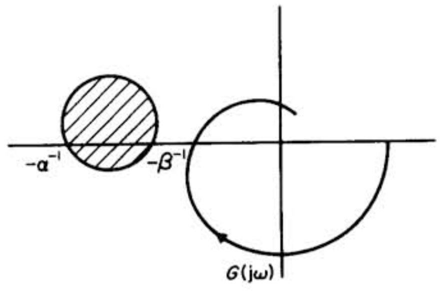

Figure 4
Off-axis circle criterion

case that α and β are positive, as in Fig. 4, it states that the feedback system is stable given that $G(s)$ is stable and there exists a disk passing through $-\alpha^{-1}$ and $-\beta^{-1}$ which is not intersected or encircled by the plot of $G(j\omega)$ for $\omega \geq 0$.

All these criteria have been generalized in some form to the multivariable case, where the signals are vectors and F and G are matrices. Graphical representations of the circle criterion can still be given, provided that $G(s)$ satisfies some form of diagonal dominance condition (Rosenbrock 1974). For the generalized Popov criterion (Jury and Lee 1965), some progress towards graphical application has been made by Shankar and Atherton (1977). In the case of the off-axis circle criterion, the multivariable version has only been shown to exclude the possibility of sustained oscillations, although it is conjectured to be sufficient for stability as well (Cook 1978).

3. Approximate Results

One of the most widely used techniques for the analysis of nonlinear feedback systems is the describing-function method (Atherton 1975). Stability criteria can be based on this approach but are unreliable because the method assumes that all signals in the feedback loop are sinusoids of the same frequency; the results can, however, be improved by taking harmonics into account (Mees and Bergen 1975).

It has also been conjectured that a nonlinear system will be stable if all the linear systems obtained by replacing the nonlinearities by all possible values within their sector bounds (Aizerman 1949) or slope bounds (Kalman 1957) are stable. These predictions can normally be relied on in practice but are not rigorously valid, and counter-examples have in fact been discovered (Willems 1970).

See also: Nonlinear Systems: Circle Criterion Extensions; Nonlinear Systems: Lyapunov Stability; Sinusoidal Input Describing Function: Evaluation and Properties; Nonlinear Systems Stability: Vector Norm Approach; Stability: Dissipativeness Concept

Bibliography

Aizerman M A 1949 On a problem concerning the stability in the large of dynamic systems. *Usp. Mat. Nauk.* **4**, 187–88

Atherton D P 1975 *Nonlinear Control Engineering.* Van Nostrand Reinhold, London

Cook P A 1978 Stability of systems containing slope-restricted nonlinearities. In: Gregson M J (ed.) *Recent Theoretical Developments in Control.* Academic Press, London, pp. 161–73

Desoer C A, Vidyasagar M 1975 *Feedback Systems: Input-Output Properties.* Academic Press, New York

Jury E I, Lee B W 1965 The absolute stability of systems with many nonlinearities. *Autom. Remote Control* **26**, 943–61

Kalman R E 1957 Physical and mathematical mechanisms of instability in nonlinear automatic control systems. *Trans. Am. Soc. Mech. Eng.* **79**, 553–66

Mees A I, Bergen A R 1975 Describing functions revisited. *IEEE Trans. Autom. Control* **20**, 473–78

Narendra K S, Taylor J H 1973 *Frequency Domain Criteria for Absolute Stability.* Academic Press, New York

Popov V M 1962 Absolute stability of nonlinear systems of automatic control. *Autom. Remote Control* **22**, 857–75

Rosenbrock H H 1974 *Computer-Aided Control System Design.* Academic Press, London

Shankar S, Atherton D P 1977 Graphical stability analysis of nonlinear multivariable control systems. *Int. J. Control* **25**, 375–88

Willems J C 1971 The generation of Lyapunov functions for input-output stable systems. *SIAM J. Control Optim.* **9**, 105–34

Willems J L 1970 *Stability Theory of Dynamical Systems.* Nelson, London

Zames G 1966 On the input–output stability of time-varying nonlinear feedback systems. *IEEE Trans. Autom. Control* **11**, 228–38; 465–76

P. A. Cook

Nonlinear Filtering and Quantum Physics

Until quite recently, the basic approach to nonlinear filtering theory was by using the "innovations method" originally proposed by Kailath in 1967 and subsequently rigorously developed by Fujisaki, Kallianpur and Kunita (1972) in their seminal paper. The difficulty with this approach is that the innovations process is not, in general, explicitly computable (except in the well-known Kalman–Bucy case). To circumvent this difficulty the construction of the filter can be divided into two parts: (a) a universal filter which is the evolution equation describing the unnormalized conditional density, the Duncan–Mortensen–Zakai (DMZ) equation, and (b) a state-output map, which depends on the statistic to be computed, where the state of the filter is the unnormalized conditional density. The reason for focusing on the DMZ equation is that it is an infinite-dimensional bilinear system driven by the incremental observation process, and a much simpler object than the conditional density equation (which is a nonlinear equation) and can be treated using geometric ideas. Moreover, it was noticed by the present author that this equation bears striking similarities to the equations arising in (Euclidean) quantum mechanics and it was felt that many of the ideas and methods used there could be used in this context. The ideas and methods referred to here are the functional integration view of Feynman (for a modern exposition see Glimm and Jaffe 1981). In many senses, this viewpoint has been remarkably successful—although the results obtained so far have been of a negative nature. Nevertheless the recent work has given us a deeper understanding of the DMZ equation which has been essential for progress in nonlinear filtering, as well as in stochastic control. The variational interpretation of nonlinear filtering given by Fleming and Mitter (1982) and Mitter (1983a,b) and the work on the partially observable stochastic control problem by Fleming and Pardoux (1982) can be considered to have arisen from the "state-space" interpretation given to the filter.

1. The Filtering Problem Considered, and the Basic Questions

We consider the signal–observation model:

$$\left.\begin{aligned} dx_t &= b(x_t)\,dt + \sigma(x_t)\,dw_t \qquad x(0)=x_0 \\ dy_t &= h(x_t)\,dt + d\eta_t \end{aligned}\right\} \quad 0 \leq t \leq 1 \tag{1}$$

where x, w and y are $\mathbb{R}^n$, $\mathbb{R}^m$ and $\mathbb{R}^p$-valued processes, and it is assumed that the drift velocities b, h and the density ρ are, respectively, vector-valued, matrix-valued and vector-valued functions which are smooth (which means C^∞-functions). It is further assumed that the stochastic differential equation (1) has a global solution in the sense of Ito. It is further assumed that x_t and η_t are independent and

$$E\int_0^1 |h(x_t)|^2\,dt < \infty$$

For much of our considerations, the function $h(\cdot)$ will be a polynomial.

