stichting
mathematisch
centrum

AFDELING MATHEMATISCHE BESLISKUNDE BW 139/81 MAART
(DEPARTMENT OF OPERATIONS RESEARCH)

R.W. BROCKETT

NONLINEAR SYSTEMS AND NONLINEAR ESTIMATION THEORY

Preprint

kruislaan 413 1098 SJ amsterdam

Printed at the Mathematical Centre, 413 Kruislaan, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

1980 Mathematics subject classification: 57R27, 60H15, 93B15, 93C10, 93E03, 93E11

Nonlinear systems and nonlinear estimation theory*)

by

R.W. Brockett**)

ABSTRACT

The estimation problem for a class of diffusion models is presented. Concepts of nonlinear system theory and Lie algebra theory are used to analyze the unnormalized conditional density equation. Examples of finite dimensional filter systems are treated. A nonexistence result is mentioned.

KEY WORDS & PHRASES: *Nonlinear systems, Lie algebra's, stochastic filtering, unnormalized conditional density equation, stochastic differential equations*

---

*) This report will be submitted for publication elsewhere.

**) Harvard University Division of Applied Sciences Cambridge, Massachusetts 02138 U.S.A.

This research was supported in part by the Army Research Office under Grant DAAG29-76-C-0139, the U.S. Office of Naval Research under the Joint Services Electronics Program Contract N00014-75-C-0648 and the National Science Foundation under Grant ENG-79-09459. The final manuscript was prepared while the author was visiting the Mathematics Center in Amsterdam using a set of lecture notes compiled and reworked by J. van Geldren, A. van der Schaft, and A. van Swieten. The hospitality of the Mathematics Center in Amsterdam, and, in particular, that of Jan van Schuppen is greatfully acknowledged.

I. NONLINEAR SYSTEMS AND LIE ALGEBRAS

1.1 Lie Algebras

Definition: A Lie algebra over a field $\mathcal{F}$ ($\mathbb{R}$ or $\mathbb{C}$ for these lectures) is a triple $(V,+,[\cdot,\cdot])$ where $(V,+)$ is a vector space over $\mathcal{F}$ and where $[\cdot,\cdot]$ is a bilinear map from $V \times V$ into V such that

i) $[v_1,v_2] = -[v_2,v_1]$ (antisymmetry)

ii) $[v_1,[v_2,v_3]]+[v_2,[v_3,v_1]]+[v_3,[v_1,v_2]] = 0$
(Jacobi identity)

Every Lie algebra we will encounter here can be thought of as a Lie algebra of linear operators with the bracket being $[v_1,v_2] = v_1v_2-v_2v_1$.

Example 1: Let $C^\infty(M)$ be the vector space of all infinitely differentiable functions defined on a differentiable manifold M. The vector space of all differential operators $A : C^\infty(M) \to C^\infty(M)$ becomes a Lie algebra if the Lie bracket $[A,B]$ is defined as $[A,B] = AB-BA$ where AB denotes the ordinary composition of the operators.

Example 2: A subclass of all differential operators on $C^\infty(M)$ is the set of all vector fields $\sum_i f_i(x) \frac{\partial}{\partial x_i}$. The Lie bracket of two vector fields turns out to be not a second order partial differential operator but another first order partial differential operator. So the set of all vector fields is a Lie algebra with addition and bracketing defined as in example 1.

Example 3: The vector space of all n×n matrices over a field with $[A,B] = AB-BA$ is a Lie algebra.

Example 4: A subclass of all vector fields on $\mathbb{R}^n$ is the set of all vector fields of the form $\sum_{i,j} a_{ij}x_j \frac{\partial}{\partial x_i}$. We call these linear vector fields. The Lie bracket of two linear vector fields is another linear vector field and the set of all linear vector fields is a Lie algebra under addition and bracketing as defined in example 2.

Example 5: A common way in which Lie algebras arise is the following. Given a smooth function $f : R^n \to R^n$, by identifying the tangent space of R^n with R^n we can think of f as defining a vector field on R^n; $\Sigma f_i \frac{\partial}{\partial x_i}$. We are, in this way, led to the following definition of the Lie bracket:

$[g,f] = \frac{\partial f}{\partial x} g - \frac{\partial g}{\partial x} f$ where $\frac{\partial f}{\partial x}$ and $\frac{\partial g}{\partial x}$ are the Jacobian matrices of f and g respectively.

