Nonlinear Filtering of Diffusion Processes

A Guided Tour

by

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

1. Introduction

In this paper we give a guided tour through the development of nonlinear filtering of diffusion processes. The important topic of filtering of point processes is not discussed in this paper.

There are two essentially different approaches to the nonlinear filtering problem. The first is based on the important idea 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 nonlinear 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. LIPSTER-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 focused 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]. A detailed account of the second viewpoint can be found in the lectures given by Kunita, Pardoux and Mitter in the CIME lecture Notes on Nonlinear Filtering and Stochastic Control [1982] and in HAZEWINKEL-WILLEMS [1981].

2. Basic Problem Formulation

To simplify the exposition we consider the situation where all processes are scalar-valued.

Let $(\Omega, \mathscr{F}, P)$ be a complete probability space and let $\mathscr{F}_t$, $t \in [0, T]$ be an increasing family of sub σ-fields of $\mathscr{F}$. Let ξ_t be an $\mathscr{F}_t$-adapted process, considered to be the signal process and consider the observation process y_t given by

$$(2.1) \qquad y_t = \int_0^t h_s ds + n_t,$$

where n_t is an $\mathscr{F}_t$-Weiner process and it is assumed that $\sigma(n_t - n_s | t \geq s)$ is independent of the past of the joint signal-observation process $\sigma(y_u, h_u | u \leq s)$. Information about the ξ-process is contained in h which satisfies

$$E\int_0^t |h_s|^2 ds < \infty \quad \forall\ t \in [0,T].$$

Let $\mathscr{F}_t^y = \sigma(y_s | s \leq t)$. Then the filtering problem consists of computing

$$E\left(\phi(\xi_t) | \mathscr{F}_t^y\right) \overset{\Delta}{=} \hat{\xi}_t,$$

where ϕ, say is a bounded, continuous function (indeed any function such that the conditional expectation makes sense).

3. The Innovations Method

The fundamental paper of Fujisaki-Kallianpur-Kunita proceeds as follows:

Define the innovations process:

$$(3.1) \qquad \nu_t = y_t - \int_0^t \hat{h}_s ds \text{ , where}$$

$$(3.2) \qquad \hat{h}_s \overset{\Delta}{=} E(h_s | \mathscr{F}_s^y).$$

Then it can be shown that

(i) ν_t is $\mathscr{F}_t^y$-Wiener

(ii) $\sigma\left\{\nu_u - \nu_t | u \geq t\right\}$ is independent of $\mathscr{F}_t^y$.

In general it is not true that:

(Innovations Property) $\mathscr{F}_t^y = \mathscr{F}_t^\nu$ (Cirelson counter-example).

If is assumed that h_t and n_t are independent then innovations property holds (cf. ALLINGER-MITTER [1981]).

However even without the innovations property holding, it can be proved:

Every square-integrable $\mathscr{F}_t^y$-martingale m_t can be represented as:

$$(3.3) \qquad m_t = m_0 + \int_0^t \phi_s d\nu_s \text{ , where}$$

ϕ_s is jointly measurable, adapted to

$\mathcal{F}_s^y$ and $\int_0^T E|\phi_s|^2 ds < \infty$.

To proceed further, let us assume that ξ_t is a continuous semimartingale

(3.4) $\xi_t = \xi_0 + \int_0^t f_s ds + v_t$, where v_t is a square integrable $\mathcal{F}_t$-martingale. Then

(3.5) $m_t = \hat{\xi}_t - \xi_0 - \int_0^t \hat{f}_s ds$ [1] is a $\mathcal{F}_t^y$- martingale and from the previous result

(3.6) $m_t = \int_0^t \phi_s d\nu_s$, where ϕ_s can be identified as

(3.7) $\phi_s = \widehat{\xi_s h_s} - \hat{\xi}_s \hat{h}_s + \hat{D}_s$, where D_s is a $\mathcal{F}_s$-predictable process and $\int_0^t D_s ds = \langle v,n\rangle_t$, where $\langle\ ,\ \rangle_t$ denotes the quadratic variation.

In case h_t and n_t are independent there is an essential simplification using the innovations property.