It is well known that the unnormalized conditional density $\rho(t,x)$ (where we have suppressed the $y(\cdot)$ and ω-dependence) satisfies the DMZ equation:

$$d\rho(t,x) = \left(\mathscr{L}^* - \frac{1}{2}\sum_{i=1}^{p} h_i^2(x)\right)\rho(t,x)\,dt + \sum_{i=1}^{p} h_i(x)\rho(t,x)\circ dy_t \tag{2}$$

where

$$\mathscr{L}^*\phi = \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}(\sigma(x)\sigma'(x))_{ij}\phi - \sum_{i=1}^{n}\frac{\partial}{\partial x_i} b_i(x)\phi \tag{3}$$

and the $\circ$ denotes the Stratanovich differential. It is imperative that we consider Eqn. (2) as a Stratanovich differential equation, since the Ito integral, because it "points to the future," is not invariant under smooth diffeomorphisms of the x space, and we want to study Eqn. (2) in an "invariant manner."

We think of $\rho(t,\cdot)$ as the "state" of the filter, previously referred to as the universal part of the filter. If ϕ, say, is a bounded, continuous function then the filter typically is required to compute $E\{\phi(x_t|\mathscr{F}_t^y\}$, where $\mathscr{F}_t^y = \sigma\{y_s; 0 \leq s \leq t\}$. If we write

$$\hat{\phi}_t \triangleq E\{\phi(x_t)|\mathscr{F}_t^y\}$$

then $\hat{\phi}_t$ is obtained from $\rho(t,x)$ by integration:

$$\hat{\phi}_t = \int_{\mathbb{R}^n}\phi(x)\rho(t,x)\,dx \Big/ \int_{\mathbb{R}^n}\rho(t,x)\,dx \tag{4}$$

$\hat{\phi}_t$ will be referred to as a "conditional statistic," and no matter what $\hat{\phi}_t$ we wish to compute, $\rho(t,x)$ serves as a "sufficient statistic."

2. Pathwise Nonlinear Filtering and Analogy of the DMZ Equation to Schrödinger Equation

The DMZ equation bears a close resemblance to the Schrödinger equation with a random potential if we formally rewrite it as

$$\frac{d\rho}{dt}(t,x) = \mathscr{L}^*\rho(t,x) - \frac{1}{2}\left[\sum_{i=1}^{p} h_i^2(x) - 2\left(\sum_{i=1}^{p} h_i(x)(\dot{y}_t)_i\right)\right]\rho(t,x) \tag{5}$$

where $\dot{y}_t$ is the formal derivative of y_t. However, since the operator $\rho(t,x) \to \Sigma_{i=1}^p h_i(x)(\dot{y}_t)_i\rho(t,x)$ is a multiplication operator we can transform this equation by utilizing a time-dependent gauge transformation. To simplify the notation, we assume y is scalar and hereafter we use subscript x to denote the partial derivative with respect to x.

This leads us to ideas of pathwise nonlinear filtering (Clark 1978, Davis 1980, Mitter 1980).

There is as yet no theory of nonlinear filtering for observations

$$Y(t) = h[x(t)] + \tilde{w}(t) \tag{6}$$

where $\tilde{w}$ is a physical wide-band noise and hence smooth. (To conform to a partial differential equation viewpoint we are writing processes as $\tilde{w}(t)$, etc., instead of $\tilde{w}_t$, etc.) Define $Y(t) = \dot{y}(t)$ and $\tilde{w}(t) = \dot{\eta}(t)$ where the overdot denotes differentiation. Then (6) can be written as

$$dy(t) = h[x(t)]\,dt + d\tilde{\eta}(t) \tag{7}$$

or

$$y(t) = \int_0^t h[x(t)]\,dt + \eta(t) \tag{8}$$

Equation (8) is a mathematical model of the physical observation (6) where the wide-band noise $\tilde{w}(t)$ has been approximated as "white noise" $\dot{\eta}(t)$, and hence $\eta(t)$ is a Wiener process.

Now, if we wish to compute

$$E\{\phi(x(t))|\mathcal{F}_t^y\}$$

$$= \text{functional of } y \text{ a.s. Wiener measure}$$

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 Eqn. (2) by a gauge transformation.

Define $q(t,x)$ by

$$\rho(t,x) = \exp[h(x)y(t)]q(t,x) \tag{9}$$

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

$$q_t = (L^y)^* q + \tilde{V}^y q \quad \text{where}$$

$$L^y\phi = L\phi - y(t)a(x)h_x(x)\phi_x, \quad a(x) = \sigma(x)\sigma'(x)$$

$$V^y(t,x) = \tfrac{1}{2}h^2(x) - y(t)Lh(x) + \tfrac{1}{2}y^2(t)h_x'(x)a(x)h_x(x) \tag{10}$$

Equation (10) is the pathwise nonlinear filtering equation and should be solved for each (physical) observation path y. Equation (10) can be written explicitly as

$$\left.\begin{aligned} & q_t = \tfrac{1}{2}\mathrm{tr}[a(x)q_{xx} + g^y(x,t)q_x + {}^y(x,t)q] \\ & q(0,x) = p^0(x), \text{ the density of } x(0) \text{ where} \\ & \left.\begin{aligned} & g^y = -b + y(t)ah_x + \gamma,\ \gamma_j = \sum_{i=1}^n \frac{\partial a_{ij}}{\partial x_j},\ j = 1,2,\ldots,n \\ & V^y = \tilde{V}^y - \mathrm{div}(b - y(t)ah_x) + \frac{1}{2}\sum_{i,j=1}^n \frac{\partial^2 a_{ij}}{\partial x_i \partial x_j} \end{aligned}\right\} \end{aligned}\right\} \tag{11}$$

Equation (11) can be considered to be a rigorous version of Eqn. (5).

3. *Schrödinger Operators, Diffusion Operators and Time Reversibility*

Under suitable hypotheses (e.g., uniform ellipticity, growth conditions on g^y, V^y bounded above) we can express the solution of (11) as a Feynman–Kac integral:

$$q(t,x) = E_x\left\{p^0(x_t)\exp(L_t)\exp\left[\int_0^t V^y(t,x_s)\,ds\right]\right\} \tag{12}$$

where

$$L_t = \int_0^t a^{-1}(x_s)g^y(x_s,s)\,dw_s - \frac{1}{2}\int_0^t |a^{-1}(x_s)g^y(x_s,s)|^2\,ds$$

and where E_x denotes expectation with respect to the path space of ξ, where ξ satisfies

$$\left.\begin{aligned} d\xi_t &= \sigma(\xi_t)\,dw_t \\ \xi_0 &= x \end{aligned}\right\} \tag{13}$$

We may ask whether the functional integration (12) can be reduced to quadratures. This leads us to consider the relation between Schrödinger operators and diffusion operators, or what is equivalent, the relation between the Feynman–Kac formula and the Girsanov formula (see Mitter 1980).