1.2 The Exponential

We will use the symbol "exp" in several ways. For a linear equation $\dot{x} = Ax$ in a linear space the unique solution which satisfies $x(0) = x_o$ can be written as $x(t) = (\exp At)x_o$ where exp At is a linear map defined by $\exp At = 1 + At \frac{1}{2!} A^2t^2 + \ldots.$ For a nonlinear equation $\dot{x} = f(x)$ we cannot in general explicitly calculate a solution. However we can still denote points on the trajectory that passes through x_o at $t = 0$ by $(\exp ft)x_o$. For a third way of using the expression "exp", suppose we have m vector fields $f_1,\ldots f_m$ defined on some open subset of R^n. Then we define $\exp\{f_1,\ldots,f_m\}x_o$ to be the set of all points reachable by following all possible integral curves of the vector fields f_i one after another in a piecewise fashion.

1.3 Controllability

In order to see why we care about Lie brackets in control theory consider the differential equation:

$$\dot{x} = u_1(t)g_1(x)+u_2(t)g_2(x), \quad x(0) = x_o \tag{1}$$

If we apply the following control:

$$\begin{array}{lll} u_1 = 1, & u_2 = 0 & \text{from } 0 \text{ to } \varepsilon \text{ units of time} \\ u_1 = 0, & u_2 = 1 & \text{from } \varepsilon \text{ to } 2\varepsilon \text{ units of time} \\ u_1 = -1, & u_2 = 0 & \text{from } 2\varepsilon \text{ to } 3\varepsilon \text{ units of time} \\ u_1 = 0, & u_2 = -1 & \text{from } 3\varepsilon \text{ to } 4\varepsilon \text{ units of time} \end{array} \tag{2}$$

then we reach at time 4ε the point

$$x(4\varepsilon) = (\exp -g_2\varepsilon)(\exp -g_1\varepsilon)(\exp g_2\varepsilon)(\exp g_1\varepsilon)x_o$$

Using the expansion

$$\begin{aligned}(\exp \varepsilon g_i)x_o = x(\varepsilon) &= x_o+ \dot{x}(0) + \frac{1}{2}\varepsilon^2 x(0)+0(\varepsilon^3) \\ &= x_o+g_i(x_o) + \frac{1}{2}\varepsilon^2(\frac{\partial g_i}{\partial x} g_i)(x_o)+0(\varepsilon^3) \\ &\qquad i = 1,2\end{aligned}$$

one finds

$$x(4\varepsilon) = x_o+\varepsilon^2[g_1,g_2]+0(\varepsilon^3) \simeq (\exp \varepsilon^2[g_1,g_2])x_o$$

Definition: The real Lie algebra generated by a pair of vector fields g_1 and g_2 is the real linear span of the following set of vector fields

$\{g_1, g_2, [g_1, g_2], [g_1, g_2], [g_2, [g_1, g_2]], \ldots\}$

This association of a Lie algebra with a pair of vector fields is preserved under a change of coordinates in the manifold; if one brackets two vector fields and then changes coordinates the result is the same as if one were to change coordinates first and then compute the bracket in the new coordinate system. Given a control system

$$\dot{x} = f(x) + \sum_{i=1}^{m} u_i g_i(x) \tag{3}$$

then we can form the Lie algebra $\{f, g_1, \ldots, g_m\}_{LA}$ generated by the fector fields f, g, ..., g_m. It is called the <u>controllability Lie algebra</u>. A theorem of Frobenious states that under a mild hypothesis there exists for each x_o in the manifold X a manifold $\exp\{f, g_1, \ldots, g_m\}_{LA} x_o$. It contains the reachable set for (3) syarting at x_o but unless special assumptions are made it does not equal the reachable set.

<u>Remark 1</u>: The circle of ideas under discussion here stems in part from Hermann [1] who, to my knowledge, was the first person to point out the connection between Lie algebras and controllability. After some work in the 60's by Hermes and Haynes. These ideas were developed extensively in the period 1970-1973 by Krener, Lobry, Sussmann, and other people [2-6].

