The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

Stat.

Netherlands school of economics

ECONOMETRIC INSTITUTE

PROCEEDINGS FILTER-DAY ROTTERDAM 1980

MICHIEL HAZEWINKEL, ed.

WITHDRAWN
GIANNINI FOUNDATION OF
AGRICULTURAL ECONOMICS
LIBRARY
JAN 21 1981

REPORT 8011/M

ERASMUS UNIVERSITY ROTTERDAM P.O. BOX 1738. 3000 DR. ROTTERDAM. THE NETHERLANDS.

PROCEEDINGS FILTER-DAY ROTTERDAM 1980

(New trends in filtering and identification of stochastic systems, 23 Jan. 1980)

by Michiel Hazewinkel (ed)

ABSTRACT.

This report contains the complete (expanded) texts of 4 of the five lectures and a 2 page introduction to the remaining lecture delivered at a one day conference on "New trends in filtering and identification of stochastic systems", which took place on Jan. 23, 1980 at Erasmus University Rotterdam. We aimed to present some of the rather striking and very promising new developments especially in recursive and/or nonlinear filtering and as such the meeting was also intended as a preliminary warm-up for the Advanced-Study-Institute "Stochastic Systems: the mathematics of filtering and identification which will take place at Les Arcs (Savoie) this summer (June 22-July 5).

Contents

1. Introduction
2. S.I. Marcus, C.H.Liu, The Lie-algebra structure of a class of finite dimensional nonlinear filters
3. P. DeWilde, H. Dym, Schur recursions, error formulas and convergence of rational estimators for stationary stochastic sequences
4. J.H. van Schuppen, The Stochastic filtering problem for point processes
5. M.I. Friedman, J.C. Willems, Smooth representation of systems with differentiated inputs.
6. C. Martin, Riccati equations and Lagrangian Grassmann manifolds.

The series is divergent, so we may be
able to do something with it.

Oliver Heaviside

1. INTRODUCTION

In fair generality the filtering problem involves two second order stochastic processes y_t, z_t, $0 \leq t < \infty$ and we desire to calculate the "best" estimate of z_t given the whole past y_s, $0 \leq s \leq t$ of the process y_t. In the mean square sense this best estimate is given by the conditional expectation $E[z_t|y^t]$, where y^t stands for the whole past y_s, $0 \leq s \leq t$. However, this theoretical answer is of limited use as it is rarely possible to calculate this expression. Much of the work in filtering concerns a set of (Itô) equations of the type

$$
\begin{aligned}
dx &= f(x)dt + g(x)dw \\
dy &= h(x)dt + dv \qquad (1) \\
z &= k(x)
\end{aligned}
$$

where $x \in \mathbb{R}^n$, $w \in \mathbb{R}^m$, $v \in \mathbb{R}^p$, $z \in \mathbb{R}^q$ with f, g, h, k vector and matrix valued functions (of the appropriate dimensions) and with w and v Wiener processes. There are of course corresponding sets of equations for the discrete time case.

In case, f, g, h and k are all linear we are dealing with a linear filtering problem. And this case is a pleasant and for applications in virtually all fields of inquiry most important exception. In this case we can calculate $E[z_t|y^t]$ and the calculation is done for us by the Kalman filter (in case z = x). Moreover this filter is itself a system like (1) driven by the observations y_t, albeit a slightly nonlinear one. This means that the filter is recursive in a very pleasant way. In the case of just noise with linear observations, i.e. in the case of the system

$$(2) \qquad dx = dw, \; dy = xdt + dv$$

the Kalman filter is

$$
\begin{aligned}
d\hat{x} &= -\hat{x}Pdt + Pdy \qquad (3) \\
dP &= (1-P^2)dt
\end{aligned}
$$

In the more dimensional case, the nonlinear part gets replaced by the socalled <u>matrix</u> Riccati differential equation. This makes the geometry and topology of these equations most important. It turns out that these equations are most naturally studied not on the space of all n x n matrices, $\mathbb{R}^{nxn}$, but on certain compactificationscalled Grassmann manifolds (among other things in order to understand various escapes to infinity). This was the subject of Clyde Martin's talk (section 6 of these proceedings).

The success and wide applicability of the Kalman filter stimulated the search for similar filters for nonlinear systems. Until recently with very modest success. One can write down a stochastic differential equation for the conditional expectation of the state $E[x_t|y^t]$ and also for the conditional density $p(x,t)$ (assuming it exist; cf. equations (2) and (3) of section 2 of these proceeding. However, these equations do not look very nice, in particular not recursive. Now a certain unnormalized version $\rho(x,t)$ of the conditional density of the state satisfies the Zakai equation

$$(4) \qquad d\rho(t,x) = \pounds\rho(t,x)dt + h(x)\rho(t,x)dy_t$$

which, apart from being infinitedimensional, looks recursive, indeed bilinear. Here $\pounds$ is the Fokker-Planek operator (cf. equation (4) of section 2 below). From experience with finite dimensional bilinear systems it now seems natural to pay some attention to the Lie algebra generated by the two differential operators $\pounds$ and $h(x)$ (= multiplication with the function $h(x)$). This philosophy of Brockett [5] has been amazingly successful and has already let to several classes of recursive nonlinear filters. One such class was discussed by Steve Marcus (section 2 of these proceedings). Several other classes will be discussed at Les Arcs [7].

Before calculating the Lie algebra $L(\Sigma)$ generated by the operators and $h(x)$ one puts (4) into Fisk-Stratonovič form (so that the ordinary rules of calculus apply). In the case of scalar observations this gives us

$$(5) \qquad d\rho(t,x) = (\pounds-\tfrac{1}{2}h(x)^2\rho(x,t)dt + h(x)\rho(t,x)dy_t$$

Now suppose that the system is the one given by (2) above. Then an easy calculation shows that the Lie-algebra $L(\Sigma)$ is this case is a four dimensional Lie algebra of some fame in physics, the socalled oscillator Lie algebra. The Lie algebra generated by the two vector-fields of the Kalman filter (3) in this case is the even more famous 3-dimensional Heisenberg (commutation relation) algebra which is isomorphic to the oscillator Lie algebra modulo its centre. These observations [5] are most stimulating and make one moreover suspect that there are deep relations between filtering and quantum theory.

And indeed, there are; cf. Sanjoy Mitter's contribution in [7] and [10]. More results pertaining to this approach to nonlinear filtering can be found in [1,2,3,4,6] and [7] and the references therein.

To return to the linear case, a modeling filter (or stochastic realization) of a real stationary zero-mean stochastic sequence with given covariance function is a linear, time invariant, causal filter, e.g. a discrete version of a "machine" like (1) above, which when driven by white noise reproduces an output with the perscribed covariance. In practice it is of importance to compute approximate (linear, discrete) machines like (1) which as the dimension increases give better and better approximations, and to do this in a recursive manner so that when the matching accuracy has to be improved this can be achieved by adding some structural elements to the already existing machine. This is the subject of Patrick DeWilde's contribution, (Section 3). This section contains only the introduction to the full paper: P. DeWilde, H. Dym, Schur recursions, error formulas and convergence of rational estimators for stationary stochastic sequences. Report 92, 1979, Delft Univ. of Techn., Dept. Electrical Engineering. (To appear Trans IEEE on Information Theory).

All of the above concerns stochastic systems with continuous observations. Yet many processes in nature (and in human affairs) concern stochastic processes with observation processes, which are jump processes, or more generally counting processes or point processes). In section 4 Jan van Schuppen addresses himself to some basic question in the theory of such stochastic systems.

Finally much of the above involves stochastic calculus; often Ito-calculus. However, Ito integrals do not transform right under coordinate changes and hence are of limited use on topologically nontrivial manifolds. Fisk-Stratonovic integrals do transform right but have their own (conceptual) difficulties. All this makes one wonder whether a strictly pathwise approach might (under suitable conditions) be possible. Several aspects related to this matter were taken up by Jan Willems in his lecture. The problems are also much related to robustness properties of stochastics systems (c.q. filters). Below we have (with thanks) reproduced the original paper on which the talk was based (it originally appeared in the IEEE Trans. AC 23 (1978)), together with a supplementary list of references. Still another reference which could be added to this list is [9]. The title of Jan Willems' talk at the conference was "Approximation of stochastic differential equations by deterministic systems".

REFERENCES

1. J. Baillieul, Families of Low Dimensional Estimation Problems, preprint.
2. J. Baras, G. Blankenship, Nonlinear Filtering of Diffusion Processes: a Generic Example, preprint
3. V.E. Benes, Exact Finite Dimensional Filters for Certain Diffusions with Nonlinear Drift, presented at the 1979 CDC, Ft Lauderdale
4. G. Blankenship, A. Haddad, Asymptotic Analysis of a Class of Nonlinear Filtering Problems, IRIA Workshop on Singular Perturbations in Control, June 1978.
5. R.W. Brockett, Remarks on Finite Dimensional Nonlinear Estimation, In: C. Lobry (ed), Journées sur les systèmes (Bordeaux 1978), Astérisque, to appear.
6. M. Hazewinkel, S.I. Marcus, On Lie Algebras and Finite Dimensional Filtering, preprint.
7. M. Hazewinkel, J.C. Willems (eds), Proc. NATO-ASI "Stochastic-Systems": the Mathematics of Filtering and Identification and Applications", Reidel Publ. Cy, to appear.
8. A.J. Krener, A Formal Approach to Stochastic Integration and Differential Equations. Stochastics 3(1979), 105-125

Michiel Hazewinkel

Krimpen a/d Ijssel, 8 June 1980

THE LIE ALGEBRAIC STRUCTURE OF A CLASS OF FINITE DIMENSIONAL NONLINEAR FILTERS.*

Chang-Huan Liu and Steven I. Marcus

Dept. of Electrical Engineering
University of Texas at Austin
Austin, Texas 78712

ABSTRACT.

We present an example of the application of Lie algebraic techniques to nonlinear estimation problems. The method relates the computation of the (unnormalized) conditional density and the computation of statistics with finite dimensional estimators. The general method is explained; for a particular example, the structures of the Lie algebras associated with the unnormalized conditional density equation and the finite dimensionally computable conditional moment equations are analyzed in detail. The relationship between these Lie algebras is studied, and the implications of these results are discussed.

*. Supported in part by the Air Force Office of Scientific Research (AFSC) under Grant. AFOSR-79-0025, in part by the National Science Foundation under Grant ENG 76-11106, and in part by the Joint Services Electronics Program under Contract F 49620-77-C-0101.

1. INTRODUCTION.

This paper is concerned with the optimal recursive estimation of the state x_t of a nonlinear stochastic system, given the past observations $z^t = \{z_s, 0 \leq s \leq t\}$. Specifically, we consider systems of the form

$$
(1)\qquad \begin{aligned} dx_t &= f(x_t)dt + G(x_t)dw_t \\ dz_t &= h(x_t)dt + R_t^{\frac{1}{2}}dv_t \end{aligned}
$$

where w and v are independent unit variance vector Wiener processes, f and h are vector-valued functions, G is a matrix-valued function, and $R > 0$. The optimal (minimum-variance) estimate is of course the conditional mean $\hat{x}_t = E[x_t | z^t]$ (also denoted $\hat{x}_{t|t}$ or $E^t[x_t]$); $\hat{x}_t$ satisfies the (Ito) stochastic differential equation [1] - [3]

$$
(2)\qquad \begin{aligned} d\hat{x}_t = [\,\hat{f}(x_t) &- (\widehat{x_t h^T} - \hat{x}_t\hat{h}^T)R^{-1}(t)\hat{h}]dt \\ &+ (\widehat{x_t h^T} - \hat{x}_t\hat{h}^T)R^{-1}(t)dz_t \end{aligned}
$$

where $\hat{}$ denotes conditional expectation given z^t and h denotes $h(x_t)$. Also, the conditional probability density $p(t,x)$ of x_t given z^t (we will assume that $p(t,x)$ exists) satisfies the stochastic partial differential equation [3], [4]

$$
(3)\qquad dp(t,x) = \pounds p(t,x)dt + (h(x)-\hat{h}(x))^T R^{-1}(t)(dz_t - \hat{h}(x)dt)p(t,x)
$$

where

$$
(4)\qquad \pounds(.) = -\sum_{i=1}^{n} \frac{\partial(.f_i)}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2(.(GG^T)ij)}{\partial x_i \partial x_j}
$$

is the forward diffusion operator.

Notice that the differential equation (2) is not recursive, and indeed appears to involve an infinite dimensional computation in general. Aside from the linear-Gaussian case in which the Kalman filter is optimal, there are very few known cases in which the optimal estimator is finite dimensional (a number of these are summarized in [5]). However, in [6] - [8] we have shown that for certain classes of nonlinear stochastic systems in continuous and discrete time, the conditional mean can be computed with a recursive filter of fixed finite dimension. The typical nonlinear system in these classes consists of a linear system with linear measurements, which feeds forward into a nonlinear system described by a certain type of Volterra series expansion or by a bilinear system satisfying certain algebraic conditions. The major purpose of this paper is to consider these estimation problems from a new perspective, and to gain much deeper insight into their structure.

The new perspective, originally proposed by Brockett [9] (see also [10], [11]), takes the following approach to the general estimation problem (1) (we assume for simplicity that z is a scalar). Instead of studying the equation (3) for the conditional density, we consider the Zakai equation for an unnormalized conditional density $\rho(t,x)$ [12]:

(5) $$d\rho(t,x) = \mathcal{L}\rho(t,x)dt + h(x)\rho(t,x)dz_t$$

where $\rho(t,x)$ is related to $p(t,x)$ by the normalization

(6) $$\rho(t,x) = \rho(t,x).(\int\rho(t,x)dx)^{-1}.$$

The Zakai equation (5) is much simpler than (3); indeed, (5) is a bilinear differential equation [13] in ρ, with z considered as the input. This is the first clue that the Lie algebraic and differential geometric techniques developed for finite dimensional systems of this type may be brought to bear here. <u>Suppose</u> that some statistic of the conditional distribution of x_t given z^t can be calculated with a finite dimensional recursive estimator of the form

(7) $$d\eta_t = a(\eta_t)dt + b(\eta_t)dz_t$$

(8) $$E[c(x_t)|z^t] = \gamma(\eta_t)$$

where η evolves on a finite dimensional manifold, and a and b are suitably smooth. Of course, this statistic can also be obtained from $\rho(t,x)$: by

(9) $E[c(x_t)|z^t] = \int c(x)\rho(t,x)dx \; (\int\rho(t,x)dx)^{-1}$

For Lie-algebraic calculations, it is more convenient to write (5) and (3) in Fisk-Stratonovich form (so that they obey the ordinary rules of calculus)

(10) $d\eta_t = \tilde{a}(\eta_t)dt + b(\eta_t)dz_t$

(11) $d\rho(t,x) = [\mathcal{L} - \frac{1}{2} h^2(x)]\rho(t,x)dt + h(x)\rho(t,x)dz_t$

where the i^{th} component $\tilde{a}_i(\eta) = a_i(\eta) - \frac{1}{2} \sum_j b_j(\eta)\frac{\partial b_i}{\partial \eta_j}(\eta)$

(Beginning with (10), all equations will be in Fisk-Stratonovich form, unless otherwise indicated). The two systems (10), (8) and (11), (9) are thus two representations of the same mapping from "input" functions z to "outputs" $E[c(x_t)|z^t]$: (11), (9) via a bilinear infinite dimensional state equation, and (10), (8) via a nonlinear finite dimensional state equation. Generalizing the results of [14], [15] to infinite dimensional state equations, the major assertion of [9] is that, under appropriate hypotheses, the Lie algebra F generated by $\tilde{a}$ and b (under the commutator $[\tilde{a},b] = \frac{\partial b}{\partial \eta}\tilde{a} - \frac{\partial \tilde{a}}{\partial \eta} b$) is a homomorphic image of the Lie algebra L generated by $A_o = \mathcal{L} - \frac{1}{2} h^2(x)$ and $B_o = h(x)$ (under the commutator $[A_o,B_o] = A_oB_o - B_oA_o$). Conversely, any homomorphism of L onto a Lie algebra generated by two complete vector fields on a finite dimensional manifold allows the computation of some information about the conditional density with a finite dimensional estimator of the form (10).

In [9], this approach is explicitly carried out and analyzed for the problem in which f, G and h (1) in are all linear. In that case, the Lie algebra L of the Zakai equation is finite dimensional and the unnormalized conditional density can in fact be computed with a finite dimensional estimator , the Kalman filter. In this paper, we carry out a similar analysis for the simplest example of the class considered in [6] - [8]. For this example, all conditional moments of the state can be computed with finite dimensional estimators; the Lie algebra L is <u>infinite</u> dimensional but has many finite dimensional homomorphic images (the Lie algebras of the finite dimensional estimators), thus yielding a very interesting structure. The example to be considered has state equations

$$dx_t = dw_t \tag{12}$$
$$dy_t = x_t^2 dt$$

with observations

$$dz_t = x_t dt + dv_t \tag{13}$$

where v and w are unit variance Wiener processes, $\{x_o, y_o, v, w\}$ are independent, and x_o is Gaussian. The computation of $\hat{x}_t$ is of course straightforward by means of the Kalman filter, but the computation of $\hat{y}_t$ requires a nonlinear estimator.

2. THE LIE ALGEBRA OF THE UNNORMALIZED CONDITIONAL DENSITY EQUATION.

For the system (12) - (13), the equation (5) in Fisk-Stratonovich form is

$$d\rho(t,x) = \left(-x^2\frac{\partial}{\partial y} + \frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}x^2\right)\rho(t,x)dt + x\rho(t,x)dz_t, \tag{14}$$

so the Lie algebra L is generated by $A_o = -x^2\frac{\partial}{\partial y} + \frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}x^2$ and $B_o = x$.

The following theorem is straightforward to prove.

