REPRINT

D. REIDEL PUBLISHING COMPANY
DORDRECHT-HOLLAND / BOSTON-U.S.A.

SOME NONLINEAR FILTERING PROBLEMS ARISING IN RECURSIVE IDENTIFICATION

P.S. Krishnaprasad* and Steven I. Marcus+

*University of Maryland, College Park, Maryland 20742
+University of Texas, Austin, Texas 78712

Abstract: In this paper, we outline an intrinsic formulation of the identification problem of linear system theory. The nonlinear filtering problems which appear in this way essentially fall into four distinct classes, distinguished by their estimation algebra. In principle, it is possible to explicitly solve the identification problem in the 'hyperbolic cases' using classical methods from the theory of partial differential equations. This is illustrated by an example which indicates the required sufficient statistics for solving the identification problem.

1. INTRODUCTION

Consider the linear stochastic differential system,

$$\begin{aligned} dx_t &= A(p)x_t dt + b_1 u_t dt + b_2 dw_t \\ dy_t &= \langle q, x_t \rangle \, dt + dv_t \end{aligned} \tag{1.1}$$

where u_t denotes a known input function, $\{w_t\}$ and $\{v_t\}$ are independent Brownian motion processes, $\{x_t\}$ and $\{y_t\}$ are respectively the state and measured output processes. For reasons of identifiability we let,

$$A(p) = \begin{bmatrix} 0 & 1 & & & \\ & & 1 & & \\ & & & \ddots & \\ & & & & 1 \\ -p_1 & , -p_2, & \cdots & & -p_n \end{bmatrix}$$

M. Hazewinkel and J. C. Willems (eds.), Stochastic Systems: The Mathematics of Filtering and Identification and Applications, 299–304.

the rational canonical form associated with $p=(p_1,\ldots,p_n)'\varepsilon\mathbb{R}^n$ and we let, $q=(q_1,\ldots,q_n)'\varepsilon\mathbb{R}^n$. The vectors b_1 and b_2 are known and fixed. When p and q are known, the state-estimation problem for (1.1) has the well-known solution -- the Kalman-Bucy filter. By the identification problem we shall mean the problem of jointly estimating the state and the parameters -- in other words it is the nonlinear filtering problem for the extended system with state $z_t=(x_t,p_t,q_t)$ defined by,

$$dx_t = A(p_t)x_t\, dt + b_1u_t dt + b_2 dw_t$$

$$dp_t = 0 \tag{1.3}$$

$$dq_t = 0 \;,$$

$$dy_t = \langle q_t,x_t\rangle dt + dv_t. \tag{1.4}$$

More precisely, in solving the identification problem one seeks a solution to the Kushner-Stratonovitch eqn, [1] satisfied by the conditional density $p(t,z)$ given the observations y_s, $0\leq s\leq t$. Although this point of view goes back to Kushner [2], progress along these lines has been impeded by the nonlinearity of the Kushner equation.

More recently, it has been recognized [3,4,5] that an understanding of the evolution equation satisfied by the so-called unnormalized conditional density, $\psi(t,z)$ is essential for further progress in nonlinear filtering theory. Knowing $\psi(t,z)$, the conditional density is determined by the normalization

$$p(t,x) = \psi(t,z)/\int\psi(t,z)\, dz. \tag{1.5}$$

In the general situation when $\{z_t\}$ is a diffusion process with observation (semimartingale) $\{y_t\}$ of the form

$$dy_t = h(z_t)\, dt + dv_t \tag{1.6}$$

it is known that, $\psi(t,z)$ satisfies a linear stochastic partial differential equation (Mortenson-Duncan-Zakai equation) of the Ito type,

$$d\psi(t,z) = \mathcal{L}\psi(t,z)\, dt + h(z)\, \psi(t,z)dy_t \tag{1.7}$$

where $\mathcal{L}$ is the Kolmogorov forward operator associated with the diffusion $\{z_t\}$. (See the expository paper by Davis and Marcus in these proceedings). Now, as regards questions related to the complexity of a nonlinear filtering problem geometric ideas play a crucial role and one looks at the Stratonovitch version of (1.7) written formally as,

$$\partial\psi/\partial t(t,z)=[\mathcal{L}-\tfrac{1}{2}h^2(z)]\psi(t,z)+h(z)dy/dt. \quad (1.8)$$

In particular, the operator Lie algebra $\tilde{G}$ generated by $\mathcal{L}-\frac{1}{2}h^2$ and h, known as the estimation algebra [3], has been emphasized by Brockett, Mitter, Ocone and others as an object of central interest. In what follows we use the estimation algebra to classify identification problems and investigate special cases.