We are also interested, for reasons that have to do with the assymetry of <u>f</u> as compared with the g_i, in an algebra which we shall write $\bar{\mathscr{L}}$.

<u>Definition</u>: $\bar{\mathscr{L}}$ is the smallest subalgebra which contains $\{g_i\}_{i=1}^{m}$ and is closed under bracketing with f.

<u>Remark 2</u>: I am using here $\bar{\mathscr{L}}$ for something which is often written $\mathscr{L}_o$ but the reason is that $\mathscr{L}_o$ is used in other ways in a stochastic setting.

<u>Example</u>:

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = -x_1 + u$$
$$\dot{x}_3 = 1$$

We see that x_3 is just the time. This system is controllable in the sense that as time goes on we can reach any point in the half-space (x_1, x_2, x_3); $x_3 > 0$. Unfortunately if we plot x_3 versus time we obtain a straight line so we cannot hit a desired point at an arbitrary point in time.

Let's compute $\mathscr{L}$ and $\overline{\mathscr{L}}$. We have $F = x_2 \frac{\partial}{\partial x_1} - x_1 \frac{\partial}{\partial x_2} + \frac{\partial}{\partial x_3}$ and $G = \frac{\partial}{\partial x_2}$ respectively. Now $\mathscr{L}$ contains F, G and $[F,G] = -\frac{\partial}{\partial x_1}$ and $[F,[F,G]] = -\frac{\partial}{\partial x_2}$. Note that $[G,[F,G]] = 0$. Thus $\mathscr{L}$ spans the tangent space of $\mathbb{R}^3$ and $\overline{\mathscr{L}}$, the two-dimensional subalgebra spanned by $\frac{\partial}{\partial x_1}$ and $\frac{\partial}{\partial x_2}$, does not. This is a manifestation of the following phenomenon. There are some systems for which one can reach every point but the time at which a certain point is reached may not be adjustable. If this is true (as in the example) then $\overline{\mathscr{L}}$ is a proper subalgebra of $\mathscr{L}$. If one wants what is sometimes called "exact time controllability" it is appropriate to focus attention on $\overline{\mathscr{L}}$. To fix ideas let's consider systems for which the control enters linearly as, for example, a deterministic version of the conditional density equation.

Example 1

$$\dot{x} = Ax + \sum_i u_i B_i x \qquad x \in R^n \tag{4}$$

Here $f = Ax$ and $g_i = B_i x$ so that $[f,g_i] = -[Ax,B_i x] = -(AB_i - B_i A)x$. Each vector field which occurs in $\mathscr{L}$ is of the form Mx for some matrix M. Because the set of all n by n matrices is an n^2-dimensional vector space we have $\dim \mathscr{L} \leqslant n^2$.

Example 2

$$\dot{x} = ux+vx^3; \qquad x \in R^1 \tag{5}$$

Here $g_1(x) = x$, $g_2(x) = x^3$, so that $[g_1,g_2] = -[x,x^3] = 2x^3 = 2g_2$. In this case the Lie algebra is two-dimensional; it is the real linear span of x and x^3.

Example 3

$$\dot{x} = u+vx^3 \qquad x \in R^1 \tag{6}$$

Now $g_1(x) = 1$ and $g_2(x) = x^3$

$$[1,x^3] = 3x^2 \qquad [x^2,x^3] = x^4 \quad \text{and so on.}$$

We generate an infinite-dimensional Lie algebra on a 1-dimensional manifold (the real line).

Example 4: Let B_n be the set of all nxn intensity matrices. An intensity matrix is defined to be a square matrix $A = (a_{ij})$ for which $\sum_{i=1}^{n} a_{ij} = 0$ for each j and $a_{ij} \geqslant 0$ for all $i \neq j$. Intensity matrices play a role as generators of finite state stochastic processes. B_n is not a vector space because the difference of two intensity matrices need not be an intensity matrix. In order to form the appropriate Lie algebra form the linear closure of B_n, that is the smallest linear space that contains B_n, and then take the Lie algebra generated by that linear span. This Lie algebra turns out to be isomorphic to the Lie algebra of the group of affine transformations on an (n-1)-dimensional vector space.