<u>Structure theorem 1</u>:

(i) The Lie algebra L generated by A_o and B_o has as basis the elements A_o and B_i, C_i, D_i, $i = 0,1,2, \ldots$, where

$$B_i = x\frac{\partial^i}{\partial y^i} \qquad i = 0,1,2,\ldots$$

$$C_i = \frac{\partial}{\partial x}\frac{\partial^i}{\partial y^i} \qquad i = 0,1,2,\ldots$$

$$D_i = \frac{\partial^i}{\partial y^i} \qquad i = 0,1,2,\ldots$$

(ii) The commutation relations are given by

$$[A_o, B_i] = C_i, \quad \forall i$$

$$[A_o, C_i] = B_i + 2B_{i+1}, \quad \forall i$$

$$[A_o, D_j] = [B_i, D_j] = [C_i, D_j] = 0, \quad \forall i,j$$

$$[B_i, C_j] = -D_{i+j}, \quad \forall i,j$$

$$[B_i, B_j] = [C_i, C_j] = 0, \quad \forall i \neq j$$

(iii) The center of L is $\{D_i, i = 0,1,2,\ldots\}$.

(iv) Every ideal of L has finite codimension; i.e., for any ideal I, the quotient L/I is finite dimensional.

(v) Let I_j be the ideal generated by B_j, with basis $\{B_i, C_i, D_i;\ i \geq j\}$. Then $I_o \supset I_1 \supset \ldots$ and $\bigcap_j I_j = \{0\}$, so that the canonical map $\pi: A \to \bigoplus_j A/I_j$ is injective.

(vi) L is the semidirect sum [18] of A_o and the nilpotent ideal I_o; hence L is solvable.

In light of the remarks in the previous section, it should be expected that many statistics of the conditional distribution can be computed with finite dimensional estimators, since there are an infinite number of finite dimensional quotients(homomorphic images) L/I . By Ado's theorem, these can be realized by bilinear systems. However, we will present a slightly different realization of the sequence of quotients in (vi) above: L/I_1 is realized by the Kalman filter for $\hat{x}_t$ (L/I_1 is the oscillator algebra [9] - [11]), and L/I_j $(j \geq 2)$ is realized by the estimator which computes $\hat{x}_t$ and $\widehat{y^i_t} = E[y^i_t | z^t]$ $(i = 1,2,\ldots,j-1)$. Of course, the dimension of L/I_j increases with j, so we will only present the estimator equations for j = 4 in the next section. Other sequences of quotients possessing the property (vi) can also be realized (e.g., those generated by the $\{C_j\}$), but those realizations do not have as natural an interpretation in terms of conditional moments.

The properties (iv) and (v) of the structure theorem are useful for an "estimation algebra" to possess, in the following sense: they basically say that L has enough finite dimensional quotients that it is determined by their direct sum. Translating this into an estimation context via the reasoning of the previous section, if we can realize all the quotients with finite dimensionally computable statistics, then these properties give us hope of being able to approximate the conditional density (or conditional characteristic function) with a convergent series of functions of these statistics, even if the conditional density cannot be computed exactly by a finite dimensional estimator.

3. THE LIE ALGEBRA OF THE FINITE DIMENSIONAL ESTIMATOR.

The method of [6] for computing the finite dimensional estimator for $\hat{y}_t$ systematically uses the estimation equation (2) and the fact that the conditional density of $x_{\cdot}$ given $z_{\cdot}$ is Gaussian to express higher order moments in terms of lower. This procedure can also be applied to obtain equations for higher order conditional moments of y for the estimation problem (12) - (13). The first three conditional moments of y_t, together with $\hat{x}_t$ and the necessary auxiliary filter states are computed recursively by the finite dimensional estimator (in Fisk-Stratonovich form, with explicit time-dependent notation omitted):

$$
(15)\quad d\begin{bmatrix} \hat{x} \\ \hat{\xi} \\ \hat{y} \\ \hat{\theta} \\ \widehat{y^2} \\ \hat{\phi} \\ \widehat{y^3} \\ \\ \\ t \end{bmatrix} = \underbrace{\begin{bmatrix} -\hat{x}P \\ \hat{x}(1-P_{12}) - \hat{\xi}P^{-1} \\ \hat{x}^2 - 2\hat{x}\hat{\xi}P + P - PP_{12} \\ \hat{x}(P_{12}-P_{13}) + \hat{\xi}P(1-P_{12}) - \hat{\theta}P^{-1} \\ 2\hat{x}^2\hat{y} + 2\hat{y}P + 8\hat{x}\hat{\xi}P + 2PP_{12} - 4\hat{x}\hat{\xi}\hat{y}P - 8\hat{x}\hat{\theta}P - 2\hat{y}PP_{12} - 4\hat{\xi}^2P^2 - 4PP_{13} \\ \hat{x}(P_{13}-P_{14}) + \hat{\xi}P(P_{12}-P_{13}) + \hat{\theta}(P-PP_{12}) - \hat{\phi}P^{-1} \\ 3\hat{x}^2\widehat{y^2} + 3\widehat{y^2}P + 24\hat{x}\hat{\xi}\hat{y}P + 48\hat{x}\hat{\theta}P + 24\hat{\xi}^2P^2 + 6\hat{y}PP_{12} \\ + 24PP_{13} - 3\widehat{y^2}PP_{12} - 48\hat{\xi}\hat{\theta}P^2 - 12\hat{y}PP_{13} - 12\hat{\xi}^2\hat{y}P^2 - 24PP_{14} \\ - 6\hat{x}\hat{\xi}y^2P - 24\hat{x}\hat{\theta}\hat{y}P - 48\hat{x}\hat{\phi}P \\ 1 \end{bmatrix}}_{a_o} dt +
$$

$$\underbrace{\begin{bmatrix} P \\ P_{12} \\ 2\hat{\xi}P \\ P_{13} \\ 4\hat{\xi}\hat{y}P + 8\hat{\Theta}P \\ P_{14} \\ 6\hat{\xi}\widehat{y^2}P + 24\hat{\Theta}\hat{y}P + 48\hat{\phi}P \\ 0 \end{bmatrix}}_{b_o} dz \quad ; \quad \begin{bmatrix} \hat{x}_o \\ \hat{\xi}_o \\ \hat{y}_o \\ \hat{\Theta}_o \\ \widehat{y^2_o} \\ \hat{\phi}_o \\ \widehat{y^3_o} \\ t_o \end{bmatrix} = \begin{bmatrix} E[x_o] \\ 0 \\ E[y_o] \\ 0 \\ E[y_o^2] \\ 0 \\ E[y_o^3] \\ 0 \end{bmatrix}$$

where the nonrandom conditional covariance equations are

$$\begin{aligned} &\dot{P} = 1 - P^2 \\ &\dot{P}_{12} = P - (P+P^{-1})P_{12} \\ (16) \qquad &\dot{P}_{13} = 2PP_{12} - PP_{12}^2 - (P+P^{-1})P_{13} \\ &\dot{P}_{14} = 2PP_{13} + PP_{12}^2P_{13} - (P+P^{-1})P_{14} \\ &P(0) = \mathrm{cov}(x_o) \neq 0;\ P_{12}(0) = P_{13}(0) = P_{14}(0) = 0 \end{aligned}$$

The estimator (15) is obtained by first augmenting the state x with auxiliary states ξ, Θ, and ϕ; then the Kalman filter for the linear system with states $[x,\xi,\Theta,\phi]$ and observations z computes $[\hat{x},\hat{\xi},\hat{\Theta},\hat{\phi}]$. In addition, $[P,P_{12},P_{13},P_{14}]$ is the first row of the Kalman filter error covariance matrix; (16) is obtained by selecting the corresponding components of the Riccati equation. Then $\hat{y}$, $\widehat{y^2}$, and $\widehat{y^3}$ are seen, after tedious calculations, to be computed by the given equations (some of the calculations are presented in the Appendix, in order to illustrate the method). The filter state is augmented with t in order to make (15) time-invariant thus facilitating the use of Lie algebraic techniques. The filter (15) can be viewed as a cascade of linear filters [19]: $[\hat{x},\hat{\xi},\hat{\Theta},\hat{\phi},t]$ satisfies a linear equation; some of these states then feed forward and can be viewed as parameters in a linear equation for $\hat{y}$; the states $\hat{x},\hat{\xi},\hat{\Theta},\hat{y},t$ then feed forward as parameters into a linear equation for $\widehat{y^2}$; etc. This structure is typical of the class of finite dimensional estimators derived in [6] - [8].

In order to study the structure of the estimation problem as discussed in section 1, we must analyze the Lie algebra F generated by a_o and b_o in (15). The structure of the class of problems of [6] is analyzed from a different point of view in [21].

<u>Structure theorem 2</u>:

(i) F has as basis the elements a_o; b_i, c_i, $i = 0,1,2,3$; d_i, $i = 1,2,3$, where a_o and b_o are given in (15) and

$$c_o = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad b_1 = \begin{bmatrix} 0 \\ 1 \\ 2\hat{x} \\ P_{12} \\ 4\hat{x}\hat{y} + 8\hat{\xi}P \\ P_{13} \\ 6\hat{x}\widehat{y^2} + 24\hat{\xi}\hat{y}P + 48\hat{\Theta}P \\ 0 \end{bmatrix}, \quad c_1 = \begin{bmatrix} 0 \\ P^{-1} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

$$b_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 8\hat{x} \\ P_{12} \\ 24\hat{x}\hat{y} + 48\hat{\xi}P \\ 0 \end{bmatrix}, \quad c_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ P^{-1} \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad b_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 48\hat{x} \\ 0 \end{bmatrix}, \quad c_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ P^{-1} \\ 0 \\ 0 \end{bmatrix}$$

$$d_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 2\hat{y} \\ 0 \\ 3\widehat{y^2} \\ 0 \end{bmatrix}, \quad d_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 3\hat{y} \\ 0 \end{bmatrix}, \quad d_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

(ii) The commutation relations are given by

$$[a_o,b_i] = c_i \quad , \quad i = 0,1,2,3$$

$$[a_o,c_i] = b_i - b_{i+1} \quad , \quad i = 0,1,2$$

$$[a_o,c_3] = b_3$$

$$[b_i,c_j] = \begin{cases} -2d_{i+j} & i+j=1 \\ -8d_{i+j} & i+j=2 \\ -48d_{i+j} & i+j=3 \\ 0 & \text{otherwise} \end{cases}$$

$$[a_o,d_j] = [b_i,d_j] = [c_i,d_j] = 0, \ \forall\ i,j$$

(iii) Let $\tilde{I}_4$ be the ideal in L with basis B_i, C_i, D_i, $i \geq 4$ <u>and</u> D_o. Then F is isomorphic to $L/\tilde{I}_4$; hence, F is also solvable.

(iv) The isomorphism ϕ between L and $F/\tilde{I}_4$ is given by:
$\phi(A_o) = a_o$; $\phi(B_i) = (-\frac{1}{2})^i b_i$,
$\phi(C_i) = (-\frac{1}{2})^i c_i$, $i = 0,1,2,3$; $\phi(D_i) = (-1)^i(i!)d_i$; $i = 1,2,3$;
$\phi(E) = 0$, $E \in \tilde{I}_4$.

(v) F is the semidirect sum of a_o and the nilpotent ideal generated by b_o.

<u>Remarks</u>:

(i) The estimator (15) is not quite a realization of L/I_4, since D_o is also in the kernel of the homomorphism (i.e., the ideal $\tilde{I}_4$). However, a finite dimensional estimator realizing L/I_4 (or L/I_j; for any;) is easily obtained by augmenting (15) with the equation for the normalization factor α_t for $\rho(t,x)$ (the denominator of (6)) which satisfies (in Ito form)

$$d\alpha_t = \hat{x}_t\, \alpha_t\, dz_t$$

or (in Fisk-Stratonovich form)

$$(17) \qquad d\alpha_t = -\frac{1}{2}(\hat{x}_t^2+P_t)\alpha_t dt + \hat{x}_t\alpha_t dt$$

If (17) is augmented at the end of (15), the Lie algebra generated by a_o and b_o has the same commutation relations as in (ii) above, except that

$$[b_o, c_o] = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \alpha \end{bmatrix} \stackrel{\Delta}{=} d_o$$

and d_o commutes with all the other elements. Thus a realization of L/I_4 is an easy modification of (15), so we will concentrate on (15).

(ii) The property (v) is typical of a cascade of linear systems.

(iii) One of the conditions in [9] for the existence of a Lie algebra homomorphism from L to the Lie algebra of a finite dimensional estimator is that the estimator be a "minimal" realization in a certain sense. If we consider the output of (15) to be $\widehat{y^3}$ and consider this realization of the input-output map from z to $\widehat{y^3}$, then it can be verified by the methods of [15] that the realization is locally weakly controllable and locally weakly observable. This implies that there is no other realization with lower dimension; it is in this sense that the statistics $\hat{\xi}$, $\hat{\theta}$, $\hat{\phi}$ are necessary for the computation of $\widehat{y^3}$.

Images of L under homomorphisms with successively larger kernels can be realized by using only certain of the equations in the finite dimensional estimator (15); that is, some subset of the equations (15) will generate a Lie algebra isomorphic to L/I. Let $\tilde{I}_j$ denote the ideal with basis D_o and B_i, C_i, D_i, $i \geq j$; we will also use the notation that, e.g., $\tilde{I}_j \oplus D_\ell$ denotes the ideal with basis the above elements and D_ℓ (which is in the center of L). Realizations of some of the many possible quotients are summarized in the following table, which gives the quotient along with the set of states of a finite dimensional estimator which realizes it (the filter states satisfy the corresponding equations in (15)). For example, $L/\tilde{I}_3$ is realized by (15) with the equations for $\hat{\phi}$ and $\widehat{y^3}$ omitted, with all the other filter states retained.

QUOTIENT	REALIZATION
$L/\tilde{I}_4$	$\hat{x}, \hat{\xi}, \hat{y}, \hat{\Theta}, \widehat{y^2}, \hat{\phi}, \widehat{y^3}, t$
$L/(\tilde{I}_4 \oplus D_3)$	$\hat{x}, \hat{\xi}, \hat{y}, \hat{\Theta}, \widehat{y^2}, \hat{\phi}, t$
$L/(\tilde{I}_4 \oplus D_2 \oplus D_3)$	$\hat{x}, \hat{\xi}, \hat{y}, \hat{\Theta}, \hat{\phi}, t$
$L/(\tilde{I}_4 \oplus D_1 \oplus D_2 \oplus D_3)$	$\hat{x}, \hat{\xi}, \hat{\Theta}, \hat{\phi}, t$
$L/\tilde{I}_3$	$\hat{x}, \hat{\xi}, \hat{y}, \hat{\Theta}, \widehat{y^2}, t$
$L/(\tilde{I}_3 \oplus D_2)$	$\hat{x}, \hat{\xi}, \hat{y}, \hat{\Theta}, t$
$L/(\tilde{I}_3 \oplus D_1 \oplus D_2)$	$\hat{x}, \hat{\xi}, \hat{\Theta}, t$
$L/\tilde{I}_2$	$\hat{x}, \hat{\xi}, \hat{y}, t$
$L/(\tilde{I}_2 \oplus D_1)$	$\hat{x}, \hat{\xi}, t$
$L/\tilde{I}_1$	$\hat{x}, t$

Table 1. Realization of some quotients.

The results of Table 1 follow from two observations: first, that if I and J are ideals of L such that $I \subset J$, then J/I is an ideal of L/I and $(L/I)/(J/I)$ is naturally isomorphic to L/J [16,p.8] (e.g., $I = \tilde{I}_4$ and $J = \tilde{I}_3$). Also, it is clear that if one defines homomorphisms from $L/\tilde{I}_4$ to the quotients in Table 1 by the canonical map, then the image can be realized by (15) with certain equations omitted. For example, it is clear that sending $d_3 \to 0$ can be accomplished by eliminating the equation for $\widehat{y^3}$; each d_i thus represents, in some sense, $\widehat{y^i}$. Notice, in particular, that $A/\tilde{I}_1$ is realized by the Kalman filter for $\hat{x}$.

Other interesting quotients are obtained by homomorphisms which send other elements of the center of $L/\tilde{I}_4$, say just d_1, to zero. However, such a quotient is more difficult to realize by an estimator, since the realization is not obtained by merely eliminating certain equations. For these quotients, the following result leads to a realization.

Proposition 1: Let F be the Lie algebra generated by two n-dimensional vector fields a and b. Assume that there is an element d in the center of F and a constant n-vector β such that $\beta' d = 1$ (prime denotes transpose). Then the mapping ϕ with $\phi(a) = a - (\beta' a)d$ and $\phi(b) = b - (\beta' b)d$ extends to a Lie algebra homomorphism with $\phi(f) = f - (\beta' f)d$ for all $f \in F$, $\phi(d) = 0$, and $\phi(F)$ isomorphic to $F/\{d\}$.

Proof: We must show that, for $f,g \in F$,

$$\phi([f,g]) = [f,g] - (\beta'[f,g])d = [\phi(f), \phi(g)].$$

Now, since $\beta' f$ and $\beta' g$ are functions (not constants),

$$\begin{aligned}[\phi(f), \phi(g)] &= [f-(\beta' f)d, g-(\beta' g)d] \\ &= [f,g] - [(\beta' f)d, g] - [f, (\beta' g)d] + [(\beta' f)d, (\beta' g)d] \\ &= [f,g] - \{(\beta' f)[d,g] - g(\beta' f)d\} \\ &- \{(\beta' g)[f,d] + f(\beta' g)d\} \\ &+ \{(\beta' f)(\beta' g)[d,d] + (\beta' f)d(\beta' g)d - (\beta' g)d(\beta' f)d\}\end{aligned}$$

Notice that, for any $f \in F$, $\beta'[f,d] = 0$ and $\partial(\beta' d)/\partial x = 0$ imply that

$$d(\beta' f) = \frac{\partial(\beta' f)}{\partial x} d = \frac{\partial(\beta' d)}{\partial x} f = 0.$$

Thus

$$\begin{aligned}[\phi(f), \phi(g)] &= [f,g] - \{-g(\beta' f) + f(\beta' g)\}d \\ &= [f,g] - (\beta'[f,g])d.\end{aligned}$$

Note finally that $\phi(d) = d - (\beta' d)d = d - d = 0$.

