\xi(t)$. The *fundamental solution* of the latter is a nonsingular, $(d \times d)$ matrix-valued function $\Phi(\cdot)$ that satisfies the matrix differential equation $\dot{\Phi}(t) = A(t)\Phi(t)$, terms of $\Phi(\cdot)$, the solutions of $(6.3)'$, and of (6.3) with $B(\cdot) \equiv 0$, are given by

$$\xi(t) = \Phi(t)\left[\xi(0) + \int_0^t \Phi^{-1}(s)a(s)ds\right]$$

and

$$X_t = \Phi(t)\left[X_0 + \int_0^t \Phi^{-1}(s)a(s)ds + \int_0^t \Phi^{-1}(s)b(s)dW_s\right],$$

respectively. In particular, if X_0 has a (d-variate) normal distribution, then X is a Gaussian process with mean vector $m(t) \triangleq EX_t$ and covariance matrix $\rho(s,t) \triangleq E[(X_s - m(s)).(X_t - m(t))^T]$ given by

$$m(t) = \Phi(t)\left[m(0) + \int_0^t \Phi^{-1}(s)a(s)ds\right]$$
$$\rho(s,t) = \Phi(s)\left[\rho(0,0) + \int_0^{s\wedge t} \Phi^{-1}(u)b(u)(\Phi^{-1}(u)b(u))^T du\right]\Phi^T(t).$$

Let us now look at a few, very special cases of (6.3), with $d = n = 1$.

(i) If $A(\cdot) = a(\cdot) = b(\cdot) \equiv 0$, the unique solution of the resulting linear equation $dX_t = B(t)X_t dW_t$ is given by

$$X_t = X_0 \cdot \exp\left\{\int_0^t B(s)\,dW_s - \frac{1}{2}\int_0^t B^2(s)\,ds\right\};$$

cf. (4.11), (4.12).

(ii) If $A(t) = -\alpha < 0,\ b(t) = b > 0,\ a(\cdot) = B(\cdot) \equiv 0$, the resulting *Langevin equation*

$$dX_t = -\alpha X_t dt + b dW_t \tag{6.4}$$

leads to the so-called *Ornstein-Uhlenbeck process* (Brownian movement with proportional restoring force). Assuming that there is a unique solution to (6.4) – cf. Theorem 6.4 below – we can find it via an "integration by parts":

$$d(X_t e^{\alpha t}) = e^{\alpha t}(dX_t + \alpha X_t dt) = b e^{\alpha t} dW_t\,, \quad \text{that is,}$$

$$X_t e^{\alpha t} = X_0 + \int_0^t b e^{\alpha s} dW_s\ , \quad t \geq 0\ . \tag{6.5}$$

Now if X_0 is independent of W and has a distribution with mean $EX_0 = \mu$ and variance $Var(X_0) = \sigma^2$, it follows easily from (6.5) that $EX_t = \mu e^{-\alpha t}$, $Var(X_t) = (\sigma^2 - \frac{b^2}{2\alpha})e^{-2\alpha t} + \frac{b^2}{2\alpha}$. Thus, the limiting (and invariant) distribution for this process is the normal $\mathcal{N}(0, b^2/2\alpha)$.

For a complete treatment of the general linear equation (6.3) in several dimensions, cf. Karatzas & Shreve (1987), §5.6.

6.2 Exercise: Introduce the *diffusion matrix* $\alpha(t,x) = \sigma(t,x)\sigma^T(t,x)$, i.e.,

$$a_{ij}(t,x) = \sum_{k=1}^n \sigma_{ik}(t,x)\sigma_{jk}(t,x)\,, \quad 1 \leq i,j \leq d \tag{6.6}$$

and the linear, second-order differential operator

$$\mathcal{A}_t\phi(x) \triangleq \sum_{i=1}^d b_i(t,x)\frac{\partial\phi(x)}{\partial x_i} + \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d a_{ij}(t,x)\frac{\partial^2\phi(x)}{\partial x_i \partial x_j}\,. \tag{6.7}$$

If the process X satisfies the equation (6.2), $f : [0,\infty) \times \mathcal{R}^d \to \mathcal{R}$ is a function of class $C^{1,2}$, and $\beta_t \triangleq e^{-\int_0^t K_u du}$ for some measurable, adapted and nonnegative process K, show that the process

$$M_t^f \triangleq \beta_t f(t,X_t) - f(0,X_0) - \int_0^t \left(\frac{\partial f}{\partial s} + \mathcal{A}_s f - K_s f\right)(s,X_s)\,\beta_s\,ds\,, \quad 0 \leq t < \infty$$

is a local martingale (square-integrable martingale, if f is of compact support) with continuous sample paths, and can be represented actually as

$$\sum_{i=1}^d \sum_{k=1}^n \int_0^t \frac{\partial f(s,X_s)}{\partial x_i}\sigma_{ik}(s,X_s)\beta_s dW_s^{(k)}.$$

6.3 Exercise: In the case of bounded, continuous drift and diffusion coefficients $b(t,x)$, and $a(t,x)$, establish their respective interpretations as local *velocity vector*

$$b_i(t,x) = \lim_{h \downarrow 0} \frac{1}{h} E[X_{t+h}^{(i)} - x_i | X_t = x]\,; \quad 1 \le i \le d \tag{6.8}$$

and local *variance-covariance matrix*

$$a_{ij}(t,x) = \lim_{h \downarrow 0} \frac{1}{h} E[(X_{t+h}^{(i)} - x_i)(X_{t+h}^{(j)} - x_j) | X_t = x]\,; \quad 1 \le i,j \le d. \tag{6.9}$$

Here is the fundamental existence and uniqueness result for the equation (6.2).

6.4 Theorem: *Suppose that the coefficients of the equation (6.2) satisfy the Lipschitz and linear growth conditions*

$$\|b(t,x) - b(t,y)\| + \|\sigma(t,x) - \sigma(t,y)\| \le K\|x - y\|, \quad \forall \;\; x, y \in \mathcal{R}^d, \tag{6.10}$$

$$\|b(t,x)\| + \|\sigma(t,x)\| \le K(1 + \|x\|), \quad \forall \;\; x \in \mathcal{R}^d, \tag{6.11}$$

for some real $K > 0$. Then there exists a unique process X that satisfies (6.2); it has continuous sample paths, is adapted to the filtration $\{\mathcal{F}_t^W\}$ of the driving Brownian motion W, is a Markov process, and its transition probability density function

$$P[X_t \in A | X_s = y] = \int_A p(t,x;s,y)dx$$

satisfies, under appropriate conditions, the backward

$$\Big(\frac{\partial}{\partial s} + \mathcal{A}_s\Big)\, p(t,x;\;\cdot,\;\cdot) = 0 \tag{6.12}$$

and forward (Fokker-Planck)

$$\Big(\frac{\partial}{\partial t} - \mathcal{A}_t^*\Big)\, p(\cdot,\;\cdot\;; s,y) = 0 \tag{6.13}$$

Kolmogorov equations. In (6.13), $\mathcal{A}_t^*$ is the adjoint of the operator of (6.7), namely

$$\mathcal{A}_t^* f(x) \triangleq \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} [a_{ij}(t,x) f(x)] - \sum_{i=1}^{d} \frac{\partial}{\partial x_i} [b_i(t,x) f(x)]. \qquad \diamond$$

The idea in the proof of the existence and uniqueness part in Theorem 6.4, is to mimic the procedure followed in ordinary differential equations, i.e., to consider the "Picard iterations"

$$X^{(0)} \equiv \eta, \quad X_t^{(k+1)} = \eta + \int_0^t b(s, X_s^{(k)})ds + \int_0^t \sigma(s, X_s^{(k)})dW_s$$

for $k = 0, 1, 2, \cdots$. The conditions (6.10), (6.11) then guarantee that the sequence of continuous processes $\{X^{(k)}\}_{k=0}^{\infty}$ converges to a continuous process X, which is the unique solution of the equation (6.2); they also imply that the sequence $\{X^{(k)}\}_{k=1}^{\infty}$ and the solution X satisfy moment growth conditions of the type

$$E\|X_t\|^{2\lambda} \leq C_{\lambda,T} \cdot \left(1 + E\|\eta\|^{2\lambda}\right) \ , \quad \forall\ 0 \leq t \leq T$$

for any real numbers $\lambda \geq 1$ and $T > 0$, where $C_{\lambda,T}$ is a positive constant depending only on λ, T and on the constant K of (6.10), (6.11).

6.5 Exercise: Let $f : [0,T] \times \mathcal{R}^d \to \mathcal{R}$ and polynomial growth condition

$$\max_{0 \leq t \leq T} |f(t,x)| + |g(x)| \leq C(1 + \|x\|^p), \quad \forall\ x \in \mathcal{R}^d \tag{6.15}$$

for some $C > 0$, $p \geq 1$, let $k : [0,T] \times \mathcal{R}^d \to [0,\infty)$ be continuous, and suppose that the *Cauchy problem*

$$\begin{aligned} \frac{\partial V}{\partial t} + \mathcal{A}_t V + f = kV, \quad &\text{in} \quad [0,T) \times \mathcal{R}^d \\ V(T,\cdot) = g, \quad &\text{in} \quad \mathcal{R}^d \end{aligned} \tag{6.16}$$

has a solution $V : [0,T] \times \mathcal{R}^d \to \mathcal{R}$ which is continuous on its domain, of class $C^{1,2}$ on $[0,T) \times \mathcal{R}^d$, and satisfies a growth condition of the type (6.15) (cf. Friedman (1975), Chapter 6 for sufficient conditions).

Show that the function V admits then the *Feynman-Kac representation*

$$V(t,x) = E\left[\int_t^T e^{-\int_t^\theta k(u,X_u)du} f(\theta, X_\theta)d\theta + g(X_T)e^{-\int_t^T k(u,X_u)du}\right] \tag{6.17}$$

for $0 \leq t \leq T$, $x \in \mathcal{R}^d$, in terms of the solution X of the stochastic integral equation

$$X_\theta = x + \int_t^\theta b(s,X_s)ds + \int_t^\theta \sigma(s,X_s)\, dW_s \ , \quad t \leq \theta \leq T. \tag{6.18}$$

We are assuming here that the conditions (6.10), (6.11) are satisfied, and are using the notation (6.7).

(*Hint*: Exploit (6.14) in conjunction with the growth conditions (6.15), to show that the local martingale M^V of Exercise (6.2) is actually a martingale.)

6.6 Important Remark: For the equation

$$X_t = \xi + \int_o^t b(s,X_s)ds + W_t \ , \quad 0 \leq t \leq T \tag{6.19}$$

of the form (6.2) with $\sigma = I_d$, the Girsanov Theorem 5.5 provides a solution for drift functions $b(t,x) : [0,T] \times \mathcal{R}^d \to \mathcal{R}^d$ which are only *bounded* and *measurable*.

Indeed, start by considering a Brownian motion B, and an independent random variable ξ with distribution F, on a probability space $(\Omega, \mathcal{F}, P_0)$. Define $X_t \triangleq \xi + B_t$, recall the exponential martingale

$$Z_t = \exp\left\{ \int_0^t b(s, \xi + B_s)\, dB_s - \frac{1}{2} \int_0^t \|b(s, \xi + B_s)\|^2 \, ds \right\} , \quad 0 \le t \le T ,$$

and define the probability measure $P(d\omega) \triangleq Z_T(\omega) P_0(d\omega)$. According to Theorem 5.5, the process

$$W_t \triangleq B_t - \int_0^t b(s, \xi + B_s)\, ds = X_t - \xi - \int_0^t b(s, X_s)\, ds , \quad 0 \le t \le T$$

is Brownian motion on $[0,T]$ under P, and obviously the equation (6.19) is satisfied. We also have for any $0 = t_0 \le t_1 \le ... \le t_n \le t$ and any function $f : \mathcal{R}^n \to [0,\infty)$:

$$Ef(X_{t_1}, ..., X_{t_n}) = E[f(\xi + B_{t_1}, ..., \xi + B_{t_n})] \tag{6.20}$$

$$= E_0 \Big[f(\xi + B_{t_1}, \cdots, \xi + B_{t_n}) \cdot \exp\Big\{ \int_0^t b(s, \xi + B_s)\, dB_s - \frac{1}{2} \int_0^t \|b(s, \xi + B_s)\|^2 ds \Big\} \Big]$$

$$= \int_{\mathcal{R}^d} E_0 \Big[f(x + B_{t_1}, \cdots, x + B_{t_n}) \cdot \exp\Big\{ \int_0^t b(s, x + B_s)\, dB_s - \frac{1}{2} \int_0^t \|b(s, x + B_s)\|^2 ds \Big\} \Big] F(dx).$$

In particular, the transition probabilities $p_t(x; z)$ are given as
(6.21)

$$p_t(x;z) dz = E_0 \left[1_{\{x + B_t \in dz\}} \cdot \exp\left\{ \int_0^t b(s, x + B_s) dB_s - \frac{1}{2} \int_0^t ||b(s, x + B_s)||^2 ds \right\} \right] .$$

6.7 Remark: Our ability to compute these transition probabilities hinges on carrying out the function-space integration in (6.21), not an easy task. In the one-dimensional case with drift $b(\cdot) \in C^1(\mathcal{R})$, we get

$$p_t(x;z)\, dz = \exp\{G(z) - G(x)\} \cdot E_0 \left[1_{\{x + B_t \in dz\}} \exp\left\{ -\frac{1}{2} \int_0^t V(x + B_s) ds \right\} \right] , \tag{6.22}$$

from (6.21), where $V = b' + b^2$ and $G(x) = \int_0^x b(u) du$. In certain special cases, the Feynman-Kac formula of (6.17) can help carry out the computation.

