LECTURES ON NONLINEAR FILTERING AND STOCHASTIC CONTROL[1]

by

Sanjoy K. Mitter

CHAPTER 1

The Basic Equations of Nonlinear Filtering

1. Introduction

This paper is concerned with the variational and geometrical interpretation of nonlinear filtering using stochastic control and Lie-theoretic ideas. The origin of these ideas can be traced back to the work of Feynman [1965].

There are two essentially different approaches to the nonlinear filtering problem. The first is based on the important ideas of innovations processes, originally introduced by Bode and Shannon (and Kolmogoroff) in the context of Wiener Filtering problems and later developed by Kailath and his students in the late sixties for non-linear filtering problems. This approach reaches its culmination in the seminal paper of Fujisaki-Kallianpur-Kunita [1972]. A detailed account of this approach is now available in book form cf. Liptser-Shiryayev [1977] and Kallianpur [1980]. The second approach can be traced back to the doctoral dissertation of Mortensen [1966], Duncan [1967] and the important paper of Zakai [1969]. In this approach attention is focussed on the unnormalized conditional density equation, which is a bilinear stochastic partial differential equation, and it derives its inspiration from function space integration as originally introduced by Kac [1951] and Ray [1954]. Mathematically, this view is closely connected to the path integral formulation of Quantum Physics due to Feynman [1965]. For an exposition of this analogy see Mitter [1980, 1981].

The relationships between non-linear filtering and stochastic control is obtained by considering the pathwise equations of non-linear filtering and via an exponential transformation giving it a stochastic control interpretation (cf. Fleming-Mitter [1982]).

2. Formulation of the Nonlinear Filtering Problem

To simplify the exposition we shall discuss the situation where the observation y is scalar.

Let (Ω, F, P) be a complete probability space and let F_t, $t \in [0,T]$ be an increasing family of sub σ-fields of F. Consider the observation process

$$y(t) = \int_0^t h(x(s))ds + \tilde{w}(t) \tag{2.1}$$

where $\tilde{w}(t)$ is an F_t-adapted Wiener process and $x(s)$, the state process is an n-dimensional F_t-adapted process and satisfies the stochastic differential equation

$$\begin{cases} dx(t) = b(x(t))dt + \sigma(x(t))dw(t), \\ x(0) = x \end{cases} \tag{2.2}$$

where $w(t)$ is an n-dimensional Wiener process which is independent of $w(t)$.

We shall make the assumption

$$E \int_0^t |h(x_s)|^2 ds < \infty \qquad \forall\, t \in [0,T] \tag{2.3}$$

and we make the following further assumptions on b, σ and h:

$$\sigma \in C_b^\infty(\mathbb{R}^n; L(\mathbb{R}^n)), \quad b \in C_b^\infty(\mathbb{R}^n; \mathbb{R}^n), \quad h \in C^\infty(\mathbb{R}^n, \mathbb{R}) \tag{2.4}$$

Let $a(x) = \sigma(x)\sigma'(x)$; assume there exists $\lambda_1, \lambda_2 > 0$ such that $\lambda_1|\xi|^2 \le \sum_{i,j=1}^n a_{ij}\xi_i\xi_j < \lambda_2|\xi|^2$, $\forall\, \xi \in \mathbb{R}^n$ (2.5)

h is a polynomial of degree m, such that $h = h_1 + h_2$ where h_1 is a homogeneous polynomial of degree m such that $\lim_{|x|\to\infty} |h_1(x)| = \infty$ and h_2 is of degree $< m$. (2.6)

The assumptions on σ and b could be relaxed.

Let $F_t^y = \sigma\{y(s), 0 \le s \le t\}$ and let $\Pi(t,dx,\omega)$ denote the conditional distribution of $x(t)$ given F_t^y:

$$\Pi(t,dx,\omega) = P\{x(t) \in dx | F_t^y\}(\omega).$$

The problem of non-linear filtering is the characterization of $\Pi(t,dx,\omega)$.

3. The Basic Equations of Nonlinear Filtering

In this section we describe the basic equations of non-linear filtering and discuss our basic point of view towards non-linear filtering. For a derivation of these equations see the lectures of Kunita, this volume.

Let $\phi \in C_b^2(\mathbb{R}^n)$ and denote by

$$(\Pi(t), \phi) = \int_{\mathbb{R}^n} \phi(x)\Pi(t,dx,\omega) \tag{3.1}$$

Then $\Pi(t)$ satisfies the non-linear stochastic partial differential equation (Kushner-Stratonovich Equation)

$$(\Pi(t),\phi) = (\Pi(0),\phi) + \int_0^t (\Pi(s), L\phi)ds + \int_0^t [(\Pi(s), h\phi) - (\Pi(s),\phi)(\Pi(s),h)](dy(s) - (\Pi(s), h)ds) \, , \tag{3.2}$$

where L is the generator of the diffusion process given by (2.2),

$$L\phi = \frac{1}{2}\,\mathrm{tr}\, a(x)\phi_{xx} + b(x)\cdot\phi_x \, ,$$

$$\phi_x = (\phi_{x_1}, \ldots, \phi_{x_n}), \quad \mathrm{tr}(a(x))\phi_{xx} = \sum_{i,j=1}^n a_{ij}(x)\phi_{x_i x_j} \, .$$

We think of the filter as a dynamical system and we think of equation (3.2) as the input-output equation of the filter, the input being the observations y(s) and the output the conditional distribution $\Pi(t)$.

Let $\rho(t,dx,\omega)$ be a continuous stochastic process with values in the set of finite positive measures on $\mathbb{R}$. Denote by

$$(\rho(t),\phi) = \int_{\mathbb{R}^n} \phi(x)\rho(t,dx,\omega) \tag{3.3}$$

Consider the equation

$$(\rho(t),\phi) = (\rho(0),\phi) + \int_0^t (\rho(s), L\phi) + \int_0^t (\rho(s), h\phi)dy_s \tag{3.4}$$

This is the weak form of the Zakai equation.

Now it can be proved:

If $\rho(t)$ is a solution of (3.3) then

$$(\Pi(t),\phi) = \frac{(\rho(t),\phi)}{(\rho(t),1)}$$

is a solution of equation (3.2)

Moreover we have the Feynman-Kac formula

$$(\rho(t),\phi) = E[\phi(x(t))\exp(\int_0^t h(x(s))dy(s) - \frac{1}{2}\int_0^t h(x(s))^2 ds)|F_t^y] \tag{3.5}$$

For later use we shall consider the weak form of the Zakai equation in Stratonvich form:

$$(\rho(t),\phi) = (\rho(0),\phi) + \int_0^t (\rho(s), [L - \frac{1}{2}h^2]\phi)ds + \int_0^t (\rho(s), h\phi)\cdot dy(s) \tag{3.6}$$

where $\cdot$ denotes the Stratonovich differential.

If $\rho(t,dx,\omega)$ has a smooth density $\rho(t,x)$ then $\rho(t,x)$ satisfies the Zakai equation:

$$d\rho(t,x) = (L^* - \frac{1}{2}h^2)\rho(t,x)dt + h\rho(t,x)\cdot dy(t) \tag{3.7}$$

where * denotes formal adjoint.

$\rho(t,x)$ has the interpretation of an unnormalized conditional density and is to be thought of as the "state" of the filter.

To compute "conditional statistics", corresponding to $\phi \in C_b^2(\mathbb{R}^n)$, we need the state-output equation

$$\widehat{\phi(x(t))} = \frac{\int_{\mathbb{R}^n}\phi(x)\rho(t,x)dx}{\int_{\mathbb{R}^n}\rho(t,x)dx} \tag{3.8}$$

The fundamental problem of non-linear filtering is the "invariant" study of equation (3.7). The analytic difficulty of this problem stems from the following:

(i) In most interesting situtions the operator $x \to h(x)$ is unbounded.

(ii) The paths of the y-process are only Holder continuous of exponent $< \frac{1}{2}$.

4. Pathwise Non-Linear Filtering

The ideas of this section are due to Clark [1978], Davis [1980] and Mitter [1980].

There is as yet no theory of non-linear filtering where the observations are:

$$y(t) = h(x(t)) + \eta(t) \tag{4.1}$$

where $\eta(t)$ is physical wide-band noise and hence smooth. Define $y(t) \Rightarrow Y(t)$ and $\eta(t) = \dot{\tilde{w}}(t)$ where $\cdot$ denotes differentiation. Then (4.1) can be written as:

$$dY(t) = h(x(t))dt + d\tilde{w}(t), \quad \text{or} \tag{4.2}$$

$$Y(t) = \int_0^t h(x(t))dt + \tilde{w}(t) \tag{4.3}$$

Equation (4.3) is a mathematical model of the physical observation (4.1) where the wide band noise $\eta(t)$ has been approximated as "white noise" $\dot{\tilde{w}}(t)$ and hence $\tilde{w}(t)$ is a Wiener process.

Now, if we wish to compute

$$E(\phi((x(t))|F_t^Y) = \text{Functional of } Y \text{ a.s. Wiener measure}$$

then this filter does not accept the physical observation y. The idea is to at least construct a suitable version of the conditional expectation so that the performance of the filter as measured by the mean-square error remains close when the physical observation 'y' is replaced by the mathematical model of the observation.

This is most easily done by eliminating the stochastic integral in (3.7) by a suitable transformation (gauge transformation in the language of physicists).

Define $q(t,x)$ by

$$\rho(t,x) = \exp(h(x)y(t))q(t,x) \tag{4.4}$$

Then $q(t,x)$ satisfies the parabolic partial differential equation

$$\begin{cases} q_t = (L^y)^* q + \tilde{V}^y q, \quad \text{where} \\ L^y\phi = L\phi - y(t)a(x)h_x(x)\cdot\phi_x \\ \tilde{V}^y(t,x) = -\frac{1}{2}h^2(x) - y(t)Lh(x) + \frac{1}{2}y^2(t)h_x(x)'a(x)h_x(x) \end{cases} \tag{4.5}$$

Equation (4.5) is the pathwise non-linear filtering equation and should be solved for each observation path y (which can be taken to be physical observation).

In this paper we shall prove that under certain assumptions, from the solution of the pathwise filtering equations, we can construct a measure which coincides with the measure given by the Feynman-Kac formula for $\rho(t,x)$, namely (3.5). This will be done using direct methods due to Sheu [1982] and using stochastic control ideas due Fleming-Mitter [1982].

We first write down equation (4.5) explicitly as

$$q_t = \frac{1}{2}\operatorname{tr} a(x)\phi_{xx} + g^y(x,t)\cdot\phi_x + V^y(x,t)q \tag{4.6}$$
$q(0,x) = p^0(x)$, the density of $x(0)$, where

$$\begin{cases} g^y = -b + y(t)ah_x + \gamma, \quad \gamma_j = \sum_{i=1}^{n} \frac{\partial a_{ij}}{\partial x_j}, \; j = 1,2,\dots n \\ V^y = \tilde{V}^y - \operatorname{div}(b - y(t)ah_x) + \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial^2 a_{ij}}{\partial x_i \partial x_j} \end{cases} \tag{4.7}$$

5. Estimates for Parabolic Equations

Consider the linear parabolic equation

$$\begin{cases} p_t = \frac{1}{2}\operatorname{tr} a(x)p_{xx} + g(x,t)\cdot p_x + V(x,t)p, \quad t>0 \\ p(0,x) = p^0(x) \in C_b(\mathbb{R}^n) . \end{cases} \tag{5.1}$$

We assume

$g, V \in C^\infty$ (H1)

$a_{ij} \in C_b^\infty$ and there exist constants $\lambda_1, \lambda_2 > 0$ such that
$$\lambda_1|\xi|^2 \le \sum_{i,j=1}^{n} a_{ij}\xi_i\xi_j \le \lambda_2|\xi|^2 \quad \forall\, \xi \in \mathbb{R}^n \tag{H2}$$

There exists $\bar{V} \in C^\infty(\mathbb{R}^n)$, $\bar{V} > 0$ such that

$|\bar{V}_x|$, $|g_j|$, $|(g_j)_x| = O(\bar{V})$ as $|x| \to \infty$

$|V_x| = O(\bar{V}^2)$ (H3)

$\bar{V} \to \infty$ as $|x| \to \infty$

$\forall T > 0$, there exists $C_1 > 0$ and C_2 such that
$$-V(t,x) \ge C_1\bar{V}^2(x) + C_2, \quad t \in [0, T] \quad . \tag{H4}$$

Consider the stochastic differential equation

$$\begin{cases} dx(t) = \sigma(x(t))dw(t) \\ x(0) = x \end{cases} \tag{5.2}$$

where $\sigma = a^{1/2}$ and w is brownian motion. Then

Theorem 5.1 (Existence)

The solution of equation (5.1) has the probabilistic representation

$$p(t,x) = E_x[p^0(x(t))\exp(\int_0^t \sigma^{-1}g(x(s))dw(s) - \frac{1}{2}\int_0^t|\sigma^{-1}g|^2ds) \; . \tag{5.3}$$
$$\exp(\int_0^t V(x(s))ds] \; ,$$

where E_x denotes expectation with respect to the measure on the path space of $x(.)$.