This result can be applied, for example, to $F = L/\tilde{I}_4$ and d_1, since d_1 is in the center and the third component of d_1 equals 1 (thus $\beta = [0\ 0\ 1\ 0\ \ldots\ 0]'$). The proposition implies that if we implement (15) with a_o, b_o replaced by $a_o - (\beta' a_o)d_1$ and $b_o - (\beta' b_o)d_1$, respectively, then the resulting estimator (call it (15')) will generate a Lie algebra isomorphic to $L/(\tilde{I}_4 \oplus D_1)$. Notice that this transformation (due to the form of d_1) eliminates the $\hat{y}$ equation, modifies the $\widehat{y^2}$ and $\widehat{y^3}$ equations, and does not affect the others. From another point of view, the right-hand side of (15) is transformed from $a_o dt + b_o dz_t$ to

$$\begin{aligned}&a_o dt + b_o dz_t - d_1[(\beta' a_o)dt + (\beta' b_o)dz_t] \\ &= a_o dt + b_o dz_t - d_1 d\hat{y}_t\end{aligned} \tag{18}$$

Denoting the statistics in this estimator which replace $\widehat{y^2}$ and $\widehat{y^3}$ by $\widetilde{y^2}$ and $\widetilde{y^3}$, respectively, we see from (18) and the form of d_1 that

$$d\widetilde{y^2_t} = d\widehat{y^2_t} - 2\hat{y}_t d\hat{y}_t = d\widehat{y^2_t} - d(\hat{y}_t)^2;$$

thus this estimator computes the conditional second central moment $E[(y_t-\hat{y}_t)^2/z^t]$, rather than the second moment $\widehat{y^2_t}$. However,

$$d\widetilde{y^3_t} = d\widehat{y^3_t} - 3\widehat{y^2_t}d\hat{y}_t$$

$$= (24\hat{x}\hat{\xi}\hat{y}P + 48\hat{x}\hat{\Theta}P + 24\hat{\xi}^2P^2 + 6\hat{y}PP_{12} + 24PP_{13} - 48\hat{\xi}\hat{\Theta}P^2$$

$$- 12\hat{y}PP_{13} - 12\hat{\xi}^2\hat{y}P^2 - 24PP_{14} - 24\hat{x}\hat{y}\hat{\Theta}P - 48\hat{x}\hat{\phi}P)dt$$

$$+ (24\hat{\Theta}\hat{y}P+48\hat{\phi}P)dz_t$$

which is not the equation for any easily recognized statistic of the conditional distribution of y_t given z^t. On the other hand, the results of [17] - [18] imply that, since there is a Lie algebra homomorphism from the Lie algebra F of (15) to that of (15') and the isotropy subalgebra of F is {0} at every point, then there is (at least locally) an analytic map λ that carries solutions of (15) into those of (15'). We have already seen that λ takes the components $\hat{x}$, $\hat{\xi}$, $\hat{\Theta}$, $\hat{\phi}$, t into themselves, $\lambda(\widehat{y^2_t}) = E[(y-\hat{y}_t)^2/z^t]$, and $\lambda(\hat{y}_t) = 0$. The image $\lambda(\widehat{y^3_t})$ is difficult to compute, although a method is given in [17]; to first order for small t, $\widetilde{y^3_t} \cong \widehat{y^3_t} - 3\widehat{y^2_o}(\hat{y}_t-\hat{y}_o)$, but more complete calculations are very involved.

4. CONCLUSIONS.

We have presented one example of the method proposed in [9] for using Lie algebraic techniques to study nonlinear estimation problems (a similar analysis can of course be done for other problems in the class discussed in [6] - [8]). This method clarifies the relationship between the computation of the (unnormalized) conditional density and the finite dimensional computation of certain statistics of the conditional distribution (in this case, the conditional moments). Although moments of any order can be computed by a finite dimensional estimator in this example, it is unresolved whether the same is true of the conditional density. That is, the Lie algebra of the Zakai equation (5) is infinite dimensional, but that certainly does not preclude its being isomorphic to a Lie algebra generated by two vector fields on a finite dimensional manifold (which would be the case if it could be computed in terms of finite dimensionally computable sufficient statistics). However, since moments of all orders can be calculated, it may be possible (modulo questions such as moment determinacy) to approximate the conditional density to any desired degree of accuracy by means of a series in the finite dimensionally computable statistics.

On the other hand, the Lie algebra of the Zakai equation may have very few ideals, in which case there may be no statistics which are "more easily" computable than the unnormalized conditional density. Examples of both types and further analysis along these lines will be presented in [22]. Finally, we should warn that Lie algebraic conditions do not always present the whole picture; as discussed in [20], one must essentially be able to "integrate" the abstract Lie algebra representations obtained in order to actually construct the estimator, and this is not always possible (see [23] for one further class of systems for which this <u>is</u> possible).

ACKNOWLEDGEMENT.

The second author would like to thank M. Hazewinkel for many stimulating discussions, and for providing him the opportunity to develop some of these ideas while visiting the Econometric Institute, Erasmus University Rotterdam, Rotterdam.

APPENDIX.

DERIVATION OF FINITE DIMENSIONAL ESTIMATOR.

First we note [6, Appendix B] that if $x = [x_1 \ \dots \ x_k]'$ is a Gaussian random vector with mean m and covariance P, then

$$E[x_1 \dots x_k] = E[x_k]E[x_1 \dots x_{k-1}] + \sum_{\alpha_1=1}^{k} P_{k\alpha_1} E[x_{\alpha_2} \dots x_{\alpha_{k-1}}]$$

$$(A.1) \qquad = m_1 \dots m_k + \sum_{(\alpha_1,\alpha_2)} P_{\alpha_1\alpha_2} m_{\alpha_3} \dots m_{\alpha_k}$$

$$+ \sum_{(\alpha_1,\alpha_2,\alpha_3,\alpha_4)} P_{\alpha_1\alpha_2} P_{\alpha_3\alpha_4} m_{\alpha_5} \dots m_{\alpha_k} + \dots$$

where each set $\{\alpha_i, \ i = 1,\dots,k\}$ is a permutation of $\{1,\dots,k\}$ and the sums in (A.1) are over all possible combinations of pairs of the $\{\alpha_i\}$. Now, x in the problem (12) - (13) is conditionally Gaussian with conditional cross-covariance defined by (for $\sigma_1, \sigma_2 \leq t$)

$$P(\sigma_1,\sigma_2,t) = E[(x_{\sigma_1}-\hat{x}_{\sigma_1|t})(x_{\sigma_2}-\hat{x}_{\sigma_2|t})|z^t],$$

where $\hat{x}_{\sigma|t} = E[x_\sigma|z^t]$; using the results of [6, section 2] it can be shown that

$$(A.2) \qquad \frac{d}{dt} P(\sigma_1,\sigma_2,t) = -P(\sigma_1,t,t)P(\sigma_2,t,t)$$

$$(A.3) \qquad P(\sigma,t,t) = K(t,\sigma)P_t$$

$$(A.4) \qquad \frac{d}{dt} K(t,\sigma) = -P_t^{-1}K(t,\sigma) \ ; \ K(\sigma,\sigma) = 1$$

where $P_t = P(t,t,t)$ is the solution of the Riccati equation (16).

The conditional mean $\hat{y}_t$ satisfies equation (2) in Ito form:

$$(A.5) \qquad d\hat{y}_t = E^t[x_t^2]dt + \{E^t[y_t x_t] - \hat{y}_t\hat{x}_t\}[dz_t-\hat{x}_t dt]$$

But $E^t[x_t^2] = \hat{x}_t^2 + P_t$, and using (A.1), (A.3), and (A.4),

$$E^t[y_t x_t] - \hat{y}_t\hat{x}_t = \int_0^t (E^t[x_s^2 x_t] - E^t[x_s^2]\hat{x}_t)ds$$

$$= \int_0^t 2P(s,t,t)\hat{x}_s ds$$

$$= 2\hat{\xi}_t P_t$$

where ξ satisfies

$$\dot{\xi}_t = x_t - \xi_t P_t^{-1}, \xi_o = 0.$$

Thus the Kalman filter for the system with states x,ξ and observations z computes $\hat{x}$, $\hat{\xi}$, and $\hat{y}$ is computed according to (A.5), thus yielding the first three equations in (15) and P, P_{12} in (16) (once they have been converted to Fisk-Stratonovich form).

Furthermore, since $dy_t^2 = 2y_t dy_t = 2y_t x_t^2 dt$, equation (2) yields

(A.6) $$d\widehat{y_t^2} = 2E^t[y_t x_t^2]dt + \{E^t[y_t^2 x_t] - \widehat{y_t^2}\hat{x}_t\}(dz_t - \hat{x}_t dt).$$

Using (A.1),

$$E^t[y_t x_t^2] = \int_0^t E^t[x_t^2 x_s^2]ds$$

$$= \hat{x}_t^2\hat{y}_t + \hat{y}_t P + 4\hat{x}_t\hat{\xi}_t P + 2PP_{12}.$$

Also, (A.1) - (A.4) imply that

$$E^t[y_t^2 x_t] - \widehat{y_t^2}\hat{x}_t$$

$$= 4\int_0^t \int_0^t P(s,t,t)E^t[x_s x_\tau^2]dsd\tau$$

$$= 4\int_0^t \int_0^t P(s,t,t)[\hat{x}_{s|\tau}E^t[x_\tau^2] + 2P(s,\tau,t)\hat{x}_{\tau|t}]dsd\tau$$

$$= 4\{(\int_0^t P(s,t,t)\hat{x}_{s|t}ds)\hat{y}_t + 2E^t[\int_0^t P(s,t,t)(\int_0^t P(s,\tau,t)x_\tau d\tau)ds]\}$$

$$= 4(\hat{\xi}_t\hat{y}_t P + 2\hat{\Theta}_t P)$$

where Θ_t satisfies

$$\dot{\Theta}_t = P_{12}x_t + \xi_t(P-PP_{12}) - P^{-1}\Theta_t; \quad \Theta_o = 0$$

The Kalman-Bucy filter for the state equations of x, ξ, and Θ with observation z computes $\hat{x}_t$, $\hat{\xi}_t$, $\hat{\Theta}_t$, and $\widehat{y_t^2}$ is computed according to (A.6). After correction terms are added, these result in the first five equations in (15), and the first three in (16). The third moment $\widehat{y^3}$ (and higher moments) are computed in a similar manner.

REFERENCES.

1. H.J. Kushner, "Dynamical Equations for Optimal Nonlinear Filtering", J. Diff. Equations, Vol. 3, 1967, pp. 179-190.

2. M. Fujisaki, G. Kallianpur, and H. Kunita", Stochastic Differential Equations for the Nonlinear Filtering Problem". Osaka J. Math., Vol. 1, 1972, pp. 19-40.

3. R.S. Liptser and A.N. Shiryayev, Statistics of Random Processes I. New York: Springer-Verlag, 1977.

4. H.J. Kushner, "On the Dynamical Equations of Conditional Probability Functions with Application to Optimal Stochastic Control Theory," J. Math. Anal. Appl., Vol. 8, 1964, pp. 332-344.

5. J.H. Van Schuppen, "Stochastic Filtering Theory: A Discussion of Concepts, Methods, and Results", in Stochastic Control Theory and Stochastic Differential Systems, M. Kohlmann and W. Vogel, Eds. New York, Springer-Verlag, 1979.

6. S.I. Marcus and A.S. Willsky, "Algebraic Structure and Finite Dimensional Nonlinear Estimation", SIAM J. Math. Anal., Vol. 9, April 1978, pp. 312-327.

7. S.I. Marcus, "Optimal Nonlinear Estimation for a Class of Discrete-Time Stochastic Systems", IEEE Trans. Automat. Contr., Vol. AC-24, April 1979, pp. 297-302.

8. S.I. Marcus, S.K. Mitter, and D. Ocone, "Finite Dimensional Nonlinear Estimation for a Class of Systems in Continuous and Discrete Time," Proc. of Int. Conf. on Analysis and Optimization of Stochastic Systems, Oxford, Sept. 6-8, 1978.

9. R.W. Brockett, "Remarks on Finite Dimensional Nonlinear Estimation, Journees sur l'Analyse des Systèmes, Bordeaux, 1978.

10. R.W. Brockett, "Classification and Equivalence in Estimation Theory," Proc. 1979 IEEE Conf. on Decision and Control, Ft. Lauderdale, December 1979.

11. S.K. Mitter, "Modeling for Stochastic Systems and Quantum Fields," Proc. 1978 IEEE Conf. on Decision and Control, San Diego, January 1979.

12. M. Zakai, "On the Optimal Filtering of Diffusion Processes," Z. Wahr. Verw. Geb., Vol. 11, 1969, pp. 230-243.

13. R.W. Brockett, "System Theory on Group Manifolds and Coset Spaces," SIAM J. Control, Vol. 10, 1972, pp. 265-284.

14. H.J. Sussmann, "Existence and Uniqueness of Minimal Realizations of Nonlinear Systems," Math. Systems Theory, Vol. 10, 1976/1977.

15. R. Hermann and A.J. Krener, "Nonlinear Controllability and Observability," IEEE Trans. Automat. Contr., Vol. AC-22, October 1977, pp. 728-740.

16. J.E. Humphreys, Introduction to Lie Algebras and Representation Theory. New York: Springer-Verlag, 1972.

17. A.J. Krener, "On the equivalence of Control Systems and the Linearization of Nonlinear Systems," SIAM J. Control, Vol.11, 1973, pp. 670-676.

18. A.J. Krener, "A Decomposition Theory for Differentiable Systems," SIAM J. Control, Vol. 15, 1977, pp. 813-829.

19. E.D. Sontag, Polynomial Response Maps. New York: Springer-Verlag, 1979.

20. S.K. Mitter, in Richerche di Automatica, to appear.

21. S.D. Chikte and J.T.-H. Lo, "Optimal Filters for Bilinear Systems with Nilpotent Lie Algebras," IEEE Trans. Automat. Contr., Vol. AC-24, December 1979, pp. 948-953.

22. M. Hazewinkel, C.-H. Liu, and S.I. Marcus, "Some Examples of Lie Algebraic Structure in Nonlinear Estimation, " to be presented at the 1980 Joint Automatic Control Conference, San Francisco, August 1980.

23. V.E. Benes, "Exact Finite Dimensional Filters for Certain Diffusions with Nonlinear Drift," presented at the 1979 IEEE Conf. on Decision and Control, Ft. Lauderdale, December 1979.

3. SCHUR RECURSIONS, ERROR FORMULAS AND CONVERGENCE OF RATIONAL ESTIMATORS FOR STATIONARY STOCHASTIC SEQUENCES

by

P. DeWilde and H. Dym

ABSTRACT.

In this paper we present an exact and an approximate realization theory for estimation and model filters of second order, stationary stochastic sequences. We use the properties of J-lossless matrices as a unifying framework. We deduce necessary and sufficient conditions for the realization of an estimator and a model filter as a submatrix of a J-lossless system. We show next that an extension of the so called Schur-algorithm provides for approximate J-lossless realizations based on partial past information about the process. We study the geometric properties of such partial realizations and their convergence. Finally, we make connections with the Nevanlinna-Pick problem, and we show how the techniques presented constitute a generalization of many aspects of the Levinson-Szegö theory of partial realizations. As a consequence generalized recursive formulas for reproducing kernels and Christoffel-Darboux formulas are obtained. The paper considers only the scalar case for easy readability. However, the matrix case may be obtained without much more difficulty.

1. INTRODUCTION

Let $x(t)$, $t = 0, \pm 1, \ldots,$ denote a real, stationary, zero mean scalar stochastic sequence with covariance function

$$r(k) = E\, x(t)x(t-k)$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ik\Theta} W(e^{i\Theta})\, d\Theta$$

and spectral density W.

A shaping (or modeling) filter for this stochastic process is a linear, time invariant, causal, filter which, when driven by white noise, produces an output with covariance equal to r(k). An innovations filter is the inverse of a shaping filter. It produces white noise when it is driven by the given stochastic sequence. In practice it is of interest to establish *recursive procedures* for computing approximate shaping and innovations filters and thus avoid having to recompute the whole filter each time the permitted complexity is increased: the additional improvement should be achieved by adding sections without changing the existing structure. The same philosophy underlies both the "ladder structure" used in modern digital filter design and the theory of expanding in terms of orthonormal functions: the initial coefficients $f_j = (f,\phi_j)$ of f with respect to an orthonormal set of functions ϕ_j, j = 1, ..., n, in a Hilbert space with inner product (,) do not have to be recomputed as the size of the set changes.

Levinson established a recursive procedure for computing a sequence of approximate autoregressive (AR) shaping filters from the covariance data which, under mild conditions, converge to the exact shaping filter. The theory of these filters is intimitely connected with the theory of polynomials on the unit circle which are orthogonal with respect to the spectral density of the underlying stochastic sequence. In the paper of DeWilde, Vieira, and Kailath it was shown that the Levinson procedure could be viewed as a special case of an algorithm for coprime factorization. In the present paper we exploit the greater flexibility afforded by the Schur algorithm to construct a recursive sequence of autoregressive moving average (ARMA) approximate filters. We shall show that the approximate filter produced are optimal in a least squares sense and shall derive explicit formulas for the approximation error. Moreover we shall show that under appropriate conditions the approximate filters converge in a strong sense to the exact filter. Finally we deduce generalized Christoffel-Darboux formulas for the appropriate reproducing kernels. The theory associated with the Levinson procedure will emerge as a special case.

The paper is organized as follows: in the next section we summarize the relevant material from the theory of J-lossless matrices and discuss the natural correspondence which exists between these matrices and a class of suitably restricted spectral densities W. The theorem (2.2) which is proved there is somewhat more general than is needed in the sequel but, we think, adds a certain amount of perspective to the ensuing discussion. The basic strategy is to approximate the given W by an appropriately chosen sequence W_n of spectral density with associated J-lossless matrices and then to allow $n \uparrow \infty$. The actual procedure followed, however, is quite concrete and is based on a recursive algorithm due to Schur which is explained in section 3. Applications to various cases are then made in sections 4 and 6. Section 5 contains a short discussion of the Adamyan-Arov-Krein representation formula.

The full paper will be publishbd in the Trans IEEE on Information theory. Meanwhile it is available as preprint no 92 (1979), Dept. Electrical Engineering, Delft Technological University, Delft, The Netherlands.