7. FILTERING THEORY

Let us place ourselves now on a probability space $(\Omega, \mathcal{F}, P)$, together with a filtration $\{\mathcal{F}_t\}$ with respect to which all our processes will be adapted. In particular, we shall consider two processes of interest:

(i) a *signal* process $X = \{X_t\,;\,0 \le t \le T\}$, which is not directly observable, and

(ii) an *observation* process $Y = \{Y_t\,;\,0 \le t \le T\}$, whose value is available to us at any time and which is suitably correlated with X (so that, by observing Y, we can say something about the distribution of X).

For simplicity of exposition and notation, we shall take both X and Y to be one-dimensional. The problem of Filtering can then be cast in the following terms: *to compute the conditional distribution*

$$P[\,X_t \in A\,|\,\mathcal{F}_t^Y\,], \quad 0 \le t \le T$$

of the signal X_t at time t, given the observation record up to that time. Equivalently, to compute the conditional expectations

$$\pi_t(f) \triangleq E[f(X_t)|\mathcal{F}_t^Y], \quad 0 \le t \le T \tag{7.1}$$

for a suitable class of test-functions $f : \mathcal{R} \to \mathcal{R}$.

In order to make some headway with this problem, we will have to assume a particular model for the observation and signal processes.

7.1 Observation Model: Let $W = \{W_t, \mathcal{F}_t; 0 \le t \le T\}$ be a Brownian motion and $H = \{H_t, \mathcal{F}_t; 0 \le t \le T\}$ a process with

$$E\int_0^T H_s^2 ds < \infty. \tag{7.2}$$

We shall assume that the observation process Y is of the form:

$$Y_t = \int_0^t H_s ds + W_t\ , \qquad 0 \le t \le T. \tag{7.3}$$

Remark: The typical situation is $H_t = h(X_t)$, a deterministic function $h : \mathcal{R} \to \mathcal{R}$ of the current signal value. In general, H and X will be suitably correlated with one another and with the process W.

7.2 Proposition: Introduce the notation

$$\widehat{\phi}_t \triangleq E(\phi_t|\mathcal{F}_t^Y) \tag{7.4}$$

and define the *innovations process*

$$N_t \triangleq Y_t - \int_0^t \widehat{H}_s ds, \quad \mathcal{F}_t^Y\,; \quad 0 \le t \le T. \tag{7.5}$$

This process is a Brownian motion.

Proof: From (7.3), (7.5) we have

$$N_t = \int_0^t (H_s - \widehat{H}_s) ds + W_t\,, \tag{7.6}$$

and with $s < t$:

$$E(N_t | \mathcal{F}_s^Y) - N_s = E\Big[\int_s^t (H_u - \widehat{H}_u)\, du + (W_t - W_s) \,\Big|\, \mathcal{F}_s^Y\Big] =$$

$$E\Big(\int_s^t \{E(H_u|\mathcal{F}_u^Y) - \widehat{H}_u\, du \,\Big|\, \mathcal{F}_s^Y\Big) + E\left[E\left(W_t - W_s \,\Big|\, \mathcal{F}_s\right) \,\Big|\, \mathcal{F}_s^Y\right] = 0$$

by well-known properties of conditional expectations. Therefore, N is a martingale with continuous paths and quadratic variation $\langle N \rangle_t = \langle W \rangle_t = t$, because the absolutely continuous part in (7.6) does not contribute to the quadratic variation. According to Theorem 4.3, N is thus a Brownian motion. $\diamondsuit$

7.3 Discussion: Since N is adapted to $\{\mathcal{F}_t^Y\}$, we have $\{\mathcal{F}_t^N\} \subseteq \{\mathcal{F}_t^Y\}$. For *linear* systems, we also have $\{\mathcal{F}_t^N\} = \{\mathcal{F}_t^Y\}$: the observations and the innovations carry the same information, because in that case there is a causal and causally invertible transformation that derives the innovations from the observations; cf. Remark 7.11. It has been a long-standing conjecture of T. Kailath, that this identity should hold in general. We know now (Allinger & Mitter (1981)) that this is indeed the case if H and W are independent, and that the identity $\{\mathcal{F}_t^N\} = \{\mathcal{F}_t^Y\}$ does *not* hold in general. However, the following positive – and extremely useful – result holds.

7.4 Theorem: *Every local martingale M with respect to the filtration $\{\mathcal{F}_t^Y\}$ admits a representation of the form*

$$M_t = M_0 + \int_0^t \Phi_s dN_s\,, \qquad 0 \le t \le T \tag{7.7}$$

where Φ is measurable, adapted to $\{\mathcal{F}_t^Y\}$, and satisfies $\int_0^T \Phi_s^2 ds < \infty$ (w.p.1). If M happens to be a square integrable martingale, then Φ can be chosen so that $E \int_0^T \Phi_s^2 ds < \infty$.

Comment: The result would follow directly from Theorem 5.3, if only the "innovations conjecture" $\{\mathcal{F}_t^N\} = \{\mathcal{F}_t^Y\}$ were true in general ! Since this is not the case, we are going to perform a change of probability measure, in order to transform Y into a Brownian motion, apply Theorem 5.3 under the new probability measure, and then "invert" the change of measure to go back to the process N.

The **Proof of Theorem 7.4** will be carried out only in the case of bounded H , which allows the presentation of all the relevant ideas with a minimum of technical fuss.

Now if H is bounded, so is the process $\widehat{H}$, and therefore

$$Z_t \triangleq \exp\Big(-\int_0^t \widehat{H}_s dN_s - \frac{1}{2}\int_0^t \widehat{H}_s^2 ds\Big),\ \mathcal{F}_t^Y\ ,\quad 0 \le t \le T \tag{7.8}$$

is a martingale (Remark 5.6); according to the Girsanov Theorem 5.5, the process $Y_t = N_t - \int_0^t (-\widehat{H}_s)\, ds$ is Brownian motion under the new probability measure

$$\widetilde{P}(d\omega) = Z_t(\omega) P(d\omega).$$

Consider also the process

$$\begin{aligned} \Lambda_t \triangleq Z_t^{-1} &= \exp\Big(\int_0^t \widehat{H}_s dN_s + \frac{1}{2}\int_0^t \widehat{H}_s^2 ds\Big) \\ &= \exp\Big(\int_0^t \widehat{H}_s dY_s - \frac{1}{2}\int_0^t \widehat{H}_s^2 ds\Big)\,,\quad 0 \le t \le T, \end{aligned} \tag{7.9}$$

and notice the "likelihood ratios"

$$\frac{d\widetilde{P}}{dP}\Bigg|_{\mathcal{F}_t^Y} = Z_t\,, \qquad \frac{dP}{d\widetilde{P}}\Bigg|_{\mathcal{F}_t^Y} = \Lambda_t$$

as well as the stochastic integral equations (cf. (4.12)) satisfied by the exponential processes of (7.8) and (7.9):

$$Z_t \;=\; 1 - \int_0^t Z_s \widehat{H}_s\, dN_s, \qquad \Lambda_t = 1 + \int_0^t \Lambda_s \widehat{H}_s\, dY_s\,. \tag{7.10}$$

Because of the so-called *Bayes rule*

$$\widetilde{E}(Q|\mathcal{F}_s^Y) = \frac{E[QZ_t|\mathcal{F}_s^Y]}{Z_s}, \tag{7.11}$$

(valid for every $s < t$ and nonnegative, $\mathcal{F}_t^Y$ – measurable random variable Q), the fact that M is a martingale under P implies that ΛM is a martingale under $\widetilde{P}$:

$$\widetilde{E}[\Lambda_t M_t|\mathcal{F}_s^Y] = \frac{E[\Lambda_t M_t Z_t|\mathcal{F}_s^Y]}{Z_s} = \Lambda_s M_s\,,$$

and vice-versa. An application of Theorem 5.3 gives a representation of the form

$$\Lambda_t M_t = \int_0^t \Psi_s dY_s = \int_0^t \Psi_s (dN_s + \widehat{H}_s \, ds) \, . \tag{7.12}$$

Now from (7.12), (7.10) and the integration by parts formula (4.14), we obtain:

$$\begin{aligned} M_t &= (\Lambda M)_t Z_t = \int_0^t (\Lambda M)_s dZ_s + \int_0^t Z_s d(\Lambda M)_s - \int_0^t \Psi_s Z_s \widehat{H}_s ds \\ &= \int_0^t \Lambda_s M_s Z_s(-\widehat{H}_s) dN_s + \int_0^t Z_s \Psi_s (dN_s + \widehat{H}_s ds) - \int_0^t \Psi_s Z_s \widehat{H}_s ds \\ &= \int_0^t \Phi_s dN_s, \qquad \text{where} \quad \Phi_t = Z_t \Psi_t - M_t \widehat{H}_t \, . \qquad \diamond \end{aligned}$$

In order to proceed further we shall need even more structure, this time on the signal process X.

7.5 Signal Proces Model: We shall assume henceforth that the signal process X has the following property: for every function $f \in C_0^2(\mathcal{R})$ (twice continuously differentiable, with compact support), there exist $\{\mathcal{F}_t\}$ - adapted processes $\mathcal{G}f$, α^f with $E \int_0^T \{|\mathcal{G}f_t| + |\alpha_t^f|\} dt < \infty$, such that

$$M_t^f \triangleq f(X_t) - f(X_0) - \int_0^t (\mathcal{G}f)_s \, ds, \; \mathcal{F}_t \; ; \quad 0 \le t \le T \tag{7.13}$$

is a martingale with

$$\langle M^f, W \rangle_t = \int_0^t \alpha_s^f \, ds. \tag{7.14}$$

7.6 Discussion: Typically $(\mathcal{G}f)_t = (\mathcal{A}_t f)(t, X_t)$, where $\mathcal{A}_t$ is a second-order linear differential operator as in (6.7). Then the requirement (7.13) imposes the Markov property on the signal process X (the famous "martingale problem" of Stroock & Varadhan (1969, 1979), which characterizes the Markov property in terms of martingales of the type (7.13)).

On the other hand, (7.14) is a statement about the correlation of the signal X with the "noise" W in the observation model.

7.7 Example: Let X satisfy the one-dimensional stochastic integral equation

$$X_t = X_0 + \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) dB_s \tag{7.15}$$

where B is a Brownian motion independent of X_0, and the functions $b : \mathcal{R} \to \mathcal{R}$, $\sigma : \mathcal{R} \to \mathcal{R}$ satisfy the conditions of Theorem 6.4. The second-order operator of (6.7) becomes then

$$\mathcal{A}f(x) = b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x), \tag{7.16}$$

and according to Exercise 6.2 we may take

$$(\mathcal{G}f)_t = \mathcal{A}f(X_t), \qquad \alpha_t^f = \sigma(X_t)f'(X_t) \cdot \frac{d}{dt}\langle B, W\rangle_t. \tag{7.17}$$

In particular, $\alpha^f \equiv 0$ if B and W (hence also X and W) are independent.

7.8 Theorem: *For the observation and signal process models of 7.1 and 7.5, we have for every $f \in C_0^2(\mathcal{R})$ and with $f_t \equiv f(X_t)$, in the notation of (7.4), the fundamental filtering equation:*

$$\widehat{f}_t = \widehat{f}_0 + \int_0^t \widehat{\mathcal{G}f_s}\, ds + \int_0^t \left(\widehat{f_s H_s} - \widehat{f}_s \widehat{H}_s + \widehat{\alpha_s^f} \right) dN_s\,, \quad 0 \le t \le T. \tag{7.18}$$ ◇

Let us try to discuss the significance and some of the consequences of Theorem 7.8, before giving its proof.