2. ESTIMATION ALGEBRAS:

The fact that the estimation algebra is invariant under change of coordinates [6], makes it useful as a classifying tool. In particular the choice of canonical forms is not crucial. Essentially there are four cases:

Case (1) $b_2=0$; $b_1=0$. Then $\mathcal{L}=-\langle Ax,\partial/\partial x\rangle-tr(A)$. Define $A_0=\mathcal{L}-\frac{1}{2}\langle q,x\rangle^2$; $A_1=\langle q,x\rangle$. Then the estimation algebra $\tilde{G}=\{A_0,A_1\}_{L.A.}$ is defined by the structure equations, $[A_0,A_j]=A_{j+1}$; $[A_j,A_k]=0$ $k,j\geq 1$ where, $A_j=(-1)^{j-1}\langle q,A^{j-1}x\rangle$ $j=1,2\ldots$. We have a sequence of abelian ideals $\tilde{G}_n=\text{span}\{A_j;j=n,n+1,\ldots\}n\geq 1$, with finite codimension (a feature of potential value in connection with approximation problems).

Case (2) $b_1=e_n$ say; $b_2=0$. The presence of deterministic inputs does not alter the structure of the estimation algebra. Note $\mathcal{L}=-\langle Ax,\partial/\partial x\rangle-tr(A)-u_t\langle b_1,\partial/\partial x\rangle$. Define $A_0=\mathcal{L}-\frac{1}{2}\langle q,x\rangle^2$ and $A_j=(-1)^{j-1}\langle q\ ,A^{j-1}x\rangle+(-1)^{j-1}\langle q,A^{j-2}b_1\rangle u_t$, $j\geq 2$. Then once again $[A_0,A_j]=A_{j+1}$; $[A_j,A_k]=0$; for $j,k\geq 1$.

Case (3) $b_1=e_n$; $b_2=0$ and the parameters p are known. $A_0,A_1,\ldots$ etc are as in the previous case. But by the Cayley-Hamilton theorem $A^n(p)=-\sum_{i=1}^{n} p_i\ A^{i-1}(p)$. Hence the operators A_k for $k>n+1$ are linearly dependent on the operators A_j for $j\leq N+1$. Then the estimation algebra $\tilde{G}=\{A_0,A_1\}_{L.A.}$ is finite dimensional. In fact using tensor products it can be shown that the underlying filtering problem is linear.

Case (4) $b_2\neq 0$. The presence of driving noise drastically alters the structure of the Lie algebra. The general situation is not unlike the example below:

$$dx_t = \alpha x_t dt + dw_t$$
$$d\alpha_t = 0$$
$$dy_t = x_t dt + dv_t$$

$$\mathcal{L}=-\alpha x\partial/\partial x-\alpha+\tfrac{1}{2}\partial^2/\partial x^2;\ A_0=\mathcal{L}-\tfrac{1}{2}x^2;A_1=x.$$

Define $A_{2n+1} = (\alpha^2+1)^n x$; $A_{2n+2} = (\alpha^2+1)^n(\partial/\partial x-\alpha x)$ with $n=0,1,2,\ldots$. Also let $B_k = -(\alpha^2+1)^k$, $k=0,1,2,\ldots$. Then the structure equations are,

$$[A_0,A_{2n+1}]=A_{2n+2};\quad [A_0,A_{2n+2}]=A_{2n+3}\ [A_{2n+1},A_{2m+1}]=0;$$
$$[A_{2n+2},A_{2m+2}]=0;\ [A_{2n+1},A_{2m+2}]=B_{m+n}.\ [B_j,A_k]=0;\ [B_j,B_k]=0.$$

It is possible to write down filtrations of $\tilde{G}$ by sequences of ideals as before. In fact in each case above the algebra $\tilde{G}$ is a profinite dimensional filtered Lie algebra (see Hazewinkel-Marcus [10]). All the known nonlinear filtering problems that admit finite dimensionally computable statistics have Lie algebras of this type.

It is however important to note that the identification problem in our formulation is tractable in the cases (1), (2), (3) above where there is no driving noise. In these cases, the Stratonovitch form of the evolution equation for the unnormalized conditional density is given by,

$$\partial\psi/\partial t = (A_0 + A_1\dot{y})\psi \tag{2.1}$$

where, $A_0 = -\langle Ax,\partial/\partial x\rangle - tr(A) - u_t\langle b_1,\partial/\partial x\rangle - \frac{1}{2}\langle q,x\rangle^2$, and $A_1 = \langle q,x\rangle$. In principle, equation (2.1) can be solved by the method of characteristics. See below.

3. AN EXAMPLE

Consider the special case of (1.3)-(1.4) given by

$$\begin{aligned} dx_t &= -\alpha x_t dt + u_t dt \\ dy_t &= x_t dt + dv_t \end{aligned} \tag{3.1}$$

Then (2.1) reduces to,

$$\partial\psi/\partial t = (\alpha x - u_t)\partial\psi/\partial x + (\alpha - x^2/2 + x\dot{y})\psi. \tag{3.2}$$