For example, one would have the representation

(3.8) $$\hat{\xi}_t = \hat{\xi}_0 + \int_0^t \frac{\partial}{\partial s} E(\xi_t \nu_s) d\nu_s + \int_0^t\int_0^{s_1} \frac{\partial^2}{\partial s_1 \partial s_2} E(\xi_t \nu_{s_1} \nu_{s_2}) d\nu_{s_1} d\nu_{s_2} + \cdots$$

To proceed further let us assume that ξ_t is a Markov diffusion process satisfying the Ito equation

(3.9) $$d\xi_t = f(\xi_t)\, dt + g(\xi_t)\, dB_t \text{, and}$$

(3.10) $$h_s = h(\xi_s) \text{ .}$$

Let us also assume that B_t and n_t are independent (assumed throughout the rest of the paper). Then using the Ito differential rule for $\phi \in C_b^2$, one gets

(1) $\hat{}$ always denotes conditional expectation with respect to $\mathcal{F}_t^y$

(3.11) $$\hat{\phi}(\xi_t) = \hat{\phi}(\xi_0) + \int_0^t \widehat{L\phi(\xi_s)}\, ds + \int_0^t \left(\widehat{\phi(\xi_s) h_s} - \hat{\phi}(\xi_s)\hat{h}_s\right) d\nu_s,$$

where L is the generator of the diffusion process.

4. <u>The Innovations Method Continued</u>: <u>Kushner-Stratonovich Equation</u>.[1]

Let $\Pi_t(dy,\omega)$ denote the conditional distribution of ξ_t given $\mathcal{F}_t^y$. Let $\phi \in C_b^2$ and denote by

(4.1) $$\Pi_t(\phi) = \int \phi(y)\, \Pi_t(dy,\omega)$$

Then Π_t satisfies the Nonlinear Stochastic Partial Differential Equation (Kushner-Stratonovich Equation)

(4.2) $$\begin{cases} \Pi_t(\phi) = \Pi_0(\phi) + \int_0^t \Pi_s(L\phi)ds \\ \qquad + \int_0^t \left[\Pi_s(\phi h_s) - \Pi_s(\phi)\Pi_s(h_s)\right]\left(dy_s - \Pi_s(h_s)ds\right). \end{cases}$$

We think of the filter as a dynamical system and we think of equation (4.2) as the <u>input-output equation</u> of the filter, the input being the observations y_s and the output the conditional distribution Π_t.

5. <u>Zakai Equation</u>

Let ρ_t be a continuous stochastic process with values in the set of finite positive measures on $\mathbb{R}$. Denote by

(5.1) $$\rho_t(\phi) = \int \phi(y)\, \rho_t\, (dy) \text{ .}$$

Consider the equation

(5.2) $$\rho_t(\phi) = \rho_0(\phi) + \int_0^t \rho_s\, (L\phi)ds + \int_0^t \rho_s\, (h_s\phi)dy_s \text{ .}$$

(weak form of Zakai equation)

Now it can be proved:

(1) The development here follows Kunita [1982].

If ρ_t is a solution of (5.2) then $\Pi_t(\phi) = \frac{\rho_t(\phi)}{\rho_t(1)}$ is a solution of equation (4.2).

Morever we have the Feynman-Kac formula:

$$(5.3)\quad \rho_t(\phi) = E\left[\phi(\xi_t)\exp\left(\int_0^t h_s\,dy_s - \frac{1}{2}\int_0^t h_s^2\,ds\right)\Big|\mathscr{F}_t^Y\right]$$

For later use (and throughout the rest of the paper) we shall consider the Weak form of the Zakai equation in Stratonovich form:

$$(5.4)\quad \rho_t(\phi) = \rho_0(\phi) + \int_0^t \rho_s\left[(L - \tfrac{1}{2}h^2)\phi\right] ds + \int_0^t \rho_s(h\phi)\circ dy_s$$

If the solution $\rho_t(dy)$ has a smooth density $\rho_t(y)$ it satisfies the Zakai equation:

$$(5.5)\qquad d\rho_t = (L^* - \tfrac{1}{2}h^2)\,\rho_t(y)\,dt + h\,\rho_t(y)\circ dy_t \quad ,$$

where * denotes formal adjoint.

$\rho_t(y)$ has the interpretation of an unnormalized density and is to be thought of as the "state" of the filter.