Moreover, $p(t,x) \in C^\infty$ for $t>0$ and it is a classical solution of (5.1).

Proof: If g is bounded and V is bounded below (5.3) is a classical result. It is obtained by using the Girsanov transformation and the Feynman-Kac formula (cf. Simon [1979]). For g and V satisfying the hypotheses of this section, it is still valid using an approximation argument on g and V (for example, approximating by C_b^∞ - functions).

The probabilistic representation is then used to obtain a weak solution for (5.1) and then a classical regularity result shows that the probabilistic solution gives a C^∞-solution of (5.1).

We now need the following probabilistic estimates:

Lemma 5.2 (Sheu [1982])

$\forall\, k>0$, $\forall T > 0$ there exists a constant C such that
$$E_x[\exp(-\int_0^t \bar{V}^2(x(s))ds)] \le C\exp(-tk\,\bar{V}(x)) \quad \forall\, t \in [0,T]. \tag{5.4}$$

Lemma 5.3 (Sheu [1982]

$\forall\, m$, $\varepsilon>0$, there exists a constant C such that
$$E_x[\bar{V}^m(x(t))\exp(-\int_0^t \bar{V}^2(x(s))ds)] \le C\exp(\varepsilon\bar{V}(x)) \; . \tag{5.5}$$

Theorem 5.3 (Sheu [1982]). If $p^0 \in C_b^\infty$ then we have the estimates:

$\forall\, k>0$, $t \in [s,T]$, $s>0$, there exists a C such that
$$|p(t,x)|,\; |p_t(t,x)|,\; |p_x(t,x)| \le C\exp(-kV(x)). \tag{5.6}$$

Proof (Sketch)

Let $\ell(t) = \exp(\int_0^t \sigma^{-1}g(x(s))dw(s) - \frac{1}{2}\int_0^t |\sigma^{-1}g(x(s))|^2ds)$ and
$m(t) = \ell_t \exp(\int_0^t V(s,x(s))ds)$.

From the Ito differential rule and the Feynman-Kac formula (5.3), we obtain
$$d(p^0(x(t))m(t)) = m(t)[\frac{1}{2}\operatorname{tr}a(x(t))p^0_{xx}(x(t)) + V(x(t))p_0(x(t))]dt \tag{5.7}$$
$$+ m(t)[\sigma'(x(t))p^0_x(x(t)) + p^0(x(t))\sigma(x(t))g(x(t))]\cdot dw(t)$$
$$\triangleq m(t)\xi(t) + m(t)\eta(t)\cdot dw(t),$$

$$p(t,x) = p^0(x) + \int_0^t E_x\{m(s)\xi(s)\}ds \qquad (5.8)$$

$$p_t(t,x) = E_x[m(t)\xi(t)] \qquad (5.9)$$

Now the estimates on p and p_t are obtained by applying the probabilistic estimates (5.4), (5.5) to (5.7) and (5.8).

The last estimate is obtained by obtaining an expression p_x from (5.8), replacing g and V by approximations $g^{(n)}$ and $V^{(n)} \in C_b^\infty$ and applying the probabilistic estimates (5.4), (5.5) and taking the limits in n.

6. Uniqueness of the Weak Zakai Equation

Consider the Zakai equation (3.4) and define a measure $\sigma^y(t)$ by

$$(\sigma^y(t), \phi) = (\rho(t), \exp(-y(t)h)\phi), \quad \forall\ \phi \in C_0^\infty(\mathbb{R}^n). \qquad (6.1)$$

Then $\sigma^y(t)$ satisfies

$$(\sigma^y(t),\phi) = (\sigma^y(0),\phi) + \int_0^t (\sigma^y(s), [L^y + \tilde{V}^y]\phi)ds \qquad (6.2)$$

where L^y and $\tilde{V}^y$ are given by (4.5).

Theorem 6.1 [Sheu, 1982]

Let $\sigma(t)$ be the solution of

$$\frac{d}{dt}(\sigma(t),\phi) = (\sigma(t), L^y\phi)\ \forall\ \phi \in C_0^\infty(\mathbb{R}^n) \qquad (6.3)$$

with $\sigma(t) \geq 0$ and $(\sigma(t), \exp(y(t)h))$ bounded on bounded intervals with a continuous y fixed.

For $\overline{\psi} \in C_0^\infty(\mathbb{R}^n)$ let $\psi(s)$ be the solution of the backward equation

$$\begin{cases} \frac{d\psi(x)}{ds} + L^y\psi(x) = 0 & s < t \\ \psi(t) = \overline{\psi} & \end{cases} \qquad (6.4)$$

Then $(\sigma(s), \psi(s))$ = constant on $[0,t]$ and $(\sigma(t),1)$ is uniformly bounded on intervals $[s,T]$, $0<s<T$.

Sketch of Proof.

If we define $\tilde{\psi}(s) = \psi(t-s)$ then we can apply the estimates of Theorem (5.2) to obtain

$$|\tilde{\psi}(s,x)|,\ |\tilde{\psi}_s(s,x)|,\ |\tilde{\psi}_x(s,x)| \leq C\exp(-k|h(x)|)$$

for $s \in [\delta,t]$, $\delta > 0$, C defending on δ. By an approximation argument equation (6.3) holds for all ϕ satisfying

$$|\phi|,\ |\phi_x|,\ |L^y\phi| \leq C_k \exp(-k|h(x)|) \quad k \geq |y|$$

Hence

$$\frac{d}{ds}(\sigma(s),\psi(s)) = (\sigma(s), L^y\psi(s)) + (\sigma(s), \frac{d\psi}{ds}) = 0 \quad s < t$$

and hence $(\sigma(s), \psi(s))$ = constant on $[0,t)$.

A simple approximation argument shows that if $\psi(s)$ is the solution of (6.4) with $\overline{\psi} = 1$ then $(\sigma(t), 1) \leq (\sigma(s),\psi(s))$, and hence using the hypotheses of the theorem and the estimates of Theorem 5.3 we get $(\sigma(t), 1)$ is uniformly bounded on $[s,T]$.

It remains to show, $(\sigma(s),\psi(s)) = (\sigma(t),\overline{\psi})$. Since $(\sigma(s),\psi(s)) = (\sigma(u),\psi(u))$ $s, u \in [0,t)$ this is proved by a limiting argument with $u \to t$.

CHAPTER 2

On the Relation Between Nonlinear Filtering and Stochastic Control

1. Introduction

In the 1960's Bryson and Frazier [1962] Cox [1964] and Mortensen [1968] suggested a heuristic approach to the non-linear filtering problem based on maximizing a likelihood functional in functional space. This topic was taken up by Hijab [1980] in his doctoral dissertation and a maximum likelihood estimate was constructed by Mitter [1980] in a special situation and the general situation was discussed by Bismut [1981] based on a suggestion of Mitter. The main purpose of this chapter is to show that these ideas can be given a rigorous and at the same time natural interpretation.

To introduce the ideas we present a heuristic discussion as in Mortensen [1978]. Suppose we have a nonlinear system described by the differential equation

$$\dot{x}(t) = f(x(t),t) \tag{1.1}$$

Measurements of the state are made according to the model

$$y(t) = h(x(t),\ t) + \varepsilon(t) \tag{1.2}$$

In (1.2), $\varepsilon(t)$ represents an error term which it is beyond our power to mitigate.

Since (1.1) involves no state noise, if we can estimate the initial state at some starting time t_0, then we at once can compute the entire state trajectory. Suppose first that we take measurements over a fixed interval $t_0 \leq t \leq T$. Also suppose that we have some imprecise prior information which suggests that the starting state $x(t_0)$ is in the neighborhood of the value a. A reasonable approach is as follows:

Let $\phi(t,t_0,x_0)$ denote the complete solution to (1.1) with initial condition $x(t_0) = x_0$,

$$\frac{\partial}{\partial t}\phi(t,t_0,x_0) = f[\phi(t,t_0,x_0),\ t] \tag{1.3}$$

$$\phi(t_0,t_0,x_0) = x_0 \tag{1.4}$$

Then, given $y(t)$, $t_0 \leq t \leq T$, we seek the value of x_0 which minimizes

$$J = \frac{\lambda}{2}(x_0 - a)^2 + 1/2\int_{t_0}^{T}\{y(t) - h[\phi(t,t_0,x_0),\ t]\}^2 dt\ ,\quad \lambda>0\ . \tag{1.5}$$

Therefore, one needs to find the root of the "likelihood equation"

$$\frac{\partial J}{\partial x_0} = 0 \tag{1.6}$$

We now use Hamilton-Jacobi theory. To do this, we replace the fixed endpoint time T in (1.5) with current time t, and we use a dummy time variable τ as the

variable of integration. Let us consider possible trajectories which arrive at state x at time t. Introduce the function

$$V(x,t) = \frac{\lambda}{2}[\phi(t_0,t,x) - a]^2 + \frac{1}{2}\int_{t_0}^{t}\{y(\tau) - h[\phi(\tau,t,x),\tau]\}^2 d\tau \tag{1.7}$$

Given the measurement record $y(\tau)$, $t_0 \leq \tau \leq t$, $V(x,t)$ is the numerical value of the penalty assigned by our least squares rationale for estimating the value x as the state at time t. Direct differentiation reveals that $V(x,t)$ obeys the partial differential equation

$$\frac{\partial}{\partial t}V(x,t) + f(x,t)\frac{\partial V}{\partial x}(x,t) = \frac{1}{2}[y(t) - h(x,t)]^2 \tag{1.8}$$

This suggests that we define the Hamiltonian function

$$H(x,p,t) = pf(x,t) - \frac{1}{2}[y(t) - h(x,t)]^2 \tag{1.9}$$

We may then make the interpretation that $V(x,t)$ satisfies the Hamilton-Jacobi equation

$$\frac{\partial}{\partial t}V(x,t) + H[x,\ \frac{\partial V}{\partial x}(x,t),\ t] = 0 \tag{1.10}$$

The boundary condition on (1.10) is, from (1.7),

$$V(x,t_0) = \frac{\lambda}{2}(x - a)^2 \tag{1.11}$$

Now define the modal trajectory estimate $\hat{x}(t)$ as the root of the equation

$$\frac{\partial V(x,t)}{\partial x} = 0 \tag{1.12}$$

Computing $\frac{d}{dt}\left(\frac{\partial V}{\partial x}\right)$ we get

$$\frac{d\hat{x}(t)}{dt} = \left.\frac{\partial H(x,p,t)}{\partial p}\right|_{\substack{x = \hat{x}(t)\\ p = 0}} + \frac{1}{\pi_2(t)}\left.\frac{\partial H(x,p,t)}{\partial x}\right|_{\substack{x = \hat{x}(t)\\ p = 0}} \tag{1.13}$$

where $\pi_2(t)$ is defined as

$$\pi_2(t) = \left.\frac{\partial^2 V(x,t)}{\partial x^2}\right|_{x = \hat{x}(t)} \tag{1.14}$$

Substituting (1.9) into (1.13), the modal trajectory filter is explicitly

$$\frac{d\hat{x}(t)}{dt} = f(\hat{x}(t),t) + \frac{1}{\pi_2(t)}h_x(\hat{x}(t),t)[y(t) - h(\hat{x}(t),t)] \tag{1.15}$$

where $h_x = \frac{\partial h}{\partial x}$. Here it is assumed that $\frac{\partial^2 V}{\partial x^2}$ is invertible at $\hat{x}(t)$, $\forall t$.