1.4 Observability

A second Lie algebra associated with a control problem arises in connection with questions of observability (or indistinguishability, etc.). Consider the control problem

$$\dot{x} = f(x) + \sum_i u_i g_i(x)$$

where x takes on values in a differentiable manifold X and suppose we observe $y = h(x(t))$ where y takes values in a differentiable manifold Y and h is a mapping from X into Y. We want to deduce information about x from the observation of y. Assuming enough smoothness we can differentiate y. If, for example, $u(\cdot) = 0$, then

$$\begin{aligned} \dot{y} &= \sum_{i=1}^{n} \frac{\partial h}{\partial x_i} f_i = h_1(x) \\ \ddot{y} &= \sum_{i,j}^{n} \frac{\partial}{\partial x_j}\left(\Sigma \frac{\partial h}{\partial x_i} f_i\right) f_j = \sum_j^n \frac{\partial h_1}{\partial x_j} f_j = h_2(x) \end{aligned} \tag{7}$$

etc. We may think of this in the following way. The vector $(h, h_1, h_2, \ldots, h_n)$ maps X into $T^{n-1}Y$, the (n-1)st jet bundle over Y.

Are two or more initial states in the manifold X compatible with these observations? If

$$\det \begin{bmatrix} \frac{\partial h}{\partial x_1} & \frac{\partial h}{\partial x_2} & \frac{\partial h}{\partial x_n} \\ \cdots & \cdots & \cdots \\ \frac{\partial h_{n-1}}{\partial x_1} & \frac{\partial h_{n-1}}{\partial x_2} & \frac{\partial h_{n-1}}{\partial x_n} \end{bmatrix} \neq 0 ,$$

then in view of the inverse function theorem we can assert that in some neighborhood of the true initial state there are no other points which give rise to the same response. However that does not preclude the possibility that there are some other points some distance away in the manifold X which give rise to exactly the same y's.

Example: Consider Newton's law for rotational motion

$$\ddot{\theta} = u \qquad y = \sin\theta \tag{8}$$

where the moment of inertia is one and u is the applied torque. A natural state space for this system is the cylinder $S^1 \times R^1$. With this state space representation the system is observable. On the other hand the equations

$$\ddot{x} = u \qquad y = \sin x$$

often appear in the literature with the interpretation that x is a real number. This corresponds to cutting the cylinder along the side, flattening it out, laying it down on R and covering R with a countable number of copies of it. As far as observability is concerned the countable number of points $x, x\pm2\pi, x\pm4\pi\ldots$ can be regarded as the same. In terms of this model for the state space the system is only locally observable and this local observability is not enough to determine the initial state uniquely.

We want to now code the information about observability in a different way, one that is compatible with the way we will be looking at the conditional density equation. The vector field associated with the free motion is $F = \sum_i f_i \frac{\partial}{\partial x_i}$. The formal adjoint of this linear operator is the operator $F^* = -\sum_i \frac{\partial}{\partial x_i} f_i$, not a differential operator. F^* can be thought of as operating on the space $C^\infty(X)$ of all infinitely differentiable functions defined on the manifold X.

A function $h(\cdot) \in C^\infty(X)$ also defines a linear operator on $C^\infty(X)$, namely "multiplication by $h(\cdot)$". This maps $\phi \in C^\infty(X)$. Thus we can form the Lie algebra of operators generated by the two operators $-\sum_i \frac{\partial}{\partial x_i} f_i$ and $h(\cdot)$. We propose to call this the <u>little observability algebra</u>. This algebra contains the commutator $[h(\cdot), -\sum_i \frac{\partial}{\partial x_i} f_i] = -\sum_i f_i \frac{\partial h}{\partial x_i} = -h_1$. It also contains $h_2, h_3, \ldots$ A sufficient condition for local observability around the free motion is that the little observability algebra contains n functions whose Jacobian is nonsingular.

We also have the <u>big observability algebra</u> associated with the controlled motion $(u \neq 0)$. It is defined as $\{\sum_i \frac{\partial}{\partial x_i} f_i, \ \sum_i \frac{\partial}{\partial x_i} g_i, \ \ldots h(x)\}_{LA}$.