To compute conditional statistics, we need the state-output equation

$$(5.6)\qquad \hat{\phi}(x_t) = \frac{\int_{\mathbb{R}}\phi(x)\rho_t(x;y)\,dx}{\int_{\mathbb{R}}\rho_t(x;y)\,dx}$$

The fundamental problem of nonlinear filtering is the "invariant" study of equation (5.5). The analytic difficulty of this problem stems from the following:

(i) In most interesting situations the operator $x \longmapsto h(x)$ is unbounded.

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

6. Pathwise Nonlinear Filtering

The ideas of this section are due to CLARK [1978], DAVIS [1980], and MITTER [1980].

There is as yet no theory of nonlinear filtering where the observations are:

$$(6.1)\qquad y_t = h(\xi_t) + \eta_t \ ,$$

where η_t is physical wide-band noise and hence smooth. Define $y_t = \dot{Y}_t$ and $\eta_t = \dot{N}_t$ where $\cdot$ denotes differentiation. Then (6.1) can be written as:

$$(6.2)\qquad dY_t = h(\xi_t)\,dt + dN_t \ , \text{ or}$$

$$(6.3)\qquad Y_t = \int_0^t h(\xi_s)\,ds + N_t \ .$$

Equation (6.3) is a mathematical model of the physical observation (6.1) where the wide band noise η_t has been approximated as "white noise" $\dot{N}_t$ and hence N_t is a Wiener process.

Now, if we wish to compute

$$E\left(\phi(\xi_t)\ |\mathscr{F}_t^Y\right) = \text{Functional of } Y \text{ a.s. Wiener measure}$$

then this filter does not accept the physical observation y. The idea is to construct a suitable version of the conditional expectation so that the performance of the filter as measured by the mean-square error remains close to 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 (5.5) by a suitable transformation (gauge transformation in the language of physicists).

Define $q_t(x;y)$ by

$$(6.4)\qquad \rho_t(x;y) = \exp\left(y_t h(x)\right) q_t(x;y) \quad .$$

Then q_t satisfies the parabolic partial differential equation

$$(6.5)\qquad \frac{\partial q}{\partial t} = \frac{1}{2}\,a(x)\,\frac{\partial^2 q}{\partial x^2} + b^y(x,t)\,\frac{\partial q}{\partial t} + V^y(x,t)\,q$$

where $\quad a(x) = g^2(x), \quad b^y = -f + y_t\,a\,\frac{dh}{dx} + \frac{da}{dx}$,

$$V^y = \frac{1}{2}h^2 - y_t Lh + \frac{1}{2}y_t^2\,a\left(\frac{dh}{dx}\right)^2 - \frac{df}{dx} + y_t a\frac{dh}{dx} + \frac{1}{2}\frac{d^2a}{dx^2}$$

Equation (6.5), the pathwise filter equation, is now needed to be solved for each y, the observation (which can be taken to be a physical observation).

7. Existence and Uniqueness Results for the Zakai Equation.

In case the observation h is bounded or linear, existence and uniqueness results for the Zakai equation has been given by PARDOUX [1982] by studying equation (5.5) directly. Existence, uniqueness and estimate of tail distributions for equation (5.5) including unbounded observations (in the scalar case all polynominal h) have been given by FLEMING-MITTER [1982], using stochastic control arguments. This approach studies equation (6.5) by transforming it into a Bellman-Hamilton-Jacobi equation (cf. paper of FLEMING this volume) using an exponential transformation and then showing that the measure constructed from q coincides with the measure given by formula (5.3). For other literature on this problem see the bibliography in FLEMING-MITTER (1982). See also PARDOUX [1981] and MITTER [1982] for an interpretation of the exponential transformation in the context of nonlinear filtering. For related variational considerations see MITTER [1980], BISMUT [1981], HIJAB [1980].

8. Geometrical Theory of Nonlinear Filtering.

How can one answer the question when two filtering problems have identical solutions? How can one decide whether the Zakai equation admits a finite-dimensional statistic?