The Stochastic Filtering Problem for Point Processes*)

by

J.H. van Schuppen

ABSTRACT

The purpose of this paper is to give a brief exposition of the stochastic filtering problem in the case of point process observations. Some examples of these problems will be presented. We discuss the modelling of point processes and formulate the associated stochastic dynamical systems. Two methods to resolve the stochastic filtering problem will be given, namely the semimartingale representation method and the measure transformation method. For several examples the solution to the stochastic filtering problem will be given. Some open questions are mentioned.

KEY WORDS & PHRASES: *Point processes, Marked point processes, stochastic analysis, stochastic dynamical systems, stochastic filtering problems.*

---

*) This report will appear in the Proceedings of the meeting on "New Trends in Filtering and Identification of Stochastic Systems", held on January 23, 1980, at the Erasmus University, Rotterdam, The Netherlands. Not for review.

1. INTRODUCTION

The purpose of this paper is to give a brief exposition of the stochastic filtering problem for point process observations. No proofs and few details will be presented. However, references to the literature have been provided.

For many practical filtering problems the observations are not continuous processes, but point processes or marked point processes. Examples of such problems arise in the areas of optical communication, nuclear medicine, urban traffic control, and operations research. Many questions in these areas may be formulated as stochastic filtering problems or stochastic control problems.

What results have been obtained for these problems? Representations for point processes have been obtained using concepts from the theory of stochastic processes. These representations may be considered to be stochastic dynamical systems. For the stochastic filtering problem for these systems two methods have been developed. Both of these methods yield general representation results that have to be applied to specific models. For several models filtering algorithms have been obtained. A brief summary of these results is presented below.

For the material on point processes we refer to the books [16, 18, 24], and for application oriented books we suggest [21, 35].

2. MOTIVATING EXAMPLES

Some examples of stochastic filtering problems with point process observations are presented below.

<u>EXAMPLE 2.1</u>. Optical Communication

The physical model is as follows. A signal modulates an optical source, which in turn generates an optical beam. The beam, after travelling a certain distance, is incident on a detector. The optimal detector produces an electric current which we regard as the observed process.

The problem then is to estimate the signal on the basis of the observed process.

Particular cases of the above problem occur in:

1) experiments in nuclear physics;
2) light scattering studies;
3) optical communication with lasers.

For references on these problems see [19, 20, 22, 33].

EXAMPLE 2.2. An Industrial Problem

The model is that of a machine that irregularly produces unit products. Initially the machine operates at full capacity. At some point in time the machine breaks down partially, with the effect that it yields output at a lower rate. One may then associate a cost function with this model, such that under normal conditions a profit is made and a loss when the machine is partially defective.

Then problem is then:

1) to estimate when the machine breaks down partially on the basis of the production data only;
2) to resolve the stochastic control problem of when to shut down the machine so as to minimize the costs.

The first problem mentioned above is also known as the Poisson disorder problem. For references see [12, 40].

EXAMPLE 2.3. Traffic Estimation

The model is that of an urban traffic network. Information on the traffic flow is obtained from detection lines. The ultimate objective is computer control of urban traffic.

The problem here is to estimate and to predict traffic intensities.

For some references on this problem see [1, 2, 36].

EXAMPLE 2.4. The Firefly Model

The model is for a swarm of flireflies. Each of the fireflies irregularly produces flashes of light. One assumes that the swarm has a Gaussian density, and that the mean of this density moves around according to a Gauss-Markov process. The authors of this model leave it to the imagination of the reader to think of other applications of this model.

The problem is to estimate the mean of the density on the basis of the observations of the light flashes [41].

EXAMPLE 2.5. Filtering in Queueing Problems

Consider a model for a waiting line, in which we distinguish an arrival and a departure process, and the queue process. One may also consider a network of queues.

The problem is to estimate the queue process on the basis of the departure process. The theory of stochastic filtering is also used to prove a certain result for so-called Jackson networks.

For references on the application of stochastic filtering to queueing problems see [7, 9].

EXAMPLE 2.6. A Miscellaneous Model

We finally mention a model that has been addressed in the literature. Here the signal process x is a finite state or denumerable state Markov process. The observed process y is denumerably valued and related to x by the equation $y_t = h(t,x_t)$.

The problem then is to estimate the signal process given past observations. For references on this problem see [25, 26]. In this paper we will not discuss this model any further.

3. MODELLING OF JUMP PROCESSES

In this section we answer the question: How to model a jump process? In the sequel we limit attention to continuous time processes. We assume given a complete probability space (Ω,F,P). In this paper we will not be very detailed about technical conditions. The reader is referred to the references [3, 10, 13, 18, 21, 23] for definitions and results.

3.1. Single Unit Jump Case

We start the modelling of jump processes with a rather elementary case. Let $x: \Omega \times T \to R$, on $T = R_+$, be a process with a single unit jump, with $x_0 = 0$. Define the jump time as

$$\tau(\omega) := \begin{cases} \inf\{t \in T \mid x_t(\omega) \neq x_0(\omega)\}, \\ +\infty \qquad \text{if } x_t(\omega) = x_0(\omega) \text{ for all } \omega \in \Omega, \end{cases}$$

and the jump distribution function as $f(t) = P(\{\tau \leq t\})$.

We then have the following characterization.

THEOREM 3.1.

(a) *Given the single unit jump process described above. Then there exists an unique process* $\bar{x}: \Omega \times T \to \overline{R_+}$ *such that*

$$x = \bar{x} + m \qquad (*)$$

where $(m_t, F_t, t \in T) \in M_1$, *meaning it is a martingale. Actually* $\bar{x}$ *is given by*

$$\bar{x}_t = \int_{(0,t\wedge\tau]} [1 - f(u-)]^{-1} f(du).$$

Here $(F_t, t \in T)$ *is an increasing* σ*-algebra family.*

(b) *The process* $\bar{x}$ *uniquely characterizes the measure* P. [11, 18:III].

The decomposition (*) above is called the special semimartingale decomposition. This decomposition will play a fundamental role throughout this discussion.

To further clarify jump processes with a single unit jump, we illustrate the classification of jump times, as introduced in [13], with some examples.

1. Totally inaccessible jump times. Example: $f(t) = 1 - \exp(-t)$. Here the distribution of the jump is diffuse on R_+.
2. Predictable jump times. Example: Jump distribution $f(t) = I_{[c,\infty)}(t)$, for some $c \in R_+$. Intuitively this jump can be predicted with certainty.

3. Accessible jump times. Example: Jump distribution $f(t) = \Sigma_{k=1}^{n} \alpha_k I_{[u_k,\infty)}(t)$ with $0 < u_1 < u_2 < \ldots < u_n < \infty$, and $\alpha_k \in R$ with $\Sigma_{k=1}^{n} \alpha_k = 1$. In this case the jump is not predictable but can only occur at certain time moments.

From the above classification one may deduce that in general one uses totally inaccessible jump times.

3.2. The Counting Process Case

The modelling of jump processes introduced above may be extended to counting processes. A stochastic process is a counting process if it starts at time zero, is piecewise constant, and has positive unit jumps. By convention it is taken to be right continuous. A counting process is also called a point process.

THEOREM 3.2.1. *Given a counting process* $(n_t, F_t, t \in T)$.

(a) *There exists an unique increasing predictable process* $a: \Omega \times T \to R_+$ *such that*

$$n = a + m \qquad (*)$$

with $(m_t, F_t, t \in T) \in M_{1uloc}$, *meaning* m *is a local martingale* [23].

(b) *If there exists a process* $\lambda: \Omega \times F \to R_+$ *such that*

$$a_t = \int_0^t \lambda_s ds$$

then we call λ *the rate process associated with* $(n_t, F_t, t \in T)$. *In this case we obtain the representation*

$$dn_t = \lambda_t dt + dm_t, \qquad n_0 = 0. \qquad (**)$$

(c) *The process* a *characterizes uniquely the measure* P *with respect to the counting process* n.

References [11, 18]. The decomposition (*) and (**) above are called special semimartingale decompositions.

EXAMPLE 3.2.2. Let $n: \Omega \times T \to R$, $\lambda: \Omega \times T \to R_+$ be processes such that

$$E[\exp(iv(n_t - n_s)) \mid F_s^n \vee F_\infty^\lambda] = \exp((e^{iv}-1)(\int_s^t \lambda_\tau d\tau)).$$

Thus, conditioned on F_∞^λ, n is a Poisson process with rate λ. Such a process is called a *doubly stochastic Poisson process*. This description is equivalent with the description

$$dn_t = \lambda_t dt + dm_t, \qquad n_0 = 0,$$

with $(m_t, F_t^n \vee F_\infty^\lambda, t \in T) \in M_{1uloc}$.

3.3. Arbitrary Jump Processes

We first mention several descriptions of jump processes.

1. The *marked point process* description: given $\{x_n, s_n, n \in Z_+\}$, where s_n represents the interarrival time between the n-1 and the n-th jump, and x_n the value or mark at the n-th jump. Let $\tau_0 = 0$, and for $n \in Z_+$ $\tau_n = \Sigma_{k=1}^n s_k$, to be called the n-th jump time.
2. The *jump process* description: given the process $x: \Omega \times T \to R$ such that

$$x_t = \sum_{n \in Z_+} x_n I_{(\tau_n \le t < \tau_{n-1})}.$$

3. The *jump measure* description: given the random measure $p: \Omega \times B_T \otimes B \to \overline{R_+}$,

$$p(\omega, A) = \sum_{n \in Z_+} I_A(\tau_n(\omega), x_n(\omega)).$$

The above descriptions can be shown to be equivalent. The sought for characterization reads then as follows.

THEOREM 3.3.1. *Given a jump measure* $p: \Omega \times B_T \otimes B \to \overline{R_+}$.

(a) *There exists an unique predictable random measure* $\bar{p}: \Omega \times B_T \otimes B \to \overline{R_+}$ *such that*

$$p(\omega, dt \times dz) = \bar{p}(\omega, dt \times dz) + q(\omega, dt \times dz)$$

with $(q(\omega, 0,t] \times A), F_t, t \in T) \in M_{1uloc}$ *for all* $A \in B$.

(b) *The random measure* $\bar{p}$ *uniquely characterizes the measure* P. [18:III].

3.4. Modelling of the Rate Process

Consider a counting process n which admits a rate process λ, say with representation

$$dn_t = \lambda_t dt + dm_t, \qquad n_0 = 0.$$

The question then is how to model the rate process. We present several models for the rate process that are used in the literature.

1. The constant rate process: $\lambda_t = \lambda_0$ for all $t \in T$, with some distribution for λ_0 specified.
2. The rate process as a finite or denumerable state Markov process.
3. The energy model: $\lambda_t = \mu_0 + \mu_1 x_t^2$ with $\mu_0, \mu_1 \in (0,\infty)$ and x a Gauss-Markov or a diffusion process.
4. The linear model: $\lambda_t = \mu_0 + \mu_1 x_t$, with $\mu_0, \mu_1 \in (0,\infty)$, and x a diffusion process. The problem with this model is that the rate process has to be stopped if it becomes negative; hence it is not a useful model.

As an illustration we present one example of a model for a counting process.

EXAMPLE 3.4.1. The Poisson-FSMP Model. Given the process, $n: \Omega \times T \to R$, with $n_0 = 0$,

$$E[\exp(iv(n_t - n_s)) \mid F_s^n \vee F_\infty^\lambda] = \exp((e^{iv}-1)(\int_s^t \lambda_\tau d\tau)),$$

for all $t > s$, $v \in R$,

and the process $\lambda: \Omega \times T \to X := \{r_1, r_2, \ldots, r_m\} \subset (0,\infty)$ which is a finite state Markov process on $(\lambda_t, F_t^n \vee F_t^\lambda, t \in T)$. Under certain differentiability conditions on the semigroup of the Markov process we may represent these processes as

$$dz_t = A(t)z_t dt + \phi(t,0)dm_{1t}, z_0,$$

$$dn_t = D\, z_t dt + dm_t, \qquad n_o = 0,$$

$$\lambda_t = D\, z_t,$$

where

$$z: \Omega \times T \to R^m, \quad z_t^i := I_{(\lambda_t = r_i)},$$

$$D := (r_1 r_2 \ldots r_m) \in R^{1\times m}, \quad \phi: T \times T \to R^{m\times m},$$

$$\phi_{ij}(t,s) := \begin{cases} E[z_t^i z_s^j]/E[z_s^j], & \text{if } E[z_s^i] > 0,\ s \le t, \\ 0, & \text{otherwise,} \end{cases}$$

$$A(t) := \lim_{s\to t} [\phi(t,s) - I]/(t-s),$$

$$(m_t, F_t^n \vee F_t^\lambda, t \in T) \in M_{1uloc}, \quad (m_{1t}, F_t^n \vee F_t^\lambda, t \in T) \in M_1.$$

3.5. Stochastic Dynamical Systems

The preceding representations of jump processes may be considered as stochastic dynamical systems. The concept of a stochastic dynamical system we define below. For a discussion of this notion and a formulation of the stochastic realization problem see [37].

DEFINITION 3.5.1. A *stochastic dynamical system* (in continuous time) is a collection

$$\{\Omega, F, P, T, Y, B_Y, \underline{Y}, X, B_X\}$$

where $\{\Omega,F,P\}$ is a complete probability space, $T \subset R$ an interval, $\{Y,B_Y\}$ a measurable space with Y a vector space, $\underline{Y} \subset \{y: T \to Y\}$, $\{X,B_X\}$ a measurable space, such that if $x: \Omega \times T \to X$, $y: \Omega \times T \to Y$ are stochastic processes then

$$({}_tF^{\Delta y} \vee {}_tF^x,\ F^{x_t},F_t^x \vee F_t^y) \in CI$$

for all $t \in T$; equivalently, if

$$E[I_{A_1} I_{A_2} \mid F_t^x \vee F_t^y] = E[I_{A_1} I_{A_2} \mid F^{x_t}]$$

for all $A_1 \in {}_tF^{\Delta y}$, $A_2 \in {}_tF^x$, $t \in T$. Here

$$F^{x_t} = \sigma(\{x_t\}), \quad F_t^x = \bigvee_{s \le t} F^{x_s}, \quad {}_tF^x = \bigvee_{s > t} F^{x_s},$$

$$ {}_tF^{\Delta y} = \sigma(\{y_s - y_t,\ \forall s > t\}),$$

$$F_t^y = \bigvee_{s \le t} F^{y_s}.$$

NOTATIONS. $(\Omega,F,P,T,Y,B_y,\underline{Y},X,B_x) \in \Sigma S$, and we call x the *state process* and y the *output process*.

In the sequel we assume that $Y = R^k$, $B_y = B_k$ the Borel σ-algebra on Y, $X = R^n$, and $B_x = B_n$.

The above definition expresses that the distribution of a future state and a future output increment conditioned on past states and past outputs, depends only on the current state. This property is the characteristic of a stochastic dynamical systems. An immediate consequence of this definition is that the state process is a Markov process. For a stochastic dynamical system one may also formulate the concepts of stochastic observability and stochastic reconstructability, see [37]. However for a number of concepts related to a stochastic dynamical system precise formulations are not yet clear.

4. THE STOCHASTIC FILTERING PROBLEM

DEFINITION 4.1. Given a stochastic dynamical system,

$$(\Omega,F,P,T,Y,B_y,\underline{Y},X,B_x) \in \Sigma S.$$

(a) The *stochastic filtering problem* for this system is to determine

$$E[\exp(iu'x_t) \mid F_t^y]$$

for all $u \in R^n$, $t \in T$.

(b) A *past output based filter system* for the stochastic filtering problem defined above is a stochastic dynamical system

$$(\Omega,F,P,T,Y,B_Y,\underline{Y},Z,B_Z) \in \Sigma S$$

such that if $y: \Omega \times T \to Y$, $z: \Omega \times T \to Z$ are the underlying processes then $F^{z_t} \subset F_t^y$ for all $t \in T$.

In part (a) above to determine the conditional characteristic function means to exhibit the analytic form of the map

$$y_{[0,t]} \to E[\exp(iu'x_t) \mid F_t^y].$$

A filter system always exists, because we can take $Z = \underline{Y}$. It is therefore of interest to find a filter state space Z which is in some sense minimal. The concept of the dimension of such a state space is not yet clear. For a stochastic filtering problem with continuous observations it has been suggested to relate the dimension of a filter system to the dimension of the Lie algebra associated with the operators occurring in the equation for the conditional density. For this issue there still are many open questions.

How to resolve the stochastic filtering problem? The general procedure is to derive an equation for $E[\exp(iu'x_t) \mid F_t^y]$, and to solve this equation. We present two methods to effect this program.

5. THE SEMIMARTINGALE REPRESENTATION METHOD

In this section we present the semimartingale representation method to resolve the stochastic filtering problem. Initially we do not work with the state process but with an arbitrary special semimartingale process. The problem we then consider is to find the special semimartingale decomposition of the projection of this process on the σ-algebra family generated by the observations. For specific stochastic dynamical systems this abstract representation can be applied to yield a partial stochastic differential equation for the conditional characteristic function. Below we first state two abstract representation results, and subsequently show how these results are applied to obtain filtering algorithms.

5.1. The Counting Process Case

We summarize the main result. For a precise statement see the references mentioned below.

MODEL 5.1.1. Given a counting process model

$$dx_t = f_t dt + dm_{1t}, \qquad x_0,$$

$$dn_t = \lambda_t dt + dm_t, \qquad n_0 = 0,$$

$$d\langle m_1, m\rangle_t = \phi_t dt,$$

where n represents a counting process, assumed to have totally inaccessible jump times, λ the rate process, and x is a semimartingale with the indicated decomposition.

PROBLEM 5.1.2. To obtain the special semimartingale representation of the projection of the process x on the σ-algebra family $(F_t^n,\ t \in T)$. For the projection we take the so-called optional projection [13], which we denote by $(\hat{x}_t,\ t \in T)$. Then it follows that $\hat{x}_t = E[x_t \mid F_t^n]$ a.s. for all $t \in T$.