To build some intuition we apply the foregoing ideas to a linear system:

$$\dot{x} = Ax + bu \qquad y = cx \tag{9}$$

The controllability algebra is $\mathscr{L} = \{Ax,b\}_{LA}$ = the linear span of $\{b,Ab,\ldots A^{n-1}b,Ax\}$. In $\bar{\mathscr{L}}$ the Ax term is missing but the other terms remain. The little observability algebra is

$$\{\sum_{i,j} \frac{\partial}{\partial x_j} a_{ij}x_j, cx\}_{LA} = \text{linear span of } \{cx, cAx, \ldots cA^{n-1}x, \sum_{i,j} \frac{\partial}{\partial x_i} a_{ij}x_j\}$$

Consider now a second class of examples which is a little closer to the conditional density equation. By <u>a real Lie group</u> we understand a real Hausdorff manifold $\mathscr{G}$ with a multiplication $\cdot : \mathscr{G} \times \mathscr{G} \to \mathscr{G}$ under which $\mathscr{G}$ is a group and such that $\cdot : \mathscr{G} \times \mathscr{G} \to \mathscr{G}$ is continuous (and hence analytic by virtue of the solution of Hilberts 5th problem). The most pedestrian type of Lie group is a matrix Lie group. These include the general linear group $GL(n,r)$ the orthogonal group with det = 1, $SO(n)$, etc. Lie groups are assumed to be finite-dimensional. This is in con trast with Lie algebras which can be finite or infinite dimensional.

A class of systems somewhat analogous to linear systems is the class defined by

$$\begin{aligned} \dot{X} &= AX + \Sigma u_i B_i X \\ y &= h(X) \end{aligned} \tag{10}$$

where X takes values in a matrix Lie group and AX and B_iX are vector fields on that Lie group. In this case the controllability Lie algebra is necessarily finite dimensional and is given by

$$\{AX, B_iX\}_{LA} = \{AX, B_iX, [A,B_i]X, \ldots\}$$

An example of a system defined on a Lie group is that of controlled rigid body motion. The state space is the tangent bundle to $SO(3)$; a manifold which admits the structure of a 6 dimensional Lie group.

1.5 System Isomorphism

Suppose we have two different linear control systems with zero initial conditions

$$\begin{aligned} \dot{x} &= Ax+Bu; \quad y = Cx; \quad x(0) = 0; \quad x(t) \in R^n \\ \dot{z} &= Fz+Gu; \quad y = Hz; \quad z(0) = 0; \quad z(t) \in R^{n'} \end{aligned} \tag{11}$$

Assume that these systems, as models for a real system with input u and output y, possess exactly the same input-output behavior. We then have the following result.

Theorem: If (A,B,C) is a controllable and observable then there exists a linear map $P : R^n \to R^n$ such that P preserves trajectories i.e. $Pz(t) = x(t)$ for every input $u(\cdot)$.

A theorem with the same hypothesis and conclusion holds for bilinear systems:

$$\begin{aligned} \dot{x} &= Ax+uBx \qquad y = Cx \qquad x(0) = 0 \\ \dot{z} &= Fz+uGz \qquad y = Hz \qquad z(0) = 0 \end{aligned} \tag{12}$$

The following result is very much in spirit of results in the literature [6, 7, 8] but probably does not appear exactly this way. Consider the control systems

$$\begin{aligned} \dot{x} &= f(x)+ug(x); \quad y = h(x); \quad x(0) = x_o; \quad x \in X \\ \dot{z} &= a(z)+ub(z); \quad y = c(z); \quad z(0) = z_o; \quad z \in Z \end{aligned}$$

where X and Z are analytic manifolds, f and g are analytic vector fields on X, and a and b are analytic vector fields on Z. We assume that f, g, a and b are complete; i.e. the integral curves of the vector fields can be continued for all time from $-\infty$ to $+\infty$. Finally, assume that the system on X is controllable and observable. Observable in this case means that we can distinguish between any two points on X provided we use the right input. Now we have the following theorem.