We now seek a differential equation for computing $\pi_2(t)$. A standard computation shows that this equation is

$$\frac{d\pi_2(t)}{dt} = \frac{\pi_3(t)}{\pi_2(t)} \frac{\partial H(x,p,t)}{\partial x}\bigg|_{\substack{x=\hat{x}(t)\\ p=0}} - \left[\frac{\partial^2 H(x,p,t)}{\partial x^2} + 2\pi_2(t)\frac{\partial^2 H(x,p,t)}{\partial x \partial p} + \right.$$

$$\left. \pi_2^2(t)\frac{\partial^2 H(x,p,t)}{\partial p^2}\right]_{\substack{x=\hat{x}(t)\\ p=0}} \tag{1.16}$$

In (1.16), $\pi_3(t)$ is defined as

$$\pi_3(t) = \frac{\partial^3 V(x,t)}{\partial x^3}\bigg|_{x=\hat{x}(t)} \tag{1.17}$$

However, this approach has the drawback that the equation for $\pi_k(t)$ involves $\pi_{k+1}(t)$ and hence we are led to an infinite set of coupled equations.

We now show how these ideas have a natural and rigorous interpretation based on the relation between linear parabolic equations and Bellman-Hamilton-Jacobi equations which is based on a logarithmic transformation, well known in the physics literature. The results in the next section are taken from Fleming-Mitter [1982].

2. The logarithmic transformation. Let us consider a linear parabolic partial differential equation of the form

$$\begin{aligned} p_t &= \frac{1}{2}\operatorname{tr} a(x)p_{xx} + g(x,t)\cdot p_x + V(x,t)p, \quad t \geq 0, \\ p(x,0) &= p^0(x). \end{aligned} \tag{2.1}$$

By solution $p(x,t)$ to (2.1) we mean a "classical" solution $p \in C^{2,1}$, i.e. with p_{x_i}, $p_{x_i x_j}$, p_t continuous, $i, j = 1,\dots,n$.

If p is a positive solution to (2.1), then $X = -\log p$ satisfies the nonlinear parabolic equation

$$\begin{aligned} S_t &= \frac{1}{2}\operatorname{tr} a(x)S_{xx} + H(x,t,S_x), \quad t \geq 0. \\ S(x,0) &= S^0(x) = -\log p^0(x), \\ H(x,t,S_x) &= g(x,t)\cdot S_x - \frac{1}{2}S_x' a(x)S_x - V(x,t). \end{aligned} \tag{2.2}$$

Conversely, if $S(x,t)$ is a solution to (2.2), then $p = \exp(-S)$ is a solution to (2.1).

This logarithmic transformation is well known. For example, if $g = V = 0$, then it changes the heat equation into Burger's equation.

We consider $0 \leq t \leq t_1$, with t_1 fixed but arbitrary. Let $Q = R^n \times [0,t_1]$. We say that a function ϕ with domain Q is a class $\mathscr{L}$ if ϕ is continuous and, for every compact $K \in R^n$, $\phi(\cdot,t)$ satisfies a uniform Lipschitz condition on K for $0 \leq t \leq t_1$. We say that ϕ satisfies a polynomial growth condition of degree r, and write $\phi \in \mathscr{P}_r$, if there exists M such that

$$|\phi(x,t)| \leq M(1+|x|^r), \quad \text{all } (x,t) \in Q .$$

Throughout this section and §3 following assumptions are made. Somewhat different assumptions are made in §'s 4,5 as needed. We assume:

$$\sigma, \sigma^{-1} \text{ are bounded, Lipschitz functions on } R^n. \tag{2.3}$$

From some $m \geq 1$

$$g \in \mathscr{L}\cap\mathscr{P}_m, \ V \in \mathscr{L}\cap\mathscr{P}_{2m} . \tag{2.4}$$

For some $\ell \geq 0$

$$S^0 \in C^2 \cap \mathscr{P}_\ell . \tag{2.5}$$

For some M_1,

$$V(x,t) \leq M_1, \quad S^0(x) \geq -M_1 . \tag{2.6}$$

We introduce the following stochastic control problem, for which (2.2) is the dynamic programming equation. The process $\xi(t)$ being controlled is n-dimensional and satisfies

$$d\xi = \underline{u}(\xi(\tau),\tau)d\tau + \sigma[\xi(\tau)]dw, \quad 0 < \tau \leq t, \quad \xi(0) = x. \tag{2.7}$$

The control is feedback, R^n-valued:

$$u(\tau) = \underline{u}(\xi(\tau),\tau). \tag{2.8}$$

Thus, the control $\underline{u}$ is just the drift coefficient in (2.7). We admit any $\underline{u}$ of class $\mathscr{L}\cap\mathscr{P}_1$. Note that $\underline{u} \in \mathscr{P}_1$ implies at most linear growth of $|\underline{u}(x,t)|$ as $|x| \to \infty$. For every admissible $\underline{u}$, equation (2.7) has a pathwise unique solution ξ such that $E||\xi||_t^r < \infty$ for every $r > 0$. Here $||\ ||_t$ is the sup norm on $[0,t]$.

Let

$$L(x,t,u) = \frac{1}{2}(u-g(x,t))'a^{-1}(x)(u-g(x,t)) - V(x,t). \tag{2.9}$$

For $(x,t) \in Q$ and $\underline{u}$ admissible, let

$$J(x,t,\underline{u}) = E_x\left\{\int_0^t L[\xi(\tau), t-\tau, u(\tau)]d\tau + S^0[\xi(t)]\right\} \tag{2.10}$$

The polynomial growth conditions in (2.4), (2.5) imply finiteness of J. The stochastic control problem is to find $\underline{u}^{op}$ minimizing $J(x,t,\underline{u})$. Under the above assumptions, we cannot claim that an admissible u^{op} exists minimizing $J(x,t,\underline{u})$. However, we recall the following result, which is a rather easy consequence of the Ito differential rule.

Verification Theorem. Let S be a solution to (2.2) of class $C^{2,1}\cap\mathscr{P}_r$, with $S(x,0) = S^0(x)$. Then

(a) $S(x,t) \le J(x,t;\ \underline{u})$ for all admissible $\underline{u}$.

(b) If $\underline{u}^{op} = g - aS_x$ is admissible, then $S(x,t) = J(x,t;\ \underline{u}^{op})$.

In §3 we use (a) to get upper estimates for $S(x,t)$, by choosing judiciously comparison controls. For $\underline{u}^{op}$ to be admissible, in the sense we have defined admissibility, $|S_x|$ can grow at most linearly with $|x|$; hence $S(x,t)$ can grow at most quadratically. By enlarging the class of admissible controls to include certain $\underline{u}$ with faster growth as $|x| \to \infty$ one could generalize (b). However, we shall not do so here, since only part (a) will be used in §3 to get an estimate for S.

In §4 we consider the existence of a solution S with the polynomial growth condition required in the Verification Theorem.

Other control problems. There are other stochastic control problems for which (2.2) is also the dynamic programming equation. One choice, which is appealing conceptually, is to require instead of (2.7) that $\xi(\tau)$ satisfy

$$d\xi = \{g[\xi(\tau),\ \tau] + \underline{u}[\xi(\tau),\tau]\}d\tau + \sigma[\xi(\tau)]dw \tag{2.11}$$

with $\xi(0) = x$. We then take

$$L(x,t,u) = \frac{1}{2}\,u'a^{-1}(x)u - V(x,t)\ . \tag{2.12}$$

The feedback control $\underline{u}$ changes the drift in (2.11) from g to $g + \underline{u}$. When a = identity, $L = \frac{1}{2}|u|^2 - V(x,t)$ corresponds to an action integral in classical mechanics with time-dependent potential $V(x,t)$.

3. Upper estimates for $S(x,t)$. In this section we obtain the following upper estimates for the growth of $S(x,t)$ as $|x| \to \infty$ in terms of the constants $m \ge 1$, $\ell \ge 0$ in (2.4), (2.5)

Theorem 3.1. Let S be a solution of (2.2) of class $C^{2,1} \cap \mathcal{P}_r$, with $S(x,0) = S^0(x)$. Then there exist positive M_1, M_2 such that:

(i) For $(x,t) \in Q$, $S(x,t) \le M_1(1+|x|^\rho)$ with $\rho = \max(m+1,\ell)$.

(ii) Let $0 < t_0 < t_1$, $m > 1$. For $(x,t) \in R^n \times [t_0,t_1]$,
$S(x,t) \le M_2(1+|x|^{m+1})$.

The constant M_1 depends on t_1, and M_2 depends on both t_0 and t_1 in the hypotheses of this theorem, $S(x,t)$ is assumed to have polynomial growth as $|x| \to \infty$ with some degree r. The theorem states that r can be replaced by ρ, or indeed by $m+1$ provided $t > t_0 > 0$. Purely formal arguments suggest that $m+1$ is best possible, and this is confirmed by the lower estimate for $S(x,t)$ made in §5.

Sketch of Proof.

Theorem 3.1 is proved by noting

$$\begin{cases} L(x,t,u) \le B_1(1 + |x|^{2m} + |u|^2) \\ S^0(x) \le B_1(1 + |x|^\ell) \end{cases} \tag{3.1}$$

which follows from (2.3) - (2.6) and (2.9).

The remainder of the proof is based on choosing the open loop control $u(\tau)$, $0 \le \tau \le t$, where $u(\tau) = \dot\eta(\tau)$ and the components $\eta_i(\tau)$ satisfy

$$\begin{cases} \dot\eta_i = -(\operatorname{sgn} x_i)|\eta_i|^m, \quad i = 1,\dots,n \\ \eta(0) = x \end{cases} \tag{3.2}$$

This choice of non-optimal control gives the requisite upper bounds.

4. An existence theorem. In this section we give a stochastic control proof of a theorem asserting that the dynamic programming equation (2.2) with the initial data S^0 has a solution S. The argument is essentially taken from Fleming-Rishel [1975]. Since (2.2) is equivalent to the linear equation (2.1), with positive initial data p^0, one could get existence of S from other results which give existence of positive solutions to (2.1). However, the stochastic control proof gives a polynomial growth condition on S used in the Verification Theorem (§2).

Let $0 < \alpha < 1$. We say that a function ϕ with domain Q is of class C_α if the following holds. For any compact Γ Q, there exists M such that (x,t), $(x', t') \in \Gamma$ imply

$$|\phi(x', t') - \phi(x,t)| \le M[|t' - t|^{\alpha/2} + |x' - x|^\alpha] \tag{4.1}$$

We say that ϕ is of class $C^{2,1}_\alpha$ if ϕ, ϕ_{x_i}, $\phi_{x_ix_j}$, ϕ_t are of class C_α, $i, j = 1,\dots, n$.

In this section the following assumptions are made. The matrix $\sigma(x)$ is assumed constant. By a change of variables in R^n we may take

$$\sigma = \text{identity} \tag{4.2}$$

For fixed t, $g(\cdot, t)$, $V(\cdot, t)$ are of class C^1 on R^n, and g, g_{x_i}, V, V_{x_i}, $i = 1,\dots,n$, are of class C_α for some $\alpha \in (0,1]$. Moreover,

$$|g(x,t)| \le \gamma_1 + \gamma_2|x|^m, \quad m > 1, \tag{4.3}$$

with γ_2 small enough that (4.8) below holds. (If $g \in \mathcal{P}_\mu$ with $\mu < m$, then we can take γ_2 arbitrarily small.) We assume that

$$a_1|x|^{2m} - a_2 \le -V(x,t) \le A(1 + |x|^{2m}) \tag{4.4}$$

for some positive a_1, a_2, A and that

$$g_x \in \mathscr{P}_m,\ V_x \in \mathscr{P}_{2m} \tag{4.5}$$

We assume that $S^0 \in C^3 \cap \mathscr{P}_\ell$ for some $\ell \geq 0$, and

$$\lim_{|x| \to \infty} S^0(x) = +\infty \tag{4.6}$$

$$|S^0_x| \leq C_1 S^0 + C_2 \tag{4.7}$$

for some positive C_1, C_2.