Let us suppose that $V:\mathbb{R}^n \to \mathbb{R}$ be measurable, bounded below and tends to $+\infty$ as $|x| \to \infty$ and consider the Schrödinger operator $H = -\Delta + V$ where Δ is the n-dimensional Laplacian. The H defines a self-adjoint operator on $L^2(\mathbb{R}^n; dx)$ which is bounded below and the lower bound λ of the spectrum of H is an eigenvalue of H. Let $\psi(x)$ be the corresponding eigenfunction of H, the so-called ground state, and assume $\psi(x) > 0$. We normalize $\psi(x)$, that is,

$$\int_{\mathbb{R}^n} |\psi(x)|^2\,dx = 1$$

and define the probability measure

$$d\mu = |\psi(x)|^2\,dx$$

and the unitary operator

$$U: L^2(\mathbb{R}^n; dx) \to L^2(\mathbb{R}^n; d\mu(x))$$

$$: f \to \psi^{-1} f$$

If we define the Dirichlet form for $f, g \in C_c^\infty(\mathbb{R}^n)$,

$$\delta(f,g) = \frac{1}{2}\int_{\mathbb{R}^n} \nabla f(x)\,\nabla g(x)\,dx \tag{14}$$

then a calculation shows

$$\delta(f,g) = (\mathcal{L}f, g)_\mu \tag{15}$$

where $(\,,)_\mu$ denotes the scalar product in $L^2(\mathbb{R}^n; d\mu)$ and $\mathcal{L}$ is the diffusion operator (self-adjoint, positive):

$$\mathcal{L}\psi = -\tfrac{1}{2}\Delta\psi + \nabla b\cdot\nabla\psi \qquad b = -\log\psi \tag{16}$$

Now since ψ satisfies

$$-\tfrac{1}{2}\Delta\psi(x) + V(x)\psi(x) = 0 \tag{17a}$$

then assuming $\lambda = 0$ we get

$$V(x) = \tfrac{1}{2}(|\nabla b(x)|^2 - \Delta b(x)] \tag{17b}$$

where differentials have to be interpreted in the sense of distributions.

Let ξ_t satisfy the stochastic differential equation

$$d\xi_t = -\nabla b(\xi_t)\,dt + dw_t \qquad \xi_0 = x \tag{18}$$

where w_t is standard Brownian motion. Define

$$\begin{aligned} L_t &= \exp\left[-\int_0^t \nabla b(\xi_s)\cdot d\xi_s - \frac{1}{2}\int_0^t \nabla b(\xi_s)|^2\,ds\right] \\ &= \exp\left[-\int_0^t \nabla b(\xi_s)\cdot dw_s + \frac{1}{2}\int_0^t \nabla b(\xi_s)|^2\,ds\right] \end{aligned} \tag{19}$$

which can be shown to be a $\mathscr{B}_t$ martingale where $\beta_t = \sigma\{w_s|0 \leqslant s \leqslant t\}$. Then if μ_ω denotes Wiener measure and if we denote a new probability measure μ_ξ on the path space of ξ by $d\mu_\xi/d\mu_\omega = L_t$, then from the Girsanov theorem ξ_t is a Brownian motion under the measure μ_ξ and hence we can write the solution of

$$\left.\begin{aligned} \frac{d\rho}{ds} + \mathscr{L}\rho = 0 \quad 0 \leqslant s \leqslant t \\ \rho(t,x) = \psi(x) \end{aligned}\right\} \tag{20}$$

as $\rho(s,x) = \bar{E}_{sx}[\psi(\xi_t)]$ where $\bar{E}$ denotes integration with respect to μ_ξ.

On the other hand, by the generalized Ito-differential rule,

$$db(\xi_t) = \nabla b(\xi_t)\,d\xi_t + \tfrac{1}{2}\Delta b(\xi_t)\,dt$$

and hence (19) reduces to

$$\begin{aligned} L_t &= \exp\left[-b(\xi_t) + b(\xi_0) - \frac{1}{2}\int_0^t |\nabla b(\xi_s)|^2\,ds \right. \\ &\qquad \left. + \frac{1}{2}\int_0^t \Delta b(\xi_s)\,ds\right] \\ &= \psi(dz_0)^{-1}\psi(\xi_t)\exp\left[-\int_0^t V(\xi_s)\,ds\right] \end{aligned}$$

and therefore

$$\psi(x) = E_x[\psi(\xi_t)\exp\left[-\int_0^t V(\xi_s)\,ds\right] \tag{21}$$

where E denotes expectation with respect to Wiener measure and we have derived the Feynman–Kac formula.

Equation (18) denotes a stationary, reversible Markov process—since $\xi(t)$ and $\xi(-t)$ are stochastically equivalent with invariant measure μ. Thus with the Hamiltonian $H = -\Delta + V$ having V satisfying the hypotheses given above, we have a unique stationary, reversible Markov process built into it.

These ideas have a bearing on nonlinear filtering (Benes 1981). Consider the scalar nonlinear filtering problem

$$\left.\begin{aligned} dx_t &= -b(x_t)\,dt + dw_t \\ dy_t &= x_t\,dt + d\eta_t \end{aligned}\right\} \tag{22}$$

and assume that

$$b(x_t) = f_x(x_t)$$

and f satisfies

$$\tfrac{1}{2}(|f_x|^2 - f_{xx}) = \tfrac{1}{2}x^2 \tag{23}$$

The DMZ equation for this problem is

$$d\rho(t,x) = (\mathscr{L}^* - \tfrac{1}{2}x^2)\rho(t,x)\,dt + x\rho(t,x)\cdot dy_t \tag{24}$$

We may write its solution as (using previous considerations)

$$\begin{aligned} \rho(t,x) = E_x\left\{\exp\left[-f(x_t) + f(x_0) - \frac{1}{2}\int_0^t x_s^2\,ds\right]\right. \\ \left. \times \exp\left[\int_0^t x_s\,dy_s - \frac{1}{2}\int_0^t x_s^2\,ds\right]\right\} \end{aligned} \tag{25}$$

Because of the quadratic nature of the potential this function space integration can essentially be reduced to Gaussian integrals (see Mitter 1983a,b for an illuminating discussion). Indeed the filter for this problem is essentially a Kalman filter.

4. Variational Interpretation of Nonlinear Filtering

We now give a stochastic variational interpretation of nonlinear filtering in the spirit of the work of Feynman (Feynman and Hibbs 1965). We do this by associating a stochastic control problem with the DMZ equation. This section provides a justification of the ideas of Kalman on the duality between filtering and control. The original ideas of this section are due to Fleming and Mitter (see Fleming and Mitter 1982, Pardoux 1981). We follow the exposition of Pardoux and for simplicity consider the scalar case.

Let us formally denote the differential dy_s as $\dot{y}_s ds$ and consider the DMZ equation

$$\left.\begin{aligned} \frac{\partial\rho}{\partial t} &= (L^* - \tfrac{1}{2})\rho(t,x) + h(x)\rho(t,x)y_s \\ \rho(0,x) &= \rho_0(x) \end{aligned}\right\} \tag{26}$$

Now $\rho(t,x)$ admits the factorization

$$\rho(t,x) = l(t,x)p(t,x)$$

where $p(t, x)$ is the density of the x process and $l(t, x)$ is the likelihood function given by

$$l(t, x) = \mathrm{E}\left\{\exp\left[-\int_0^t h(x_s)\dot{y}_s\,\mathrm{d}s - \frac{1}{2}\int_0^t |h(x_s)|^2\,\mathrm{d}s\right]\bigg| x_t = x\right\} \quad (27)$$

Then a calculation shows that l satisfies the equation

$$\left.\begin{aligned}\frac{\partial l}{\partial t} &= (\bar{L} - \tfrac{1}{2}h^2)l + h\dot{y}_s l\\ l(0) &= 1\end{aligned}\right\} \quad (28)$$

where

$$\bar{L} = \frac{1}{2}a(x)\frac{\partial^2}{\partial x^2} + \bar{b}\frac{\partial}{\partial x}$$

with

$$\bar{b} = \frac{1}{p}\frac{\partial}{\partial x}[a(x)p] - b$$

Here $\bar{L}$ is the infinitesimal generator of the time-reversed x process with $x_t = x$.