7.9 Example: Suppose that the signal process X satisfies the stochastic equation (7.15) with B independent of W and X_0, and

$$H_t = h(X_t)\,, \quad 0 \le t \le T, \tag{7.19}$$

where the function $h : \mathcal{R} \to \mathcal{R}$ is continuous and satisfies a linear growth condition. It is not hard then to show that (7.2) is satisfied, and that (7.18) amounts to

$$\pi_t(f) = \pi_0(f) + \int_0^t \pi_s(\mathcal{A}f)\, ds + \int_o^t \{\pi_s(fh) - \pi_s(f)\pi_s(h)\}\, dN_s \tag{7.20}$$

in the notation of (7.1) and (7.16). Furthermore, let us assume that the conditional distribution of X_t, given $\mathcal{F}_t^Y$, has a density $p_t(\cdot)$, i.e., $\pi_t(f) = \int_{\mathcal{R}} f(x)p_t(x)dx$. Then (7.20) leads, via integration by parts, to the *stochastic partial differential equation*

$$dp_t(x) = \mathcal{A}^* p_t(x)\, dt + p_t(x) \left\{ h(x) - \int_{\mathcal{R}} h(y)p_t(y)\, dy \right\} dN_t\,. \tag{7.21}$$

Notice that, if $h \equiv 0$ (i.e., if the observations consist of pure independent white noise), (7.21) reduces to the Fokker-Planck equation (6.13).

You should not fail to notice that (7.21) is a formidable equation, since it has all of the following features:

(i) it is a second-order partial differential equation,

(ii) it is *nonlinear*,

(iii) it contains the *nonlocal* (functional) term $\int_{\mathcal{R}} h(x)p_t(x)dx$, and

(iv) it is *stochastic*, in that it is driven by the Brownian motion N.

In the next section we shall outline an ingenuous methodology that removes, gradually, the "undesirable" features (ii)-(iv).

Let us turn now to the proof of Theorem 7.8, which will require the following auxiliary result.

7.10 Exercise: Consider two $\{\mathcal{F}_t\}$ - adapted processes V, C with $E|V_t| < \infty$, $\forall\ 0 \le t \le T$ and $E\int_0^T |C_t|dt < \infty$. If $V_t - \int_0^t C_s ds$ is an $\{\mathcal{F}_t\}$ - martingale, then

$$\widehat{V}_t - \int_0^t \widehat{C}_s ds \qquad \text{is an } \{\mathcal{F}_t^Y\} - \text{martingale.}$$

Proof of Theorem 7.8: Recall from (7.13) that

$$f_t = f_0 + \int_0^t \mathcal{G}f_s\, ds + M_t^f, \tag{7.22}$$

where M^f is an $\{\mathcal{F}_t\}$ - local martingale; thus, in conjunction with Exercise 7.10 and Theorem 7.4, we have

$$\widehat{f}_t - \widehat{f}_0 - \int_0^t \widehat{\mathcal{G}f_s}\, ds = (\{\mathcal{F}_t^Y\} - \text{local martingale}) = \int_0^t \Phi_s\, dN_s \tag{7.23}$$

for a suitable $\{\mathcal{F}_t^Y\}$ – adapted process Φ with $\int_0^t \Phi_s^2 ds < \infty$ (w.p.1). The whole point is to compute Φ "explicitly", namely, to show that

$$\Phi_t = \widehat{f_t H_t} - \widehat{f}_t \widehat{H}_t + \widehat{\alpha_t^f}. \tag{7.24}$$

This will be accomplished by computing $E[f_t Y_t|\mathcal{F}_t^Y] = Y_t \widehat{f}_t$ in *two* ways, and then comparing the results.

On the one hand, we have from (7.22), (7.3) and the integration-by-parts formula (4.14) that

$$\begin{aligned} f_t Y_t &= \int_0^t f_s(H_s ds + dW_s) + \int_0^t Y_s(\mathcal{G}f_s.ds + dM_s^f) + \int_0^t \alpha_s^f ds \\ &= \int_0^t \left(f_s H_s + Y_s.\mathcal{G}f_s + \alpha_s^f\right) ds \ + \ (\{\mathcal{F}_t\} - \text{local martingale}), \end{aligned}$$

whence from Exercise 7.10:

$$\widehat{f_t Y_t} = \widehat{f}_t\, Y_t = \int_0^t \{\widehat{f_s H_s} + Y_s \cdot \widehat{\mathcal{G} f_s} + \widehat{\alpha_s^f}\}\, ds \;+\; (\{\mathcal{F}_t^Y\} - \text{local martingale}). \tag{7.25}$$

On the other hand, from (7.23), (7.5) and the integration-by-parts formula (9.14), we obtain,

$$\begin{aligned} \widehat{f}_t Y_t &= \int_0^t \widehat{f}_s (dN_s + \widehat{H}_s ds) + \int_0^t Y_s \left(\widehat{\mathcal{G} f_s} ds + \Phi dN_s\right) + \int_0^t \Phi_s ds \\ &= \int_0^t \left(\widehat{f}_s \widehat{H}_s + Y_s . \widehat{\mathcal{G} f_s} + \Phi_s\right) ds \;+\; (\{\mathcal{F}_t\} - \text{local martingale}). \end{aligned} \tag{7.26}$$

Comparing (7.25) with (7.26), and recalling that a continuous martingale of bounded variation is constant (Corollary 2.8), we conclude that (7.24) holds.

7.9 Example (Cont'd): With $h(x) = cx$ (linear observations) and $f(x) = x^k$; $k = 1, 2, \cdots$ we obtain from (7.20):

$$\widehat{X}_t = \widehat{X}_0 + \int_0^t \widehat{b(X_s)} ds + c \int_0^t \left\{ \widehat{X_s^2} - (\widehat{X}_s)^2 \right\} dN_s\,, \tag{7.27}$$

$$\begin{aligned} \widehat{X_t^k} = \widehat{X_0^k} &+ k \int_0^t \left\{ \frac{k-1}{2} \widehat{\sigma^2(X_s) X_s^{k-2}} + \widehat{b(X_s) X_s^{k-1}} \right\} ds \\ &+ c \int_0^t \left\{ \widehat{X_s^{k+1}} - \widehat{X}_s \widehat{X_s^k} \right\} dN_s\,; \quad k = 2, 3, \cdots . \end{aligned} \tag{7.28}$$

The equations (7.27), (7.28) convey the basic difficulty of nonlinear filtering: in order to solve the equation for the k^{th} conditional moment, one needs to know the $(k+1)^{st}$ conditional moment (as well as $\pi(f)$ for $f(x) = x^{k-1} b(x)$, $f(x) = x^{k-2} \sigma^2(x)$, etcetera). In other words, the computation of conditional moments cannot be done by induction (on k) and the problem is inherently *infinite dimensional*, except in the linear case !

7.11 The Linear Case, when $b(x) = ax$ and $\sigma(x) \equiv 1$:

$$\begin{aligned} dX_t &= aX_t dt + dB_t, \quad X_0 \sim \mathcal{N}(\mu, v) \\ dY_t &= cX_t dt + dW_t, \quad Y_0 = 0 \end{aligned} \tag{7.29}$$

with X_0 independent of the two-dimensional Brownian motion (B, W). As in Example 6.1, the $\mathcal{R}^2$−valued process (X, Y) is Gaussian, and thus the conditional distribution of X_t given $\{\mathcal{F}_t^Y\}$ is *normal*, with mean $\widehat{X}_t$ and variance

$$V_t \;=\; E\left[\left(X_t - \widehat{X}_t \right)^2 \,\middle|\, \mathcal{F}_t \right] \;=\; \widehat{X_t^2} - (\widehat{X}_t)^2. \tag{7.30}$$

The problem then becomes, to find an algorithm (preferably recursive) for computing the *sufficient statistics* $\hat{X}_t$, V_t from their initial values $\widehat{X}_0 = \mu$, $V_0 = v$.

From (7.27), (7.28) with $k = 2$ we obtain

$$d\widehat{X}_t = a\widehat{X}_t dt + cV_t dN_t, \tag{7.31}$$

$$d\widehat{X_t^2} = \left(1 + 2a\widehat{X_t^2}\right) dt + c\left(\widehat{X_t^3} - \widehat{X}_t\widehat{X_t^2}\right) dN_t. \tag{7.32}$$

But now, if $Z \sim \mathcal{N}(\mu, \sigma^2)$, *we have for the third moment*:

$$EZ^3 = \mu(\mu^2 + 3\sigma^2),$$

whence $\widehat{X_t^3} = \widehat{X}_t[(\widehat{X}_t)^2 + 3V_t]$ and:

$$\widehat{X_t^3} - \widehat{X}_t\widehat{X_t^2} = \widehat{X}_t[(\widehat{X}_t)^2 + 3V_t - \widehat{X_t^2}] = 2V_t\widehat{X}_t.$$

From this last equation, (7.31), (7.32) and the chain rule (4.2), we obtain

$$dV_t = d\left(\widehat{X_t^2} - (\widehat{X}_t)^2\right) = \left(1 + 2a\widehat{X_t^2}\right)dt + 2cV_t\widehat{X}_t dN_t - c^2V_t^2 dt$$
$$- 2\widehat{X_t}[a\widehat{X}_t dt + cV_t dN_t],$$

which leads to the (nonstochastic) *Riccati equation*

$$\dot{V}_t = 1 + 2aV_t - c^2V_t^2, \quad V_0 = v. \tag{7.33}$$

In other words, the conditional variance V_t is a *deterministic function* of t, and is given by the solution of (7.33); thus there is really only one sufficient statistic, the conditional mean, and it satisfies the linear equation

$$\begin{aligned} d\widehat{X}_t &= a\widehat{X}_t dt + c\,V_t\,dN_t \\ &= (a - c^2V_t)\widehat{X}_t dt + cV_t dY_t, \quad \widehat{X}_0 = \mu. \end{aligned} \tag{7.31$'$}$$

The equation (7.31)$'$ provides the celebrated **Kalman-Bucy filter**.

In this particular (one-dimensional) case, the Riccati equation can be solved explicitly; if $a > 0$, $-\beta$ are the roots of $-cx^2 + 2ax + 1$, and $\lambda = c^2(\alpha + \beta)$, $\gamma = (v + \beta)/(\alpha - v)$, then

$$V_t \equiv \frac{\alpha\gamma e^{\lambda t} - \beta}{\gamma e^{\lambda t} - 1} \underset{t\uparrow\infty}{\longrightarrow} \alpha.$$

Everything goes through in a similar way for the multidimensional version of the Kalman-Bucy filter, in a signal/observation model of the type

$$dX_t = [A(t)X_t + a(t)]dt + b(t)dB_t, \quad X_0 \sim \mathcal{N}(\mu, v)$$
$$dY_t = H(t)X_t dt + dW_t, \quad Y_0 = 0$$

and $\langle W^{(i)}, B^{(j)}\rangle_t = \int_0^t \alpha_{ij}(s)\, ds$, for suitable deterministic matrix-valued functions $A(\cdot)$, $H(\cdot)$, $b(\cdot)$ and vector-valued function $a(\cdot)$. The joint law of the pair (X_t, Y_t) is multivariate normal, and thus the conditional distribution of X_t given $\mathcal{F}_t^Y$ is again multivariate normal, with *mean vector* $\hat{X}_t$ and non-random *variance - covariance matrix* $V(t)$. In the special case $\alpha(\cdot) = a(\cdot) = 0$, $V(\cdot)$ satisfies the matrix Riccati equation

$$\dot{V}(t) = A(t)V(t) + V(t)A^T(t) - V(t)H^T(t)H(t)V(t) + b(t)b^T(t), \qquad V(0) = v$$

(which, unlike its scalar counterpart (7.33), does not admit in general an exacit solution), and $\hat{X}$ is then obtained as the solution of the Kalman-Bucy filter equation

$$d\widehat{X}_t = A(t)\widehat{X}_t dt + V(t)H^T(t)[dY_t - H(t)\widehat{X}_t dt], \quad \widehat{X}_0 = \mu.$$

7.12 Remark: It is not hard to see that *the "innovations conjecture"* $\{\mathcal{F}_t^N\} = \{\mathcal{F}_t^Y\}$ *holds for linear systems.*

Indeed, it follows from Theorem 6.4 that the solution $\widehat{X}$ of the equation (7.31) is adapted to the filtration $\{\mathcal{F}_t^N\}$ of the driving Brownian motion N, i.e., $\{\mathcal{F}_t^{\widehat{X}}\} \subseteq \{\mathcal{F}_t^N\}$. From $Y_t = N_t - c\int_0^t \widehat{X}_s ds$ it develops that Y is adapted to $\{\mathcal{F}_t^N\}$, i.e., $\{\mathcal{F}_t^Y\} \subseteq \{\mathcal{F}_t^N\}$. Because the reverse inclusion holds anyway, the two filtrations are the same.