Example. Suppose that $V(x,t) = -kV_0(x) + V_1(x,t)$ with $V_0(x)$ a positive, homogeneous polynomial of degree $2m$, $k > 0$, and $V_1(x,t)$ a polynomial in x of degree $\leq 2m-1$ with coefficients Hölder continuous functions of t. Suppose that $g(x,t)$ is a polynomial of degree $\leq m-1$ in x, with coefficients Hölder continuous in t, and $S^0(x)$ is a polynomial of degree ℓ satisfying (4.6). Then all of the above assumptions hold.

From (2.9), (4.2), $L = \frac{1}{2}|u-g|^2 - V$. If γ_2 in (4.3) is small enough, then

$$\beta_1(|u|^2 + |x|^{2m}) - \beta_2 \leq L(x,t,u) \leq B(1 + |u|^2 + |x|^{2m}) \tag{4.8}$$

for suitable positive β_1, β_2, B. Moreover,

$$L_x = -g_x(u-g) - V_x,$$

$$|L_x| \leq \frac{1}{2}|u|^2 + |g_x|^2 + \frac{1}{2}|g|^2 + |V_x|,$$

where $|g_x|$ denotes the operator norm of g_x regarded as a linear transformation on R^n. From (4.3), (4.5), (4.8)

$$|L_x| \leq C_1 L + C_2 \tag{4.9}$$

for some positive C_1, C_2 (which we may take the same as in (4.7).)

Theorem 4.1. Let $r = \max(2m, \ell)$. Then equation (2.2) with initial data $S(x,0) = S^0(x)$ has a unique solution $S(x,t)$ of class $C^{2,1} \cap \mathscr{P}_r$, such that $S(x,t) \to \infty$ as $|x| \to \infty$ uniformly for $0 \leq t < t_1$.

Sketch of Proof of Theorem 4.1.

The existence part of the proof is contained in Fleming [1969] and Fleming-Rishel [1975].

To obtain uniqueness, $p = \exp(-S)$ is a $C^{2,1}_\alpha$ solution of (2.1), with $p(x,t) \to 0$ as $|x| \to \infty$ uniformly for $0 \leq t \leq t_1$. Since $V(x,t)$ is bounded above, the maximum principle for linear parabolic equations implies that $p(x,t)$ is unique among solutions to (2.1) with these properties, and with initial data $p(x,0) = p^0(x) = \exp[-S^0(x)]$. Hence, S is also unique, proving theorem 4.1.

It would be interesting to remove the restriction that σ = constant made in this

section.

5. A lower estimate for $S(x,t)$. To complement the upper estimates in Theorem 3.1, let us give conditions under which $S(x,t) \to +\infty$ as $|x| \to \infty$ at least as fast as $|x|^{m+1}$, $m \geq 1$. This is done by establishing a corresponding exponential rate of decay to 0 for $p(x,t)$. In this section we make the following assumptions. We take $\sigma \in C^2$ with

$$\sigma,\ \sigma^{-1},\ \sigma_{x_i} \text{ bounded},\ \sigma_{x_i x_j} \in \mathscr{P}_r,\ i, j = 1,\dots,n\ , \tag{5.1}$$

for some $r > 0$. For each t, $g(\cdot,t) \in C^2$. Moreover,

$$g \in \mathscr{P}_\mu,\ \mu < m,\ g_{x_i} \in \mathscr{P}_r,\ g_{x_i x_j} \in \mathscr{P}_r, \tag{5.2}$$

and g, g_{x_i}, $g_{x_i x_j}$ are continuous on Q. For each t, $V(\cdot,t) \in C^2$. Moreover, V satisfies (4.4),

$$V_{x_i} \in \mathscr{P}_r,\ V_{x_i x_j} \in \mathscr{P}_r, \tag{5.3}$$

and V, V_{x_i}, $V_{x_i x_j}$ are continuous on Q. We assume that $p^0 \in C^2$ and there there exist positive β, M such that

$$\exp[\beta|x|^{m+1}]\,[p^0(x) + |p^0_{x_i}(x)| + |p^0_{x_i x_j}(x)|] \leq M \tag{5.4}$$

Theorem 5.1. Let $p(x,t)$ be a $C^{2,1}$ solution to (2.1) such that $p(x,t) \to 0$ as $|x| \to \infty$, uniformly for $0 \leq t \leq t_1$. Then there exists $\delta > 0$ such that $\exp[\delta|x|^{m+1}]p(x,t)$ is bounded on W.

Proof. Let

$$\psi(x) = (1+|x|^2)^{\frac{m+1}{2}}, \quad \pi(x,t) = \exp[\delta\psi(x)]p(x,t).$$

Then π is a solution to

$$\pi_t = \frac{1}{2}\operatorname{tr} a\pi_{xx} + \bar{g}\cdot\pi_x + \bar{V}\pi\ , \tag{5.5}$$

$$\bar{g} = g - \delta a\psi_x$$

$$\bar{V} = V - \delta g\cdot\psi_x + \frac{1}{2}(\delta^2 a\psi_x\cdot\psi_x - \delta\operatorname{tr} a\psi_{xx})$$

The proof is now completed by using Theorem 5.1 of Chapter 1 and the maximum principle for parabolic equations.

Corollary: For some positive δ, δ_1

$$S(x,t) \geq \delta|x|^{m+1} - \delta_1. \tag{5.8}$$

6. Connection with the Pathwise Filter Equations

Consider now the pathwise filter equations (4.5) of Chapter 1. Using the techniques of this section we get existence and uniqueness of a classical solution of equation (4.5). Moreover defining

$$S^y = -\log q^y$$

we get the bounds

$$S^y(x,t) \le M_1(1 + |x|^{\rho}), \quad 0 < t \le t_1, \quad \rho = \max(m+1, \ell) \tag{6.1}$$

$$S^y(x,t) \le M_2(1 + |x|^{m+1}), \quad 0 < t_0 \le t \le t_1, \quad m > 1 \tag{6.2}$$

where M_1 and M_2 depend on y. For $q^0 = \exp(-S^0)$ we need $\ell > m+1$.

From the corollary to Theorem 5.1 we get the lower bound

$$S^y(x,t) \ge \delta|x|^{m+1} - \delta_1, \quad 0 \le t \le t_1 \tag{6.3}$$

Finally defining

$$p = \exp(y(t)h)q \tag{6.4}$$

and defining a positive measure

$$\tilde{\Lambda}_t(\phi) = \int_{\mathbb{R}^n} \phi(x)p(t,x)\,dx, \qquad \phi \in C_b, \tag{6.5}$$

using Theorem 6.1 of Chapter 1, one can show that $\tilde{\Lambda}_t = \Lambda_t$ is given by the Kallianpar-Striebel Formula, (3.5) of Chapter 1.

The consequences of these results are important. It shows that using the pathwise filter equations one can compute the unnormalized conditional measure of the filtering problem given by the Kallianpar-Striebel formula (and indeed using the bounds one can see that the normalized conditional measure can be computed). This does give the strongest possible robustness result. Furthermore, the upper and lower bounds (6.1) - (6.3) gives us the means of evaluating approximation schemes and also bounds on estimation errors. We conjecture that these results will give sharper results on lower bounds for non-linear filtering (cf. Galdos [1975], [1979]).

7. Remarks on the Previous Results

The existence and uniqueness results for the filtering equations do not cover the following situation:

Consider the non-linear filtering problem:

$$\begin{cases} dx_1(t) = dw_1(t) \\ dx_2(t) = dw_2(t) \end{cases} \tag{7.1}$$

with the observation equation

$$dy(t) = (x_1^3(t) + x_2^3(t))dt + d\eta(t), \tag{7.2}$$

where w_1, w_2 and η are independent Brownian motions. The pathwise filtering equations are:

$$\begin{cases} q_t = \frac{1}{2}\Delta q + g^y(t,x)\cdot q_x + V^y(t,x)q \\ q(0,x) = \rho^0(x) > 0, \quad \text{where} \end{cases} \tag{7.3}$$

$$g^y(t,x) = \begin{pmatrix} y(t)\ 3x_1^2 \\ y(t)\ 3x_2^2 \end{pmatrix}$$

and

$$V^y = -\frac{1}{2}(x_1^3 + x_2^3)^2 + \frac{1}{2}y^2(t)(9x_1^4 + 9x_2^4) + \frac{1}{2}y(t)(36x_1^3 + 36x_2^3)$$

The difficulty with handling this situation is that V^y is not bounded above along the direction $x_1 = -x_2$. Nevertheless it is possible to prove that a weak solution exists in a suitable weighted Sobolev space (cf. Mitter [1982]. However, it is not known whether the unconditional measure can be constructed using the pathwise filtering equations.

8. Construction of a Filter

A filter can now be unstructed using the ideas of Section 1 of this Chapter but working with the equation:

$$\begin{cases} S_t = \frac{1}{2}\mathrm{tr}\, a(x)S_{xx} + H(x,t,S_x), \quad t \ge 0 \\ S(x,0) = S^0(x) = -\log \rho^0(x), \quad \text{where} \end{cases} \tag{8.1}$$

$$H(t,x,S_x) = g^y(t,x)\cdot S_x - \frac{1}{2}S_x'\,a(x)S_x - V^y(t,x) \tag{8.2}$$

and g^y and V^y are given by equation (4.7) of Chapter 1.

If we now make the assumption that S_{xx} is invertible at $\hat{x}(t)$ where $\hat{x}(t)$ is obtained by solving $S_x = 0$, then by the Morse Lemma [Lang, pp. 174, (1969)], there exists a suitable coordinate system in which S is a quadratic in the neighborhood of $\hat{x}(t)$.

In this way one gets an approximate filter which has the structure of an extended Kalman Filter, see equation (1.15) of this Chapter. $E(S_{xx})$ is the analog of the Fisher Information Matrix and the requirement that it is invertible corresponds to requiring an observability condition on the nonlinear system. It is interesting to remark that in examples where the Extended Kalman filter is known not to work S_{xx} turns out to be singular.

9. Remarks

The Stochastic Maximum Principle (see the lectures of Bensoussan this volume) could be applied to the stochastic control problem (8.1)-(8.2) to obtain the bi-characteristics (stochastic) corresponding to equation (8.1).

CHAPTER 3

A Path Integral Calculation for a Class of Filtering Problems

1. Introduction

In this chapter we present a path-integral calculation for a class of filtering problems which gives a new derivation of results of Benes [1981] for this problem and at the same time shows the relationship of this class of problems to certain ideas of Feynman (cf. Feynman-Hibbs 1965). The ideas of this section in their present form are due to Mitter and Ocone (unpublished).

Feynman was concerned with calculating the Green function $G(x,x',t)$ for the Schrödinger equation of a particle in a potential field $U(x)$ from function integral representations. Roughly spreaking

$$G(x,x',t) = \int e^{i\int_0^t \mathscr{L}(x(s),\dot{x}(s))ds}\, d\mu(x(\cdot))$$

where μ is a measure, formally defined as a limit of Gaussian measures, on the space of paths $\{x(s)\,|\,x(0) = x', x(t) = x\}$. $\mathscr{L}$ is the Lagragian of the physical system under consideration. Feynman thought of the paths of the underlying function space as the possible trajectories of the particle. He proposed representing these paths as perturbations about the classical motion $x_c(t)$ of the particle in the field $U(x)$,

$$x(t) = x_c(t) + z(t) .$$

In the simple case of quadratic $U(x)$, corresponding to the harmonic oscillator, this substitution neatly accounts for the x and x' dependencies in G by a clever use of the dynamical equation for $x_c(t)$.

We show how these ideas ca be adapted to a class of filtering problems first considered by Benes.