Now consider the transformation

$$\bar{L}(t, x) = \exp[-S(t, x)]$$

Then S satisfies the Bellman equation

$$\left.\begin{aligned}\frac{\partial S}{\partial t} &= \bar{L}S - \frac{1}{2}a\left(\frac{\partial S}{\partial x}\right)^2 + \tfrac{1}{2}|h - \dot{y}_s|^2 - \dot{y}_s^2\\ S(0) &= 0\end{aligned}\right\} \quad (29)$$

Denote by $\bar{x}$ the reverse x Markov process conditioned on $x_t = x$. Then (29) corresponds to the following stochastic control problem:

$$\left.\begin{aligned}\mathrm{d}\bar{x}_s + [\bar{b}(\bar{x}_s) + u_s]\mathrm{d}s + \sigma(\bar{x}_s)\oplus\mathrm{d}\bar{w}_s = 0, \quad s \leq t\\ \bar{x}_t = x\end{aligned}\right\} \quad (30)$$

where $\oplus$ denotes the backward Ito differential and, where the control u_s is to be chosen as a Markovian feedback control to minimize the cost function

$$\mathrm{E}\left\{\int_0^t \left[a^{-1}(\bar{x}_s)|u_s|^2 + \tfrac{1}{2}|h(\bar{x}_s) - \dot{y}_s|^2 - \dot{y}_s^2\right]\mathrm{d}s\right\} \quad (31)$$

and $S(t, x)$ is the optimal value function of this stochastic control problem.

In the situation that the dynamics are linear and the observation map is linear we have a linear filtering problem and the stochastic control problem of Eqns. (30–31) corresponds to a linear quadratic Gaussian problem with full observations. But the theory of this problem is essentially the same as the linear quadratic deterministic optimal control problem. This explains in a clear manner the duality principle first enunciated by Kalman.

This stochastic variational interpretation can be effectively used to construct maximum *a posteriori* density filters and maximum likelihood filters and allow us to give a derivation of the extended-Kalman filter (Mitter 1983a, b).

5. Geometric Theory of Nonlinear Filtering

In the introduction we have suggested that the fact that the DMZ equation is in infinite-dimensional bilinear equation allows us to develop a geometrical theory of nonlinear filtering. This geometrical theory, originally independently suggested by Brockett and Clark (1980), Brockett (1980) and Mitter (1980) was motivated by the desire to measure the complexity of nonlinear filters and to discover whether finite-dimensional filters existed for nonlinear problems. The present exposition follows Mitter (1983b).

To proceed further, we need to make a definition. By a *finite-dimensional* filter for a conditional statistic $\hat{\phi}_t$ we mean a stochastic dynamical system derived by the observation

$$\mathrm{d}\xi_t = \alpha(\xi_t)\,\mathrm{d}t + \beta(\xi_t)\circ\mathrm{d}y_t$$

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 $\hat{\phi}_t = \gamma(\xi_t)$, which computes the conditional statistic. Equation (5) is to be interpreted in the Stratanovich sense for reasons we have mentioned above. We shall also assume that the stochastic dynamical system of Eqns. (5, 6) is minimal in the sense of Sussmann (1977).

For the definition and properties of Lie algebras and Lie groups the reader is referred to the Appendix (see *Lie Brackets; Lie Groups: Controllability*).

5.1 Lie Algebra of Operators Associated with the Filtering Problem

Consider the Lie algebra generated by the unbounded operators

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

where the operators $\mathscr{L}$ and $h_i(x)$, with the h_i considered as multiplication operators $\phi(x) \to h_i(x)\phi(x)$, act on some common dense invariant domain $\mathscr{D}$, say $\mathscr{D} = C_0^\infty(\mathbb{R}^n)$ or $\mathscr{S}(\mathbb{R}^n)$.

This Lie algebra contains important information, and if it is finite dimensional it is a guide that a finite-dimensional universal filter for computing $\rho(t, x)$ may exist.

Care should be taken in interpreting this statement. First, referring to the definition of a finite-dimensional filter in Eqn. (5), there is a Lie algebra of vector fields associated with it which in general is infinite dimensional. Therefore, the fact that the Lie algebra $\mathscr{L}\mathrm{A}\{L, h_1, \ldots, h_p\}$ is infinite dimensional does not preclude the filtering problem having a finite-dimensional solution. Secondly, even if $\mathscr{L}A\{L, h_1, \ldots, h_p\}$ is finite-dimensional it does not mean that a finite-dimensional filter exists. The reason for this is that constructing the filter requires integrating the Lie algebra, and it is a well-known fact from the theory of unitary representations of Lie groups that not all Lie algebra representations extend to a group representation (see the Appendix). However, it is still a good question to ask whether examples of filtering problems exist where the Lie algebra $\mathscr{L}A\{L, h_1, \ldots, h_p\}$ is finite dimensional and also how big is this class. The answer to the first part of this question is positive but the answer to the second part appears to be that this class is small.

EXAMPLE 1. The case of Kalman filtering is now discussed

$$\left.\begin{aligned} dx_t &= Ax_t\,dt + b\,dw_t \quad && A = n\times n \text{ matrix} \\ & && b = n\times 1 \text{ matrix} \\ dy_t &= c'x_t\,dt + dW_t \quad && c' = n\times 1 \text{ matrix} \end{aligned}\right\} \tag{32}$$

Then

$$\left.\begin{aligned} \mathscr{L}^* &= \frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2}{\partial x_i\partial x_j}Q_{ij} - \sum_{\partial x_i}^{n}\frac{\partial}{\partial x_i}(Ax)_i, \qquad Q = bb' \\ L &= \mathscr{L}^* - \tfrac{1}{2}(c'x)^2 \end{aligned}\right\} \tag{33}$$

Define the Hamiltonian matrix

$$E = \begin{bmatrix} -A' & cc' \\ bb' & A \end{bmatrix}$$

the vector

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

and the controllability matrix

$$W = [\alpha : E\alpha \ldots E^{2n-1}\alpha]$$

and assume that W is nonsingular. Define $Z_1 = c'x$ and

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

Then one can show that

$$Z_i = \sum_{j=1}^{n}(E^{i-1}\alpha)_j x_j + \sum_{j=1}^{n}(E^{i-1}\alpha)_{j+n}\frac{\partial}{\partial x_j} \tag{34}$$

$$[Z_i, Z_j] = [E^{i-1}\alpha]\begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix}[E^{j-1}\alpha] \tag{35}$$

and $\mathscr{F} = \mathrm{span}\ \{L,\ Z_1, \ldots, Z_{2n}, I\}$, 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 $2n+2$ (see the Appendix).