Theorem: Under the above hypothesis, if both systems generate the same input-output map then there exists an analytic mapping $\Phi : Z \to X$ that preserves trajectories i.e. $\phi(z(t)) = x(t)$ for every input $u(\cdot)$.

Example: Consider Newton's law for rotational motion. Let $X = S^1 \times R^1$ and $Z = R^2$ exactly as was done above.

We have a mapping $\phi : \begin{cases} x+2\pi k \ \ (k \text{ integer}) \longmapsto \theta = x \\ \dot{x} \longmapsto \dot{\theta} = \dot{x} \end{cases}$

The mapping $\Phi : Z \to X$ induces a mapping Φ_* from the tangent bundle of Z into the tangent bundle of X. Because of the preservation of trajectories under Φ we must have $\Phi_*: a \longmapsto f$ and $\Phi_*: b \longmapsto g$. Now we assert that Φ_* extends to a homomorphism of the Lie algebra generated by a and b into the Lie algebra generated by f and g: $\Phi_* : \{a,b\}_{LA} \to \{f,g\}_{LA}$. In general Φ_* is not an isomorphism, it may have a nontrivial kernel because $\dim\{f,g\}_{LA}$ may be smaller than $\dim\{a,b\}_{LA}$.

Without going into detail, we mention one important construction. Suppose we start the systems on X and Z at x_o and z_o, respectively, and suppose we apply some input in both systems so that we arrive at x_1 and z_1 respectively. At that point we begin experimenting with the input. For instance if we have systems with controls u_1 and u_2 and choose them as indicated by equation (2) then we take both systems around in a loop. From this kind of construction we get a relationship between the Lie algebras for the X-system and that of the Z-system. In more detail; if we have two systems without a drift term

$$\begin{aligned} \dot{x} &= vf + ug \\ \dot{z} &= va + ub \end{aligned} \tag{13}$$

then, choosing v and u as before (2), we get

$$\Phi((\exp -a\varepsilon)(\exp -b\varepsilon)(\exp a\varepsilon)(\exp b\varepsilon)z_o)$$
$$= (\exp -f\varepsilon)(\exp -g\varepsilon)(\exp f\varepsilon)(\exp g\varepsilon)x_o$$

$$\Phi(\exp[a,b]\varepsilon^2 z_o) = \exp[f,g]\varepsilon^2 x_o$$

1.6 The Wei-Norman Equations

There exist representations of solutions of differential equations that will let us establish a connection between the unnormalized conditional density equation and a certain Lie algebra. This material is most explicit in Wei-Norman [9] an earlier paper by Chen [10] covers similar ground and the basic ideas could probably be traced back at least to Lie and Cartan.

To begin with, consider the finite dimensional linear equation:

$$\dot{x} = (uA+vB)x \tag{14}$$

with u and v functions from R^1 to R^1 and A and B constant n n matrices. Naively one might expect to find that the fundamental solution $\Phi(\cdot)$ is

$$\begin{aligned} \Phi(t) &= e^{\int_0^t u(\sigma)d\sigma)A+(\int_0^t v(\sigma)d\sigma)B} \\ &= Z+(\int_0^t u(\sigma)d\sigma)A+(\int_0^t v(\sigma)d\sigma)B+ \\ &\qquad \frac{1}{2}((\int_0^t u(\sigma)d\sigma)A+(\int_0^t v(\sigma)d\sigma)B)^2+\ldots \end{aligned} \tag{15}$$

But this would imply that

$$\begin{aligned} \dot{\Phi}(t) &= (uA+vB) + \frac{1}{2}(uA+vB) \\ &\qquad \int_0^t (uA+vB)d + \frac{1}{2}(\int_0^t (uA+vB)d\ )(uA+vB)+\ldots \end{aligned} \tag{16}$$

It is not possible to factor out (uA+uB) from this expression because in general A and B do not commute and thus the above expression for Φ does not work. However, for n×n matrices A and B we can use the identity

$$\begin{aligned} e^{-A}Be^{A} &= (1+A+\frac{1}{2!}A^2+\ldots)B(1-A+\frac{1}{2}A^2\ \ldots) \\ &= B+AB-BA+\frac{1}{2}(A^2B-2ABA+BA^2) \\ &= B+[A,B]+\frac{1}{2}[A,[A,B]]+\ldots \end{aligned} \tag{17}$$