7.13 Remark: For the signal and observation model

$$\begin{aligned} X_t &= \xi + \int_0^t b(X_s)ds + W_t \\ Y_t &= \int_0^t h(X_s)ds + B_t \end{aligned} \tag{7.32}$$

with $b \in C^1(\mathcal{R})$ and $h \in C^2(\mathcal{R})$, which is a special case of Example 7.9, we have from Remark 6.6 and the Bayes rule:

$$\pi_t(f) = E[f(X_t)|\mathcal{F}_t^Y] = \frac{E_0[f(\xi + w_t)\Theta_t|\mathcal{F}_t^Y]}{E_0[\Theta_t|\mathcal{F}_t^Y]}. \tag{7.33}$$

Here

$$\begin{aligned} \Theta_t &\triangleq \exp\left\{\int_0^t b(\xi + w_s)dw_s + \int_0^t h(\xi + w_s)\, dY_s - \frac{1}{2}\int_0^t [b^2(\xi + w_s) + h^2(\xi + w_s)]ds\right\} \\ &= \exp\Big\{G(\xi + w_t) - G(\xi) + Y_t\ h(\xi + w_t) - \int_0^t Y_s h'(\xi + w_s)\, dw_s \\ &\quad - \frac{1}{2}\int_0^t (b' + b^2 + h^2 + Y_s \cdot h'')(\xi + w_s)ds\Big\}, \end{aligned}$$

$P_0(d\omega) = \Theta_T^{-1}(\omega)P(d\omega)$, $G(x) = \int_0^x b(u)\, du$. Here (w, Y) is, under P_0, an $\mathcal{R}^2$-valued Brownian motion, independent of the random variable ξ.

For every continuous $f : \mathcal{R}^d \to [0,\infty)$ and $y : [0,T] \to \mathcal{R}$, let us define the quantity

(7.34)

$$\rho_t(f;y) \triangleq E_0\Big[f(\xi+w_t).\exp\Big\{G(\xi+w_t)-G(\xi)+y(t)h(\xi+w_t) - \int_0^t y(s)h'(\xi+w_s)dw_s - \frac{1}{2}\int_0^t \big(b'+b^2+h^2+y(s)h''\big)(\xi+w_s)ds\Big\}\Big]$$

$$= \int_{\mathcal{R}^d} E_0\Big[f(x+w_t).\exp\Big\{G(x+w_t)-G(x)+y(t)h(x+w_t) - \int_0^t y(s)h'(x+w_s)dw_s - \frac{1}{2}\int_0^t \big(b'+b^2+h^2+y(s)h''\big)(x+w_s)ds\Big\}\Big]F(dx),$$

where F is the distribution of the random variable ξ. Then (7.33) takes the form

$$\pi_t(f) = \frac{\rho_t(f;y)}{\rho_t(1;y)}\bigg|_{y=Y(\omega)}. \tag{7.35}$$

The formula (7.34) simplifies considerably if $h(\cdot)$ is linear, say $h(x)=x$; then

(7.36)

$$\rho_t(f;y) = \int_{\mathcal{R}^d} E_0\Big[f(x+w_t).\exp\Big\{G(x+w_t)-G(x)+y(t)(x+w_t) - \int_0^t y(s)dw_s - \frac{1}{2}\int_0^t V(x+w_s)ds\Big\}\Big]\,F(dx),$$

where

$$V(x) \triangleq b'(x)+b^2(x)+x^2. \tag{7.37}$$

Whenever this potential is quadratic, i.e.,

$$b'(x)+b^2(x) = \alpha x^2+\beta x+\gamma; \qquad \alpha > -1, \tag{7.38}$$

then the famous result of Beneš (1981) shows that the integration in (7.36) can be carried out explicitly, and leads in (7.35) to a distribution with a finite number of sufficient statistics; these latter obey recursive schemes (filters).

Notice that (7.38) is satisfied by linear functions $b(\cdot)$, but also by genuinely nonlinear ones like $b(x) = \tanh(x)$.

8. ROBUST FILTERING

In this section we shall place ourselves in the context of Exanple 7.9 (in particular, of the filtering model consisting of (7.3), (7.15), (7.19) and $\langle B, W \rangle = 0$), and shall try to simplify in several regards the equations (7.20), (7.21) for the conditional distribution of X_t, given $\mathcal{F}_t^Y = \sigma\{Y_s;\ 0 \le s \le t\}$.

We start by recalling the probability measure $\widetilde{P}$ in the proof of Theorem 7.4, and the notation introduced there. From the Bayes rule (7.11) (with the rôles of P and $\tilde{P}$ interchanged, and with Λ playing the rôle of Z):

$$\pi_t(f) = E[f(X_t)|\mathcal{F}_t^Y] = \frac{\widetilde{E}[f(X_t)\Lambda_t|\mathcal{F}_t^Y]}{\Lambda_t} = \frac{\sigma_t(f)}{\sigma_t(1)} \tag{8.1}$$

where

$$\sigma_t(f) \triangleq \widetilde{E}[f(X_t)\Lambda_t|\mathcal{F}_t^Y]. \tag{8.2}$$

In other words, $\sigma_t(f)$ is an *unnormalized conditional expectation* of $f(X_t)$, given $\mathcal{F}_t^Y$.

What is the stochastic equation satisfied by $\sigma_t(f)$? From (8.1), we have

$$\sigma_t(f) = \Lambda_t \pi_t(f),$$

and from (7.10), (7.20):

$$d\Lambda_t = \pi_t(h)\Lambda_t\, dY_t\,,$$

$$d\pi_t(f) = \pi_t(\mathcal{A}f)dt + \{\pi_t(fh) - \pi_t(f)\pi_t(h)\}(dY_t - \pi_t(h)dt).$$

Now an application of the integration by parts formula (4.10) leads easily to

$$\sigma_t(f) = \sigma_0(f) + \int_0^t \sigma_s(\mathcal{A}f)ds + \int_0^t \sigma_s(fh)\, dY_s. \tag{8.3}$$

Again, if this unnormalized conditional distribution has a density $q_t(\cdot)$, i.e.

$$\sigma_t(f) = \int_{\mathcal{R}} f(x)q_t(x)dx$$

and $p_t(x) = q_t(x)/\int_{\mathcal{R}} q_t(x)dx$, then (8.3) leads, at least formally, to the equation

$$dq_t(x) = \mathcal{A}^* q_t(x)dt + h(x)q_t(x)dY_t \tag{8.4}$$

which is still a stochastic, second-order partial differential equation (PDE) of parabolic type, but *without* the drawbacks of nonlinearity and nonlocality.

To make matters even more impressive, (8.4) can be written equivalently as a *non-stochastic* second-order partial differential equation of the parabolic type, with the randomness (the observation $Y_t(\omega)$ at time t) appearing only *parametrically in the coëfficients.* We shall call this a ROBUST (or pathwise) form of the filtering equation, and will outline the clever method of B.L. Rozovskii, that leads to it.

The idea is to introduce the function

$$z_t(x) \triangleq q_t(x) \cdot \exp\{-h(x)Y_t\}\,, \qquad 0 \le t \le T, \quad x \in \mathcal{R}. \tag{8.5}$$

Because $\lambda_t(x) = \exp\{-h(x)Y_t\}$ satisfies

$$d\lambda_t(x) \;=\; \lambda_t(x)\,\left[-h(x)\,dY_t + \frac{1}{2}h^2(x)dt\right],$$

the integration-by-parts formula (4.10) leads, in conjunction with (8.4), to the nonstochastic equation

$$\frac{\partial}{\partial t} z_t(x) = \lambda_t(x)\,\mathcal{A}^*\big(z_t(x)/\lambda_t(x)\big) - \frac{1}{2}h^2(x)z_t(x). \tag{8.6}$$

In our case, we have

$$\mathcal{A}^* f(x) \;=\; \frac{1}{2}\frac{\partial}{\partial x}\left(\sigma^2(x)\frac{\partial f(x)}{\partial x}\right) - \frac{\partial}{\partial x}[b(x)f(x)],$$

the equation (8.6) leads – after a bit of algebra – to

$$\frac{\partial}{\partial t} z_t(x) \;=\; \frac{1}{2}\sigma^2(x)\,\frac{\partial^2}{\partial x^2} z_t(x) + B(t,x,Y_t)\,\frac{\partial}{\partial x} z_t(x) + C(t,x,Y_t)\,z_t(x), \tag{8.7}$$

where $B(t,x,y) = y\sigma^2(x)h'(x) + \sigma(x)\sigma'(x) - b(x)$,

$$C(t,x,y) = \frac{1}{2}\sigma^2(x)[h''(x)y + (h'(x)y)^2] + yh'(x)[\sigma(x)\sigma'(x) - b(x)] - \left(b'(x) + \frac{1}{2}h^2(x)\right).$$

The equation (8.7) is of the form that was promised: a *linear* second-order partial differential equation of parabolic type, with the randomness $Y_t(\omega)$ appearing only in the drift and potential terms. Obviously this has significant implications, of both theoretical and computational nature.

8.1 Example: Assume now the same observation model, but let X be a continuous-time Markov chain with finite state-space S and given Q-matrix. With the notation

$$p_t(x) = P\left[\, X_t = x \,|\, \mathcal{F}_t^Y \,\right]\,, \qquad x \in S$$

we have the analogue of equation (7.21):

$$dp_t(x) = (Q^* q_t)(x)\,dt + p_t(x)\Big[h(x) - \sum_{\xi\in S} h(\xi)p_t(\xi)\Big]\,dN_t, \tag{8.8}$$

with $(Q^*f)(x) \stackrel{\triangle}{=} \sum_{y\in S} q_{yx} f(y)$. On the other hand, the analogue of (8.4) for the unnormalized probability mass function $q_t(x) = \widetilde{E}\left[1_{\{X_t=x\}}\,\Lambda_t \,|\, \mathcal{F}_t^Y\right]$ is

$$dq_t(x) = (Q^* q_t)(x)\,dt + h(x)q_t(x)\,dY_t\,, \tag{8.9}$$

and $z_t(x) \stackrel{\triangle}{=} q_t(x)\exp\{-h(x)Y_t\}$ satisfies again the analogue of equation (8.6), namely:

$$\begin{aligned}\frac{\partial}{\partial t} z_t(x) \;&=\; e^{-h(x)Y_t}\;Q^*\left[z_t(\cdot)e^{h(\cdot)Y_t}\right](x) \;-\; \frac{1}{2}h^2(x)z_t(x)\\ &=\; \sum_{y\in S} q_{yx} z_t(y) e^{[h(y)-h(x)]Y_t} \;-\; \frac{1}{2}h^2(x)z_t(x)\,, \quad x\in S\,.\end{aligned} \tag{8.10}$$

This equation (a nonstochastic ordinary differential equation, with the randomness $Y(\omega)$ appearing parametrically in the coëfficients) is widely used – for instance, in real-time speech and pattern recognition.

9. STOCHASTIC CONTROL

Let us consider the following stochastic integral equation

$$X_\theta \;=\; x + \int_t^\theta b(s,X_s,U_s)ds + \int_t^\theta \sigma(s,X_s,U_s)dW_s\;, \qquad t\le\theta\le T. \tag{9.1}$$

This is a "controlled" version of the equation (6.18), the process U being the element of control. More precisely, let us suppose throughout this section that the real-valued functions $b = \{b_i\}_{1\le i\le d}$, $\sigma = \{\sigma_{ij}\}_{\substack{1\le i\le d\\ 1\le j\le n}}$ are defined on $[0,T]\times\mathcal{R}^d\times A$ (where the *control space* A is a compact subset of some Euclidean space) and are bounded, continuous, with bounded and continuous derivatives of first and second order in the argument x.

9.1 Definition: An *admissible system* $\mathcal{U}$ consists of

(i) a probability space $(\Omega,\mathcal{F},P)$, $\{\mathcal{F}_t\}$, and on it

(ii) an adapted, $\mathcal{R}^n$ – valued Brownian motion W and

(iii) a measurable, adapted, A - valued process U (the *control process*).

Thanks to our conditions on the coëfficients b and σ, the equation (9.1) has a unique (adapted, continuous) solution X for every admissible system $\mathcal{U}$. We shall call occasionally X the "state process" corresponding to this system.

Now consider two other bounded and continuous functions, namely $f : [0,T] \times \mathcal{R}^d \times A \to \mathcal{R}$ which plays the rôle of a *running cost* on both the state and the control, and $g : \mathcal{R}^d \to \mathcal{R}$ which is a *terminal cost* on the state. We assume that both f, g are of class C^2 in the spatial argument. Thus, corresponding to every admissible system $\mathcal{U}$, we have an associated expected total cost

$$J(t,x;\mathcal{U}) \triangleq E\Big[\int_t^T f(\theta, X_\theta, U_\theta)d\theta + g(X_T)\Big] . \tag{9.2}$$

The control problem is to minimize this expected cost over all admissible systems $\mathcal{U}$, to study the *value function*

$$Q(t,x) \triangleq \inf_{\mathcal{U}} J(t,x;\ \mathcal{U}) \tag{9.3}$$

(which can be shown to be measurable), and to find ε - optimal admissible systems (or even optimal ones, whenever these exist).

9.2 Definition: An admissible system is called

(i) $\varepsilon-$*optimal* for some given $\varepsilon > 0$, if

$$Q(t,x) \le J(t,x;\ \mathcal{U}) \le Q(t,x) + \varepsilon\ ; \tag{9.4}$$

(ii) *optimal*, if (9.4) holds with $\varepsilon = 0$.