5.2 Invariance Properties of the Lie Algebra and the Benes Problem

The filter algebra is invariant under certain transformations, namely, diffeomorphisms on the x space and gauge transformations to be discussed below. These ideas are best discussed in an example.

Consider the filtering problem

$$\left.\begin{aligned} x_t &= w_t \\ dy_t &= x_t\,dt + d\eta_t \end{aligned}\right\} \tag{36}$$

A basis for the filter algebra $\mathscr{F}$ is $\{L, x, d/dx, I\}$ where

$$L = \frac{1}{2}\frac{d^2}{dx^2} - \tfrac{1}{2}x^2$$

and this is the four-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, and two filtering problems with isomorphic Lie algebras should have the same filter.

EXAMPLE 2. The following was first treated by Benes (1981).

$$\left.\begin{aligned} dx_t &= f(x_t)\,dt + dw_t \\ dy_t &= x_t\,dt + d\eta_t \end{aligned}\right\} \tag{37}$$

where f is the solution of the Riccati equation

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

and the coefficients a, b and 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 (37) to one which is isomorphic to the four-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 note that for (37)

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

where the brackets are computed on $C_0^\infty(\mathbb{R})$.

Now consider the commutative diagram

$$\begin{array}{ccc} C_0^\infty(\mathbb{R}) & \xrightarrow{\ \frac{d}{dx}\ } & C_0^\infty(\mathbb{R}) \\ \Psi\downarrow & & \downarrow\Psi \\ C_0^\infty(\mathbb{R}) & \xrightarrow[\ \frac{d}{dx}-f\ ]{} & C_0^\infty(\mathbb{R}) \end{array}$$

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

$$\mathcal{L}^* = \frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{\partial}{\partial x} f$$

transforms to

$$\frac{1}{2}\frac{\partial^2}{\partial x^2} - \tfrac{1}{2}V(x)$$

where

$$V(x) = \frac{df}{dx} + f^2$$

It is easy to see that the filter algebra $\mathcal{F}$ is isomorphic to the Lie algebra with generators

$$\frac{1}{2}\frac{d^2}{dx^2} - \tfrac{1}{2}V(x) - \tfrac{1}{2}x^2 \qquad x$$

We now see that if $V(x)$ is a quadratic, then this Lie algebra is essentially the four-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 DMZ equation, and what we have shown is that the filter algebra is invariant under this isomorphism.

However, for the class of scalar models considered in (11) with general drifts f, the Benes problem is the only one with a finite-dimensional Lie algebra (we restrict ourselves to diffusions defined on the whole real line). For further details on this point the reader should consult Ocone (1980).

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

5.3 The Weyl Algebras and the Cubic Sensor Problem

The Weyl algebra W_n is the algebra of all polynomial differential operators

$$\mathbb{R}\left(x_1, \ldots, x_n, \frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\right)$$

A basis for W_n consists of all monomial expressions

$$\frac{X^\alpha \partial^\beta}{\partial x^\beta} = X_1^{\alpha_1} \ldots X_n^{\alpha_n} \frac{\partial^{\beta_1}}{\partial x_1^{\beta_1}} \cdots \frac{\partial^{\beta_n}}{\partial x_n^{\beta_n}}$$

where α, β range over all multi-indices

$$\alpha = (\alpha_1, \ldots, \alpha_n), \qquad \beta = (\beta_1, \ldots \beta_n)$$

W_n can be endowed with a Lie algebra structure in the usual way. The center of W_n, that is the ideal $\mathcal{J} = \{Z \in W_n | [x, Z] = 0,\ \forall x \in w_n$, 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

$$x_t = W_t, \qquad dy_t = x_t^3\, dt + d\eta_t$$

Then the filter algebra $\mathcal{F}$ generated by the operators

$$L = \frac{1}{2}\frac{d^2}{dx^2} - \tfrac{1}{2}x^6, \qquad \mathcal{L}_1 = x^3$$

is the Weyl algebra $W_1/\mathbb{R}$. A proof of this can be constructed by performing calculations similar to that of Avez and Heslot (1979).

5.4 Example with Pro-Finite-Dimensional Lie Algebra

Consider the filtering problem (Hazewinkel and Marcus 1982)

$$x_t = W_t$$
$$d\xi_t = x_t^2\, dt$$
$$dy_t = x_t\, dt + dv_t$$

It can be shown that all conditional moments of ξ_t can be computed using recursive filters. For this problem $\mathcal{F}$ is generated by

$$-x^2\frac{\partial}{\partial \xi} + \frac{1}{2}\frac{\partial}{\partial x^2} - \tfrac{1}{2}x^2 = \mathcal{L} \qquad x = \mathcal{L}_1$$

A basis for $\mathcal{F}$ is given by $\mathcal{L}$ and

$$x\frac{\partial^i}{\partial \xi^i}, \qquad \frac{\partial\partial^i}{\partial\partial \xi^i}, \qquad \frac{\partial^i}{\partial \xi_i}, \quad i = 0, 1, \ldots$$

Defining $\mathcal{J}_i$ to be the ideal generated by $x(\partial^i/\partial\xi^i)$, $i = 0, 1, 2, \ldots$ it can be shown that $\mathcal{F}$ is a pro-finite-dimensional filtered Lie algebra which is solvable, and that $\mathcal{F}/\mathcal{J}_i$ is finite dimensional and can be realized in terms of finite-dimensional filters corresponding to conditional statistics.

REMARK 1. Other examples of finite-dimensional filters can be constructed by combining the attributes of the Benes example considered in Sect. 5.2 and the above example. Thus, in the above example the process x_t may be replaced by

$$dx_t = f(x_t)\, dt + dw_t$$

where f satisfies

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

and a, b and c are chosen so that this equation has a global solution. Then it can be shown that all conditional moments of ξ_t can be computed using finite-dimensional recursive filters (Ocone *et al.* 1982).

REMARK 2. The Lie algebraic and representation approach to the filtering problem is really concerned with the "classification" question for filters. The actual construction of the filter can apparently be achieved using probabilistic techniques.

5.5 Existence and Nonexistence of Finite-Dimensional Filters and the Homomorphism Ansatz of Brockett

Earlier we have given the definition of a finite-dimensional filter. We would consider this definition to be the description of a control system with inputs y_t and output $\hat{\phi}_t$. Furthermore, as we have said we may assume that this representation is minimal in the sense of Sussmann. We thus have two ways of computing $\hat{\phi}_t$—one via the DMZ equation and the other via the control system. 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 exist a nontrivial antihomomorphism between the filter algebra $\mathcal{F}$ and $L(\Sigma)$ such that $L \to \alpha$ and $h_i \to \beta_i$ where β_i is the ith 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.