RESULT 5.1.3. The solution to the above formulated problem is given by

$$d\hat{x}_t = \hat{f}_t dt + [\hat{\Sigma}_{t-}^{x\lambda} + \hat{\phi}_{t-}]\hat{\lambda}_{t-}^{-1}(dn_t - \hat{\lambda}_t dt), \qquad \hat{x}_0 = E(x_0)$$

$$\hat{\Sigma}_t^{x\lambda} = E[(x_t - \hat{x}_t)(\lambda_t - \hat{\lambda}_t) \mid F_t^n].$$

References [6, 21, 27, 29, 30, 31, 38, 39].

5.2. The Jump Process Case

MODEL 5.2.1. Given the processes

$$x = x_0 + a + m$$

$$p(\omega, dt\times dz) = h(\omega,t,z)\mu(\omega,dt\times dz) + q(\omega,dt\times dz),$$

$$d\langle m, q(\omega,(0,t]\times A)\rangle_t = \int_A \psi(\omega,s,z)\mu(\omega,dt\times dz),$$

where x is a semimartingale, the second line a jump measure description of a jump process, and the third line represents the relation between the jump process and the semimartingale. We assume that the jump times are totally inaccessible.

PROBLEM 5.2.2. To determine the semimartingale representation of the projection of the process x on the σ-algebra family generated by the jump process.

RESULT 5.2.3. The solution to the above formulated problem is given by the representation

$$\hat{x}_t = \hat{x}_0 + \bar{a}_t + \int_0^t \int_R \hat{k}(\omega,s-,z)\bar{q}(\omega,ds\times dz),$$

$$\hat{q}(\omega,dt\times dz) = p(\omega,dt\times dz) - \hat{h}(\omega,s,z)\mu(\omega,dt\times dz),$$

$$\hat{k}(\omega,t,z) = [E[(x_t-\hat{x}_t)(h(\omega,t,z) - \hat{h}(\omega,t,z)) \mid F_t^y]$$

$$+ E[\psi(\omega,t,z)h(\omega,t,z) \mid F_t^y]]\hat{h}^{-1}(\omega,t,z).$$

For a precise statement of this result see [3]. Related references are [8, 17, 41].

5.3. Examples

We present the solutions to stochastic filtering problems for certain examples.

EXAMPLE 5.3.1. The Poisson-FSMP model. This model has been formulated in 3.4.1 and reads as follows,

$$dz_t = A(t)z_t dt + \phi(t,0)dm_{1t}, \qquad z_0,$$

$$dn_t = D\, z_t dt + dm_t, \qquad n_0 = 0.$$

$$\lambda_t = D\, z_t.$$

The solution to the stochastic filtering problem for this model is

$$E[\exp(iu\lambda_t) \mid F_t^n] = \sum_{j=1}^{m} \exp(iu_j r_j)\hat{z}_t^j,$$

$$d\hat{z}_t = A(t)\hat{z}_t dt + K(\hat{z}_{t-})(D\hat{z}_{t-})^{-1}(dn_t - D\hat{z}_t dt), \qquad \hat{z}_0 = E(z_0),$$

$$K(\hat{z}_t) = [\text{diagonal}(\hat{z}_t) - \hat{z}_t\hat{z}_t']D',$$

$$\hat{\lambda}_t = D\hat{z}_t.$$

Reference [28]. The above result can be extended to the case where λ is a denumerable state Markov process. An application of this model is in the estimation problem for the industrial model formulated in (2.2).

EXAMPLE 5.3.2. The Poisson-Gamma model. This model is rather elementary. We use it to illustrate the solution procedure for the stochastic filtering problem by the semimartingale representation method. The model is specified by

$$dx_t = \alpha x_t dt, \qquad x_0,$$

$$dn_t = x_t dt + dm_t, \qquad n_0 = 0,$$

where n is a counting process, x its associated rate process, $\alpha \in (-\infty,0)$, $x_0: \Omega \to R_+$ a random variable with a Gamma distribution function, with density function $p(v) = \beta^{-r}v^{r+1}e^{-v/\beta}/\Gamma(r)$, $r, \beta \in (0,\infty)$. Of course $x_t = \exp(\alpha t)x_0$.

We sketch the solution procedure. Set

$$c: \Omega \times T \times R \to C,$$

$$c_t(u) := \exp(iux_t).$$

Then

$$dc_t(u) = iu\alpha x_t c_t(u)dt, \qquad c_0(u).$$

Applying the semimartingale representation result to the process $c(u)$ we obtain

$$d\hat{c}_t(u) = u\alpha\hat{c}'_t(u)dt + [-i\hat{c}'_t(u) + i\hat{c}'_{t-}(0)\hat{c}_{t-}(u)](-i\hat{c}'_{t-}(0))^{-1} (dn_t + i\hat{c}'_{t-}(0)dt),$$

$$\hat{c}_0(u) = E[\exp(iux_0)] = (1 - iu\beta)^{-r},$$

$$\hat{c}'_t(u) = \partial\hat{c}_t(u)/\partial u.$$

This is a partial stochastic differential equation, driven by a counting process.

The solution to the stochastic filtering problem for the above model then is

$$E[\exp(iux_t) \mid F^n_t] = (1 - iu\beta(t))^{-(n_t+r)},$$

$$\dot{\beta}(t) = \alpha\beta(t) - \beta^2(t), \qquad \beta(0) = \beta.$$

Then it follows that

$$\hat{x}_t = E[x_t \mid F^n_t] = \beta(t)(n_t + r),$$

satisfies the stochastic differential equation,

$$d\hat{x}_t = \alpha\hat{x}_t dt + \beta(t)(dn_t - \hat{x}_t dt), \qquad \hat{x}_0 = r\beta.$$

References [15, 27].

<u>EXAMPLE 5.3.3</u>. The firefly model: according to (2.4) we have

$$dx_t = \alpha x_t dt + \beta dv_t, \qquad x_0,$$

$$p(\omega, dt\times dz) = (2\pi\sigma(t)^2)^{-\frac{1}{2}}\exp(-(z-x_t)^2/2\sigma(t)^2)\,dtdz + q(\omega, dt\times dz),$$

where $\alpha \in R_-$, $\beta \in R$, v is a standard Brownian motion process, $x_0 \colon \Omega \to R$ is a random variable with a Gaussian distribution. The solution to the stochastic filtering problem is

$$E[\exp(iux_t) \mid F_t^y] = \exp(iu\hat{x}_t - \frac{1}{2}u^2\hat{r}_t),$$

$$d\hat{x}_t = \alpha\hat{x}_t dt + k_{t-}\int_R (z-\hat{x}_{t-})\bar{q}(\omega, dt\times dz),$$

$$d\hat{r}_t = [2\alpha\hat{r}_t + \beta^2 - \hat{r}_t^2(\sigma(t)^2 + \hat{r}_t)^{-1}]dt + k_{t-}\hat{r}_{t-}\int_R \bar{q}(\omega, dt\times dz),$$

$$\bar{q}(\omega, dt\times dz) = p(\omega, dt\times dz) - (2\pi(\sigma(t)^2+\hat{r}_t))^{-\frac{1}{2}}\exp(-(z-\hat{x}_t)^2/$$

$$/(2\pi(\sigma(t)^2 + \hat{r}_t)))\,dtdz,$$

$$k_t = \hat{r}_t(\sigma(t)^2 + \hat{r}_t)^{-1}.$$

Reference [41]. The Gaussian density and the Gaussian distributions are essential here. The filter system has some analogy with the Kalman-Bucy filter system. It differs from this filter system in that the equation for the conditional covariance depends directly on the observations. A serious difficulty is the integration over the jump measures.

<u>EXAMPLE 5.3.4</u>. A model for optical communication:

$$dx_t = \alpha x_t dt + \beta dv_t, \qquad x_0,$$

$$dn_t = (\mu_0 + \mu_1 x_t^2)\,dt + dm_t, \qquad n_0 = 0,$$

$$\lambda_t = (\mu_0 + \mu_1 x_t^2),$$

where v is a standard Brownian motion process, and $x_0\colon \Omega \to R$ is a Gaussian distributed random variable. The solution to the stochastic filtering problem for this model is only known locally, between jump times,

$$E[\exp(iu\lambda_t) \mid F_t^n] I_{(\tau_k \le t < \tau_{k+1})}$$
$$= \exp(iu\mu_0) \sum_{j=0}^{k} a_{kj}(t)[1 - iug(t)]^{-j-\frac{1}{2}} / [\sum_{j=0}^{k} a_{kj}(t)],$$

where $(a_{kj}(t),\ F_t^n,\ t \in T)$ are adapted stochastic processes for which equations are known, and $g\colon T \to R$ is a determinstic function. Reference [4]. The characteristic function above is locally a convex combination of characteristic functions of gamma type. The resulting filter system does not seem to be finite dimensional. Aprpoximations to the solution may be attempted.

6. THE MEASURE TRANSFORMATION METHOD

6.1. Introduction

A second method to resolve the stochastic filtering problem is the measure transformation method. To introduce this method we first consider an elementary example.

EXAMPLE 6.1.1. Consider the model with random variables $n\colon \Omega \to R_+$, $\lambda\colon \Omega \to R_+$ such that

$$E[\exp(iun) \mid F^\lambda] = \exp(\lambda(e^{iu}-1)),$$

λ has a Gamma distribution, with density function $p(v) = \beta^{-r} v^{r+1} e^{-v/\beta}/\Gamma(r)$, with $\beta, r \in (0,\infty)$. Thus conditioned on λ, n has a Poisson distribution. The problem is to determine

$$E[\exp(iu\lambda) \mid F^n].$$

We sketch the method. Let $\rho = \lambda^n \exp(-\lambda+1)$. Define a new measure P_0 on (Ω,F) by $P_0(A) = E[I_A \rho^{-1}]$. Then it can be shown that: 1. P_0 is absolute continuous with respect to the original measure P; 2. $E_0[\exp(iun)] = \exp(e^{iu}-1)$; 3. $(F^n, F^\lambda) \in I(P_0)$, or F^n, F^λ are independent under P_0; 4. $P_0 = P_1$ on F^λ; 5.

$$E[\exp(iu\lambda) \mid F^n] = E_0[\exp(iu\lambda)\rho \mid F^n]/E_0[\rho \mid F^n]$$

$$= (1 - iu\beta/(\beta+1))^{-(n+r)}.$$

The above procedure may be considered as a measure theoretic formulation of the Bayes method. The calculation in point five above is straightforward because under P_0, F^n and F^λ are independent.

We can now formulate the measure transformation method to resolve the stochastic filtering problem. It consists of the steps:

1. to perform a measure tranformation such that under the new measure the state process and the observed process are independent;
2. to obtain a semimartingale representation for the unnormalized conditional characteristic function with respect to the new measure. This representation will be in the form of a partial stochastic differential equation.

The advantage of this method is that the calculations under the new measure are easier than under the old measure. Below we briefly sketch the measure transformation method for the case of counting process observations.

6.2. The Counting Process Case

THEOREM 6.2.1. *Given the processes, with respect to a measure* P_1,

$$x: \Omega \times T \to R,$$

$$dn_t = \lambda(x_t)dt + dm_t, \qquad n_0 = 0,$$

where n *is a counting process.*

(a) *Then there exists a probability measure* P_0 *on* (Ω,F) *such that:*

1. $P_1 \ll P_0$ *with* $\rho_t := E_0[dP_1/dP_0 \mid F_t^n \vee F_\infty^x] =$
   $= \exp(\int_0^t \ell n(\lambda(x_s))dn_s - \int_0^t [\lambda(x_s) - 1]ds)$;
2. *under* P_0, n *is a standard Poisson process;*
3. $(F_\infty^n, F_\infty^x) \in I(P_0)$;
4. $P_1 = P_0$ *on* F_∞^x;
5. $E_1[\exp(iux_t) \mid F_t^n] = E_0[\exp(iux_t)\rho_t \mid F_t^n]/E_0[\rho_t \mid F_t^n]$ (*)

(b) *If in addition* x *has the representation*

$$dx_t = f(x_t)dt + dm_{1t}, \qquad x_0,$$

$$d\langle m_1, m_1\rangle_t = \phi_t dt,$$

with $(m_{1t}, F_t, t \in T) \in M_1^c$ *and suitable conditions on* x_0, f, ϕ, *then we have the partial stochastic differential equation*

$$dE_0[\exp(iux_t)\rho_t \mid F_t^n]$$

$$= E_0[iu\rho_t f(x_t)\exp(iux_t) - \frac{1}{2} u^2 \rho_t \phi_t \exp(iux_t) \mid F_t^n]dt$$

$$+ E_0[\rho_t[\lambda(x_t) - 1]\exp(iux_t) \mid F_{t-}^n](dn_t - dt), \quad E_0[\exp(iux_0)].$$

References [3, 5]. We call the process $E_0[\exp(iux_t)\rho_t \mid F_t^n]$ the unnormalized conditional characteristic function. Note that if it is known, then by setting u = 0 one obtains the denumerator in (*), and thus the desired expression via again (*)

7. OPEN QUESTIONS

We mention some open questions for the stochastic filtering problem in the case of jump process observations.

1. Which stochastic dynamical systens lead to stochastic filter systems which are, in some sense, finite dimensional? One difficulty is that the concept of finite dimensionality is not yet clear. However the examples of 5.3 indicate what is understood by this term. The importance of the finite dimensionality of a filter system is clear from the viewpoint of applications. One would hope to obtain sufficient conditions for a stochastic dynamical system such that the associated filter system is finite dimensional. It is possible that differential geometric concepts play a role in answering this question.
2. Can stochastic realization be useful in resolving the stochastic filtering problem? For second order processes stochastic realization theory has provided new insights for the stochastic filtering problem. For jump processes this approach is still undeveloped.
3. How useful is the approximation of a rate process by a finite state Markov process? From a viewpoint of applications the most useful result is the filtering algorithm 5.3.1, for the case where the state process is a finite state Markov process. For the application of this result to concrete problems there are several questions of modelling and implementation.

8. REFERENCES

[1] BARAS, J.S., W.S. LEVINE & T.L. LIN, *Discrete-time point processes in urban traffic queue estimation,* IEEE Trans. Automatic Control, 24 (1979), pp. 12-27.

[2] BARAS, J.S., A.J. DORSEY & W.S. LEVINE, *Estimation of Traffic Platoon Structure from Headway Statistics,* IEEE Trans. Automatic Control, 24 (1979), pp. 553-559.

[3] BOEL, R., P. VARAIYA & E. WONG, *Martingales on Jump Processes; I: Representation Results; II: Applications,* SIAM J. Control, 13 (1975), pp. 999-1061.

[4] BOEL, R.K. & V.E. BENES, *Recursive nonlinear estimation of a diffusion acting as the rate of an observed Poisson process,* IEEE Trans. Information Theory, to appear.

[5] BRÉMAUD, P., *A Martingale Approach to Point Processes*, Ph.D. thesis, Univ. of California, Berkeley, Ca., August, 1972.

[6] BRÉMAUD, P., *The matingale theory of point processes over the real half line admitting the intensity*, Proc. IRIA Coll. Control Theory, June, 1974, Lecture Notes Econ. Math. Systems 107, Springer, Berlin, pp. 519-542.

[7] BRÉMAUD, P., *Estimation de l'état d'une file d'attente et du temps de panne d'une machine par la méthode des semi-martingales*, Adv. in Appl. Prob., 7 (1975), pp. 845-863.

[8] BRÉMAUD, P., *La Méthode des Semi-Martingales en Filtrage quand l'observation est un Processu Ponctuel marque*, in *"Séminaire Probabilités X"*, Lecture Notes in Math., volume 511, Springer-Verlag, Berlin, 1975, pp. 1-18.

[9] BRÉMAUD, P., *On the output theorem of queueing, via filtering*, J. Appl. Prob., 15 (1978), pp. 397-405.

[10] BRÉMAUD, P. & J. JACOD, *Processus Punctuels et Martingales: Revue des Résultats récents sur la modélisation et le filtrage*, Adv. in Appl. Prob., 9 (1977), pp. 362-416.

[11] DAVIS, M.H.A., *The Structure of Jump Processes and Related Control Problems*, in *"Stochastic Systems: Modelling, Identification and Optimization II"*, Math. Programming Studies, no. 6, North-Holland Publ. Co., Amsterdam, The Netherlands, 1976, pp. 2-14.

[12] DAVIS, M.H.A., *A note on the Poisson disorder problem*, Banach Center Publications, 1 (19), pp. 65-72.

[13] DELLACHERIE, C. & P.A. MEYER, *Probabilités et Potentiel*, Hermann, Paris, 1975.

[14] FISHMAN, P.M. & D.L. SNYDER, *The statistical analysis of space-time point processes*, IEEE Trans. Information Theory, 22 (1976), pp. 257-274.

[15] FROST, P.A., *Examples of linear solutions to nonlinear estimation problems*, Proc. 5th annual Princeton Conf. on Information sciences and systems, 1971, pp. 20-24.

[16] GRANDELL, J., *Doubly stochastic Poisson processes*, Lecture Notes in Math., 529, Springer-Verlag, Berlin, 1976.

[17] HADJIEV, D.I., *On the filtering of semimartingales in case of observations of point processes*, Th. Probab. Appl., 24 (1979), pp. 169-178.

[18] JACOD, J., *Calcul Stochastique et problèmes de martingales*, Lecture Notes in Math., volume 714, Springer-Verlag, Berlin, 1979.

[19] KARP, S., E.L. O'NEILL & R.M. GAGLIARDI, *Communication Theory for the free-space Optical Channel*, Proc. IEEE, 58 (1970), pp. 1611-1626.

[20] KARP, S. & J.R. CLARK, *Photon counting: a problem in classical noise theory*, IEEE Trans. Info-th 16, 6, (Nov. 1970), pp. 672-680.

[21] LIPSTER, R.S. & A.N. SHIRYAYEV, *Statistics of Random Processes, I General Theory; II Applications*, Springer-Verlag, Berlin, 1977.

[22] MACCHI, O. & B.C. PICINBONO, *Estimation and Detection of Weak Optical Signals*, IEEE Trans. Information Theory, 18 (1972), pp. 562-573.

[23] MEYER, P.A., *Un cours sur les intégrales stochastiques*, in *"Séminaire de Probabilités X"*, Lecture Notes in Math., volume 511, Springer-Verlag, Berlin, 1975, pp. 245-400.