9.3 Definition: A *feedback control law* is a measurable function $\alpha : [0,T] \times \mathcal{R}^d \to A$, for which the stochastic integral equation with coëfficients $b \circ \alpha$ and $\sigma \circ \alpha$, namely

$$X_\theta = x + \int_t^\theta b(s, X_s, \alpha(s, X_s))ds + \int_t^\theta \sigma(s, X_s, \alpha(s, X_s))dW_s\ , \quad t \le \theta \le T$$

has a solution X on some probability space $(\Omega, \mathcal{F}, P)$, $\{\mathcal{F}_t\}$ and with respect to some Brownian motion W on this space.

Remark: This is the case, for instance, if α is continuous, or if the diffusion matrix

$$a(t,x,u) = \sigma(t,x,u)\sigma^T(t,x,u) \tag{9.5}$$

satisfies the strong nondegeneracy condition

$$\xi\ a(t,x,u)\ \xi^T \ge \delta||\xi||^2; \qquad \forall \quad (t,x,u)\ \in\ [0,T] \times \mathcal{R}^d \times A \tag{9.6}$$

for some $\delta > 0$; see Karatzas & Shreve (1987), Stroock & Varadhan (1979) or Krylov (1973). $\diamond$

Quite obviously, to every feedback control law corresponds an admissible system with $U_t \equiv \alpha(t, X_t)$; it makes then sense to talk about "ε - optimal" or "optimal" *feedback* laws. The constant control law $U_t \equiv u$, for some $u \in A$, has the associated expected cost

$$J^u(t,x) \equiv J(t,x;u) = E\Big[\int_t^T f^u(\theta, X_\theta)d\theta + g(X_T)\Big]$$

with $f(\cdot,\cdot,u) \stackrel{\triangle}{=} f^u(\cdot,\cdot)$, which satisfies the Cauchy problem

$$\begin{aligned} \frac{\partial J^u}{\partial t} + \mathcal{A}_t^u J^u + f^u = 0; \quad &\text{in} \quad [0,T) \times \mathcal{R}^d \\ J^u(T,\cdot) = g; \quad &\text{in} \quad \mathcal{R}^d \end{aligned} \tag{9.7}$$

with the notation

$$\mathcal{A}_t^u \phi = \frac{1}{2}\sum_{i=1}^d \sum_{j=1}^d a_{ij}(t,x,u)\frac{\partial^2 \phi}{\partial x_i \partial x_j} + \sum_{i=1}^d b_i(t,x,u)\frac{\partial \phi}{\partial x_i}$$

(cf. Exercise 6.5). Since Q is obtained from $J(\cdot,\cdot;\mathcal{U})$ by minimization, it is natural to ask whether Q satisfies the "minimized" version of (9.7), that is, the *HJB (Hamilton-Jacobi-Bellman)* equation:

$$\begin{aligned} \frac{\partial Q}{\partial t} + \inf_{u \epsilon A}[\mathcal{A}_t^u Q + f^u] = 0, \quad &\text{in} \quad [0,T) \times \mathcal{R}^d, \\ Q(T,\cdot) = g, \quad &\text{in} \quad \mathcal{R}^d. \end{aligned} \tag{9.8}$$

We shall see that this is indeed the case, provided that (9.8) is interpreted in a suitably *weak sense.*

9.4 Remark: Notice that (9.8) is, in general, a *strongly nonlinear* and *degenerate* second-order equation. With the notation $D\phi = \{\partial\phi/\partial x_i\}_{1 \le i \le d}$ for the gradient, and $D^2\phi = \{\partial^2\phi/\partial x_i \partial x_j\}_{1 \le i \le d}$ for the Hessian, and with

$$F(t,x,\xi,M) = \inf_{u \epsilon A}\Big[\frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d a_{ij}(t,x,u)M_{ij} + \sum_{i=1}^d b_i(t,x,u)\xi_i + f(t,x,u)\Big] \tag{9.9}$$

($\xi \in \mathcal{R}^d$; $M \in S^d$, where S^d is the space of symmetric $(d \times d)$ matrices), the HJB equation (9.8) can be written equivalently as

$$\frac{\partial Q}{\partial t} + F(t,x,DQ,D^2Q) = 0. \tag{9.10}$$

We call this equation *strongly nonlinear*, because the nonlinearity F acts on both the gradient DQ and the higher-order derivatives D^2Q.

On the other hand, if the diffusion coëfficients do not depend on the control variable u, i.e., if we have $a_{ij}(t,x,u) \equiv a_{ij}(t,x)$, then (9.10) is transformed into the *semilinear equation*

$$\frac{\partial Q}{\partial t} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} a_{ij}(t,x)D_{ij}^2 Q + H(t,x,DQ) = 0, \tag{9.11}$$

where the nonlinearity

$$H(t,x,\xi) \triangleq \inf_{u\epsilon A}\Big[\,\xi^T b(t,x,u) + f(t,x,u)\Big] \tag{9.12}$$

acts only on the gradient DQ, and the higher-order derivatives enter linearly. For this reason, (9.11) is in principle a much easier equation to study than (9.10). $\diamond$

Let us quit the HJB equation for a while, and concentrate on the fundamental characterization of the value function Q via the so-called *Principle of Dynamic Programming* of R. Bellman (1957):

$$Q(t,x) \;=\; \inf_{\mathcal{U}}\; E\Big[\int_t^{t+h} f(\theta, X_\theta, U_\theta)d\theta + Q(t+h, X_{t+h})\Big] \tag{9.13}$$

for $0 \le h \le T-t$. This says roughly the following: Suppose that you do not know the optimal expected cost at time t, but you *do* know how well you can do at some later time $t+h$; then, in order to solve the optimization problem at time t, compute the expected cost associated with the policy of

"applying the control U during $(t, t+h)$, and behaving optimally from $t+h$ onward",

and then minimize over $\mathcal{U}$.

9.5 Theorem: Principle of Dynamic Programming.

(i) *For every stopping time σ of $\{\mathcal{F}_t\}$ with values in the interval $[t,T]$, we have*

$$Q(t,x) = \inf_{\mathcal{U}} E\Big[\int_t^{\sigma} f(\theta, X_\theta, U_\theta)d\theta + Q(\sigma, X_\sigma)\Big]. \tag{9.14}$$

(ii) *In particular, for every admissible system $\mathcal{U}$, the process*

$$M_\theta^{\mathcal{U}} = \int_t^{\theta} f(s, X_s, U_s)ds + Q(\theta, X_\theta), \qquad t \le \theta \le T \tag{9.15}$$

is a submartingale; it is a martingale if and only if $\mathcal{U}$ is optimal. $\diamond$

The technicalities of the **Proof of (9.14)** are awesome (and will not be produced here), but the basic idea is fairly simple: take an ε - optimal admissible system $\mathcal{U}$ for (t,x); it is clear that we should have

$$E\Big[\int_\sigma^T f(\theta, X_\theta, U_\theta)d\theta + g(X_T)\;\Big|\; \mathcal{F}_\sigma\Big] \ge Q(\sigma, X_\sigma), \quad w.p.\,1$$

(argue this out!), and thus from (9.4):

$$
\begin{aligned}
Q(t,x)+\varepsilon \;\geq\; J(t,x;\mathcal{U}) &= E\Big[\int_t^\sigma f(\theta,X_\theta,U_\theta)d\theta + E\Big\{\int_\sigma^T f(\theta,X_\theta,U_\theta)d\theta + g(X_T)\,\Big|\,\mathcal{F}_\sigma\Big\}\Big] \\
&\geq E\Big[\int_t^\sigma f(\theta,X_\theta,U_\theta)d\theta + Q(\sigma,X_\sigma)\Big] \\
&\geq \inf_{\mathcal{U}} E\Big[\int_t^\sigma f(\theta,X_\theta,U_\theta)d\theta + Q(\sigma,X_\sigma)\Big].
\end{aligned}
$$

Because this holds for every $\varepsilon > 0$, we are led to

$$
Q(t,x) \;\geq\; \inf_{\mathcal{U}} E\Big[\int_t^\sigma f(\theta,X_\theta,U_\theta)d\theta + Q(\sigma,X_\sigma)\Big].
$$

In order to obtain an inequality in the reverse direction, consider an arbitrary admissible system $\mathcal{U}$ and an admissible system $\mathcal{U}^{\varepsilon,\sigma}$ which is $\varepsilon-$optimal at (σ, X_σ), i.e.,

$$
E\Big[\int_\sigma^T f(\theta,X_\theta^{\varepsilon,\sigma},U_\theta^{\varepsilon,\sigma})d\theta + g(X_T^{\epsilon,\sigma})\,\Big|\,\mathcal{F}_\sigma\Big] \;\leq\; Q(\sigma,X_\sigma)+\varepsilon.
$$

Considering the "composite" control process $\widetilde{U}_\theta = \left\{ \begin{array}{ll} U_\theta & ;\ t \leq \theta \leq \sigma \\ U_\theta^{\varepsilon,\sigma} & ;\ \sigma < \theta \leq T \end{array} \right\}$ and the associated admissible system $\widetilde{\mathcal{U}}$ (there is a lot of hand-waving here, because the two systems may not be defined on the same probability space), we have

$$
\begin{aligned}
Q(t,x) \;\leq\; E\Big[\int_t^T f(\theta,\widetilde{X}_\theta,\widetilde{U}_\theta)d\theta + g(\widetilde{X}_T)\Big] \;=\; E\Big[\int_t^\sigma f(\theta,X_\theta,U_\theta)d\theta \,+ \\
+\int_\sigma^T f(\theta,X_\theta^{\varepsilon,\sigma},U_\theta^{\varepsilon,\sigma})d\theta + g(X_T^{\varepsilon,\sigma})\Big] \;\leq\; E\Big[\int_t^\sigma f(\theta,X_\theta,U_\theta)d\theta + Q(\sigma,X_\sigma)\Big] + \varepsilon.
\end{aligned}
$$

Taking the infimum on the right-hand side over $\mathcal{U}$, and noting the arbitrariness of $\varepsilon > 0$, we arrive at the desired inequality.

On the other hand, the **Proof of (i) $\Rightarrow$ (ii)** is straightforward; for an arbitrary $\mathcal{U}$, and stopping times $\tau \leq \sigma$ with values in $[t,T]$, the extension

$$
Q(\tau,X_\tau) \;\leq\; \inf_{\mathcal{U}} E\left[\int_\tau^\sigma f(\theta,X_\theta,U_\theta)d\theta + Q(\sigma,X_\sigma)\,|\,\mathcal{F}_\tau\right]
$$

of (9.14) gives $E\left(M_\tau^{\mathcal{U}}\right) \leq E\left(M_\sigma^{\mathcal{U}}\right)$, and this leads to the submartingale property.

If $M^{\mathcal{U}}$ is a martingale, then obviously

$$
Q(t,x) = E\left(M_0^{\mathcal{U}}\right) = E\left(M_T^{\mathcal{U}}\right) = E\Big[\int_t^T f(\theta,X_\theta,U_\theta)d\theta + g(X_T)\Big]
$$

whence the optimality of $\mathcal{U}$; if $\mathcal{U}$ is optimal, then $M^{\mathcal{U}}$ is a submartingale of constant expectation, thus a martingale. $\diamond$

Now let us convince ourselves that the HJB equation follows, "in principle", from the Dynamic Programming condition (9.13).

9.6. Proposition: *Suppose that the value function Q of (9.3) is of class $C^{1,2}([0,T]\times R^d)$. Then Q satisfies the HJB equation*

$$\begin{aligned} \frac{\partial Q}{\partial t} + \inf_{u \epsilon A}[A_t^u Q + f^u] = 0, \quad &\text{in} \quad [0,T)\times\mathcal{R}^d, \\ Q(T,\cdot) = g, \quad &\text{in} \quad \mathcal{R}^d. \end{aligned} \tag{9.8}$$

Proof: For such a Q we have from Itô's rule (Proposition 4.4) that

$$Q(t+h, X_{t+h}) = Q(t,x) + \int_t^{t+h} \Big(\frac{\partial Q}{\partial s} + \mathcal{A}^{U_s} Q\Big)(x, X_s)ds \; + \; \text{martingale};$$

back into (9.13), this gives

$$\inf_{\mathcal{U}} \frac{1}{h}\, E\int_t^{t+h} \Big\{ f(s, X_s, U_s) + \mathcal{A}^{U_s} Q(s, X_s) + \frac{\partial Q}{\partial s}(s, X_s)\Big\} ds \;=\; 0\,,$$

and it is not hard to derive (using the $C^{1,2}$ regularity of Q) that

$$\inf_{\mathcal{U}} \; E\,\frac{1}{h}\int_t^{t+h} \Big(\frac{\partial Q}{\partial s} + \mathcal{A}^{U_s} Q + f^{U_s}\Big)(s,x)ds \;\underset{h\downarrow 0}{\longrightarrow}\; 0. \tag{9.16}$$

Choosing $U_t \equiv u$ (a constant control), we obtain

$$\Lambda^u \;\triangleq\; \frac{\partial Q}{\partial t}(t,x) + \mathcal{A}^u Q(t,x) + f^u(t,x) \;\geq\; 0\,, \qquad \text{for every} \;\; u \;\in\; A\,,$$

whence: $\inf(C) \geq 0$ with $C \triangleq \{\Lambda^u; u\,\epsilon\, A\}$.