The Brockett ansatz suggests a possible strategy for obtaining finite-dimensional filters for computing certain conditional statistics. Suppose we are in the situation of the example given in Sect. 5.4, that is, the Lie algebra $\mathcal{F}$ is pro-finite-dimensional. Since $\mathcal{F}/\mathcal{J}_i$ is finite-dimensional it has a faithful finite-dimensional representation (by Ado's theorem) and hence can be realized with linear vector fields on a finite-dimensional manifold which may give rise to a bilinear filter computing some conditional statistic. However, what statistic this filter computes is in general difficult to determine, and one has to resort to indirect and probabilistic techniques for this determination. One should also remark again that $\mathcal{F}$ (or any of its quotients) need not be finite-dimensional for a finite-dimensional filter to exist.

5.6 The Kalman Filter Revisited

It is instructive to view the Kalman filter in the light of the above discussion and solve explicitly the corresponding DMZ equation. We shall consider the special case where the filter Lie algebra is generated by

$$\left\{\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} - \tfrac{1}{2}x^2, \frac{\mathrm{d}}{\mathrm{d}x}, x, I\right\}$$

For a rigorous justification of the calculations which follow see Ocone (1980).

The basic idea is to do the following formal calculation which needs to be justified.

Suppose that we want to solve the evolution equation

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = L_1\rho + u(t)L_2\rho \tag{38}$$

where L_1 and L_2 are in general unbounded linear operators and $u(t)$ is a given continuous function. Let us assume that the Lie algebra of operators $A\{L_1, L_2\}$ has a finite set of generators $\{L_1, L_2, \ldots, L_d\}$. We try a solution

$$\rho(t) = \exp(g_1(t)L_1)\exp(c_2(t)L_2)\ldots\exp(g_d(t)L_d)\rho(0) \tag{39}$$

where $\rho(0)$ is the initial condition. For ideas similar to this in the context of ordinary stochastic differential equations, see Kunita (1981).

Differentiating the above, we obtain

$$\begin{aligned}\frac{\mathrm{d}\rho}{\mathrm{d}t} &= \dot{g}_1(t)L_1\rho + \dot{g}_2(t)\exp(g_1(t)L_1)L_2 \\ &\quad\times \exp(g_2(t)L_2)\ldots\exp(g_d(t)L_d)\rho(0) \\ &\quad+ \dot{g}_d(t)\exp(q_1(t)L_1)\ldots L_d\exp(q_d(t)L_d)\rho(0)\end{aligned}$$

Now, we use the Campbell–Baker–Hausdorff formula: for $1 \leqslant i, j \leqslant d$,

$$\exp(tL_j)L_i = \sum_{m=1}^{d} c_m^{i,j}(t)L_m\exp(tL_j)$$

repeatedly to obtain

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = F_1(g(t),\dot{g}(t))L_1\rho + \cdots + F_d(g(t),\dot{g}(t))L_d\rho \tag{40}$$

for some nonlinear functions F_i of $g(t) = (g_1(t), \ldots, g_d(t))$ and $\dot{g}(t)$. For (40) to define a solution of (38), we need

$$\begin{aligned}&F_1(g(t), g(t)) = 1\\ &F_2(g(t), \dot{g}(t)) = u(t)\\ &F_j(g(t), \dot{g}(t)) = 0 \quad \text{for } j > 2\end{aligned}$$

For the Kalman filter problem considered, one obtains (formally)

$$\begin{aligned}\dot{g}_1(t) &= 1\\ \dot{y}(t) &= \dot{g}_2(t)\cosh g_1(t) + \dot{g}_3(t)\sinh g_1(t)\\ 0 &= \dot{g}_2(t)\sinh g_1(t) + \dot{g}_3(t)\cosh g_1(t)\\ 0 &= \dot{g}_4(t) - \dot{g}_3(t)g_2(t)\\ g_i(0) &= 0 \quad i = 1, 2, 3, 4\end{aligned}$$

One can explicitly solve the above set of equations to obtain

$$\begin{aligned}g_2(t) &= \int_0^t \cosh(s)\,\mathrm{d}y(s)\\ g_3(t) &= -\int_0^t \sinh(s)\,\mathrm{d}y(s)\\ g_4(t) &= \int_0^t \sinh(s)\cosh(s)\,\mathrm{d}s - \int_0^t g_2(s)\sinh(s)\,\mathrm{d}y(s)\end{aligned}$$

where we have now used stochastic integrals.

Substituting the above in (39) and using

$$\exp(tL_1)\phi(s) = \int_{-\infty}^{\infty} G(x,y,t)\phi(y)\,\mathrm{d}y, \quad t \geqslant 0$$

where

$$G(x,y,t) = (2\pi \sinh t)^{-1/2} \times \exp[-\tfrac{1}{2}\coth(x^2+y^2) + xy/\sinh t)]$$

one obtains

$$\rho(x,t) = \int_{-\infty}^{\infty} k(z,t) \times \exp(-\tfrac{1}{2}p^{-1}(t)[x-m(t)]^2)\rho_0(z)\,\mathrm{d}z$$

where $p(t) = \tanh t$ and

$$m(t) = \frac{z}{\cosh t} + \int_0^t \frac{\sinh s}{\cosh t}\,\mathrm{d}y(s)$$

(and $k(z,t)$ is a function which can be computed), which is the familiar Kalman filter solution.

The essential point in proving the above results rigorously is to note that

$$-\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \tfrac{1}{2}x^2$$

generates a positively-preserving hypercontractive semigroup and that the operators

$$-\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + \tfrac{1}{2}x^2, \quad x, \quad \frac{\mathrm{d}}{\mathrm{d}x}$$

have a common dense set of analytic vectors.

Finally, since the Lie algebra corresponding to the Kalman filter is solvable, Eqn. (39) is a global representation for the solution. We remark that the Benes problem considered in Sect. 3.1 can be integrated in exactly the same fashion. Note also that this method computes the fundamental solution of the DMZ equation and hence these ideas can be applied to solve Kalman filtering problems with non-Gaussian initial conditions.

5.7 Nonexistence of Finite-Dimensional Filters

In an earlier part of this section we have suggested a strategy for obtaining finite-dimensional filters when the Lie algebra of the filter has a "good" ideal structure using the Brockett homomorphism ansatz. We have also remarked how the same ansatz may lead to negative results.

Now, in Sect. 3.2 we have shown that for the cubic-sensor problem the Lie algebra of the filter is isomorphic to the $W_1/\mathbb{R}$. Hazewinkel and Marcus (1982) have shown that $W_1/\mathbb{R}$ cannot be realized as the Lie algebra of vector fields with smooth coefficients on a finite-dimensional smooth manifold. On the other hand, Sussmann (1981) has shown that if there is a finite-dimensional filter for a conditional statistic then there exists a nonzero homomorphism of Lie algebras according to the Brockett prescription. Some further work combining these two ideas shows that no conditional statistic for the cubic-sensor problem can be computed using finite-dimensional filters.