2. Problem Formulation and Results

We consider a filtering problem where $x(t)$ will be an $\mathbb{R}^n$-valued process satisfying

$$dx(t) = f(x(t))dt + db(t) \tag{2.1}$$

$$x(t) = x_0$$

where

$$f(x) = \nabla V(x) \tag{2.2}$$

for some twice continuously differentiable $V:\mathbb{R}^n \to \mathbb{R}$ and x_0 is a point of $\mathbb{R}^n$ (and not a random variable). Likewise

$$dy(t) = Hx(t)dt + d\eta(t) \tag{2.3}$$

where $y(t)$ is an $\mathbb{R}^p$-valued process and H a $p\times n$ matrix. Of course, $b(\cdot)$ and $\eta(\cdot)$ are independent Brownian motions of appropriate dimensions.

We assume further that, for some $T>0$,

$$P[\int_0^T \langle f(x(s)), f(x(s))\rangle ds<\infty] = 1$$

$$P[\int_0^T \langle f(x_0+b(s)), f(x_0+b(s))\rangle ds<\infty] = 1$$

$$P[\int_0^T \langle Hx(s), Hx(s)\rangle ds<\infty] = 1$$

We shall work always with $t\leq T$.

The specific form for $f(x)$ is important because it allows a useful representation for $\int_0^t \langle f(x(s)), dx(s)\rangle$. Indeed, by Ito's differentiation rule

$$\int_0^t \langle f(x(s)), dx(s)\rangle = V(x(t)) - V(x(0)) - \frac{1}{2}\int_0^t \nabla^2 V(x(s))ds \tag{2.4}$$

Our hypotheses guarantee that the unnormalized conditional density makes sense. From the vector version of the Kallianpur-Stiaebel formula, this is

$$p(x,t) = E_0\{\exp[\int_0^t \langle Hx(s), dy(s)\rangle - \frac{1}{2}\int_0^t \langle Hx(s), Hx(s)\rangle ds]\,|\,F_t^y, x(t) = x\} \times q(x,t). \tag{2.5}$$

In the above E_0 represents expectation with respect to the transformed measure. We show the existence of the density $q(x,t)$ of $x(t)$ in the next lemma.

Lemma 2.1. $x(t) - x_0$ is a Brownian motion under P_1,

$$\frac{dP_1}{dP_0} = \exp[-\int_0^t \langle f(x(s)), db(s)\rangle - \frac{1}{2}\int_0^t ||f(x(s))||^2 ds]$$

and

$$\frac{dP_0}{dP_1} = \exp[\int_0^t \langle f(x(s)), dx(s)\rangle - \frac{1}{2}\int_0^t ||f(x(s))||^2 ds] \tag{2.6}$$

$q(x,t)$ exists and

$$q(x,t) = [2\pi t]^{-n/2} e^{-\frac{1}{2t}||x-x_0||^2} E_1\{\frac{dP_0}{dP_1}\,|\,x(t) = x\} \tag{2.7}$$

Lemma 2.1 provides the tool for rewriting (2.5) as a Gaussian integral.

Theorem 2.2.

$$p(x,t) = (2\pi t)^{n/2} e^{-\frac{1}{2t}||x-x_0||^2} e^{V(x)-V(x_0)}$$

$$\times E_1 \exp[\int_0^t \langle Hx(s), dy(s)\rangle - \frac{1}{2}\int_0^t R(x(s))ds]\,|\,F_t^y, x_t = x\}$$

$$R(x) := ||Hx||^2 + ||f(x)||^2 + \nabla^2 V(x)$$

Sketch of Proof

The proof of Theorem 2.2 is accomplished by using the rule for changing measures under conditional expectation and (2.4). □

In the next theorem, we will apply the Feynman idea mentioned before to the conditional expectation

$$E_1\{\exp[\int_0^t \langle Hx(s),\ dy(s)\rangle - \frac{1}{2}\int_0^t R(x(s))ds] | F_t^y,\ x_t = x\} \tag{2.8}$$

To do this we need the analogue of the classical motion in the Feynman path integral. To motivate our choice we recall that the path of a classical particle in a potential field $U(t,x)$ is given by

$$\ddot{x}_c(t) = -\nabla U(t, x_c(t))$$

(we assume the particle has unit mass.). If $v(t) := \dot{x}_c(t)$ we can reexress this as a two-dimensional first-order system

$$\begin{pmatrix}\dot{x}_c(t)\\ \dot{v}(t)\end{pmatrix} = \begin{pmatrix}v(t)\\ -\nabla U(t,x_c(t))\end{pmatrix}$$

For (2.8), the appropriate potential is formally

$$U(t,x) = \langle Hx, \frac{dy}{dt}\rangle - R(x)$$

(so that (2.8) equals

$$E_1\left\{e^{\int_0^t U(s,x(s))ds} \Big| F_t^y,\ x_t = x\right\}.)$$

The corresponding classical motion is then given by

$$\begin{pmatrix}\dot{x}_c(t)\\ \dot{v}(t)\end{pmatrix} = \begin{pmatrix}v(t)\\ -H^T\frac{dy}{dt} + \frac{1}{2}\nabla R(x_c(t))\end{pmatrix}$$

which, in Ito form, is

$$\begin{pmatrix}x_c(t)\\ v(t)\end{pmatrix} = \begin{pmatrix}V(t)\\ \frac{1}{2}\nabla R(x_c(t))\end{pmatrix}dt + \begin{pmatrix}0\\ -H^t dy(t)\end{pmatrix} \tag{2.9}$$

We now state the main result.

Theorem 2.3. Suppose that

$$R(x) = x^T Ax + b^T x + c$$

and $A>0$. Let $(x_c(s),\ v(s))$ satisfy the equations

$$dx_c(s) = v(s)ds \tag{2.10}$$
$$dv(s) = [Ax_c(s) + \frac{b}{2}]ds - H^T dy(s) \qquad 0 \le s \le t$$

with boundary conditions $x_c(0) = x_0$, $x_c(t) = x$. Then

$$p(x,t) = (2\pi t)^{-\frac{n}{2}} e^{v(x)-v(x_0) - 1/2\, ct}$$
$$\times \exp[\frac{1}{2}\int_0^t \langle x_c(s),\ H^T dy(s)\rangle - \frac{1}{4}\int_0^t \langle b, x_c(s)\rangle ds - \frac{1}{2}\langle x_c(s), v(s)\rangle|_0^t\ ]$$
$$\times E_2\{\exp[-\frac{1}{2}\int_0^t \langle z(s),\ Az(s)\rangle] | z(t) = 0\} \tag{2.11}$$

in which

$$z(s) = x(s) - x_c(s)$$

and E_2 is expectation with respect to the measure P_2, defined below, under which $z(\cdot)$ is Brownian.

Proof. The first step is to represent the conditional expectation term of the expression for $p(x,t)$ in theorem (2.2) in terms of yet another measure. This involves another Girsanov measure transformation. Since

$$z(s) = x(s) - x_0 - \int_0^s v(r)dr\ ,$$

$x(s) - x_0$ is Brownian and independent of $y(\cdot)$ under P_1, and $\int_0^t v^2(s)ds < \infty$ a.s., $z(s)$ is Brownian under P_2,

$$\frac{dP_2}{dP_1} = \exp \int_0^t \langle v(s),\ dx(s)\rangle - \frac{1}{2}\int_0^t \langle v(s),\ v(s)\rangle ds$$

Further, $z(\cdot)$ and $y(\cdot)$ are independent under P_2 and

$$\frac{dP_1}{dP_2} = \exp[-\int_0^t \langle v(s),\ dz(s)\rangle \cdot \frac{1}{2}\int_0^t \langle v(s),\ v(s)\rangle ds]$$

In evaluating $\frac{dP_1}{dP_2}$ it is useful to observe that

$$\int_0^t \langle v(s),\ dz(s)\rangle = \langle v(s),\ z(s)\rangle|_0^t - \int_0^t \langle z(s),\ dv(s)\rangle$$
$$= \langle v(s),\ z(s)\rangle|_0^t - \int_0^t \langle z(s),\ Ax_c(s) + b/2\rangle ds$$
$$+ \int_0^t \langle z(s),\ H^T dy(s)\rangle \tag{2.12}$$

The first equality above is an application of Ito's differentiation rule. Now rewrite the conditional expectation term of $p(x,t)$ in Theorem 2.2 in terms of the P_2 measure;

$$E_1\{\exp \int_0^t \langle Hx(s),\ dy(s)\rangle - \frac{1}{2}\int_0^t R(x(s))ds \mid F_t^y,\ x_t = x\}$$
$$= \frac{E_2\{J(t)\mid F_t^y,\ x_t = x\}}{E_2\left\{\frac{dP_1}{dP_2}\mid F_t^y,\ x_t = x\right\}} \tag{2.13}$$

$$\ln J(t) = \int_0^t \langle Hx(s),\ dy(s)\rangle - \frac{1}{2}\int_0^t R(x(s))ds - \int_0^t \langle v(s),\ dz(s)\rangle$$
$$- \frac{1}{2}\int_0^t \|v(s)\|^2 ds \tag{2.14}$$

The next step is to analyze the numerator of (2.13). If $x(s) = x_c(s) + z(s)$ and (2.12) are substituted in (2.14), it follows that

$$\ln J(t) = \int_0^t \langle Hx_c(s),\ dy(s)\rangle - \frac{1}{2}\int_0^t [\langle x_c(s),\ Ax_c(s)\rangle + \langle v(s), v(s)\rangle]ds - \frac{1}{2}ct$$
$$- \frac{1}{2}\int_0^t \langle b,\ x_c(s)\rangle - \langle v(s),\ z(s)\rangle\Big|_0^t - \frac{1}{2}\int_0^t \langle z(s),\ Az(s)\rangle \tag{2.15}$$

This is the crucial calculation; $\ln J(t)$ contains no term other than $\langle v(s), z(s)\rangle\big|_0^t$ in which $z(\cdot)$ multiplies an F_t^y -measurable random variable. Since the integrations in (2.13) are effectively over the path space of $z(\cdot)$, the $z(\cdot)$ and $y(\cdot)$ dependencies are thereby separated. (2.15) can be further simplified by noting that

$$\langle x_c(s),\ Ax_c(s)\rangle ds + \langle v(\dot{s}),\ v(s)\rangle ds =$$
$$d\langle x_c(s),\ v(s)\rangle - \frac{1}{2}\langle b, x_c(s)\rangle ds + \langle x_c(s),\ H^T dy(s)\rangle\ ,$$

which follows by applying Ito's differentiation rule to $\langle x_c(s),\ v(s)\rangle$. Then, we derive from (2.15))

$$\ln J(t) = \frac{1}{2}\int_0^t \langle Hx(s),\ dy(s)\rangle - \frac{1}{2}\langle x_c(s),\ v(s)\rangle\Big|_0^t - 1/4\int_0^t \langle b, x_c(s)\rangle ds$$
$$- \frac{1}{2}\langle v(s), z(s)\rangle\Big|_0^t - 1/2\int_0^t \langle z(s),\ Az(s)\rangle ds\ .$$

The remainder of the calculation follows by noting that (i) $z(0) = z(t) = 0$, (ii) z and y are independent under P_2 and (iii) a Gaussian integral calculation to show that

$$e^{-\frac{1}{2t}\|x-x_0\|^2}\ E_2\left\{\frac{dP_1}{dP_2}\Big| F_t^y,\ x_t = x\right\} = 1.$$

Theorem 2.3 reduces the original functional integration to solving the two-point boundary value problem (2.10) and evaluating

$$\frac{1}{2}\int_0^t \langle x_c(s),\ H^T dy(t)\rangle - \frac{1}{4}\int_0^t \langle x_c(s),\ b\rangle ds - \frac{1}{2}[\langle x, v(t)\rangle - \langle x_0,\ v(0)\rangle] \tag{2.16}$$