On the other hand, (9.16) gives $\inf(\overline{co}(C)) \leq 0$, and we conclude because $\inf(C) = \inf(\overline{co}(C))$. Here $\overline{co}(C)$ is the closed convex hull of C. $\diamond$

We give now a fundamental result in the reverse direction.

9.7 Verification Theorem: *Let the function P be bounded and continuous on $[0,T]\times\mathcal{R}^d$, of class $C_b^{1,2}$ in $[0,T)\times\mathcal{R}^d$, and satisfy the HJB equation*

$$\begin{aligned} \frac{\partial P}{\partial t} + \inf_{u \epsilon A}[\mathcal{A}^u P + f^u] = 0, \quad &\text{in} \quad [0,T)\times\mathcal{R}^d \\ P(T,\cdot) = g, \quad &\text{in} \quad \mathcal{R}^d. \end{aligned} \tag{9.17}$$

Then $P \equiv Q$.

On the other hand, let $u^*(t,x,\xi,M) : [0,T] \times \mathcal{R}^d \times \mathcal{R}^d \times S^d \to A$ *be a measurable function that achieves the infimum in (9.9), and introduce the feedback control law*

$$\alpha^*(t,x) \triangleq u^*\Big(t,x,DP(t,x),D^2P(t,x)\Big) \;:\; [0,T]\times\mathcal{R}^d \to A. \tag{9.18}$$

If this function is continuous, or if the condition (9.6) holds, then α^* *is an optimal feedback law.*

Proof: We shall discuss only the second statement (and establish the identity $P = Q$ only in its context; for the general case, cf. Safonov (1977) or Lions (1983.a)). For an arbitrary admissible system $\mathcal{U}$, we have under the assumption $P \in C_b^{1,2}([0,T)\times\mathcal{R}^d)$ by the chain rule (4.6):

$$\begin{aligned} g(X_T) - P(t,x) &= \int_t^T \Big\{\frac{\partial}{\partial t}P(\theta,X_\theta) + \sum_{i=1}^d b_i(\theta,X_\theta,U_\theta)\frac{\partial}{\partial x_i}P(\theta,X_\theta) + \\ &\quad + \frac{1}{2}\sum_{i=1}^d\sum_{j=1}^d a_{ij}(\theta,X_\theta,U_\theta)\frac{\partial^2}{\partial x_i\partial x_j}P(\theta,X_\theta)\Big\}d\theta + (M_T - M_t) \\ &\geq -\int_t^T f(\theta,X_\theta,U_\theta)d\theta + (M_T - M_t) \end{aligned} \tag{9.19}$$

thanks to (9.17), where M is a martingale; by taking expectations we arrive at $J(t,x;\mathcal{U}) \geq P(t,x)$. On the other hand, under the assumption of the second statement, there exists an admissible system $\mathcal{U}^*$ with $\mathcal{U}^*_\theta = \alpha^*(\theta,X_\theta)$ (recall the Remark following Definition 9.3). For this system, (9.19) holds as an equality and leads to $P(t,x) = J(t,x;\mathcal{U}^*)$. We conclude that $P = Q = J(\,\cdot\,;\mathcal{U}^*)$, i.e., that $\mathcal{U}^*$ is optimal.

9.8 Remark: It is not hard to extend the above results to the case where the functions f, g (as well as their first and second partial derivatives in the spatial argument) satisfy polynomial growth conditions in this argument. Then of course the value function $Q(t,x)$ also satisfies similar polynomial growth conditions, rather than being simply bounded; Proposition 9.6 and Theorem 9.7 have then to be rephrased accordingly.

9.9 Example: With $d = 1$, let $b(t,x,u) = u$, $\sigma = 1$, $f = 0$, $g(x) = x^2$ and take the control set $A = [-1,1]$. It is then intuitively obvious that the optimal law should be of the form $\alpha^*(t,x) = -\mathrm{sgn}(x)$. It can be shown that this is indeed the case, since the solution of the relevant HJB equation

$$Q_t + \frac{1}{2}Q_{xx} - |Q_x| = 0$$
$$Q(T,x) = x^2$$

can be computed explicitly as

$$Q(t,x) = \frac{1}{2} + \sqrt{\frac{t}{2\pi}}(|x|+t-1)\cdot\exp\Big\{-\frac{(|x|-t)^2}{2t}\Big\}$$
$$+\Big\{(|x|-t)^2 + t - \frac{1}{2}\Big\}.\Phi\Big(\frac{|x|-t}{\sqrt{t}}\Big)$$
$$+e^{2|x|}\Big(|x|+t-\frac{1}{2}\Big)\Big[1-\Phi\Big(\frac{|x|+t}{\sqrt{t}}\Big)\Big]$$

where $\Phi(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-x^2/2}dx$, and satisfies: $\alpha^*(t,x) = -\text{sgn}\big(Q_x(t,x)\big) = -\text{sgn}(x)$; cf. Karatzas & Shreve (1987), section 6.6.

9.10 Example: The one-dimensional linear regulator. Consider now the case with $d=1$, $A=\mathcal{R}$ and $b(t,x,u) = a(t)x+u$, $\sigma(t,x,u) = \sigma(t)$, $f(t,x,u) = c(t)x^2 + \frac{1}{2}u^2$, $g \equiv 0$, where a, σ, c are bounded and continuous functions on $[0,T]$. Certainly the assumptions of this section are violated rather grossly, but formally at least the function of (9.12) takes the form

$$H(t,x,\xi) = a(t)x\xi + c(t)x^2 + \min_{u\in\mathcal{R}}\Big(u\xi + \frac{1}{2}u^2\Big)$$
$$= a(t)x\xi + c(t)x^2 - \frac{1}{2}\xi^2,$$

the minimization is achieved by $u^*(t,x,\xi) = -\xi$, and (9.18) becomes $a^*(t,x) = u^*(t,x,Q_x(t,x)) = -Q_x(t,x)$. Here Q is the solution of the HJB (semilinear parabolic, possibly degenerate) equation

$$Q_t + \frac{1}{2}\sigma^2(t)Q_{xx} + a(t)xQ_x + c(t)x^2 - \frac{1}{2}Q_x^2 = 0 \qquad (9.20)$$
$$Q(T,\cdot) = 0.$$

It is checked quite easily that the $C^{1,2}$ function

$$Q(t,x) = A(t)x^2 + \int_t^T A(s)\sigma^2(s)ds \qquad (9.21)$$

solves the equation (9.20), provided that $A(\cdot)$ is the solution of the Riccati equation

$$\dot{A}(t) + 2a(t)A(t) - 2A^2(t) + c(t) = 0$$
$$A(T) = 0.$$

The eminently reasonable conjecture now is that the admissible system $\mathcal{U}^*$ with

$$U_t^* = -Q_x(t,X_t^*) = -2A(t)X_t^*,$$
$$dX_t^* = [a(t) - 2A(t)]X_t^*dt + \sigma(t)dW_t,$$

is optimal.

9.11 Exercise: In the context of Example 9.10, show that $J(t,x;\mathcal{U}^*) = Q(t,x) \leq J(t,x;\mathcal{U})d\theta < \infty$ holds for any admissble system for which $E\int_t^T U_\theta^2\, d\theta < \infty$.
(*Hint:* It suffices to show that

$$Q(\theta, X_\theta) + \int_t^\theta \left\{ c(s)X_s^2 + \frac{1}{2}U_s^2 \right\} ds\ , \qquad t \leq \theta \leq T$$

is a submartingale for any admissible $\mathcal{U}$ with the above property, and is a martingale for $\mathcal{U}^*$; to this end, you will need to establish that holds for every such $\mathcal{U}$.)

The trouble with the Verification Theorem 9.7 is that it assumes a lot of smoothness on the part of the value function Q. The smoothness requirement $Q \in C^{1,2}$ was satisfied in both Examples 9.9 and 9.10 – and, more generally, is satisfied by solutions of the semi-linear parabolic equation (9.11) under nondegeneracy assumptions on the diffusion matrix $a(t,x)$ and reasonable smoothness conditions on the nonlinear function $H(t,x,\xi)$; cf. Chapter VI in Fleming & Rishel (1975). But in general, this will not be the case. In fact, as one can see quite easily on deterministic examples, the value function can fail to be even once continuously differentiable in the spatial variable; on the other hand, the fully nonlinear (and possibly degenerate, since we allow for $\sigma \equiv 0$) equation (9.10) may even fail to have a solution!

All these remarks make plain the need for a new, *weak* notion of solutions for fully nonlinear, second-order equations like (9.10), that will be met by the value function $Q(t,x)$ of the control problem. Such a concept was developed by P.L. Lions (1983.a,b,c) under the rubric of "viscosity solution", following up on the work of that author with M.G. Crandall on first-order equations. We shall sketch the general outlines of this theory, but drop the term "viscosity solution" in favor of the more intelligible one "weak solution".

Thus, let us consider a continuous function

$$F(t,x,u,\xi,M)\ :\ [0,T] \times \mathcal{R}^d \times \mathcal{R} \times \mathcal{R}^d \times S^d\ \rightarrow\ \mathcal{R}$$

which satisfies the analogue

$$A \geq B\ \Rightarrow\ F(t,x,u,\xi,A) \geq F(t,x,u,\xi,B) \tag{9.22}$$

of the classical ellipticity condition (for every $t \in [0,T]$, $x \in \mathcal{R}^d$, $u \in \mathcal{R}$, $\xi \in \mathcal{R}^d$ and A, B in S^d). Plainly, (9.22) is satisfied by the function $F(t,x,\xi,A)$ of (9.9).

We would like to introduce a weak notion of solvability for the second-order equation

$$\frac{\partial u}{\partial t} + F(t,x,u,Du,D^2u) = 0 \tag{9.23}$$

that requires only *continuity* (and no differentiablility whatsoever) on the part of the solution $u(t,x)$.

9.12 Definition: A continuous function $u : [0,T] \times \mathcal{R}^d \to \mathcal{R}$ is called a

(i) *weak supersolution* of (9.23), if for every $\psi \in C^{1,2}((0,T) \times \mathcal{R}^d)$ we have

$$\frac{\partial \psi}{\partial t}(t_0,x_0) + F\Big(t_0,x_0,u(t_0,x_0),D\psi(t_0,x_0),D^2\psi(t_0,x_0)\Big) \geq 0 \tag{9.24}$$

at every local maximum point (t_0,x_0) of $u-\psi$ in $(0,T) \times \mathcal{R}^d$;

(ii) *weak subsolution* of (9.23), if for every ψ as above we have

$$\frac{\partial \psi}{\partial t}(t_0,x_0) + F\Big(t_0,x_0,u(t_0,x_0),D\psi(t_0,x_0),D^2\psi(t_0,x_0)\Big) \leq 0 \tag{9.25}$$

at every local minimum point (t_0,x_0) of $u-\psi$ in $(0,T) \times \mathcal{R}^d$;

(iii) *weak solution* of (9.23), if it is both a weak supersolution and a weak subsolution.

9.13 Remark: It can be shown that "local" extrema can be replaced by "strict local", "global" and "strict global" extrema in Definition 9.12.

9.14 Remark: *Every classical solution is also a weak solution.* Indeed, let $u \in C^{1,2}([0,T) \times \mathcal{R}^d)$ satisfy (9.23), and let (t_0,x_0) be a local maximum of $u-\psi$ in $(0,T) \times \mathcal{R}^d$; then necessarily

$$\frac{\partial u}{\partial t}(t_0,x_0) = \frac{\partial \psi}{\partial t}(t_0,x_0)\,, \quad Du(t_0,x_0) = D\psi(t_0,x_0) \quad \text{and} \quad D^2u(t_0,x_0) \leq D^2\psi(t_0,x_0)$$

so that (9.22), (9.23) lead to:

$$\begin{aligned} 0 &= \frac{\partial u}{\partial t}(t_0,x_0) + F\left(t_0,x_0,u(t_0,x_0),Du(t_0,x_0),D^2u(t_0,x_0)\right) \\ &\leq \frac{\partial \psi}{\partial t}(t_0,x_0) + F\left(t_0,x_0,u(t_0,x_0),D\psi(t_0,x_0),D^2\psi(t_0,x_0)\right). \end{aligned}$$

In other words, (9.24) is satisfied and thus u is a weak supersolution; similarly for (9.25).