The starting point of this analysis is the Zakai euqation (5.5) in Stratonovich form. Consider the two operators:

$$\mathscr{L}_0 = L^* - \frac{1}{2} h^2$$

$$\mathscr{L}_1 = h$$

where the two operators are considered as formal differential operators on $C_0^\infty(\mathbb{R})$. Denote by $\mathscr{F}$ the Lie algebra of operators generated by $\mathscr{L}_0$ and $\mathscr{L}_1$ under the standard bracket operation. This Lie algebra is invariant under (i) smooth change of coordinates $x \mapsto \phi(x)$ and (ii) gauge transformations $\rho \longmapsto \phi\rho$ where ϕ is a C^∞-function which is invertible. Indeed, if the "invariance group" of the Zakai equation

is suitably defined then the above constitutes the largest invariance group of the Lie algebra $\mathscr{F}$. The insight here is that two filtering problems with isomophic Lie algebras are likely to have the same filters. We say likely, since for a proof, analytic considerations such as the existence of a common dense set of analytic vectors must come into play. For a rigorous analysis in specific situations see OCONE [1980].

By a finite dimensional filter for the conditional statistic $\hat{\phi}(x_t)$ we mean a stochastic dynamical system.

$$d\psi_t = \alpha(\psi_t)dt + \beta(\psi_t) \circ dy_t \tag{8.1}$$

where α and β are smooth vector fields on some finite-dimensional smooth manifold, and a state-output equation

(8.2) $\quad \hat{\phi}(x_t) = \gamma(\psi_t)$, with γ a smooth real-valued function.

The idea of studying the Lie algebra is independently due to BROCKETT [1981] (cf. the bibliography of earlier Brockett papers cited there) and MITTER [1980] (cf. the bibliography of earlier paper of Mitter cited there) and the idea of implementing the Lie algebra $\mathscr{F}$ as a Lie Algebra of vector fields is due to Brockett. The first examples of finite-dimensional filters for nonlinear filtering problems were constructed by BENES [1981] using functional integral methods. For a generalisation of Benes results see OCONE-BARAS-MARCUS [1982].

In most situations the Lie algebra F is infinite dimensional (cf. IGUSA [1981]) and in many situations simple. If the Lie algebra F is infinite dimensional it does not necessarily mean that a finite-dimensional filter does not exist. For a precise result in this direction see HAZEWINKEL-MARCUS [1982], SUSSMANN [1981].

9. Final Remarks.

There are other topics of interest and the theory represented by the second point of view is far from complete.

(i) Asymptotic Expansions: see BLANKENSHIP-LIU-MARCUS [1982]. Much work remains to be done here.

(ii) Lower Bounds on Nonlinear Filtering: see article of BOBROVSKY-ZAKAI in HAZEWINKEL-MARCUS [1981], (and the bibliography cited there). The stochastic control interpretation of FLEMING-MITTER [1982] should be important here.

(iii) Filtering on Manifolds: see DUNCAN [1977] for a beginning.

(iv) Smoothness of densities: see KUNITA [1982].

(v) For a partial solution to the vector "cubic-sensor" problem, see DELFOUR-MITTER [1982].

REFERENCES

1. Allinger D., and Mitter, S.K. (1981) New Results on the Innovations Problem of Nonlinear Filtering, Stochastics, vol 4, pp. 339-348.

2. Beneš, V.E., (1981) Exact Finite Dimensional Filters for Cetain Diffusions with Nonlinear Drift, Stochastics, vol 5, pp. 65-92.

3. Bismut, J.M., (1981) Mecanique Aléatoire, Springer Lecture Notes in Mathematics, vol. 866.

4. Blankenship, G.L., Liu, C., and Marcus, S., (1982) Asymptotic Expansions and Lie Algebras for Some Nonlinear Filtering Problems, to appear IEEE Trans on Auto. Control.

5. Brockett, R.W., (1981) Paper in Hazewinkel-Willems, loc. cit.

6. Clark, J.M.C., (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.

7. CIME Lectures on Nonlinear Filtering and Stochastic Control (1982), to be published as Springer Lecture Notes, eds. S.K. Mitter & A. Moro.

8. CIME Lectures on Nonlinear Filtering and Stochastic Control [1982], to be published as Springer Lecture Notes, ed. A. Moro.

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

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

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

12. Duncan, T.E., [1977] Some Filtering Results on Riemannian Manifolds, Information and Control, 35, pp. 182-195.

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

14. Fleming, W.H., and Mitter, S.K., [1982] Optimal Control and Nonlinear Filtering for Nondegenerate Diffusion Processes, to appear Stochastics.

15. Fujisaki, M., Kallianpur, G., and Kunita, H., [1972] Stochastic Differential Equations for the Nonlinear Filtering Problem, Osaka J. of Mathematics, Vol. 9, pp. 19-40.