The term $E_2\ \exp[-\frac{1}{2}\int_0^t \langle z(s),\ Az(s)\rangle ds]\,|z(t) = 0\}$ need not concern us further since it is independent of x and hence will disappear when $p(x,t)$ is normalized to derive the true conditional distribution. It turns out that the evaluation of (2.16) is quite easy, if we solve (2.10) by a standard substitution from the control theory of the linear regulator. Write $x_c(s)$ as

$$x_c(s) = P(s)v(s) + m(s)$$

for some nxn matrix function $P(s)$ and vector function $m(s)$. It then must happen that

$$v(s)ds = dx_c(s)$$
$$= \dot{P}(s)v(s)ds + P(s)[A[P(s)v(s) + m(s)] + \frac{b}{2}]ds$$
$$-P(s)H^T dy(s) + dm(s) \tag{2.17}$$

(2/17) will be satisfied if

$$dm(s) = [-P(s)Am(s) - P(s)\frac{b}{2}]ds + P(s)H^T dy(s) \tag{2.18}$$

$$\dot{P}(s) = I - P(s)AP(s) \tag{2.19}$$

The boundary conditions $x_c(0) = x_0$, $x_c(t) = x$ will hold if, in addition,

$$P(0) = 0,\quad m(0) = x_0 \tag{2.20}$$

$$P(t)v(t) = x-m(t) \tag{2.21}$$

It can be shown that the solution $P(s)$ of (2.19) with $P(0) = 0$ is a positive definite matrix for $s>0$. Hence a solution $v(t) = P^{-1}(t)[x-m(t)]$ to (38) exists. We then state

Theorem 2.4

$$\frac{1}{2}\int_0^t \langle x_c(s), H^T dy(s)\rangle - \frac{1}{4}\int_0^t \langle x_c(s),\ b\rangle ds - \frac{1}{2}[\langle x, v(t)\rangle - \langle x_0,\ v_0\rangle]$$
$$= -\frac{1}{2}\langle x-m(t),\ P^{-1}(t)(x-m(t))\rangle + \theta(t)$$

where m and P solve (2.19) and (2.20) with G.C.'s (2.20) and $\theta(t)$ is a random term with no x dependence.

Remark. $\theta(t)$ is again of no importance since it drops out of $p(x,t)$ after normalization. Because of theorem (2.4) a Gaussian density term

$$e^{-\frac{1}{2}\langle x-m(t),\ P^{-1}(t)(x-m)\rangle}$$

appears in $p(x,t)$. $m(t)$ is the conditional mean and $P(t)$ the conditional covariance of this part.

Finally it can be shown that (2.10) are the necessary conditions of an appropriate deterministic optimal control problem which can be given meaning using the ideas of Bismut [1981]. We thus see that in this special situation we can write the unnormalized conditional density as exp(-S(t,x)) where S is the Hamilton-Jacobi value function of a deterministic optimal control problem.

CHAPTER 4

Geometrical Theory of Nonlinear Filtering

1. Introduction

The starting point of the geometrical theory of non-linear filtering is the Zakai equation (cf. Chapter 1 equation 3.7) in Stratanovich form:

$$d\rho(t,x) = (L^* - \frac{1}{2}h^2)\rho(t,x)dt + h\rho(t,x) \cdot dy(t) \tag{1.1}$$

where L is the operator

$$L\phi = \frac{1}{2}\operatorname{tr} a(x)\phi_{xx} + b(x) \cdot \phi_x \tag{1.2}$$

denotes formal adjoint.

We have previously alluded to the fact that the invariant study of this equation problem of non-linear filtering. We try to make this remark more precise.

For this purpose we introduce the Lie algebra of operators

$$\mathscr{L}A\{L^* - \frac{1}{2}h^2, h\}$$

that is the Lie algebra generated by the two operators $L^* - \frac{1}{2}h^2$ and h considered as formal differential operators. We claim that the structure of this Lie algebra and its invariance under an approriate group of transformations exhibits the structure of the filtering problem.

The ideas of this section are due to Brockett and Mitter (for an exposition cf. the article of Brockett with its list of references in Hazewinkel-Willems [1981], the article of Mitter in the same volume and Mitter [1980] and the list of references cited there).

2. Preliminaries. (On Lie Algebras, Lie Groups and Representations)

We shall say that a vector space $\mathscr{L}$ over $\mathbb{R}$ is a real Lie algebra, if in addition to its vector space structure it posseses a product $\mathscr{L} \times \mathscr{L} \to \mathscr{L}: (X,Y) \to [X,Y]$ which has the following properties:

(i) it is bilinear over $\mathbb{R}$

(ii) it is skew commutative: $[X,Y] + [Y,X] = 0$

(iii) it satisfies the Jacobi identity:

$$[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0 .$$

(for all) $X,Y,Z \in \mathscr{L}$

Example: $M_n(\mathbb{R})$ = algebra of n x n matrices over $\mathbb{R}$.

If we denote by $[X,Y] = XY - YX$, where XY is the usual matrix product, then this commutator defines a

Lie algebra structure on $M_n(\mathbb{R})$.

Example: Let $\mathfrak{X}(M)$ denote the C^∞-vector fields on a C^∞-manifold M. $\mathfrak{X}(M)$ is a vector space over $\mathbb{R}$ and a $C^\infty(M)$ module.

(Recall, a vector field X on M is a mapping: $M \to T_p(M)$: $p \mapsto x$ where $p \in M$ and $T_p(M)$ is the tangent space to the point p at M). We can give a Lie algebra structure to $\mathfrak{X}(M)$ by defining:

$$\mathcal{J}_p f = (XY - YX)_p f = X_p(Yf) - Y_p(Xf), \quad f \in C^\infty(p)$$

(the C^∞-functions in a neighborhood of p), and

$$[X,Y] = XY - YX.$$

Both these examples will be useful to us later on.

Let $\mathscr{L}$ be a Lie algebra over $\mathbb{R}$ and let $\{X_1, \ldots, X_n\}$ be a basis of $\mathscr{L}$ (as a vector space). There are uniquely determined constants $c_{rsp} \in \mathbb{R}$. $(1 \le r,s,p \le n)$ such that

$$[X_r, X_s] = \sum_{1 \le p \le n} c_{rsp} X_p$$

The c_{rsp} are called the structure constants of $\mathscr{L}$ relative to the basis $\{X_1, \ldots, X_n\}$. From the definition of a Lie algebra:

(i) $c_{rsp} + c_{srp} = 0 \quad (1 \le r,s,p \le n)$

(ii) $\sum_{1 \le p \le n} (c_{rsp}c_{ptu} + c_{stp}c_{pru} + c_{trp}c_{psu}) = 0 \quad (1 \le r,s,t,u \le n)$.

Let $\mathscr{L}$ be a Lie algebra over $\mathbb{R}$. Given two linear subspaces M, N of $\mathscr{L}$ we denote by [M,N] the linear space spanned by [X,Y], $X \in M$ and $Y \in N$. A linear subspace K of $\mathscr{L}$ is called a sub-algebra if $[K,K] \subseteq K$, an ideal if $[\mathscr{L},K] \subseteq K$.

If $\mathscr{L}$ and $\mathscr{L}'$ are Lie algebras over $\mathbb{R}$ and $\pi: \mathscr{L} \to \mathscr{L}'$: $X \mapsto \pi(X)$, a linear map, π is called a homomorphism if it preserves brackets:

$$[\pi(X), \pi(Y)] = \pi([X,Y]) \qquad (X,Y \in \mathscr{L}).$$

In that case $\pi(\mathscr{L})$ is a subalgebra of $\mathscr{L}'$ and ker π is an ideal in $\mathscr{L}$. Conversely, let $\mathscr{L}$ be a Lie algebra over $\mathbb{R}$ and K an ideal of $\mathscr{L}$. Let $\mathscr{L}' = \mathscr{L}/K$ be the quotient vector space and $\pi: \mathscr{L} \to \mathscr{L}'$ the canonical linear map.

For $X' = \pi(X)$ and $Y' = \pi(Y)$, let

$$[X',Y'] = \pi([X,Y]).$$

This mapping is well-defined and makes $\mathscr{L}'$ a Lie algebra over $\mathbb{R}$ and π is then a homomorphism of $\mathscr{L}$ into $\mathscr{L}'$ with K as the kernel. $\mathscr{L}' = \mathscr{L}/K$ is called the quotient of $\mathscr{L}$ by K.

Let $\mathscr{U}$ be an algebra over $\mathbb{R}$, whose multiplication is bilinear but not necessarily associative. An endomorphism D of $\mathscr{U}$ (considered as a vector space) is called a derivation if

$$D(ab) = (Da)b + a\,(Db) \quad a, b \in$$

If D_1 and D_2 are derivations so is $[D_1,D_2] = D_1D_2 - D_2D_1$. The set of all derivations on $\mathscr{U}$ (assumed finite dimensional) is a subalgebra of gl($\mathscr{U}$), the Lie algebra of all endomorphisms of $\mathscr{U}$.

For us the notion of a representation of a Lie algebra is very important.

Let $\mathscr{L}$ be a Lie algebra over $\mathbb{R}$ and V a vector space over $\mathbb{R}$, not necessarily finite dimensional. By a representation of $\mathscr{L}$ in V we mean a map.

$\pi : X \to \pi(X) : \mathscr{L} \to gl(V)$ (all endomorphisms of V), such that

(i) π is linear

(ii) $\pi([X,Y]) = \pi(X)\pi(Y) - \pi(Y)\pi(X)$.

For any $X \in \mathscr{L}$ let adX denote the endomorphism of $\mathscr{L}$

$$\text{ad}X : Y \to [X,Y] \qquad (Y \in \mathscr{L})$$

ad X is a derivation of $\mathscr{L}$ and $X \to$ ad X is a representation of $\mathscr{L}$ in $\mathscr{L}$, called the adjoint representation.

Let G be a topological group and at the same time a differentiable manifold. G is a Lie group if the mapping $(x,y) \mapsto xy : G \times G \to G$ and the mapping $x \mapsto x^{-1}$: $G \to G$ are both C^∞-mappings.

Given a Lie group G there is an essentially unique way to define its Lie algebra. Conversely, every finite-dimensional Lie algebra is the Lie algebra of some simply connected Lie group.

In filtering theory some special Lie algebras seem to arise. We give the basic definitions for three such Lie algebras.

A Lie algebra $\mathscr{L}$ over $\mathbb{R}$ is said to be nilpotent if ad X is a nilpotent endomorphism of $\mathscr{L}$, $\forall X \in \mathscr{L}$. Let the dimension of $\mathscr{L}$ be m. Then there are ideals $\mathscr{I}_j$ of $\mathscr{L}$ such that (i) dim $\mathscr{I}_j = m-j$, $0 \le j \le m$.

(ii $\mathscr{I}_0 = \mathscr{L} \supseteq \mathscr{I}_1 \supseteq \ldots \mathscr{I}_m = 0$ and (iii) $[\mathscr{L},\mathscr{I}_j] \subseteq \mathscr{I}_{j+1}$, $0 \le j \le m-1$.

Let g be a Lie algebra of finite-dimension over $\mathbb{R}$ and write $\mathscr{D}_g = [g,g]$. $\mathscr{D}g$ is a sub-algebra of g called the derived algebra. Define $\mathscr{D}^p g$ $(p \ge 0)$ inductively by

$$\mathscr{D}^0 g = g$$

$$\mathscr{D}^p g = \mathscr{D}(\mathscr{D}^{p-1} g) \quad (p \ge 1).$$

We then get a sequence $\mathscr{D}^0 g \supseteq \mathscr{D}^1 g \supseteq \ldots$ of sub-algebras of g, g is said to be solvable if $\mathcal{D}^p g = o$ for some $p \geq 1$.

Examples

(i) Let $n \geq o$ and let $(p_1,\ldots,p_n,q_1,\ldots q_n, z)$ be a basis for a real vector space $\mathcal{V}$. Define a Lie algebra structure on $\mathcal{V}$ by $[p_i,q_i] = -[q_i,p_i] = Z$, the other brackets being zero. This nilpotent Lie algebra $\mathcal{N}$ is the so-called Heisenberg algebra.