Let the initial condition be given by $\psi(t,x,\alpha)|_{t=0} = \psi_0(x,\alpha)$. In the 4-dimensional (t,x,α,z) space, we want to pass an integral hyper-surface S: $z=\psi(t,x,\alpha)$ of the equation (3.2) through the 2-dimensional manifold Γ (Cauchy data) given parametrically by $x=s_1$, $\alpha=s_2$, $t=0$, $z=\psi_0(s_1,s_2)$. The characteristics passing through points (s_1,s_2) in Γ sweep out S and are given by the system of (characteristic) differential equations:

$$\begin{aligned} dx/d\tau &= -(\alpha x - u_t) \\ d\alpha/d\tau &= 0 \\ dt/d\tau &= 1 \\ dz/d\tau &= (\alpha - x^2/2 + x\dot{y}_t)z \end{aligned} \tag{3.3}$$

Solving (3.3), we obtain a parametric representation of the characteristic curves; $\alpha=s_2$; $t=\tau$; $x=X(s_1,s_2,\tau)= e^{-s_2\tau}s_1 + \int_0^\tau e^{-s_2(\tau-\sigma)}u_\sigma d\sigma$ and finally,

$$z=\psi_0(s_1,s_2)\cdot\exp(s_2\tau)\cdot\exp\left(\int_0^\tau X(s_1,s_2,\sigma)\dot{y}_\sigma d\sigma - \tfrac{1}{2}\int_0^\tau X^2(s_1,s_2,\sigma)d\sigma\right). \tag{3.4}$$

Equation (3.4) for z is nothing but a parametric representation of the solution ψ we are seeking. It is easy to see that given x,t, and α the parameters s_1,s_2 and τ can be eliminated and

$$X(s_1,s_2,\sigma) = e^{\alpha(t-\sigma)}x - \int_{\sigma}^{t} e^{-\alpha(\sigma-\theta)}u_\theta \cdot d_\theta . \qquad (3.5)$$

Substitution in (3.4) gives the explicit representation of $\psi(t,x,\alpha)$ for given input and output functions. The last exponential factor in (3.4) corresponds to the well-known likelihood ratio formula ([7], [8]).

4. SUFFICIENT STATISTICS

In Eqn. (3.4), only the term $\int_0^t X(s_1,s_2,\sigma)\dot{y}_\sigma d_\sigma$ inside the exponential depends on measured outputs and explicitly,

$$\int_0^t X(s_1,s_2,\sigma)\dot{y}_\sigma d_\sigma = x\sum_{k=0}^{\infty} \alpha^k \int_0^t \frac{(t-\sigma)^k}{k!} dy_\sigma - \sum_{k=0}^{\infty} (-\alpha)^k \int_0^t \gamma_k(t,\sigma)dy_\sigma \qquad (4.1)$$

where $\gamma_k(t,\sigma) = \int_{\sigma}^{t} (\sigma-\theta)^k/k! u_\theta d_\theta$. The two sequences

(a) $\beta_k(t) = \int_0^t (t-\sigma)^k/k!\ dy_\sigma \quad k = 0,1,2 ..$

and

(b) $\omega_k(t) = \int_0^t \gamma_k(t,\sigma)dy_\sigma \quad k = 0,1,2, \ldots$

may be viewed as sufficient statistics for the problem. Each β_k can be computed as the output of a finite dimensional system driven by y_t. The same holds true for the ω_k's. We must mention that the statistics ω_k are similar to the sufficient statistics determining the likelihood ratio given by Giorgio Picci [9].

ACKNOWLEDGEMENT

This work was supported in part by Air Force Office of Scientific Research grant AFOSR 79-0025 and by the Department of Energy Contract DEAC01-79-ET-29363.

REFERENCES

[1] H.J. Kushner, "Dynamical Equations for Optimal nonlinear filtering", J. Differential Equations, Vol. 3, 1967, pp. 179-190.

[2] H.J. Kushner, "On the differential equations satisfied by conditional probability densities of Markov processes, with applications," RIAS Tech. Report No. 64-6, March 1964.

[3] R.W. Brockett and J.M.C. Clark, "The geometry of the conditional density equation," Proc. of the Oxford Conference

on Stochastic Systems, Oxford, England, 1978.
[4] S.K. Mitter, "On the analogy between mathematical problems of nonlinear filtering and quantum physics," Richerche di Automatica, 1980.
[5] E. Pardoux, "Stochastic partial differential equations and filtering of diffusion processes," Stochastics, Vol. 3, 1979, pp. 127-167.
[6] R.W. Brockett, "Classification and equivalence in estimation theory," Proc. IEEE Conference on Decision and Control, Dec. 1979, Ft. Lauderdale.
[7] T.T. Kadota and L.A. Shepp, "Conditions for absolute continuity between a certain pair of probability measures," Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 16, (1970) pp. 250-260.
[8] T. Kailath and M. Zakai, "Absolute Continuity and Radon-Nikodym derivatives for certain measures relative to the Wiener measure", Ann. of Math. Statistics, 42 (1971), pp. 130-140.
[9] G. Picci, "Some connections between the theory of sufficient statistics and the identifiability problem," SIAM J. Appl. Math., Vol. 33, No. 3, 1977, pp. 383-398.
[10] H. Hazewinkel and S.I. Marcus, "On the relationship between Lie algebras and nonlinear estimation," Proc. of the IEEE Conference on Decision and Control, Dec. 1980, Albuquerque.