[24] NEVEU, J., *Processus Ponctuels*, in *"Ecole d'été de Probabilités de Saint-Flour VI - 1976"*, Lecture Notes in Mathematics, volume 598, Springer-Verlag, Berlin, 1977, pp. 250-447.

[25] RUDEMO, M., *State estimation for partially observed Markov chains*, J. Math. Anal. Appl., 44 (1973), pp. 581-611.

[26] RUDEMO, M., *Prediction and Smoothing for Partially Observed Markov Chains*, J. Math. Anal. Appl., 49 (1975), pp. 1-23.

[27] SEGALL, A., *A martingale approach to modelling, estimation and detection of jump processes*, Ph.D. thesis, Stanford Univ., Stanford, Cal., Aug. 1973.

[28] SEGALL, A., *Dynamic File Assignment in a Computer Network*, IEEE Trans. Automatic Control, 21 (1976), pp. 161-173.

[29] SEGALL, A. & T. KAILATH, *The modelling of randomly modulated jump processes,* IEEE Trans. Information Theory, 21 (1975), pp. 135-143.

[30] SEGALL, A., M.H.A. DAVIS & T. KAILATH, *Nonlinear filtering with counting observations,* IEEE Trans. Information Theory, 21 (1975), pp. 143-149.

[31] SNYDER, D.L., *Filtering and detection for doubly stochastic Poisson processes,* IEEE Trans. Information Theory, 18 (1972), pp. 91-102.

[32] SNYDER, D.L., *Smoothing for doubly stochastic Poisson processes,* IEEE-Trans. Information Theory, 18 (1972), pp. 558-562.

[33] SNYDER, D.L., *Statistical analysis of dynamic tracer data,* IEEE Trans. Biomedical Engineering, 20 (1973), pp. 11-20.

[34] SNYDER, D.L. & P.M. FISHMAN, *How to track a swarm of fireflies by observing their flashes,* IEEE-T-IT 21, 6 (1975), pp. 692-695.

[35] SNYDER, D.L., *Random point processes,* Wiley, N.Y., 1975.

[36] MAARSEVEEN, M.F.A.M. VAN, *Estimation of counting processes and application to traffic flow - A martingale approach,* Report, Onderafd. Toegepaste Wiskunde, Technische Hogeschool Twente, Enschede, The Netherlands, Sept. 1976.

[37] PUTTEN, C. VAN & J.H. VAN SCHUPPEN, *On stochastic dynamical systems,* Proc. Fourth Int. Symposium on Mathematical Theory of Networks and Systems, July 1979, Delft, The Netherlands, pp. 350-355.

[38] SCHUPPEN, J.H. VAN, *Estimation theory for continuous-time processes, a martingale approach,* Ph.D. thesis, University of California, Berkeley, Ca., Sept. 1973.

[39] SCHUPPEN, J.H. VAN, *Filtering, Prediction and Smoothing for Counting Process Observations, a Martingale Approach,* SIAM J. Appl. Math., 32 (1977), pp. 552-570.

[40] WAN, C.B. & M.H.A. DAVIS, *The General Point Process Disorder Problem,* IEEE Trans. Information Theory, 23 (1977), pp. 538-540.

[41] SNYDER, D.L. & P.M. FISHMAN, *How to track a swarm of fireflies by observing their flashes,* IEEE Trans. Information Theory, 21, (1975), pp. 692-695.

Smooth Representation of Systems with Differentiated Inputs

M. I. FREEDMAN AND JAN C. WILLEMS, MEMBER, IEEE

Abstract—Conditions are derived under which the nonlinear input–output system $\dot{y}=f(y,u,\dot{u})$ can be represented in the form $\dot{z}=g(z,u)$; $y=h(z,u)$. Various implications of the results to stochastic differential equations are considered.

I. Introduction

In this paper we will derive conditions under which the system described by the vector differential equation $\dot{x}=f(x,\dot{u})$, where u denotes the input, can be represented by a system of the form $\dot{z}=g(z,u)$; $x=h(z,u)$ which does not contain the derivative of the input. As we will see, necessary and sufficient conditions for this to be possible are that i): $f(\sigma_1,\sigma_2)$ is affine in σ_2, i.e., $f(x,\dot{u})=a(x)+b(x)\dot{u}$, and ii): the vector fields defined by the columns of b commute (if b_i denotes the ith column of b then this requires that $(\partial b_k/\partial x)b_l=(\partial b_l/\partial x)b_k$ for all k,l).

It is a well known basic result of linear "realization theory" (see, for example [1]) that for $P(s)=s^n+P_{n-1}s^{n-1}+\cdots+P_0$ and $Q(s)=Q_ns^n+\cdots+Q_0$ given matrix polynomials in s, the linear system $y^{(n)}+P_{n-1}\cdot y^{(n-1)}+\cdots+P_0y=Q_nu^{(n)}+Q_{n-1}u^{(n-1)}+\cdots+Q_0u$ (with u the input and y the output) can always be represented by a system of the type $\dot{x}=Ax+Bu$; $y=Cx+Du$. The problem treated in the present paper may be seen as a first step in obtaining an analogous theory for nonlinear systems of the type $y^{(n)}=f(y^{(n-1)},\cdots,y,u^{(n)},u^{(n-1)},\cdots,u)$. In fact, we treat the case $\dot{y}=f(y,u,\dot{u})$. It turns out that such a representation is only possible provided $f(y,u,\dot{u})=a(y,u)+b(y,u)\dot{u}$ and b satisfies $(\partial b_k/\partial u_l)+(\partial b_k/\partial y)b_l=(\partial b_l/\partial u_k)+(\partial b_l/\partial y)b_k$ for all k, l.

An interesting application of the problem treated in this paper is the sample pathwise interpretation of stochastic differential equations. Recently some interesting results in this direction have been obtained by Sussmann [2], [3] and our conditions are in fact identical to his. However, the approach taken by Sussmann is quite different: he is, in essence, looking for a smoothness result and we are looking for a specific representation. These issues are, of course, very much related but since our results are more specific they admit in the end stronger smoothness claims than those which have been obtained or, for that matter, considered in [2], [3]. Specifically, whereas Sussmann proves a certain map to be continuous in the C^0-topology, our results imply that that same map is also continuous in the $\mathcal{L}_p$-topologies for $1\leqslant p\leqslant\infty$. The resulting system thus admits input spaces which are closed under concatenation.

The final problem considered in this paper is to find conditions on a given Markov process $x(t)$ such that it can be represented as the output of a differential dynamical system which has, as its input, a Wiener process. We will show that this requires a particular Riemannian metric to be "flat" (i.e., of curvature zero) which contains as a special case all Markov process described by stochastic differential equations of the type $dx=a(x)dt+b(x)dw$ provided b satisfies the commutation conditions mentioned above.

Manuscript received February 21, 1977; revised August 5, 1977. Paper recommended by J. Davis, Chairman of the Stability, Nonlinear, and Distributed Systems Committee.

M. I. Freedman was on leave at the Mathematics Institute, University of Groningen, Groningen, The Netherlands. He is now with the Department of Mathematics, Boston University, Boston, MA 02115.

J. C. Willems is with the Mathematics Institute, University of Groningen, Groningen, The Netherlands.

II. Smooth Representation of Systems with Differentiated Inputs

The natural definition of equivalent systems which is used in system theory is basically the following: Consider the systems Σ_i: $\dot{x}_i=f_i(x_i,u)$; $y=g_i(x_i,u)$ with state spaces X_i, $i=1,2$. Then Σ_1 and Σ_2 are said to be *equivalent* if there exists a bijection $r:X_1\to X_2$ such that for all inputs there holds

i) $x_1(t)$ satisfies $\dot{x}_1(t)=f_1(x_1(t),u(t))$ if and only if $x_2(t)=r(x_1(t))$ satisfies $\dot{x}_2(t)=f_2(x_2(t),u(t))$,

ii) $g_1(x_1,u)=g_2(r(x_1),u)$.

This implies their input/output behavior to be the same. In the problem considered here we will, in addition to adding some smoothness requirements, specialize this definition to the systems:

$$\Sigma_1: \dot{x}=f(x,u,v),\quad \dot{u}=v;\quad y=x, \text{ and}$$

$$\Sigma_2: \dot{z}=g(z,u),\quad \dot{u}=v;\quad y=h(x,u).$$

Throughout the paper *smooth* will mean C^∞ and uniformly Lipschitz (thus C^∞ with bounded first derivatives). The formal definition of equivalent systems which is natural to the problem under consideration is thus the following.

Definition 1: The systems Σ_1: $\dot{x}=f(x,u,\dot{u})$ and Σ_2: $\dot{z}=g(z,u)$; $x=h(z,u)$, with f, g, and h smooth, and $x\in\mathbf{R}^n$, $u\in\mathbf{R}^m$, are said to be equivalent if there exists a smooth C^∞-diffeomorphism $r:\mathbf{R}^n\times\mathbf{R}^m\to\mathbf{R}^n\times\mathbf{R}^m$ with $r(x,u)=(r_1(x,u),u)$ such that $\Sigma_1(x_0,u_0)=\Sigma_2(r(x_0,u_0))$ for all (x_0,u_0)

$\in \mathbf{R}^n \times \mathbf{R}^m$. Here $\Sigma_1(x_0,u_0)$ denotes the map which takes[1] $v(\cdot)\in \mathcal{L}_{1oc}$ to $x(\cdot)$ via $\dot{x}=f(x,u,v)$, $\dot{u}=v$; $x(0)=x_0$, $u(0)=u_0$ and $\Sigma_2(z_0,u_0)$ is similarly defined by $\dot{z}=g(z,u)$, $\dot{u}=v$, $z(0)=z_0$, $u(0)=u_0$.

In the remainder of the paper we will only consider systems of the form $\dot{x}=f(x,\dot{u})$. It is easily verified that this constitutes no loss of generality since $\dot{x}=f(x,u,\dot{u})$ may always be written this form by augmenting x to (x,u).

In order to express the conditions of Theorem 1 succinctly we introduce now the following standard concept [4].

Definition 2: Let $r,s:\mathbf{R}^n\to\mathbf{R}^n$, $i=1,2$, be smooth. Then the vector fields $X_r=\Sigma_{i=1}^n r_i(x)\partial/\partial x_i$ and $X_s=\Sigma_{i=1}^n s_i(x)\partial/\partial x_i$ are said to *commute* if $[X_t,X_s]=0$. Here $[X_r,X_s](x)$ denotes the Lie bracket $\Sigma_{i=1}^n\Sigma_{j=1}^n\cdot(r_j(x)\partial s_i/\partial x_j(x)-s_j(x)\partial r_i/\partial x_j(x))\partial/\partial x_i$. The requirement is thus that $(\partial r/\partial x)s=(\partial s/\partial x)r$ which may be viewed simply as a condition on the first partials of r and s which admits a differential geometric interpretation.

The main result of the paper is the following.

Theorem 1: There exists a system $\dot{z}=g(z,u)$; $x=h(z,u)$ with g,h smooth which is equivalent to the given system $\dot{x}=f(x,\dot{u})$ with f smooth if and only if:

i) *$f(x,\dot{u})$ is affine in $\dot{u}$, i.e., there exist a,b such that $f(x,\dot{u})=a(x)+b(x)\dot{u}$;*
and

ii) *The m vector fields defined by the columns of b commute pairwise.*

Proof: The proof of this result is based on the following lemma which is of some interest in its own right:

Lemma 1: Consider the system of partial differential equations:$(\partial h/\partial u)(z,u)=b(h(z,u))$, with[2] $b:\mathbf{R}^n\to\mathcal{L}(\mathbf{R}^n,\mathbf{R}^m)$ smooth. Then there exists a smooth solution $h:\mathbf{R}^n\times\mathbf{R}^m\to\mathbf{R}^n$ such that $\hat{h}:(z,u)\to(h(z,u),u)$ is a smooth C^∞-diffeomorphism if and only if the vector fields $X_i=\Sigma_{j=1}^m b_{ji}(x)\partial/\partial x_j$, $i=1,2,\cdots,m$ commute pairwise.

Proof: see Appendix.

Proof of Theorem 1: ($\Rightarrow$): Let $\dot{z}=g(z,u)$; $x=h(z,u)$ be the equivalent representation. Thus $(\partial h/\partial z)(z,u)\dot{z}+(\partial h/\partial u)(z,u)\dot{u}=\dot{x}=f(h(z,u),\dot{u})$. Thus, f is affine in $\dot{u}$ and $(\partial h/\partial u)(z,u)=b(h(r,u))$. This implies by Lemma 1 that the vector fields defined by the columns of b commute pairwise.

($\Leftarrow$): Let h be the solution of $(\partial h/\partial u)(z,u)=b(h(z,u))$ as given by Lemma 1, and define g by $g(z,u)=((\partial h/\partial z)(z,u))^{-1}a(h(z,u))$. It is now easily verified that the map $r:(x,u)\to(z,u)$ with z such that $x=h(z,u)$ establishes the equivalence of $\dot{x}=a(x)+b(x)\dot{u}$ and $\dot{z}=g(z,u)$; $x=h(z,u)$. □

Some consequences of Theorem 1 are:

Corollary 1: Consider the system $\dot{x}=a(x)+b(x)\dot{u}$ with a,b smooth and u scalar valued (thus $m=1$). Then there exists an equivalent system $\dot{z}=g(z,u)$; $x=h(z,u)$.

[1] $\mathcal{L}_{1oc}$ denotes the functions integrable on finite intervals.
[2] $\mathcal{L}(\mathbf{R}^n,\mathbf{R}^m)$ denotes the real $(m\times n)$ matrices.

Proof: see Theorem 1: there is only one vector field. □

Corollary 2: Consider the system $\dot{x}=a(x)+b(x)\dot{u}$ where a,b satisfy the conditions of Theorem 1. Let F be defined on $C^1[0,T]$ as the map which takes $u(\cdot)$ to $x(\cdot)$ defined by this differential equation with fixed initial condition $x(0)=x$. Then F is Lipschitz continuous in the $C^0[0,T]$ topology and in the $\mathcal{L}_p[0,T]$ topology for $1\leqslant p\leqslant\infty$. Thus F admits continuous extensions to $C^0[0,T]$ and $\mathcal{L}_p[0,T]$.

Proof: This follows from the equivalent representation derived in Theorem 1 and from standard properties of the dependence of the solution of a differential equation on its right hand side. □

Corollary 2 constitutes a sharpening of Sussmann's results. In particular it shows that the system can always be extended by continuity to include other than C^1 inputs, also C^0 inputs (this is already known from [2,3]) and—important in applications—piecewise continuous inputs. In fact, the closure of the input space under concatenation is often an axiom on the very definition of a dynamical system and the extension to C^0 alone leaves one with this inconvenience, which is bypassed in our approach.

Examples:

1) The system $\dot{x}=a(x)+b\dot{u}$ with b constant obviously always satisfies the conditions of Theorem 1. In fact if there exists a smooth bijective map $m:\mathbf{R}^n\to\mathbf{R}^n$ such that $x_1=m(x_2)$ takes $\dot{x}_1=a_1(x_1)+b_1(x_1)\dot{u}$ into $\dot{x}_2=a_2(x_2)+b_2\dot{u}$ with b_2 constant then the first system will satisfy condition ii) of Theorem 1. The conditions on b_1 are however not equivalent to this condition although they are locally equivalent [4]. Presumably the conditions are in fact equivalent if $b_1(x)$ has constant rank.

It is easy to see which transformation will do the job for the system $\dot{x}=a(x)+b\dot{u}$. Indeed, by letting $z=x-bu$, we obtain $\dot{z}=a(z+bu)$; $x=z+bu$ which is the desired representation. The result of Theorem 1 is less obvious when $b(x)\neq$constant.

2) Assume that φ is differentiable and in $\mathcal{L}(0,\infty)$. Consider the differential equation $\dot{x}_N=a(x_N)+b(x_N)\dot{\varphi}(Nt)$ with $x_N(0)=x_0$. Then $\lim_{N\to\infty}x_N(t)=x(t)$ for all $t\geqslant 0$ with $x(t)$ the solution of $\dot{x}=a(x)$; $x(0)=x_0$, provided $b(x)$ satisfies the condition of Theorem 1. That the commutation of the vector fields is essential may be seen by considering instead the system $\dot{x}_1=x_2\dot{u}_1$, $\dot{x}_2=\dot{u}_2$, and examining the response with $u_1(t)=\sin Nt$ and $u_2(t)=\cos Nt$.

It is easily derived from Theorem 1 that $\dot{y}=f(y,u,\dot{u})$ may be represented m the form $\dot{z}=g(z,u)$; $y=h(z,u)$ provided $f(y,u,\dot{u})=a(y,u)+b(y,u)\dot{u}$ and b satisfies $(\partial b_k/\partial u_l)+(\partial b_k/\partial y)b_l=(\partial b_l/\partial u_k)+(\partial b_l/\partial y)b_k$ for all k,l.

III. Stochastic Differential Equations

Consider the Itô differential equation for $t\geqslant 0$: $dx=a(x)dt+b(x)dw$, $x(0)=x_0$ with $w(t)$ a Wiener process having $w(0)=0$ defined on a suitable probability space $\{\Omega,\mathcal{C},P\}$. The solution process $x(t)$ is defined as a limit-in-the-mean and, although this produces a nonanticipating map $F: w_\omega(\cdot) \to x_\omega(\cdot)$, this map is constructed by considering the ensemble properties of the Wiener process. The question thus arises whether it is not possible to give a specific representation of this map for example as a differential dynamical system which accepts as its inputs the realizations w_ω of the Wiener process and produces as its outputs the corresponding realizations x_ω of the Markov diffusion process generated by the Itô equation.

Whatever be the merit or the physical basis of obtaining such an interpretation of the random phenomenon modeled by the stochastic differential equation in question, it is of interest to identify the situations in which it is possible to do this. We show that Theorem 1 gives indeed one such result. It gives a sample pathwise interpretation of some stochastic differential equations as differential systems in which only the input is a function of the basic randomness expressed by its dependence on the chance variable ω.