(ii) The real Lie algebra with basis $(h,p_1,\ldots,p_n,q_1,\ldots,q_n,Z)$ satisfying the bracket relations

$[h,p_i] = q_i$, $[h,q_i] = p_i$, $[p_i,q_i] = Z$, the other brackets being zero is a solvable Lie algebra, the so-called Oscillator algebra. Its derived algebra is the Heisenberg algebra $\mathcal{N}$.

A Lie algebra is called simple if it has no non-trivial ideas. An infinite dimensional Lie algebra $\mathcal{G}$ is called pro-finite dimensional and filtered if there exists a sequence of ideal $\mathcal{I}_1 \supset \mathcal{I}_2 \ldots$ such $\mathcal{G}/\mathcal{I}_2$ is finite-dimensional for all i and $\cap \mathcal{I}_i = \{o\}$

2.1 Infinite-Dimensional Representations

Let g be a finite dimensional Lie algebra and G its associated simply connected Lie group. Let H be a complex Hilbert space (generally infinite-dimensional). We are interested in representations of g by means of linear operators on H with a common dense invariant domain $\mathcal{D}$. Let π denote this representation.

Similarly, we are also interested in representations of G as bounded linear operators on H. Let τ be such a representation. That is, $\tau : G \to L(H)$ satisfies

$$\tau(g_1 g_2) = \tau(g_1)\tau(g_2), \quad g_1, g_2 \in G.$$

The following problem of Group representation has been considered by Nelson [1959] and others. Given a representation π of g on H, when does X there exist a group representation (strongly continuous) τ of G on H such that

$$\tau(\exp(tX)) = \exp(t\pi(X)) \quad \forall X \in G$$

Here $\exp(t\pi(X))$ is the strongly continuous group generated by $\pi(X)$ in the sense that

$$\frac{d\cdot}{dt} \exp(t\pi(x))\phi = \pi(X)\phi \quad \forall \phi \in \mathcal{D}$$

and $\exp(tX)$ is the exponential mapping, mapping the Lie algebra g into the Lie group G.

Let $X_1,\ldots,X_d$ be a basis for g. A method for constructing τ locally is to define

$$\tau(\exp(t_1 x_1) \ldots \exp(t_d x_d)) = \exp(t_1\pi(x_1)) \ldots \exp(t_d(x_d))$$

A sufficient condition for this to work is that the operator identity

$$\begin{cases} \exp(tA_j)A_i = \sum_{n=0}^{\infty} \frac{t^n}{n!} [adA_j]^n A_i \exp(tA_j) \\ \text{holds for } A_j = \pi(X_j),\ 1 \leq j,\ j \leq d\ . \end{cases} \tag{2.1}$$

It is a well known fact, that many Lie algebra representations do not extend to Group representations. An example is the representation of the Heisenberg algebra consisting of three basis elements by the operators $\{-ix, \frac{d\cdot}{dx}, -i\}$ on $L^2(\mathbb{R}_+)$ with domain $C_0^\infty(\mathbb{R}_+)$ which does not extend to a unitary representation (since essential self-adjointness fails).

Although in filtering theory we are not interested in unitary group representations, nevertheless these ideas will serve as a guide for integrating the Lie algebras arising in filtering theory.

3. Lie Algebra of Operators Associated With the FIltering Problem

Consider the unbounded operators

$$L = L_o^* - \frac{1}{2} \sum_{i=1}^{p} h_i^2(x) \text{ and } h_i(x),\ i = i,\ldots,p,$$

where the operators $h_i(x)$ are considered as multiplication operators $\phi(x) \to h_i(x)\phi(x)$, on some common dense invariant domain $\mathcal{D}$ (say $\mathcal{D} = C_o^\infty(\mathbb{R}^n)$ or $\mathcal{S}(\mathbb{R}^n)$).

This Lie algebra contains important information and if it is finite-dimensional then it is a guide that a finite dimensional universal filter for computing $\rho(t,x)$ may exist. (It is not being said that if this Lie algebra is infinite-dimensional that no finite-dimensional filter exists).

Therefore, the first question that arises is: are there examples of non-linear filtering problems with finite dimensional filter algebras? The second question is: How large is this class? The answer to the first question is - yes -, but the answer to the second question appears to be is that this class is small.

Example 1. (Kalman Filtering)

$$\begin{cases} dx_t = Ax_t dt + bdw_t & A = n \times n \text{ matrix} \\ & b = n \times 1 \text{ matrix} \\ dy_t = c'x_t dt + d\eta_t & c = n \times 1 \text{ matrix} \end{cases} \tag{3.1}$$

Then

$$\begin{cases} L_0^* = \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_k} Q_{ij} - \sum_{i=1}^{n} \frac{\partial}{\partial x_i} (Ax)_i, \text{ and} \\ L = L_o^* - \frac{1}{2}(c'x)^2, \text{ where} \\ Q = bb' \end{cases} \tag{3.2}$$

Define the Hamiltonian matrix

$$E = \begin{pmatrix} -A' & cc' \\ bb' & A \end{pmatrix}, \text{ and the vector}$$

$$\alpha = \begin{pmatrix} c \\ 0 \end{pmatrix} \in \mathbb{R}^{2n}$$

and the controllability matrix

$W = [\alpha \; \vdots \; E\alpha \; \vdots \; \cdots\cdots \vdots E^{2n-1}\alpha]$ and assume that

W is non-singular.

Define $Z_1 = c'x$ and

$Z_i = [\mathrm{ad}\, L_o]^{i-1} Z_1$.

Then one can show that

$$Z_i = \sum_{j=1}^{n} (E^{i-1}\alpha)_j \, x_j + \sum_{j=1}^{n} (E^{i-1}\alpha)_{j+n} \frac{\partial}{\partial x_j}, \text{ and} \tag{3.3}$$

$$[Z_i, Z_j] = (E^{i-1}\alpha)' \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} (E^{j-1}\alpha) \tag{3.4}$$

We can then conclude that the filter algebra

$\mathscr{F} = \mathrm{span}\{L_o, Z_1, \ldots, Z_{2n}, I\}$,

where the $Z_1, \ldots, Z_{2n}$ are independent by hypothesis. Hence, $\mathscr{F}$ has dimension $2n+2$ and this algebra is isomorphic to the oscillator algebra of dimension $2n+2$.

3.1 Invariance Properties of the Lie Algebra and the Benes Problem

The filter algebra is invariant under certain transformations, namely, diffeomorphisms of the x-space and gauge transformations to be discussed below. When suitably defined this is the largest invariance group for the filter algebra. These ideas are best discussed on an example.

Consider the filtering problem:

$$\begin{cases} x_t = W_t \\ dy_t = x_t dt + d\eta_t \end{cases} \tag{3.5}$$

A basis for the filter algebra $\mathscr{F}$ is

$$\left\{ L_o, \; x, \; \frac{d}{dx}, \; I \right\},$$

where $L_o = \frac{1}{2}\frac{d^2}{dx^2} - \frac{1}{2}x^2$ and this is the 4-dimensional oscillator algebra. It is easy to see that if we perform a smooth change of coordinates $x \to \phi(x)$ then the Filter algebra gives rise to an isomorphic Lie algebra.

Now consider the example first treated by Benes [1981],

$$\begin{cases} dx_t = f(x_t)dt + dw_t \\ dy_t = x_t dt + d\eta_t, \end{cases} \tag{3.6}$$

where f is the solution of the Riccati equation:

$$\frac{df}{dx} + f^2 = ax^2 + bx + c,$$

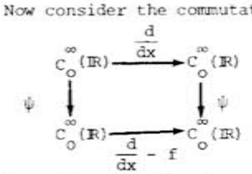
and the coefficients a,b,c are so chosen that the equation has a global solution on all of $\mathbb{R}$. We want to show that by introducing gauge transformations, we can transform the filter algebra of (3.6) to one which is isomorphic to the 4-dimensional oscillator algebra. Hence, the Benes filtering problem is essentially the same as the Kalman filtering problem considered in example 1.

To see this, first not that for (4.6)

$$[L_o, x] = \frac{d}{dx} - f,$$

where the brackets are computed on $C_o^\infty(\mathbb{R})$.

Now consider the commutative diagram:



Here Ψ is the multiplication operator $\phi(x) \to \Psi(x)\phi(x)$ and it is assumed that Ψ is invertible. Then it is easy to see that $\Psi(x) = \exp \int_0^x f(z)dz$.

Under the transformation Ψ, the operator $L_o^* = \frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{\partial}{\partial x} f$ transforms to $\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}V(x)$, where $V(x) = f_x + f^2$.

It is easy to see that the Filter algebra is isomorphic to the Lie algebra with generators

$$\left\{ \frac{1}{2}\frac{d2}{dx2} - \frac{1}{2}V(x) - \frac{1}{2}x^2, \; x \right\}.$$

We now see that if $V(x)$ is a quadratic, then this Lie algebra is essentially the 4-dimensional oscillator algebra corresponding to the Kalman Filter in Example 1.

What we have done is to introduce the gauge transformation

$$\rho(t,x) \rightarrow \Psi^{-1}(x)\rho(t,x),$$

where $\rho(t,x)$ is the solution of the Zakai equation and what we have shown is that the Filter algebra is invariant under this isomorphism.

However, for the class of models considered in (3.6) with general drifts f, the Benes problem is the only one with a finite-dimensional Lie algebra (diffusions defined on the whole real line).

There is no difficulty in generalizing these conditions to the vector case, provided f is a gradient vector field.

3.2 The Weyl Algebras and the Cubic Sensor Problem

The Weyl algebra W_n is the algebra of all polynomial differential operators $\mathbb{R}\langle x_1, \ldots, x_n, \frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\rangle$.

A basis for W_n consists of all monomial expressions

$$x^{\alpha}\frac{\partial^{\beta}}{\partial x^{\beta}} = x_1^{\alpha_1} \cdots x_n^{\alpha_n}\frac{\partial^{\beta_1}}{\partial x^{\beta_1}} \cdots \frac{\partial^{\beta_n}}{\partial x^{\beta_n}}$$

where α, β range over all multiindices $\alpha = (\alpha_1, \ldots, \alpha_n)$, $\beta = (\beta_1, \ldots, \beta_n)$. W_n can be endowed with a Lie algebra structure in the usual way. The centre of W_n, that is the ideal $\mathcal{J} = \{Z \in W_n \mid [x,Z] = 0, \ \forall X \in W_n\}$ is the one-dimensional space $\mathbb{R} \cdot 1$ and the Lie algebra $W_n/\mathbb{R} \cdot 1$ is simple.

Consider the cubic sensor filtering problem:

$$\begin{cases} x_t = w_t \\ dy_t = x_t^3 dt + d\eta_t. \end{cases} \tag{3.7}$$

Then the filter algebra generated by the operators (cf. Mitter [81], Marcus-Hazewinkel [82]) $L_o = \frac{1}{2}\frac{d^2}{dx^2} - \frac{1}{2}x^6$, and $L_1 = x^3$ is the Weyl algebra W_1.

3.3 Example with Pro-finite-dimensional Lie Algebra. (Marcus-Hazewinkel [82])

Consider the filtering problem:

$$\begin{cases} x_t = w_t \\ d\xi_t = x_t^2 dt \\ dy_t = x_t dt + dv_t \end{cases} \tag{3.8}$$

In Marcus-Mitter-Ocone [1980] it was shown that all conditional moments of ξ_t can be computed using recursive filters. For this problem $\mathcal{F}$ is generated by

$-x^2\frac{\partial \cdot}{\partial \xi} + \frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}x^2 = L_o$ and $x = L_1$. A basis for $\mathcal{F}$ is given by L_o and $x\frac{\partial^i}{\partial \xi^i}$, $\frac{\partial}{\partial x}\frac{\partial^i}{\partial \xi^i}$, $\frac{\partial^i}{\partial \xi_i}$, $i = 0,1,\ldots$. Defining $\mathcal{J}_i$ to be the ideal generated by $x\frac{\partial^i}{\partial \xi^i}$, $i = 0,1,2, \ldots$ it can be shown $\mathcal{F}$ is a pro-finite-dimensional filtered Lie algebra, solvable and $\mathcal{F}/\mathcal{J}_i$ is finite-dimensional and can be realized in terms of finite-dimensional filters corresponding to conditional statistics.