We conjecture that essentially similar results can be proved for the following class of filtering problems:

$$\mathrm{d}x_t = f(x_t)\,\mathrm{d}t + \mathrm{d}w_t$$
$$\mathrm{d}y_t = x_t\,\mathrm{d}t + \mathrm{d}y_t$$

Suppose that f satisfies

$$\frac{\mathrm{d}f}{\mathrm{d}x} + f^2 = V(x)$$

where $V(x)$ is an even-positive polynomial. Then the Lie algebra for this filtering problem is an algebra which is isomorphic to the Weyl algebra $W_1/\mathbb{R}$, and hence all the above results of this section will hold.

5.8 Some Positive Results

There have been some recent positive results using the Lie algebra formalism. One such result is concerned with the asymptotic expansion in ϵ of the unnormalized conditional density for the filtering problem

$$\mathrm{d}x_t = ax_t\,\mathrm{d}t + \mathrm{d}w_t$$
$$\mathrm{d}y_t = [x_t + \epsilon(x_t)^k]\,\mathrm{d}t + \mathrm{d}y_t \quad k \geqslant 1$$
$$y_0^\epsilon = 0 \quad \cdot\rho_0(x) \text{ Gaussian}$$

where ϵ is some small positive answer.

For this class of problems it has been shown (Sussmann 1982) that the various terms in the formal asymptotic expansion of $\rho^\epsilon(t,x)$ can be computed by finite-dimensional filters using the ideas developed in this section.

We close this section with a remark on the identification problem for linear stochastic dynamical systems. These problems can be viewed as nonlinear filtering problems and lead to Lie algebras which are known as "current algebras" in mathematical physics. The integration of these Lie algebras in a rigorous manner has recently been performed Hazewinkel *et al.* (1982).

6. Acknowledgement

This work has been supported by the US Air Force Office of Scientific Research under Grant AFOSR-82-0135A.

Appendix A.

A.1 On Lie Algebras, Lie Groups and Representations.

For most of this article, the C^∞-manifold that is of interest is $\mathbb{R}^n$ (which is covered by a single coordinate system).

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

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

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

(c) it satisfies the Jacobi identity

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

where $X, Y, Z \in \mathscr{L}$.

EXAMPLE 3. Let $M_n(\mathbb{R})$ be the algebra of $n \times n$ matrices over $\mathbb{R}$. If we denote $[X, Y]$ by $XY - YX$, where XY is the usual matrix product, then this commutator defines a Lie algebra structure on $M_n(\mathbb{R})$.

EXAMPLE 4. Let $\mathscr{X}(M)$ denote the C^∞ vector fields on a C^∞ manifold M. $\mathscr{X}(M)$ is a vector space over $\mathbb{R}$ and a $C^\infty(M)$ module. (Recall, a vector field X on M is a mapping: $M \to T_p(M)$: $p \to x$ where $p \in M$ and $T_p(M)$ is the tangent space to the point p in M). We can give a Lie algebra structure to $\mathscr{X}(M)$ by defining

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

with the C^∞ functions in a neighborhood of p, and

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

Both of 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 $c_{rsp} \in \mathbb{R}$ $(1 \leq r, s, p \leq n)$ such that

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

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,

(a) $c_{rsp} + c_{srp} = 0 \qquad (1 \leq r, s, p \leq n)$

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

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 subalgebra if $[K, K] \subseteq K$, and ideal if $[\mathscr{L}, K] \subseteq K$.

If $\mathscr{L}$ and $\mathscr{L}'$ are Lie algebras over $\mathbb{R}$ and $\pi: \mathscr{L} \to \mathscr{L}': X \to \pi(X)$, 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 $\ker \pi$ 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 any 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

$$D(ab) = (Da)b + a(Db), \qquad a, b \in \mathscr{U}$$

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 $\mathscr{U}$. The notion of a representation of a Lie algebra is very important.

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

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

(a) π is linear,

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

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

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

ad X is a derivation of $\mathscr{L}$ and $X \to \operatorname{ad} X$ 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

$$(x, y) \to xy: G \times G \to G$$

and the mapping $x \to x^{-1}: G \to G$ 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

(a) $\dim \mathscr{I}_j = m - j$, $0 \leq j \leq m$,

(b) $\mathscr{I}_0 = \mathscr{L} \supseteq \mathscr{I}_1 \supseteq \cdots \mathscr{I}_m = 0$ and

(c) $[\mathscr{L}, \mathscr{I}_j] \subseteq \mathscr{I}_{j+1}$, $0 \leq j \leq m - 1$.

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

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

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

We then get a sequence $\mathcal{D}^0 g \supseteq \mathcal{D}^1 g \supseteq \cdots$ of subalgebras of g; g is said to be solvable if $\mathcal{D}^p g = 0$ for some $p \geq 1$.

EXAMPLE 5. Let $n \geq 0$ 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.

EXAMPLE 6. The real Lie algebra with basis

$$(h, p_1, \ldots, p_n, q_1, \ldots q_n, Z)$$

satisfying the bracket relations

$$[h, p_i] = q_i \qquad [h, q_i] = p_i \qquad [p_i q_i] = Z$$

(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 nontrivial ideals. An infinite-dimensional Lie algebra $\mathcal{L}$ is called pro-finite-dimensional and filtered if there exists a sequence of ideals $\mathcal{I}_1 \supset \mathcal{I}_2 \cdots$ such $\mathcal{L}/\mathcal{I}_i$ finite-dimensional for all i and $\cap \mathcal{I}_i = \{0\}$.

A.2 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 τ 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, $\tau: G \rightarrow L(H)$ satisfies

$$\tau(g_1 g_2) = \tau(g_1)\tau(g_2) \qquad g_1, g_2 \in G$$

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

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

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

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

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

$$\tau(\exp(t_1 X_1) \ldots \exp(t_d S_d)) = \exp(t_1 \pi(X_1)) \ldots \exp(t_d \pi(x_d))$$

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

$$\exp(tA_j)A_i = \sum_{n=0}^{\infty} \frac{t^n}{n!} [\mathrm{ad} A_j]^n A_i \exp(tA_j)$$

holds for $A_j = \pi(X_j)$, $1 \leq j, j \leq d$.

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 $\{-ix, \mathrm{d}(\cdot)/\mathrm{d}x, \ -i\}$ 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 a unitary group representation, nevertheless these ideas will serve as a guide for integrating the Lie algebras arising in filtering theory.