Theorem 2: Assume that a *and* b *are smooth and that* b *satisfies condition ii) of Theorem* 1. *Then there exists a* C^∞*-diffeomorphism which is a unique solution of the partial differential equation* $(\partial h/\partial u)(z,u) = b(h(z,u))$ *with* $h(z,0) = z$. *Let*

$$g(z,u) = \left(\frac{\partial h}{\partial z}(z,u)\right)^{-1}\left(a - \frac{1}{2}\sum_{i=1}^{m}\frac{\partial b_i}{\partial x}b_i\right)(h(z,u)).$$

Then the solution of the deterministic differential system with stochastic inputs $\dot{z} = g(z,w)$, *with* $z(0)$ *such that* $x_0 = h(z(0),0)$ *satisfies* $h(z(\cdot),w(\cdot)) = x(\cdot)$ *almost surely.*

Proof: see Appendix.

The point of Theorem 2 is that, since $w(\cdot)$ is with probability one continuous, it gives a sample path interpretation on how to integrate the stochastic differential equation provided, however, the condition on b is satisfied. If this condition is not satisfied then, in view of the fact that the map $F: w(\cdot) \to x(\cdot)$ may then not be continuous in the C^0-topology, there is no hope of obtaining such a result.

Remarks:

1) If the stochastic integral involved in the definition of the solution process of the stochastic differential equation is interpreted in the sense of Stratonovich, then the result of Theorem 2 need only be modified by choosing $g(z,u) = ((\partial h/\partial z)(z,u))^{-1}a(h(z,u))$, i.e., without the notorious "correction term".

2) A great deal has been written on the question of the approximation of Itô differential equations, i.e., on the question in what sense the solution of the ordinary differential equation: $\dot{x}_n = a(x) + b(x)\dot{w}_n$. $x_n(0) = x_0$, approaches the solution of the Itô differential equation: $dx = (a - \Sigma_{i=1}^m(\partial b_i/\partial x)b_i(x))dt + b(x)dw$, $x(0) = x_0$) when $w_n \to w$ in some sense. In fact, the situation is pretty much settled in the case $m=1$ whereas in the case $m>1$ no clear picture on this approximation problem has yet emerged. Theorems 1 and 2 shed light on this problem. For instance, if condition ii) of Theorem 1 is satisfied (and hence whenever $m=1$), and if w_n is family of piecewise differentiable stochastic processes satisfying for some $1 \leqslant p \leqslant \infty$

$$\int_0^T \|w_n(t) - w(t)\|^p\,dt \underset{n\to\infty}{\to} 0$$

for almost all ω, then Theorem 2 shows that

$$\int_0^T \|x_n(t) - x(t)\|^p\,dt \underset{n\to\infty}{\to} 0$$

for almost all ω. All sorts of other approximations are easily obtained from Theorem 1. These conditions appear to be a bit simpler than what is available in the literature [5]. It would also appear that if b does not satisfy the commutation condition ii) of Theorem 1 that in view of the potential smoothness of the map $w \to x$ there is not much hope in obtaining strong approximation results (see also Remark 5).

3) Theorem 2 also shows how one could interpret the class of stochastic differential equations when b satisfies the commutation conditions of Theorem 1 but when w is not necessarily a Wiener process. The only requirement is that the input stochastic process $w(\cdot)$ is locally integrable for almost all ω. This includes, among many others, Poisson jump process.

4) Theorem 2 is close to giving a necessary condition as well. In order to do that, one needs to require that the map $h(\cdot,w)$ should be a bijection and that the measure induced by $x(t)$ should be absolutely continuous with respect to Lebesgue measure.

5) Let x be defined by the stochastic differential equation $dx = a(x)dt + b(x)dw$ but assume that b does not satisfy condition ii) of Theorem 1. Then $x_\omega(\cdot)$ is still defined as a nonanticipating map from $w(\cdot)$. Whether this map, viewed as a map from C^0 (C^0 is the support of Wiener measure) into C^0 is continuous in the C^0-topology is unclear. What we do know is that it is not the continuous extension of the map $\dot{x} = a(x) + b(x)\dot{u}$ (defined on C^1) to C^0, because this map is then not continuous in the C^0-topology. This remark is obviously relevant to the discussion of how and when to interpret Itô differential equations "physically".

IV. Representations of Diffusion Processes

In this section we will consider the problems of when a given Markov process can be represented (in a sense to be defined instantly) as the output of a deterministic differential system driven by a Wiener process. We will consider the following class of Markov processes.

Definition 3: Let $x(t)$ $(t \geqslant 0)$ be an n-dimensional Markov process with (stationary) transition function $\rho(y,t,B) = P(x(s+t) \in B \mid x(s) = y)$ for $t,s \geqslant 0$, $y \in \mathbf{R}^n$ and $B \subset \mathbf{R}^n$ Borel. Then $x(t)$ is called a *diffusion* process if:

i) for all $\epsilon > 0$ and $x \in \mathbf{R}^n$

$$\lim_{h\to 0^+}\frac{1}{h}\int_{\|x-z\|>\epsilon}P(x,h,dz)=0,$$

and

ii) there exists $a:\mathbf{R}^n\to\mathbf{R}^n$ and $\rho:\mathbf{R}^n\to\mathcal{L}(\mathbf{R}^n,\mathbf{R}^n)$ such that

$$\lim_{h\to 0^+}\frac{1}{h}\int_{\|x-z\|<\epsilon}(z-x)P(x,h,dz)=a(x)$$

$$\text{and }\lim_{h\to 0}\frac{1}{h}\int_{\|x-z\|<\epsilon}(z-x)(z-x)^TP(x,t,dz)=\rho(x).$$

The vector a is called the *local drift* and ρ is called the *local covariance*.

It is well known [6] that the solution of the Itô differential equation $dx=a(x)dt+b(x)dw$; $x(0)=x_0$, defines a diffusion with local drift a and local covariance bb^T. We will use the following notion of equivalence for diffusions.

Definition 4: Let x_i, $i=1,2$, be diffusion processes with local drifts a_i and local covariances ρ_i. Then they are said to be *equivalent* if $a_1=a_2$ and $\rho_1=\rho_2$.

Two equivalent diffusion processes have the same finite-dimensional distributions $P(x_i(t_1),x_i(t_2),\cdots,x_i(t_n))$. Thus, for many applications, we do not need to distinguish between them.

In view of the previous comments the following definition is a natural one.

Definition 5: Let $x(t)$ be a diffusion with local drift a and local covariance ρ. Assume that $x(0)=x_0$. Then it is said to be *represented* by the deterministic differential system with stochastic inputs: $\dot{z}=g(z,w)$; $\hat{x}=h(z,w)$, $z(0)=z_0$ if there exists a z_0 such that for $w(t)$ a Wiener process with $w(0)=0$ the process $\hat{x}(t)$ thus obtained is a diffusion process equivalent to $x(t)$.

The problem which we consider in this section is to derive conditions such that a diffusion process can be represented by a differential system with a Wiener process as input as defined in Definition 5. The following lemma shows that the problem is one of representing the diffusion process by a special type of stochastic differential equation (the problem of what diffusion processes may be represented by Itô equations is treated, e.g., *m* [7]; our problem requires a representation as in Definition 5).

Lemma 2: Let $x(t)$ be a diffusion process with local drift a and local covariance ρ and with $x(0)=x_0$. Assume that a is smooth. Then $x(t)$ may be represented by deterministic differential system as in Definition 5 with g,h smooth if there exists $b:\mathbf{R}^n\to\mathcal{L}(\mathbf{R}^n,\mathbf{R}^n)$, b smooth, such that:

i) $bb^T=\rho$,

and

ii) the n vector fields defined by the columns of b commute pairwise.

Proof: This follows immediately from Theorem 1. □

As shown in Lemma 2 the representation problem of Definition 5 reduces to a factorization problem on the local covariance matrix ρ. It is easily seen that such a factorization is unique up to right multiplication of $b(x)$ by an orthogonal matrix-valued function $s(x)$. We are thus looking for a matrix $s(x)$ such that $ss^T(x)=I$ and such that $b=\sqrt{\rho}\,s$ satisfies condition ii) of Lemma 2. Interestingly enough this question admits an answer in terms of some standard concepts from differential geometry.

Definition 6: Let $M:\mathbf{R}^n\to\mathcal{L}(\mathbf{R}^n,\mathbf{R}^n)$ be C^∞ and $M(x)=M^T(x)>0$. Then M is said to define a *C^∞-Riemannian metrie* on $\mathbf{R}^n$ (given by the 2-form $\langle dx,M(x)dx\rangle=\Sigma_{i,j=1}^n M_{ij}(x)dx_i dx_j$). A C^∞-Riemannian metric is said to be *flat* (or to have *zero curvature*) if there exists, locally around every point, a C^∞-diffeomorphism $r:\mathbf{R}^n\to\mathbf{R}^n$ such that the induced 2-form

$$\langle r_y(y)dy, M(r(y))r_y(y)dy\rangle \text{ equals } \langle dy,dy\rangle=\sum_{i=1}^n (dy_i)^2.$$

The main theorem of this section shows what diffusion processes can be represented as the output of a differential equation by a Wiener process:

Theorem 3: Let $x(t)$ be a diffusion process with local drift a and local covariance ρ and with $x(0)=x_0$. Assume that a is smooth, that ρ is C^∞, and that there exists $\epsilon>0$, such that $\rho(x)\geqslant\epsilon I$. Then a necessary condition for b as defined in Lemma 2 to exist is that the Riemannian metric defined by the 2-form $\langle dx,(\rho(x))^{-1}dx\rangle$ be flat.

Proof: see Appendix.

Remarks:

6) It is plausible that one could actually make a somewhat stronger claim, namely, that given the diffusion process $x(t)$ then there will exist, under the conditions of Theorem 3, a Wiener process $w(t)$ with $w(0)=0$, smooth functions g,h and a $z_0\in\mathbf{R}^n$ such that, for almost all $\omega, x(t)$ is given through $\dot{z}=g(z,w)$, $x=h(z,w)$.

7) The converse of Lemma 2 and Theorem 3 leads to the considerations indicated in Remark 4.

8) Even though Theorem 3 only ensures the existence of a local change of variables there are many situations in which it also works global, as the following example shows.

Example 3: Consider the diffusion process defined by

$$dx=a_1(x,y)\,dt+\sigma(x,y)\,dW_1 \qquad x(0)=x_0$$

$$dy=a_2(x,y)\,dt+\sigma(x,y)\,dW_2 \qquad y(0)=y_0$$

with a_1, a_2 and σ smooth. We have:

Theorem 4: Consider the diffusion process defined by the above differential equation. Then it satisfies the conditions of Lemma 2 if and only if $\log|\sigma|$ is a harmonic function. Consequently, if $\log|\sigma|$ is harmonic this diffusion process may be represented by a deterministic differential system as in Definition 5.

Proof: see Appendix.

V. Conclusions

A large class of Markov processes described by stochastic differential equations may be viewed as the output of a system described by an ordinary differential equation with a Wiener process as input. This has to do with the commutation of the vector fields defined by the columns of the input matrix of the noise terms. In such systems, which include many of those encountered in practice, a stochastic calculus is hardly necessary. The resulting processes are diffusion processes in which the inverse of the local covariance matrix defines a flat Riemannian metric. It appears in fact that this property gives exactly the class of diffusion processes which may be interpreted in this way.

Appendix

Proof of Lemma 1: ($\Rightarrow$): Let h be the solution of $(\partial h/\partial u)(r,u)=b(h(z,u))$. A simple calculation shows that this implies $\Sigma_{j=1}^{n} b_{jk}(h(z,u))(\partial b_{il}/\partial x_j)(h(z,u)) = (\partial/\partial u_k)(\partial h_i/\partial u_l)(z,u)$. Let X_i denote the vector field defined by the ith column of b. Then $\partial/\partial u_k \partial h_i/\partial u_l = \partial/\partial u_l \partial h_i/\partial u_k$ and the above equality yield $[X_k, X_l](h(z,u))=0$ for all z,u. Bit h is surjective. Hence, $[X_k,X_l]=0$ as claimed.

($\Leftarrow$): This part of the proof is "constructive". We will show that the required solution h may be constructed as follows: Consider the differential equation $\dot{r}=b(r)u$, with $r(0)=z$. Take the solution at $t=1$. Obviously this element depends on z and u. Denote this function by $h(z,u)$. We will now show that this is the function which we are looking for. To show that $(\partial h/\partial u)(z,u)=b(h(z,u))$ we will prove that the solution of $\dot{r}=b(r)\dot{f}(t)$; $r(0)=z$ at $t=1$ depends on the values taken on by $f(t)$ at $t=0$ and $t=1$ only. From there it follows, by considering f's of the type $f(t)=\mathrm{col}(u_1,\cdots,u_{i-1},u_i t,u_{i+1},\cdots,u_m)$, that $(\partial h/\partial u_i)(z,u) = b_i(h(z,u))$. To prove this path independence, take $f_0(t), f_1(t)$ with $f_0(0)=f_1(0)=0$ and $f_0(1)=f_1(1)=u$, and consider the linear homotopy $f_\alpha=(1-\alpha)f_0+\alpha f_1$. Let $r_\alpha(t)$ be defined by $\dot{r}_\alpha=b(r_\alpha)((1-\alpha)\dot{f}_0+\alpha\dot{f}_1)$, $r_\alpha(0)=z$. Let $s_\alpha= dr_\alpha/d\alpha$. We will have the path independence if we demonstrate that $s_\alpha(1)=0$. Clearly $\dot{s}_\alpha=\Sigma_{i=1}^{n}\partial b/\partial x_i(r_\alpha)(s_\alpha)_i[(1-\alpha)\dot{f}_0+\alpha\dot{f}_1]+b(r_\alpha)(\dot{f}_1-\dot{f}_0)$ with $s_\alpha(0)=0$. A straightforward but involved calculation using the commutation of the vector fields shows that $s_\alpha=b(r_\alpha)(f_1-f_0)$ is the unique solution of this differential equation. Hence, $s_\alpha(1)=0$ which was to be proven. The fact that $(z,u)\rightarrow(h(z,u),u)$ is smooth and has a smooth inverse follows from standard properties of the solutions of the differential equation (which defines h) on the parameters u. We omit these details.

Proof of Theorem 2: The existence of h and g follows from Lemma 1. Let x be defined by $\dot{z}=g(z,w)$, $x= h(z,w)$, and $z(0)$ such that $x_0=h(z(0),0)$. Then, by Itô's differentiation rule, $dx = (\partial h/\partial z)(z,w)\,dz + (\partial h/\partial z)(z,w)\,dw+\frac{1}{2}(\Sigma_{i=1}^{m}\partial^2 h/\partial^2 w_i)(z,w)\,dt$. By the definition of h this last term equals $\frac{1}{2}\Sigma_{i=1}^{m}(\partial b_i/\partial x\, b_i)(h(z,w))\,dt$ so that dx indeed satisfies $dx=a(x)\,dt+b(x)\,dw$ with $x(0)=x_0$, as required.

Proof of Theorem 3: Assume b exists. Then using the full rank condition in an argument very similar to the one used in the proof of Lemma 1 it follows that there exists a C^∞-diffeomorphism $x=r(z)$ satisfying $\partial r/\partial z=b(r)$ and $r(0)=0$. Thus $(\partial r/\partial z)^{-1}\, b(r)=I$ and consequently $(\partial r/\partial z)^{-1}\, b(r)\, b^T(r)\, ((\partial r/\partial z)^{-1})^T=I$. Thus the conditions of Theorem 3 assure the existence of a change of variables such that $(\partial r/\partial z)^{-1}\,\rho(r)\,((\partial r/\partial z)^{-1})^T=I$. It is easily seen from here that this r is the change of variable which shows flatness.

Proof of Theorem 4: ($\Rightarrow$): by computing the curvature using the formulas from differential geometry [4] it follows that $\log|\sigma|$ is harmonic. We omit the explicit calculation.

($\Leftarrow$): It is well-known that $\log|\sigma|$ harmonic implies the existence of a harmonic function $\theta(x,y)$ such that $\hat{\sigma}=\sigma e^{i\theta}$ is analytic as a function of $z=x+iy$.

Consequently, the process defined by:

$$dx=a_1(x,y)\,dt+\sigma(x,y)(\cos\theta(x,y)\,dw_1-\sin\theta(x,y)\,dw_2)$$

$$dy=a_2(x,y)\,dt+\sigma(x,y)(\sin\theta(x,y)\,dw_1+\cos\theta(x,y)\,dw_2)$$

defines an equivalent diffusion process. This process may be written in complex variable notation as

$$dz=a(z,\bar{z})\,dt+\hat{\sigma}(z)\,dw$$

where $\bar{z}=x-iy$, $a=a_1+ia_2$, and $w=w_1+iw_2$. The analytic transformation $z=r(s)$ with $dr/ds=\hat{\sigma}(r)$ transforms the above system into the equivalent one of the form:

$$ds=a^*(s,\bar{s})\,dt+dw;\quad z=r(s).$$

Letting $s-w=v$ we obtain the desired representation

$$\dot{v}=a^*(v+w,\bar{v}+\bar{w}),\quad z=r(v+w).$$

It is obvious how this can be transformed into a real variable equation.

Acknowledgment

The authors would like to thank F. Takens for some stimulating and very helpful discussions regarding some aspects of the problems treated in this paper.

References

[1] R. W. Brockett, *Finite Dimensional Linear Systems*. New York: Wiley, 1971.
[2] H. J. Sussmann, "An interpretation of stochastic differential equations as ordinary differential equations which depend on the sample point," *Bull. of the AMS*, vol. 83, no. 2, pp. 296–298, 1977.
[3] ——, "On generalized inputs and white noise," *Proceedings IEEE Conf. on Decision and Control*, 1976.
[4] M. Spivak, *A Comprehensive Introduction to Differential Geometry*. Publish or Perish, Inc., 1975.

[5] E. Wong, *Stochastic Processes in Information and Dynamical Systems*. New York: McGraw-Hill, 1971.
[6] W. H. Fleming, and R. W. Rishel, *Deterministic and Stochastic Optimal Control*. New York: Springer, 1975.
[7] D. Stroock, and S. R. S. Varadhan, "Diffusion processes with continuous coefficients," *Comm. Pure and Appl. Math*, vol. 22, pp. 345–400 and 479–530, 1969.

Boston, MA. His research interests lie in perturbation methods of control and systems theory.