4. The Homomorphism Ansatz of Brockett.

To proceed further, we need to make a definition. By a finite-dimensional filter for a conditional statistic $\hat{\phi}_t$. (cf. Chapter 1), we mean a stochastic dynamical system driven by the observations:

$$d\xi_t = \alpha(\xi_t)dt + \beta(\xi_t)\circ dy_t \tag{4.1}$$

defined on a finite-dimensional manifold M, so that $\xi_t \in M$ and $\alpha(\xi_t)$ and $\beta(\xi_t)$ are smooth vector fields on M, together with a smooth output map

$$\hat{\phi}_t = \gamma(\xi_t), \tag{4.2}$$

which computes the conditional statistic. Equation (4.1) is to be interpreted in the Stratanovich sense. We would consider (4.1) and (4.2) as the description of a control system with inputs $\dot{y}_t$ and output $\hat{\phi}_t$. Furthermore, we may assume that (4.1) - (4.2) is minimal in the sense of Sussmann [1977]. We thus have two ways of computing $\hat{\phi}_t$ - one via the Zakai equation and the other via (4.1)-(4.2). The ansatz of Brockett says: Suppose there exists a finite-dimensional filter and consider the Lie algebra of vector fields generated by $\alpha(\xi_t)$ and $\beta(\xi_t)$ and call this Lie algebra $L(\Sigma)$. Then there must exists a non-trivial homomorphism between the filter algebra $\mathcal{F}$ and $L(\Sigma)$ such that $L_o \rightarrow \alpha$ and $h_i \rightarrow \beta_i$ where β_i is the i^{th} row of β.

Conversely, suppose that the Lie algebra $\mathcal{F}$ cannot be generated as the Lie algebra of vector-fields with smooth coefficients on some finite-dimensional manifold, then there exists no such homomorphism and hence no conditional statistic can be computed using a finite-dimensional filter.

4.1 The Kalman Filter Revisited.

It is instructive to view the Kalman Filter in the light of Brockett's Ansatz and to solve explicitly the Zaki equation. We write the solution as:

$$\rho(t,x) = [(\exp(tL_o)\exp(g_1(t)Z_1)\ldots\exp(g_{2n}(t)Z_{2n})\exp g_{2n+1}(t))\rho_o](x) \tag{4.1.1}$$

and solve for the g_i (t).

We rewrite (4.1) in a more convenient form using:

$$\exp(sZ_i f\cdot)(x) = \exp[\tfrac{1}{2}(E^{i-1}\alpha)_1'(E^{i-1}\alpha)_2 s^2 + ((E^{i-1}\alpha\cdot)_1' x)s] \times f(x+(E^{i-1}\alpha)_2 s) \qquad (4.1.2)$$

where $(E^{i-1}\alpha\cdot)_1$ are the first n entries of $E^{i-1}\alpha$ and $(E^{i-1}\alpha\cdot)_2$ are the remaining n entries.

Using this we get:

$$\text{L.H.S. of } (4.1) = \exp(\ell(t))\exp(\sum_{i=1}^{2n} g_i(t)(E^{i-1}\alpha)_1' x)\rho_o(x + \sum_{i=1}^{2n} g_i(t)(E^{i-1}\alpha\cdot)_2)$$

where

$$\ell(t) = g_{2n+1}(t) - \frac{1}{2}\sum_{i=1}^{2n} g_i^2(t)(E^{i-1}\alpha)_1'(E^{i-1}\alpha)_2$$

$$\sum_{i=1}^{2n}\sum_{j=i+1}^{2n} g_i(t)g_j(t)(E^{j-1}\alpha)_1'(E^{j-1}\alpha\cdot)_2$$

Thus, if $v(t) = \sum_{i=1}^{2n} g_i(t)E^{i-1} = \begin{pmatrix} v_1(t) \\ v_2(t) \end{pmatrix}$, where

$$v_1(t) = \sum_{i=1}^{2n} g_i(t)\ (E^{i-1}\alpha)_1$$

$$v_2(t) = \sum_{i=1}^{2n} g_i(t)(E^{i-1}\alpha)_2 \quad ,$$

we can write

$$\rho(t,x) = (\exp(t\,L_o)[\exp(\ell(t))\exp(v_1'(t)y)\rho_o(y+v_2(t))])(x) \quad . \qquad (4.1.3)$$

Now, differential equations for $g_i(t)$ can be obtained by using the Baker-Campbell-Hausdorff formula and formal differentiation of both sides of (4.1) and equating coefficients. Also, we see that

$$z(t) = \begin{pmatrix} t \\ v(t) \\ \ell(t) \end{pmatrix} \quad \text{is a sufficient statistic .}$$

Finally, using the differential equations for $g_i(t)$ we obtain:

$$\dot{z}(t) = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \dot{y}_t \begin{pmatrix} 0 \\ \exp(-Z_1(t)E)\alpha\ . \\ -[Z_{n+2}(t)\ \ldots\ Z_{2n+1}(t)\ \ (\exp(-Z_1(t)E))_1 \end{pmatrix}$$

This system computes $\rho(t,x)$.

For later use: define

$$F = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad G_1 = \begin{pmatrix} 0 \\ \exp(-Z_1(t)E)\alpha \\ -[Z_{n+2}(t)\ ---\ Z_{2n+1}(t)](\exp(-Z_1(t)E))_1 \end{pmatrix}$$

considered as vector fields. Define $G_i = [\mathrm{ad}F]^{i-1}G_1$.

The required homomorphism is:

$$L_o \to F$$

$$Z_i \to G_i \quad 1 \le i \le 2n$$

$$I \to \frac{\partial}{\partial Z_{2n+2}}$$

which in fact is an isomorphism.

As we have mentioned in Section (2.1) the crucial question in making the above results rigorous is the Campbell-Baker-Hausdorff formula (2.1) for operators. For the problem considered, this can be made rigorous by using the properties of the semi-group e^{tL_o} which has strong contractive properties (cf. Ocone [1980]).

We have referred to the homomorphism theorem as an Ansatz, because it has not been rigorously proved, in general (cf. Ocone [1980] for a discussion).

Finally, since the Lie algebra of the Kalman filter is solvable, the representation (2.1) is global. It should be emphasized the method advocated in this section provides a finite-dimensional statistic for the fundamental solution of the Zakai equation and, hence, in a certain sense the Kalman Filtering problem has a finite-dimensional sufficient statistic even for non-gaussian initial conditions.

The Benes filtering problems, being "gauge equivalent" to the Kalman Filtering problem is amenable to rigorous treatment using these same ideas.

4.2 Non-existence of Finite Dimensional Filters

Hazewinkel and Marcus [1982] have rigorously shown that the Weyl algebra W_n cannot be realized as the Lie algebra of vector fields with smooth coefficients on a finite dimensional manifold. For the cubic sensor problem it has been shown that if there exists a finite dimensional filter for a conditional statistic then there exists a homomorphism according to the Brockett Ansatz.

Combining these two results one can conclude: there exists no finite dimensional filter for computing any conditional statistic for the cubic sensor problem.

REFERENCES

1. V.E. Benes [1981], Exact Finite Dimensional Filters for Certain Diffusions with Nonlinear Drift, Stochastics, 5, pp. 65-92.

2. J.M. Bismut [1981] Mecanique Aleatoire, Springer Lecture Notes in Mathematics, Vol. 866.

3. A.E. Bryson and M. Frazier, "Smoothing for Linear and Nonlinear Dynamic Systems," Proceeding of the Optimum System Synthesis Conference, Wright-Patterson Air Force Base, Ohio, September, 1962. AST-TDR-63-119.

4. J.M.C. Clark [1978], The Design of Robust Approximations to the Stochastic Differential Equations of Nonlinear Filtering, in Communication Systems and Random Process Theory: ed. J.K. Skwirzynski, Sithoff and Noordhoff.

5. H. Cox, "Estimation of State Variables via Dynamic Programming," Proceedings of the 1964 Joint Automatic Control Conference, Stanford, CA, June, 1964, pp. 376-381.

6. M.H.A. Davis [1980], On a Multiplicative Functional Transformation Arising in Nonlinear Filtering Theory, Z. Wahr, Verw Gebiete 54, 125-139.

7. M.C. Delfour and S.K. Mitter, to appear.

8. T.E. Duncan [1967], Doctoral Dissertation, Department of Electrical Engineering, Stanford University.

9. R.P. Feynman and A.R. Hibbs [1965], Quantum Mechanics and Path Integrals, McGraw Hill, New York.

10. W. Fleming and S.K. Mitter, Optimal Stochastic Control and Pathwise Filtering on Non-Degenerate Diffusions, to appear in Stochastics.

11. W.H. Fleming and R.W. Rishel, Deterministic and Stochastic Optimal Control, Springer-Verlag, 1975.

12. M. Fujisaki, G. Kallianpur, H. Kunita, Stochastic Differential Equations of Non-linear Filtering, Osaka J. of Math., Vol. 9, 1972, 19-40.

13. Galdos, J.I. [1975], Doctoral Dissertation, M.I.T., Information and Distortion in Filtering Theory.

14. Galdos, J.I., A Lower Bound on Filtering Error with Application to Phase Demodulation, IEEE Transactions on Information Theory, July 1979, Vol. IT-24, No. 4.

15. M. Hazewinkel and S. Marcus [1982], On Lie Algebras and Finite Dimensional Filtering, to appear in Stochastics.

16. M. Hazewinkel and J.C. Willems (eds.) [1981], Stochastic Systems: The Mathematics of Filtering and Identification and Applications.

17. O. Hijab [1980], Minimum Energy Estimation, Doctoral Dissertation, University of CA, Berkeley.

18. M. Kac [1951], On Some Connections Between Probability Theory and Differential and Integral Equations, Proc. 2nd Berkeley Symposium Math. Stat. and Prob., pp. 189-215.

19. G. Kallianpur, Stochastic Filtering Theory, Application of Math. Springer Verlag, 1980, Vol. 13.

20. S. Lang [1969], Real Analysis, Academic Press, New York.

21. R.S. Lipster and A.N. Shiryayev [1977], Statistics of Random Processes, Springer-Verlag, New York.

22. S. Marcus and S.K. Mitter, and D. Ocone, Finite Dimensional Nonlinear Estimation in Continuous and Discrete Time, in Analysis and Optimization of Stochastic Systems, eds. O.L.R. Jacobs, et al., Acad. Press, 1980.

23. S.K. Mitter [1980], On the Analogy Between Mathematical Problems on Nonlinear Filtering and Quantum Physics, Richerche di Automatica, Vol. 10, No. 2, pp. 163-216.

24. S.K. Mitter [1982], Remarks on Pathwise Nonlinear Filtering, to appear in Proceedings of the Santa Margherita Conference on Optimization. Proceedings to be published by Springer-Verlag.

25. R.E. Mortensen [1968], "Maximum Likelihood Recursive Nonlinear Filtering," Journal of Optimization Theory and Applications, Vol. 2, pp. 386-394.

26. R.E. Mortensen, [1966], Doctoral Dissertation, Department of Electrical Engineering, University of California, Berkeley, CA.

27. R.E. Mortensen [1978], Nonstatistical Nonlinear Filtering, Proceedings 1978 IEEE Decision and Control Conference, San Diego, California.

28. E. Nelson, Analytic Vectors, Annals of Math., 70, 1959, 572-615.

29. D.B. Ray [1949], On Spectra of Second Order Differential Operators, Trans. Am. Math. Soc., 77, pp. 299-321.

30. S. Sheu, Solution of Certain Parabolic Equations with Unbounded Coefficients and its Application to Nonlinear Filtering, to appear.

31. B. Simon [1979], Functional Integration and Quantum Physics, Academic Press, New York.

32. H. Sussmann [1981], Rigorous Results on the Cubic Sensor Problem in Willems-Hazewinkel, Loc. cit.

33. H. Sussman: Existence and Uniqueness of Minimal Realizations of Non-Linear Systems, Math. Systems Th., 10, 1977, 263-284.

Sanjoy K. Mitter

Department of Electrical Engineering and Computer Science
and
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139