The new concept relates well to the notion of weak solvability in the Sobolev sense. In particular, we have the following result (cf. Lions (1983.c):

9.15 Theorem: (i) *Let $u \in W^{1,2,p}_{loc}$ $(p > d+1)$ be a weak solution of (9.23); then*

$$\frac{\partial u}{\partial t}(t,x) \;+\; F\left(t,x,u(t,x),Du(t,x),D^2u(t,x)\right) \;=\; 0 \tag{9.26}$$

holds at a.e. point $(t,x) \in (0,T) \times \mathcal{R}^d$.

(ii) *Let* $u \in W^{1,2,p}_{loc}$ $(p > d+1)$ *satisfy (9.26) at a.e. point* $(t,x) \,\epsilon\, (0,T) \times \mathcal{R}^d$. *Then* u *is a weak solution of (9.23).* ◇

On the other hand, stability results for this new notion are almost trivial consequences of the definition.

9.16 Proposition: *Let* $\{F_n\}_{n=1}^{\infty}$ *be a sequence of continuous functions on* $[0,T] \times \mathcal{R}^{2d+1} \times S^d$*, and* $\{u_n\}_{n=1}^{\infty}$ *a sequence of corresponding weak solutions of*

$$\frac{\partial u_n}{\partial t} + F_n\big(t,x,u_n,Du_n,D^2u_n\big) = 0, \qquad \forall \quad n \geq 1.$$

Suppose that these sequences converge to the continuous functions F *and* u*, respectively, uniformly on compact subsets of their respective domains. Then* u *is a weak solution of*

$$\frac{\partial u}{\partial t} + F\big(t,x,u,Du,D^2u\big) = 0.$$

Proof: Let $\psi \in C^{1,2}([0,T) \times \mathcal{R}^d)$, and let $u - \psi$ have strict local maximum at (t_0,x_0) in $(0,T) \times \mathcal{R}^d$; recall Remark 9.13. Suppose that $\delta > 0$ is small enough, so that we have $(u-\psi)(t_0,x_0) > \max_{\partial B((t_0,x_0),\delta)}(u-\psi)(t,x)$; then for n $(= n(\delta) \to \infty$, as $\delta \downarrow 0)$ large enough, we have by continuity

$$\max_{\bar{B}}(u_n - \psi) > \max_{\partial B((t_0,x_0),\delta)}(u_n - \psi),$$

where $B \triangleq B((t_0,x_0),\delta)$. Thus, there exists a point $(t_\delta,x_\delta) \in B((t_0,x_0),\delta)$ such that

$$\max_{\bar{B}}(u_n - \psi) = (u_n - \psi)(t_\delta,x_\delta).$$

Now from

$$\frac{\partial}{\partial t}u_n(t_\delta,x_\delta) + F_n\Big(t_\delta,x_\delta,u_n(t_\delta,x_\delta),Du_n(t_\delta,x_\delta),D^2u_n\psi(t_\delta,x_\delta)\Big) \geq 0,$$

we let $\delta \downarrow 0$, to obtain (observing that $(t_\delta,x_\delta) \to (t_0,x_0)$, $u_n(t_\delta,x_\delta) \to u(t_0,x_0)$, $D^j\psi(t_\delta,x_\delta) \to D^j\psi(t_0,x_0)$ for $j = 1,2$, because $\psi \in C^{1,2}$, and recalling that F_n converges to F uniformly on compact sets):

$$\frac{\partial u}{\partial t}(t_0,x_0) + F\Big(t_0,x_0,u(t_0,x_0),D\psi(t_0,x_0),D^2\psi(t_0,x_0)\Big) \geq 0.$$

It follows that u is a weak supersolution; similarly for the weak subsolution property. ◇

Finally, here is the result that connects the concept of weak solutions with the control problem of this section.

9.17 Theorem: *If the value function Q of (9.3) is continuous on the strip* $[0,T] \times \mathcal{R}^d$, *then Q is a weak solution of*

$$\frac{\partial Q}{\partial t} + F(t, x, DQ, D^2 Q) = 0 \tag{9.27}$$

in the notation of (9.9).

Proof: Let (t_0, x_0) be a global maximum of $Q - \psi$, for some fixed test-function $\psi \in C^{1,2}((0,T) \times \mathcal{R}^d)$, and without loss of generality assume that $Q(t_0, x_0) = \psi(t_0, x_0)$. Then the Dynamic Programming condition (9.13) yields

$$\psi(t_0, x_0) = Q(t_0, x_0) = \inf_{\mathcal{U}} E\Big[\int_{t_0}^{t_0+h} f(\theta, X_\theta, U_\theta)d\theta + Q(t_0 + h, X_{t_0+h})\Big]$$
$$\leq \inf_{\mathcal{U}} E\Big[\int_{t_0}^{t_0+h} f(\theta, X_\theta, U_\theta)d\theta + \psi(t_0 + h, X_{t_0+h})\Big].$$

But now the argument used in the proof of Proposition 9.6 (applied this time to the smooth test-function $\psi \in C^{1,2}$) yields:

$$0 \leq \frac{\partial \psi}{\partial t}(t_0, x_0) + \inf_{u \epsilon A}[\mathcal{A}^u_{t_0}\psi(t_0, x_0) + f^u(t_0, x_0)] =$$
$$= \frac{\partial \psi}{\partial t}(t_0, x_0) + F\Big(t_0, x_0, Q(t_0, x_0, Q(t_0, x_0), D\psi(t_0, x_0), D^2\psi(t_0, x_0)\Big).$$

Thus Q is a weak supersolution of (9.27), and its weak subsolution property is proved similarly. ◇

We shall close this section with an example of a stochastic control problem arising in financial economics.

9.18 Consumption/Investment Optimization: Let us consider a financial market with $d+1$ assets; one of them is a risk-free asset called *bond* with interest rate r (and price $B_0(t) = e^{rt}$), and the remaining are risky *stocks*, with prices-per-share $S_i(t)$ given by

$$dS_i(t) = S_i(t)\Big[b_i dt + \sum_{j=1}^{d} \sigma_{ij} dW_j(t)\Big], \qquad 1 \leq i \leq d.$$

Here $W = (W_1, ..., W_d)^T$ is an $\mathcal{R}^d$ - valued Brownian motion which models the uncertainty in the market, $b = (b_1, ..., b_d)^T$ is the vector of *appreciation rates*, and $\sigma = \{\sigma_{ij}\}_{1 \leq i,j \leq d}$ is the *volatility matrix* for the stocks. We assume that both σ, σ^T are invertible.

It is worthwhile to notice that the discounted stock prices $\tilde{S}_i(t) = e^{-rt}S_i(t)$ satisfy the equations

$$d\widetilde{S}_i(t) = \widetilde{S}_i(t).\sum_{j=1}^{d} \sigma_{ij}\, d\widetilde{W}_j(t), \qquad 1 \le i \le d,$$

where $\widetilde{W}(t) = W(t) + \theta t$, $0 \le t \le T$ and $\theta = \sigma^{-1}(b - r1)$.

Now introduce the probability measure

$$\widetilde{P}(A) = E[Z(T)1_A] \ \text{ on } \ \mathcal{F}_T\,, \quad \text{with } \ Z(t) = \exp\left\{-\theta^T W(t) - \frac{1}{2}||\theta||^2 t\right\}\ ;$$

under $\widetilde{P}$, *the process* $\widetilde{W}$ *is Brownian motion and the* $\widetilde{S}_i$*'s are martingales on* $[0,T]$ (cf. Theorem 5.5 and Example 4.5).

Suppose now that an investor starts out with an initial capital $x > 0$, and has to decide – at every time $t \in (0,T)$ – at what rate $c(t) \ge 0$ to withdraw money for consumption and how much money $\pi_i(t), 1 \le i \le d$ to invest in each stock. The resulting *consumption* and *portfolio* processes c and $\pi = (\pi_1, ..., \pi_d)^T$, respectively, are assumed to be adapted to $\mathcal{F}_t^W = \sigma(W_s; 0 \le s \le t)$ and to satisfy

$$\int_0^T \left\{c(t) + \sum_{i=1}^{n} \pi_i^2(t)\right\} dt < \infty, \qquad \text{w.p. 1}\,.$$

Now if $X(t)$ denotes the investor's wealth at time t, the amount $X(t) - \sum_{i=1}^d \pi_i(t)$ is invested in the bond, and thus $X(\cdot)$ satisfies the equation

(9.28)

$$\begin{aligned} dX(t) &= \sum_{i=1}^{d} \pi_i(t)\Big[b_i dt + \sum_{j=1}^{d} \sigma_{ij} dW_j(t)\Big] + \Big(X(t) - \sum_{i=1}^{d} \pi_i(t)\Big) r dt - c(t)dt \\ &= [rX(t) - c(t)]dt + \sum_{i=1}^{d} \pi_i(t)\Big[(b_i - r)dt + \sum_{j=1}^{d} \sigma_{ij} dW_j(t)\Big] \\ &= [rX(t) - c(t)dt + \pi^T(t)[(b - r1)dt + \sigma dW(t)] = [rX(t) - c(t)]dt + \pi^T(t)\sigma d\widetilde{W}(t). \end{aligned}$$

In other words,

$$e^{-ru}X(u) = x - \int_0^u e^{-rs}c(s)ds + \int_0^u e^{-rs}\pi^T(s)\sigma d\widetilde{W}(s); \qquad 0 \le u \le T\,. \tag{9.29}$$

The class $\mathcal{A}(x)$ of *admissible* control process pairs (c, π) consists of those pairs for which the corresponding wealth process X of (9.29) remains nonnegative on $[0,T]$ (i.e., $X(u) \ge 0, \ \forall\ 0 \le u \le T$) w.p.1.

It is not hard to see that for every $(c, \pi) \in \mathcal{A}(x)$, we have

$$\widetilde{E} \int_0^T e^{-rt}c(t)dt \ \le\ x\,. \tag{9.30}$$

Conversely, *for every consumption process c which satisfies (9.30)*, it can be shown that *there exists a portfolio process* π *such that* $(c, \pi) \in \mathcal{A}(x)$.

(Exercise: Try to work this out ! The converse statement hinges on the fact that every $\widetilde{P}-$martingale can be represented as a stochastic integral with respect to $\widetilde{W}$, thanks to the representation Theorems 5.3 and 7.4.)

The control problem now is to maximize the expected discounted utility from consumption

$$J(x; c, \pi) \triangleq E \int_0^T e^{-\beta t} U(c(t))\, dt$$

over pairs $(c, \pi) \in \mathcal{A}(x)$, where $U : (0, \infty) \to \mathcal{R}$ is a C^1, strictly increasing and strictly concave *utility function*, with $U'(0+) = \infty$ and $U'(\infty) = 0$. We denote by $I : [0, \infty] \overset{onto}{\longrightarrow} [0, \infty]$ the inverse of the strictly decreasing function U'.

More generally, we can pose the same problem on the interval $[t, T]$ rather than on $[0, T]$, for every fixed $0 \le t \le T$, look at admissible pairs $(c, \pi) \in \mathcal{A}(t, x)$ for which the resulting wealth process $X(\cdot)$ of

$$(9.29)' \qquad e^{-ru} X(u) = x e^{-rt} - \int_t^u e^{-rs} c(s) ds + \int_t^u e^{-rs} \pi^T(s) \sigma d\widetilde{W}(s); \qquad t \le u \le T$$

is nonnegative w.p.1, and study the *value funtion*

$$(9.31) \qquad Q(t, x) \triangleq \sup_{(c,\pi) \in \mathcal{A}(t,x)} E \int_t^T e^{-\beta s} U(c(s))\, ds\,; \qquad 0 \le t \le T, \quad x \in (0, \infty)$$

of the resulting control problem. By analogy with (9.8), and in conjunction with the equation (9.28) for the wealth process, we expect this value function to satisfy the *HJB equation*

$$(9.32) \qquad \frac{\partial Q}{\partial t} + \max_{\substack{\pi \in \mathcal{R}^d \\ c \in [0,\infty)}} \Big[\frac{1}{2} ||\pi^* \sigma||^2 Q_{xx} + \{(rx - c) + \pi^T(b - r1)\} Q_x + e^{-\beta t} U(c) \Big] = 0$$

as well as the terminal and boundary conditions
(9.33)

$$Q(T, x) = 0, \quad 0 < x < \infty \quad \text{and} \quad Q(t, 0+) = \frac{e^{-\beta t}}{\beta} \Big(1 - e^{-\beta(T-t)}\Big) U(0+), \quad 0 \le t \le T.$$

Now the maximizations in (9.32) are achieved by $\hat{c} = I(e^{\beta t} Q_x)$ and $\hat{\pi} = -(\sigma^T)^{-1} \theta Q_x / Q_{xx}$, and thus the HJB equation becomes

$$(9.34) \qquad \frac{\partial Q}{\partial t} + e^{-\beta t} U\big(I(e^{\beta t} Q_x)\big) - Q_x \cdot I(e^{\beta t} Q_x) - \frac{||\theta||^2}{2} \frac{Q_x^2}{Q_{xx}} + rx Q_x = 0\,.$$

This is a *strongly nonlinear* equation, unlike the ones appearing in Examples 9.9 and 9.10. Nevertheless, it has a classical solution which, quite remarkably, can be written down in closed form for very general utility functions U; cf. Karatzas, Lehoczky & Shreve (1987), section 7, or Karatzas (1989), §9 for the details.