This is sometimes called the Baker-Campbell-Hausdorff formula. Introduce the notation

$$\mathrm{ad}_A^k B = \underbrace{[A,[A,[A.\ldots[A,B]]\ldots]}_{k\ \text{times}}; \qquad k \geqslant 1$$

$$\mathrm{ad}_B^0 B = B \tag{18}$$

($\mathrm{ad}_A^k B$ is an operator taking a pair of matrices into a single matrix) and define

$$\begin{aligned} \exp\ \mathrm{ad}_A B &= \mathrm{ad}_A^0 B + \mathrm{ad}_A^1 B + \frac{1}{2!}\mathrm{ad}_B^2 B + \ldots \\ &= B + [A,B] + \frac{1}{2!}[A,[A,B]] + \ldots \end{aligned} \tag{19}$$

and write

$$e^{A}Be^{-A} = \exp\ \mathrm{ad}_A B \tag{20}$$

Wei and Norman investigated the differential equation

$$\dot{x} = (\Sigma u_i A_i)x \tag{21}$$

by looking for a solution $\Phi(t)x_o$ which can be represented as a product of exponentials

$$\Phi(t)x_o = e^{g_1A_1}e^{g_2A_2}\ldots e^{g_mA_m}x_o \tag{22}$$

in which $g_1\ldots g_m$ are real valued functions of time. Differentiating gives:

$$\frac{d}{dt}(e^{g_1A_1}\ldots e^{g_mA_m}) = \dot{g}_1A_1e^{g_1A_1}\ldots e^{g_mA_m}+e^{g_1A_1}\dot{g}_2A_2e^{g_2A_2}\ldots e^{g_mA_m} + \ldots e^{g_1A_1}\ldots e^{g_{m-1}A_{m-1}}\dot{g}_mA_me^{g_mA_m}$$

Inserting exponentials and their inverses we can transform this into an expression in which all terms have a common factor $e^{g_1A_1}\ldots e^{g_mA_m}$ on the right. Apart from this common factor, we get the expression:

$$\dot{g}_1A_1+e^{g_1A_1}\dot{g}_2A_2e^{-g_1A_1}+\ldots \tag{23}$$

Applying the Baker-Campbell-Hausdorff-formula (23) can be written as:

$$\dot{g}_1A_1+\dot{g}_2(A_2+g_1[A_1,A_2]+\frac{1}{2}g_1^2[A_1,A_2]]+\ldots)+\ldots$$

$$\ldots+\dot{g}_m \text{ (an expression containing matrices in } \{A_i\}_{LA} \text{ and } g_i\text{'s)} \quad (24)$$

Suppose now that the A_i's in (22) are not the matrices A_i from the differential equation (21) but suppose that we first construct the Lie algebra generated by the A_i's in the differential equation (21), then pick a basis from that Lie algebra, and use that basis in (22). Under these circumstances the coefficients of the $\dot{g}_i$ in (24) are linear combinations of the A_i's.

Assuming that the A_i are a basis of the Lie algebra associated with (21) we have

$$[A_i,A_j] = \sum_k \gamma_{ijk}A_k \quad (25)$$

with certain coefficients γ_{ijk}, the so-called structure constants of the Lie algebra. In order that $\Phi(t)x_o$ satisfies the differential equation (21) we must have

$$\dot{g}_1A_1+\dot{g}_2(A_2+g_1[A_1,A_2]+\frac{1}{2}g_1^2[A_1,[A_1,A_2 22+\ldots)+\ldots$$

$$+\ldots\dot{g}_n(\ldots) = u_1A_1+\ldots u_mA_m \quad (26)$$

and because the A_i are independent as vectors in $R^{n\times n}$ we get, on equating coefficients a set of equations of the form

$$\begin{aligned}
\dot{g}_1 &= f_1(g_1,\ldots g_m, \dot{g}_2\ldots\dot{g}_m) + u_1\\
\dot{g}_2 &= f_2(g_1,\ldots g_m, \dot{g}_2\ldots\dot{g}_m) + u_2\\
&\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\;\cdot\\
\dot{g}_m &= f_m(g_1,\ldots g_m, \dot{g}_2\ldots\dot{g}_m) + u_m
\end{aligned} \quad (27)$$