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

17. Hazewinkel, M., and Willems, J. C., [1982] Stochastic Systems: The Mathematics of Filtering and Identification and Applications: eds. M. Hazewinkel and J. C. Willems, pp. 479-503, D. Reidel Publishing Company.

18. Hijab, O, [1980] Minimum Energy Estimation, Doctoral Dissertation, University of Calfornia, Berkeley.

19. Igusa, J.I., [1981] On Lie Albegras Generated by Two Differential Operators in Lie Groups and Manifolds, pp. 187-195. Birkhauser Verlag Boston.

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

21. Kallianpur, G., [1980] Stochastic Filtering Theory, Application of Math. Springer Verlag, vol. 13.

22. Kunita, H., [1982] Stochastic Partial Differential Equations Connected with Nonlinear Filtering, to appear in Proceedings on CIME School on Nonlinear Filtering and Stochastic Control.

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

24. Mitter, S.K., [1980] On the Analogy Between Mathematical Problems of Nonlinear Filtering and Quantum Physics, Ricerche di Automatica, vol. 10, No. 2, pp. 163-216.

25. Mitter, S.K., [1981] Non-linear Filtering and Stochastic Mechanics in Stochastic Systems: The Mathematics of Filtering and Identifca-tion and Applications: eds. M. Hazewinkel and J.C. Willems, pp. 479-503. D. Reidel Publishing Company.

26. Mitter, S.K.,[1982] Lectures Given at CIME School on Nonlinear Filtering and Stochastic Control, Cortona, Italy, July 1981. Proceedings to be published by Springer-Verlag.

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

28. Ocone, D., [1980] Topics in Nonlinear Filtering Theory, Doctoral Dissertation, M.I.T.

29. Ocone, D., Baras, J.S., and Marcus, S., [1982] Explicit Filters for Certain Diffusions with Nonlinear Drift, to appear Stochastics.

30. Pardoux, E. [1981] The Solution of the Nonlinear Filtering Equation as a Likelihood Function, Proceedings of 20th CDC Conference, San Diego, California.

31. Pardoux, E. [1982] Equations du Filtrage non lineaire de la Prediction et du Lissage, to appear in Stochastics.

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

33. Sussmann, H., [1981] Rigorous Results on the Cubic Sensor Problem in WILLEMS-HAZEWINKEL, loc. cit.

Note on uniqueness of semigroup associated with Bellman operator

Makiko Nisio

Department of Mathematics, Kobe University

Kobe, 657, Japan

1. Introduction. As a mathematical formulation of Bellman principle, we define a nonlinear semigroup, using the value function of stochastic optimal control. [2, 4, 7] The generator of this semigroup is the Bellman operator. The purpose of this note is to consider the uniqueness of semigroup with generator of Bellman operator, appealing to results of integral solution (Benilan solution)[1] of Cauchy problem.

First we recall our nonlinear semigroup. Let Γ be a compact convex subset of R^k, called a control region. $B(t)$, $t \geq 0$, denotes an n-dimensional Brownian motion on a probability space (Ω, F, P). Any Γ-valued $\sigma_t(B)$-progressible measurable process is called an admissible control. $\mathfrak{A}$ denotes the totality of admissible controls. For $U \in \mathfrak{A}$, we consider the following controlled stochastic differential equation

(1. 1) $$dX(t) = \alpha(X(t), U(t))dB(t) + \gamma(X(t), U(t))dt$$
$$X(0) = x(\in R^n).$$

we assume that $n \times n$ symmetric non-negative definite matrix $\alpha(x, u)$ and n-vector $\gamma(x, u)$ satisfy the following conditions (A1) and (A2).

(A1) $|h(x, u)| \leq b$ for any x, u

(A2) $|h(x, u) - h(y, v)| \leq K|x - y| + \rho(|u - v|)$

where b and K are constants and ρ continuous on $[0, \infty)$ with $\rho(0) = 0$. Then there exists a unique solution $X(t) = X(t; x, U)$, called the response for U. The value function is defined as follows

(1. 2) $$S(t, x, \phi) = \sup_{U \in \mathfrak{A}} E_x \int_0^t e^{-\int_0^s c(X(\theta), U(\theta))d\theta} f(X(s), U(s))ds$$