For instance, in the special case $U(c) = c^\delta$ with $0 < \delta < 1$, the solution of (9.34) is

$$Q(t,x) = e^{-\beta t}(p(t))^{1-\delta}x^\delta,$$

with

$$p(t) = \left\{ \begin{array}{ll} \frac{1}{k}[1-e^{-k(T-t)}] & ; \quad k \neq 0 \\ T-t & ; \quad k = 0 \end{array} \right\}, \qquad k = \frac{1}{1-\delta}\left[\beta - r\delta - \frac{\delta||\theta||^2}{2(1-\delta)}\right]$$

and the optimal consumption and portfolio rules are given by

$$\hat{c}(t,x) = \frac{x}{p(t)}, \qquad \hat{\pi}(t,x) = (\sigma^T)^{-1}\,\frac{x}{1-\delta}\,\theta\,,$$

respectively, in feedback form on the current level of wealth.

10. NOTES:

Section 1 & 2: The material here is standard; see, for instance, Karatzas & Shreve (1987), Chapter 1 (for the general theory of martingales, and the associated concepts of filtrations and stopping times) and Chapter 2 (for the construction and the fundamental properties of Brownian motion, as well as for an introduction to Markov processes).

The term "martingale" was introduced in probability theory (from gambling!) by J. Ville (1939), although the concept was invented several years earlier by P. Lévy in an attempt to extend the basic theorems of probability from independent to dependent random variables. The fundamental theory for processes of this type was developed by Doob (1953).

Sections 3 & 4: The construction and the study of stochastic integrals started with a seminal series of articles by K. Itô (1942, 1944, 1946, 1951) for Brownian motion, and continued with Kunita & Watanabe (1967) for continuous local martingales, and with Doléans-Dade & Meyer (1971) for general local martingales. This theory culminated with the course of Meyer (1976), and can be studied in the monographs by Liptser & Shiryaev(1977), Ikeda & Watanabe (1981), Elliott (1982), Karatzas & Shreve (1987), and Rogers & Williams (1987).

Theorem 4.3 was established by P. Lévy (1948), but the wonderfully simple proof that you see here is due to Kunita & Watanabe (1967). Condition (4.9) is due to Novikov(1972).

Section 5: Theorem 5.1 is due to Dambis (1965) and Dubins & Schwarz (1965), while the result of Exercise 5.2 is due to Doob (1953). For complete proofs of the results in this section, see §§ 3.4, 3.5 in Karatzas & Shreve (1987).

Section 6: The field of stochastic differential equations is now vast, both in theory and in applications. For a systematic account of the basic results, and for some applications, cf. Chapter 5 in Karatzas & Shreve (1987). More advanced and/or specialized treatments appear in Stroock & Varadhan (1979), Ikeda & Watanabe (1981), Rogers & Williams (1987), Friedman (1975/76).

The fundamental Theorem 6.4 is due to K. Itô (1946, 1951). Martingales of the type M^f of Exercise 6.2 play a central rôle in the modern theory of Markov processes, as was discovered by Stroock & Varadhan (1969, 1979). See also §5.4 in Karatzas & Shreve (1987), and the monograph by Ethier & Kurtz (1986).

For the kinematics and dynamics of random motions with general drift and diffusion coëfficients, including elaborated versions of the representations (6.8) and (6.9), see the monograph by Nelson (1967).

Section 7: A systematic account of filtering theory appears in the monograph by Kallianpur (1980), and a rich collection of interesting papers can be found in the volume edited by Hazewinkel & Willems (1981). The fundamental Theorems 7.4, 7.8 as well as the equations (7.20), are due to Fujisaki, Kallianpur & Kunita (1972); the equation (7.21) for the conditional density was discovered by Kushner (1967), whereas (7.31), (7.33) constitute the ubiquitious Kalman & Bucy (1961) filter. Proposition 7.2 is due to Kailath (1971), who also introduced the "innovations approach" to the study of filtering; see Kailath (1968), Frost & Kailath (1971).

We have followed Rogers & Williams (1987) in the derivation of the filtering equations.

Section 8: The equations (8.3), (8.4) for the unnormalized conditional density are due to Zakai (1969). For further work on the "robust" equations of the type (8.6), (8.7), (8.10) see Davis (1980, 1981, 1982), Pardoux (1979), and the articles in the volume edited by Hazewinkel & Willems (1981).

Section 9: For the general theory of stochastic control, see Fleming & Rishel (1975), Bensoussan & Lions (1978), Krylov (1980) and Bensoussan (1982). The notion of weak solutions, as in Definition 9.12, is due to P.L. Lions (1983). For a very general treatment of optimization problems arising in financial economics, see Karatzas, Lehoczky & Shreve (1987) and the survey paper by Karatzas (1989).

11. REFERENCES

ALLINGER, D. & MITTER, S.K. (1981) New results on the innovations problem for non-linear filtering. *Stochastics* **4**, 339-348.

BACELIER, A. (1900) Théorie de la Spéculation. *Ann. Sci. École Norm. Sup.* **17**, 21-86.

BELLMAN, R. (1957) *Dynamic Programming.* Princeton University Press.

BENEŠ, V.E. (1981) Exact finite-dimensional filters for certain diffusions with nonlinear drift. *Stochastics* **5**, 65-92.

BENSOUSSAN, A. (1982) *Stochastic Control by Functional Analysis Methods.* North-Holland, Amsterdam.

BENSOUSSAN, A. & LIONS, J.L. (1978) *Applications des inéquations variationnelles en contrôle stochastique.* Dunod, Paris.

DAMBIS, R. (1965) On the decomposition of continuous supermartingales. *Theory Probab. Appl.* **10**, 401-410.

DAVIS, M.H.A. (1980) On a multiplicative functional transformation arising in nonlinear filtering theory. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **54**, 125-139.

DAVIS, M.H.A. (1981) Pathwise nonlinear filtering. In *Hazewinkel & Willems (1981)*, pp. 505-528.

DAVIS, M.H.A. (1982) A pathwise solution of the equations of nonlinear filtering. *Theory Probab. Appl.* **27**, 167-175.

DOLÉANS-DADE, C. & MEYER, P.A. (1971) Intégrales stochastiques par rapport aux martingales locales. *Séminaire de Probabilités IV*, *Lecture Notes in Mathematics* **124**, 77-107.

DOOB, J.L. (1953) *Stochastic Processes.* J. Wiley & Sons, New York.

DUBINS, L. & SCHWARZ, G. (1965) On continuous martingales. *Proc. Nat'l Acad. Sci. USA* **53**, 913-916.

EINSTEIN, A. (1905) Theory of the Brownian Movement. *Ann. Physik.* **17**.

ELLIOTT, R.J. (1982) *Stochastic Calculus and Applications.* Springer-Verlag, New York.

ETHIER, S.N. & KURTZ, T.G. (1986) *Markov Processes: Characterization and Convergence.* J. Wiley & Sons, New York.

FLEMING, W.H. & RISHEL, R.W. (1975) *Deterministic and Stochastic Optimal Control.* Springer-Verlag, New York.

FRIEDMAN, A. (1975/76) *Stochastic Differential Equations and Applications* (2 volumes). Academic Press, New York.

FROST, P.A. & KAILATH, T. (1971) An innovations approach to least-squares estimation, Part III. *IEEE Trans. Autom. Control* **16**, 217-226.

FUJISAKI, M., KALLIANPUR, G. & KUNITA, H. (1972) Stochastic differential equations for the nonlinear filtering problems. *Osaka J. Math.* **9**, 19-40.

GIRSANOV, I.V. (1960) On transforming a certain class of stochastic processes by an absolutely continuous substitution of measure. *Theory Probab. Appl.* **5**, 285-301.

HAZEWINKEL, M. & WILLEMS, J.C., Editors (1981) *Stochastic Systems: The Mathematics of Filtering and Identification.* Reidel, Dordrecht.

IKEDA, N. & WATANABE, S. (1981) *Stochastic Differential Equations and Diffusion Processes.* North-Holland, Amsterdam and Kodansha Ltd., Tokyo.

ITÔ, K. (1942) Differential equations determining Markov processes (in Japanese). *Zenkoku Shijō Sūgaku Danwakai* **1077**, 1352-1400.

ITÔ, K. (1944) Stochastic integral. *Proc. Imp. Acad Tokyo* **20**, 519-524.

ITÔ, K. (1946) On a stochastic integral equation. *Proc. Imp. Acad. Tokyo* **22**, 32-35.

ITÔ, K. (1951) On stochastic differential equations. *Mem. Amer. Math. Society* **4**, 1-51.

KAILATH, T. (1968) An innovations approach to least-squares estimation. Part I: Linear filtering in additive white noise. *IEEE Trans. Autom. Control* **13**, 646-655.

KAILATH, T. (1971) Some extensions of the innovations theorem. *Bell System Techn. Journal* **50**, 1487-1494.

KALMAN, R.E. & BUCY, R.S. (1961) New results in linear filtering and prediction theory. *J. Basic Engr. ASME (Ser. D)* **83**, 85-108.

KALLIANPUR, G. (1980) *Stochastic Filtering Theory.* Springer-Verlag, New York.

KARATZAS, I. (1989) Optimization problems in the theory of continuous trading. Invited survey paper, *SIAM Journal on Control & Optimization*, to appear.

KARATZAS, I., LEHOCZKY, J.P., & SHREVE, S.E. (1987) Optimal portfolio and consumption decisions for a "small investor" on a finite horizon. *SIAM Journal on Control & Optimization* **25**, 1557-1586.

KARATZAS, I. & SHREVE, S.E. (1987) *Brownian Motion and Stochastic Calculus.* Springer-Verlag, New York.

KRYLOV, N.V. (1980) *Controlled Diffusion Processes.* Springer-Verlag, New York.

KRYLOV, N.V. (1974) Some estimates on the probability density of a stochastic integral. *Math. USSR (Izvestija)* **8**, 233-254.

KUNITA, H. & WATANABE, S. (1967) On square-integrable martingales. *Nagoya Math. Journal* **30**, 209-245.

KUSHNER, H.J. (1964) On the differential equations satisfied by conditional probability densities of Markov processes. *SIAM J. Control* **2**,106-119.

KUSHNER, H.J. (1967) Dynamical equations of optimal nonlinear filtering. *J. Diff. Equations* **3**, 179-190.

LÉVY, P. (1948) *Processus Stochastiques et Mouvement Brownien.* Gauthier-Villars, Paris.

LIONS, P.L. (1983.a) On the Hamilton Jacobi-Bellman equations. *Acta Applicandae Mathematicae* **1**, 17-41.

LIONS, P.L. (1983.b) Optimal control of diffusion processes and Hamilton-Jacobi-Bellman equations, Part I: The Dynamic Programming principle and applications. *Comm. Partial Diff. Equations* **8**, 1101-1174.

LIONS, P.L. (1983.c) Optimal control of diffusion processes and Hamilton-Jacobi-Bellman equations, Part II: Viscosity solutions and uniqueness. *Comm. Partial Diff. Equations* **8**, 1229-1276.

LIPTSER, R.S. & SHIRYAEV, A.N. (1977) *Statistics of Random Processes, Vol. I: General Theory.* Springer-Verlag, New York.

MEYER, P.A. (1976) Un cours sur les intégrales stochastiques. *Séminaire de Probabilités X, Lecture Notes in Mathematics* **511**, 245-400.

NELSON, E. (1967) *Dynamical Theories of Brownian Motion.* Princeton University Press.

NOVIKOV, A.A. (1972) On moment inequalities for stochastic integrals. *Theory Probab. Appl.* **16**, 538-541.

PARDOUX, E. (1979) Stochastic partial differential equations and filtering of diffusion processes. *Stochastics* **3**, 127-167.

ROGERS, L.C.G. & WILLIAMS, D. (1987) *Diffusions, Markov Processes, and Martingales, Vol. 2: Itô Calculus.* J. Wiley & Sons, New York.

SAFONOV, M.V. (1977) On the Dirichlet problem for Bellman's equation in a plane domain, I. *Math. USSR (Sbornik)* **31**, 231-248.

STROOCK, D.W. & VARADHAN, S.R.S. (1969) Diffusion processes with continuous coëfficients, I & II. *Comm. Pure Appl. Math.* **22**, 345-400 & 479-530.

STROOCK, D.W. & VARADHAN, S.R.S. (1979) *Multidimensional Diffusion Processes.* Springer-Verlag, Berlin.

VILLE, J. (1939) *Étude Critique de la Notion du Collectif.* Gauthier-Villars, Paris.

ZAKAI, M. (1969) On filtering for diffusion processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **11**, 230-243.