He is a member of SIAM and the American Math Society.

P.S.

The purpose of this P.S. is to draw attention to some recent and some older references related to the issues discussed in this paper.

J. Lamperti, "A simple construction of certain diffusion processes", *J. Math. Kyoto* vol. 4, pp. 161-170, 1954.

D. Elliot, *Controllable Nonlinear Systems Driven by White Noise*, doctoral dissertation, U.C.L.A., 1959.

E.J. McShane, "On the use of stochastic differential equations in models of random processes", *Proc. Sixth Berkely Symp. Math. Statist. Prob.* vol. 3, pp. 263-294, 1972.

H.J. Sussmann, "On the gap between deterministic and stochastic ordinary differential equations", *Ann. Prob.* Vol. 6, pp. 19-41, 1978.

M. Fliess, "Intégrales itérées de K.T. Chen, bruit blanc gaussien et filtrage non linéaire", *C.R. Acad. Sc. Paris, Sér.* A- vol. 284, pp. 459-462, 1977.

M. Fliess, "Quelques remarques sur le "paradoxe" de Wong et Zakai; application a l'identification non linéaire", *Proceedings 17th. Coll. Signaux et Appl.* Nice, 1979.

H. Doss, "Liens entre équations différentielles stochastiques et ordinaires", *Ann. Inst. Henri Poincaré*, vol. 13, pp. 99-125, 1977.

H. Doss and E. Lenglart, "Sur l'existence, l'unicité et le comportement asymptotique des solutions d'équations différentielles stochastiques", *Ann. Inst. Henri Poincaré*, vol. 14, pp. 189-214, 1978.

P. Protter, "On the existence, uniqueness, convergence and explosions of solutions of systems of stochastic integral equations", *Ann. Prob.*, vol. 5, pp. 243-261, 1977.

P. Protter, "H^p stability of solutions of stochastic differential equations", *Z.f. Wahrscheinlichkeitsrechnung*, vol. 44, pp. 337-353, 1978.

H.J. Kushner, "Jump-diffusion approximations for ordinary differential equations with wide-band random right-hand sides", *SIAM J. Control and Opt.*, vol. 17, pp. 729-744, 1979.

S.I. Marcus, "Modeling and approximation of stochastic differential equations driven by semimartingales", submitted to *Stochastics*.

A.J. Krener, "A formal approach to stochastic integration and differential equations", to appear in *Stochastics*.

A.J. Krener and C. Lobry, "The complexity of stochastic differential equations", submitted to *Stochastics*.

J.M.C. Clark, "The design of robust approximations to the stochastic differential equations of nonlinear filtering", in *Communication Systems and Random Processes Theory*, ed. J.K. Skwirzynski, NATO Advanced Study Institute Series, Alphen aan den Rijn: Sijthoff and Noordhoff, 1978.

M.H.A. Davis, "Pathwise solutions and multiplicative functionals in nonlinear filtering", *Proc. 18-th IEEE CDC*, FC. Lauterdale, 1979.

M.H.A. Davis, "A pathwise solution of the equations of nonlinear filtering", submitted to *Teoria Veroyatnostei i ee Primeneniya*.

M.H.A. Davis, "On a multiplicative functional transformation arising in nonlinear filtering theory", submitted to *Z.f. Wahrscheinlichkeitsrechnung*.

E. Pardoux, "Backward and forward stochastic differential equations associated with a nonlinear filtering problem", to appear in the *IEEE Trans. Automatic Control*.

RICCATI EQUATIONS AND LAGRANGIAN MANIFOLDS

by

Clyde Martin*
Case Institute of Technology
Case Western Reserve University
Cleveland, Ohio 44106

This paper is essentially the content of a lecture presented at "Filterdag Rotterdam, 1980." The work reported here is partially joint work with Robert Hermann and a complete joint paper will appear elsewhere. The main purpose of this paper is to announce the following theorem.

<u>Theorem 1</u>. Let M be the class of linear symplectic vector field on the manifold of Lagrangian subspaces of dim n in $\mathbb{R}^{2n}$. The subclass M_o of Morse-Smale vectors fields is dense in M.

The matrix Riccati differential equation arises in many contexts, and has been widely studied in the control theory literature. Recall that it arises in control theory in the following way. Let a linear controllable system be given by

$$\dot{x} = Ax + Bu \qquad x(0) = x_0 \tag{1}$$

and a quadratic integral performance index be given by

$$J(u) = \int_0^T x'Qx + u'Ru. \qquad R > 0 \qquad Q \geq 0 \tag{2}$$

*Supported in part by NASA Grant #2384. The author wises to express his appreciation to Erasmus University for their hospitality while this paper was being written.

A standard calculus of variations argument reduces the minimization to the solution of the two point boundary value problem

$$\begin{pmatrix} \dot{x} \\ x \end{pmatrix} = \begin{pmatrix} A & -BR^{-1}B' \\ -Q & -A' \end{pmatrix} \begin{pmatrix} x \\ \lambda \end{pmatrix} \qquad \begin{matrix} x(0) = x_0 \\ \lambda(T) = 0 \end{matrix} \quad . \tag{3}$$

The two point boundary value problem can be reduced by the transformation $\lambda = P(t)x$ to the nonlinear differential equation

$$\begin{aligned} \dot{P}(t) &= A'P(t) + P(t)A - P(t)BR^{-1}B'P(t) + Q \\ P(T) &= 0 \ . \end{aligned} \tag{4}$$

It is well known that this equation describes in local coordinates the flow generated by a one parameter subgroup of the symplectic group acting on the Grassmannian manifold of n planes in $\mathbb{R}^{2n}$ [HE]. If we choose the standard symplectic two form on $\mathbb{R}^{2n}$,

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix} \tag{5}$$

the manifold $LG(\mathbb{R}^{2n})$ is invariant with respect to e^{Ht}. (It is, of course, actually a homogeneous space of the symplectic group). We can realize $LG(\mathbb{R}^{2n})$ as a compactification of the space of symmetric matrices which is justification for studying the Riccati equation of optimal control in this context.

We recall the following theorem [MA 1].

<u>Theorem 2</u>: The subspace X is an equilibrium point of the e^{Ht} iff $HX \subseteq X$.

This theorem reduces the study of equilibrium solutions to a study of invariant subspaces. The invariant subspace problem is greatly simplified by the following assumption:

- *The eigenvalues of* H *are distinct.*

Throughout this paper we will work under that assumption, even though there will be times when it can be ignored or relaxed. We will use the notation $e(\lambda)$ for an eigenvector associated with the eigenvalue λ. Every equilibrium point has a basis of eigenvectors. Recall that if λ is eigenvalue of H then so is $\bar{\lambda}$, $-\lambda$ and $-\bar{\lambda}$. We are interested in real Lagrangian equilibria. The following theorem describes this class.

<u>Theorem 3</u>: The subspace X spanned by $\{e(\lambda_i):i = 1,\ldots,n\}$ is real and Lagrangian iff

1) $e(\lambda_i) \in X \Rightarrow e(\bar{\lambda}_i) \in X$

2) $e(\lambda_i), e(\lambda_0) \in X \Rightarrow \lambda_i + \lambda_j \neq 0$.

Let X be a real Lagrangian equilibrium point for e^{Ht}. Construct Y be letting Y be spanned by the n-eigenvectors not in X. Note that $X \oplus Y = \mathbb{R}^{2n}$ and hence a canonical chart for $LG(\mathbb{R}^{2n})$ is defined and, with respect to this chart, H is block diagonal. Since X and Y are symplectic H is symplectic with respect to this decomposition and, hence, H has the form

$$\begin{pmatrix} H_1 & 0 \\ 0 & -H_1' \end{pmatrix} \tag{7}$$

Note that the Riccati equation so defined is given by

$$\dot{P}(t) = H_1' P(t) + P(t) H_1 \tag{8}$$

which is linear. The stable and unstable manifold of the equilibrium point X with respect to e^{Ht} are completely described by stable and unstable manifold of $\mathcal{O}$ with respect to the linear equation (8).

The eigenvalues of the linear operator in (8) are constructed by addition from the eigenvalues of H. But since X is invariant with respect to H, the eigenvalues are given by $\lambda_1,\ldots,\lambda_n$ where the λ_i satisfy the conditions of Theorem 3. The following theorem is thus obtained.

Theorem 4: The operator of equation 8 has no zero eigenvalues. The stable and unstable submanifolds of X with respect to e^{Ht} are transversal iff for all i and j, $\lambda_i + \lambda_j$ is not pure imaginary.

Thus the Riccati equations are well behaved with respect to equilibrium points. It was observed in [MA 2] that for $n=2$ Riccati equations can be periodic solutions. The construction presented did not generalize nor did the one communicated to the author by Professor Jan Willems generalize. The following construction is due in principle (but not in detail) to Robert Hermann.

Let H have a complex eigenvalue λ (not pure imaginary). A simple evaluation shows that there is a smallest positive to such $e^{\lambda t_o}$ is real. If $e^{\lambda t_o}$ is real, then so is $e^{\bar{\lambda} t_o}$, $e^{-\lambda t_o}$, and $e^{-\bar{\lambda} t_o}$. Let X be an equilibrium point in $LG(\mathbb{R}^{2n})$ that contains $e(\lambda)$ and $e(\bar{\lambda})$. Let $R(\lambda) = e(\lambda) + e(\bar{\lambda})$ and $I(\lambda) = i(e(\lambda)-e(\bar{\lambda}))$ and let X be spanned by $<e(\lambda_1),\ldots,e(\lambda_{n-2}),i(\lambda),R(\lambda)>$. Define $X(\theta,\gamma)$ to be the space spanned by $<e(\lambda_1),\ldots,e(\lambda_{n-2}),\cos\theta\, R(\lambda) + \sin\theta\, I(\lambda),\sin\theta\, I(-\lambda) + \cos\theta\, R(-\lambda)>$. Now note that

$$e^{Ht_o}X(\theta,\gamma) = X(\theta,\gamma)$$

for all θ and γ. Now $e^{Ht}X(\theta,\gamma) = X(\theta,\gamma)$ would imply that the space spanned by

$$\{\cos\theta\, R(\lambda) + \sin\theta\, I(\lambda),\ \sin\theta\, I(-\lambda) + \cos\theta\, R(-\lambda)\}$$

is invariant for all t which implies that there are two eigenvectors in the space. It is easy to see that the eigenvectors must be either $e(\lambda)$ and $e(\bar{\lambda})$ or $e(-\lambda)$ and $e(-\bar{\lambda})$--both impossibilities. Thus $X(\theta,\gamma)$ is not invariant. However, the space spanned by $R(\lambda)$ and $I(\lambda)$ is invariant as is the space spanned by $R(-\lambda)$ and $I(-\lambda)$ and so the set of all $X(\theta,\gamma)$ is invariant. The set of all such $X(\theta,\gamma)$ is obviously a simple two-dimensional torus in the Grassmannian manifold of n-planes in $\mathbb{R}^{2n}$. Let $T(\lambda)$ denote this torus. Note that $T(\lambda)$ is not in $LG(\mathbb{R}^{2n})$ because $X(\pi/2,\pi/2)$ is not in $LG(\mathbb{R}^{2n})$. The question is: Does $T(\lambda)$ intersect $LG(\mathbb{R}^{2n})$ and does the intersection contain an equilibrium point? To show that $LG(\mathbb{R}^{2n}) \cap T(\lambda)$ is not empty, it is really only necessary to show that the corresponding three dimensional Lagrangian manifold and the torus intersect in the four dimensional Grassmannian manifold. Either topological or computational methods work satisfactorily. The result is that $LG(\mathbb{R}^{2n}) \cap T(\lambda) \approx \mathbf{P}^1(\mathbb{R})$. The intersection contains no equilibria since $e^{Ht}X(\theta,\gamma) \neq X(\theta,\gamma)$ and, hence, the intersection must be a periodic orbit.

Now suppose there is a periodic orbit that contains the space X_0. Then there is a smallest t_0 such that $e^{Ht_o}X_0 = X_0$. Suppose e^{Ht_o} has distinct eigenvalues. Then X_0 is spanned by eigenvectors of e^{Ht_o} which are also eigenvectors of H and, hence, we have $e^{Ht}X_0 = X_0$ and the orbit has only one point. Thus we can assume that e^{Ht_o} has multiple eigenvalues.

We make an assumption now on H. We assume that if λ and γ are eigenvalues with equal real parts, then $\lambda = \gamma$ or $\lambda = \bar{\gamma}$. We've seen in Theorem 4 that this assumption is necessary for transversality and so is natural condition.

Since e^{Ht_o} has multiple eigenvalues there is a λ and γ such that $e^{\lambda t_o} = e^{\gamma t_o}$ and this implies that $\lambda = \gamma$ or $\lambda = \bar{\gamma}$ by our assumption. Assume for the moment that $e^{\tau t_o}$ real implies that τ is real of $\tau = \lambda$. If X_0 contains the space spanned by $R(\lambda)$ and $I(\lambda)$ then X_0 is an equilibrium point. Thus, there is a θ such that $\cos\theta\, R(\lambda) + \sin\theta\, I(\lambda)$ is in X_0. The other basis element for X_0 must come from the space spanned by $R(-\lambda)$ and $I(-\lambda)$ and hence there is a γ such that $\cos\theta\, R(-\lambda) + \sin\theta\, I(-\lambda)$ is in X_0. We have shown that $X_0 = X(\theta,\gamma)$ for some θ and γ and, hence, the periodic orbit is the same as the one previously constructed.

If there are distinct complex eigenvalues such that $e^{\tau t_o} = e^{\lambda t_o}$ and t_o is minimal for both then $\tau = \lambda$ or $\tau = \bar{\lambda}$ (under the assumption of distinct real parts for nonconjugates). If period of the orbit for τ is less than λ this can be constructed as before, but the intersection will be a line that wraps around the torus a number of times determined by the integer λ/τ. This situation is somewhat more detailed than the other so we exclude it for the purposes of this paper.

The above construction thus produces all periodic orbits under the three assumptions:

I. H has distinct eigenvalues

II. λ,γ eigenvalues and real λ = real $\gamma \Rightarrow \lambda = \gamma$ or $\lambda = \bar{\gamma}$

III. λ,γ eigenvalues and $e^{\lambda t_0}$ and $e^{\gamma t_0}$ both real implies $t_0 = 0$, $\lambda = \gamma$ or $\lambda = \bar{\lambda}$.

In order to determine the Morse-Smale systems we must calculate the Poincaré map of the periodic orbits. It suffices to consider the case of $n = 2$. In this the periodic orbit can be shown to have the form

$$\begin{pmatrix} -\tan\gamma & 1 & 0 & 0 \\ 0 & 0 & 1 & \tan\gamma \end{pmatrix}$$

where γ ranges from 0 to π. A suitable two dimensional transversal submanifold has coordinates

$$\begin{pmatrix} 0 & 1 & 0 & x \\ y & 0 & 1 & 0 \end{pmatrix} .$$

The Hamiltonian is block diagonal with respect to these coordinates and let it have the form

$$\begin{pmatrix} F & 0 \\ 0 & -F' \end{pmatrix} .$$

Then

$$e^{Ft} = e^{\lambda t}\begin{pmatrix} \cos\theta t & -\sin\theta t \\ \sin\theta t & \cos\theta t \end{pmatrix}$$

where λ is the real part of the eigenvalue and θ is the imaginary part. A routine, but long calculation, shows that the first return occurs at $t = \pi/\theta$. (This is well defined since periodic orbits only occur if $\theta \neq 0$.) Thus e^{Ht} reduces to

$$\begin{pmatrix} -e^{\lambda\pi/\theta}I & 0 \\ 0 & -e^{-\lambda\pi/\theta}I \end{pmatrix}$$

Thus the Poincaré map is linear with reciprocal eigenvalues unless $e^{\lambda\pi/\theta} = 1$. Again this is an algebraic condition and so Theorem 1 is proven.

REFERENCES

[HE] Hermann, Robert. Cartanian Geometry, Nonlinear Waves, and Control Theory, Part A. Brookline, MA: Math Sci Press, 1979.

[MA 1] Martin, Clyde. "Grassmann Manifolds and Global Properties of the Riccati Equation," Proceedings of the International Symposium on Operator Theory of Networks and Systems, Vol. 2, August 17-19, 1977, Lubbock, Texas, pp. 82-85.

[MA 2] Martin, Clyde. "Grassmannian Manifolds, Riccati Equations and Feedback Invariants of Linear Systems," Proceedings of the NATO Advanced Study Institute and AMS Summer Seminar in Applied Mathematics on Algebraic and Geometric Methods in Linear System Theory. To appear.

LIST OF REPORTS 1980

8000 "List of Reprints, nos 241-260, Abstracts of Reports Second Half 1979".

8001/O "A Stochastic Method for Global Optimization", by C.G.E. Boender, A.H.G. Rinnooy Kan, L. Stougie and G.T. Timmer.

8002/M "The General Linear Group of Polynomial Rings over Regular Rings", by A.C.F. Vorst.

8003/O "A Recursive Approach to the Implementation of Enumerative Methods", by J.K. Lenstra and A.H.G. Rinnooy Kan.

8004/E "Linearization and Estimation of the Add -Log Budget Allocation Model", by P.M.C. de Boer and J. van Daal.

8005/O "The Complexity of the Constrained Gradient Method for Linear Programming", by J. Telgen.

8006/S "On Functions with Small Differences", by J.L. Geluk and L. de Haan.

8007/O "Analytical Evaluation of Hierarchical Planning Systems", by M.A.H. Dempster, M.L. Fisher, L. Jansen, B.J. Lageweg, J.K. Lenstra and A.H.G. Rinnooy Kan.

8008/S "Looking at Multiway Tables (continued)", by A.P.J. Abrahamse and W.M. Lammerts van Bueren.

8009/O "An Introduction to Multiprocessor Scheduling", by J.K. Lenstra and A.H.G. Rinnooy Kan.

8010/M "On Families of Systems: Pointwise-Local-Global Isomorphism Problems, by M. Hazewinkel and A.M. Perdon.

8011/M "Proceedings Filter-Day Rotterdam 1980 (New Trends in Filtering and Identification of Stochastic Systems, 23 jan. 1980), by M. Hazewinkel (ed).