See also: Nonlinear Filtering: Iterated Ito Integral Expansions; Nonlinear Estimation; Realization Theory, Nonlinear; Nonlinear Discrete-Time Systems: Algebraic Theory; Lie Brackets

Bibliography

Avez A, Heslot A 1979 L'algèbre de Lie des polynômes en les coordonnées canoniques munie du crochet de Poisson. *C.R. Hebd. Seances Acad. Sci. Paris Ser. A* **288**, 831–33

Benes V E 1981 Exact finite dimensional filters for certain diffusions with non-linear drift. *Stochastics* **5**, 65–92

Brockett R W 1980 Remarks on finite dimensional nonlinear estimation. In: Analyse des systemes. *Asterisque,* **75–76**, 199–205

Brockett R W, Clark J M C 1980 The geometry of the conditional density equation. In: Jacobs O L R *et al.* (eds.) *Analysis and Optimization of Stochastic Systems.* Academic Press, New York, pp. 299–309

Clark J M C 1978 The design of robust approximations to the stochastic differential equations of nonlinear filtering. In: Skwirzynski J K (ed.) *Communication Systems and Random Process Theory.* Sijthoff and Noordhoff, Rockville

Davis M H A 1980 On a multiplicative functional transformation arising in nonlinear filtering theory. *Z. Wahr. Verw. Geb.* **54**, 125–39

Feynman R P, Hibbs A 1965 *Quantum Mechanics and Path Integrals.* McGraw-Hill, New York

Fleming W H, Mitter S K 1982 Optimal control and nonlinear filtering for nondegenerate diffusion processes. *Stochastics* **8**, 63–78

Fleming W H, Pardoux E 1982 Existence of optimal controls for partially-observed diffusions. *SIAM j. control* **20**, 261–85

Fujisaki M, Kallianpur G, Kunita H 1972 Stochastic differential equations for the nonlinear filtering problem. *Osaka J. Math.* **1**, 19–40

Glimm J, Jaffe A 1981 *Quantum Physics.* Springer, Berlin

Hazewinkel M, Krishnaprasad P A, Marcus S I 1982 Current algebras and the identification problem. Technical Report; University of Maryland

Hazewinkel M, Marcus S I 1982 On Lie algebras and finite dimensional filtering. *Stochastics* **7**, 26–62
Kunita H 1981 On the decomposition of solutions of stochastic differential equations. In: *Stochastic Integrals*, Springer Lecture Notes in Mathematics Vol. 851. Springer, Berlin
Marcus S I, Mitter S K, Ocone D 1980 Finite dimensional nonlinear estimation for a class of systems in continuous and discrete time. In: Jacobs O L R *et al.* (eds.) *Analysis and Optimization of Stochastic Systems*. Academic Press, New York, pp. 387–406
Mitter S K 1980 Filtering theory and quantum fields. In: Analyse des Systemes. *Astérisque* **75–76**, 199–205
Mitter S K 1981 Non-linear filtering and stochastic mechanics. In: Hazewinkel M, Willems J C (eds.) *Stochastic Systems: The Mathematics of Filtering and Identification and Applications*. Reidel, Dordrecht, pp. 479–503
Mitter S K 1983a Approximations to nonlinear filtering. In: Bucy R S, Moura J (eds.) *Nonlinear Stochastic Systems*. Reidel, Dordrecht
Mitter S K 1983b Geometric theory of nonlinear filtering. In: *Methodes et outils de mathematiques en theorie des systems*.
Mitter S K, Moro A (eds.) 1983 *Nonlinear Filtering and Stochastic Control*, Springer Lecture Notes in Mathematics Vol. 972. Springer, Berlin
Ocone D 1980 Topics in nonlinear filtering theory. Ph.D. dissertation, MIT
Ocone D, Baras J S, Marcus S I 1982 Explicit filters for diffusions with certain nonlinear drifts. *Stochastics* **8**, 1–16
Pardoux E 1981 The solution of the nonlinear filtering equation as a likelihood function. In: *Proc. 20th IEEE Conf. Decision and Control*. IEEE, New York
Sussmann H J 1977 Existence and uniqueness of minimal realizations of nonlinear systems. *Meth. Syst. Theory* **1**, 263–84
Sussmann H J 1981 Rigorous results on the cubic sensor problem. In: Hazewinkel M, Willems J C (eds.) *Stochastic Systems: The Mathematics of Filtering and Identification and Applications*, Nato Advanced Study Institute Series. Reidel, Dordrecht
Sussmann H J 1982 Approximate finite-dimensional filters for some nonlinear problems. *Stochastics* **7**, 183–204

S. K. Mitter

Nonlinear Filtering: Iterated Ito Integral Expansions

Iterated (or multiple) stochastic integrals (see *Stochastic Integrals and Stochastic Calculus*) of a Brownian motion process were originally defined by Ito (1951) and Wiener (1938) in a study of the structure of functionals of a Brownian motion (see *Brownian Motion*). The Ito–Wiener expansion theorem, their main result, shows that a functional $\Phi(B(\cdot))$, where Φ maps continuous functions on $[0, \infty)$ to the reals and $B(\cdot) = (B(t): t \geq 0)$ represents Brownian motion paths, may be expressed by an infinite series whose terms are iterated integrals. In this expansion, the iterated integrals play a role analogous to that of polynomials in Taylor series expansions and (because iterated integrals of different orders are statistically orthogonal) to that of an orthogonal Hilbert-space basis. As a result, the iterated integral technique provides a clear and usable methodology for expanding functionals of Brownian motion.

This article shows how this method of iterated Ito integral expansions applies to the additive Brownian noise-filtering model. In this problem we are required to construct estimates of a signal process $X(t)$ based on observing $(Y(s): 0 \leq s \leq t)$, where

$$Y(t) = \int_0^t h[X(s)]\,\mathrm{d}s + W(t)$$

and $W(t)$ is a Brownian motion. Such estimates are just functionals $\Phi(Y(s): 0 \leq s \leq t)$ of the $Y(\cdot)$ process; hence construction of estimates is best achieved within a framework for representation and expansion of functionals. Because $Y(t)$ is an Ito process closely related to Brownian motion (see Sect. 2), iterated Ito integrals provide an appropriate technique.

The first part of this article presents the fundamentals of the theory of multiple stochastic integrals. The second section treats their application to the additive noise model, and the third section their application to filtering polynomials of a Gauss–Markov process. The ideal goal of filtering is to calculate the conditional expectation

$$\hat{\phi}(t) = \mathrm{E}\{\phi[X(t)] \mid Y(s): 0 \leq s \leq t\}$$

which is the estimate minimizing mean square error. Failing this, one desires a broadly applicable method for constructing suboptimal estimates. The results of Sect. 2 show that multiple integral expansions help to address both issues. They can be used both to provide an explicit representation of $\hat{\phi}(t)$ and to build a theory of suboptimal estimation using finite, multiple integral expansions.

The additive noise model involves an essentially nonstationary observation process. Katznelson and Gould (1962) study the case in which signal and observation are stationary, ergodic processes and employ Volterra, rather than Ito, iterated integral expansions. Schetzen (1980) may be consulted for general information on the use of functional expansions.

1. Iterated Integrals and Functional Expansions

In this section, $B(t, \omega)$ denotes a Brownian motion on the probability space $(\Omega, \mathcal{F}, P)$, and $\mathcal{F}(t)$ is the σ algebra generated by $B(s)$, $0 \leq s \leq t$. The Ito–Wiener theory studies the space

$$L^2[\mathcal{F}(t)] = \{\Phi: \Phi(\omega) \text{ is an } \mathcal{F}(t)\text{-measurable r.v. and } \mathrm{E}\Phi^2 < \infty\}$$

Since an $\mathcal{F}(t)$-measurable Φ may be written in the form $\Phi(\omega) = \psi[B(s, \omega)\ 0 \leq s \leq t]$ for some functional ψ, the space $\mathrm{L}^2[\mathcal{F}(t)]$ consists of the square-integrable