We will refer to these as the Wei-Norman equations associated with the differential equation (21) and a particular ordering of the exponential factors. Because $\Phi(0) = I$ we have initial conditions $g_1(0) = g_2(0) = \ldots = g_m(0)$. An analysis shows that the Wei-Norman equations can always be solved on some interval $|t| \leqslant \varepsilon$ however in most cases the solution cannot be continued for all time. A significant point is that the functions $f_1,\ldots,f_m$ only depend on the structure constants γ_{ijk}. That is, regardless of the representation of the Lie algebra we get the same Wei-Norman equations. We have here a situation such that by solving one set of nonlinear differential equations we simultaneously solve a whole family of linear evolution equations.

Example

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} a(t) & c(t) \\ 0 & b(t) \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

$$= \left[a(t) \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b(t) \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + c(t) \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \right] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{28}$$

Let $\mathscr{L}$ be the Lie algebra generated by

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

We choose an ordered basis for $\mathscr{L}$:

$$A_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \text{ and } A_3 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The differential equations can now be written

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = (A_1\eta_1 + A_2\eta_2 + A_3\eta_3) \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{29}$$

with $\eta_1 = a-b$, $\eta_2 = c$, $\eta_3 = b$. If we look for a fundamental solution having the form

$$\phi(t) = e^{A_1g_1} e^{A_2g_2} e^{Ig_3}$$

Then

$$\dot{\Phi} = (A_1\dot{g}_1 + e^{A_1g_1} A_2\dot{g}_2 e^{-A_1g_1} + \dot{g}_3 I) e^{A_1g_1} e^{A_2g_2} e^{Ig_3}$$

Because we have a basis with $A_3 = I$ the expression for $\dot{\Phi}$ contains the term $\dot{g}_3\Phi$ instead of a term

$$e^{A_1g_1} e^{A_2g_2} \dot{g}_3 A_3 e^{-g_2A_1}$$

that is

$$\dot{\Phi} = (A_1\dot{g}_1 + \dot{g}_2(A_2 + g_1[A_1,A_2] + \tfrac{1}{2}g_1^2[A_1,[A_1,A_2]] + \ldots) + \dot{g}_3 I) e^{A_1g_1} e^{A_2g_2} e^{Ig_3}$$

Now $[A_1,A_2] = A_2$ so that

$$\dot{} = (A_1\dot{g}_1 + \dot{g}_2(A_2 + g_1A_2 + \tfrac{1}{2} g_1^2 A_2 + \ldots) + \dot{g}_3 I) e^{g_1A_1} e^{g_2A_2} e^{g_3I}$$

$$= (A_1\dot{g}_1 + \dot{g}_2 e^{g_1} A_2 + \dot{g}_3 I) e^{g_1A} e^{g_2A_2} e^{g_3I}$$

So
$$A_1\dot{g}_1+\dot{g}_2e^{g_1}A_2+\dot{g}_3I = \eta_1A+\eta_2A_2+\eta_3I \tag{31}$$

The Wei-Norman equations become

$$\begin{aligned}\dot{g}_1 &= \eta_1 = a-b\\ \dot{g}_2 &= e^{-g_1}\eta_2 = ce^{-g_1}\ .\\ \dot{g}_3 &= \eta_3 = b\end{aligned} \tag{32}$$

The differential equations can be solved directly. By solving them we do not only find a fundamental solution of the particular set of equations (28) but also a fundamental solution of any family of operators that commute according to the same commutation relations.

1.7 The Covering Group

We can think about the Wei-Norman equation in another way. Consider the pair of linear equations

$$\dot{x} = Ax + \sum_i u_iB_ix\ , \qquad x \in \mathbb{R}^n \tag{33}$$

$$\dot{z} = Fz + \sum_i u_iG_iz, \qquad z \in \mathbb{R}^m \tag{34}$$

and suppose that there exists a map ϕ which maps A into F, B_i into G_i, and which extends to a Lie algebra isomorphism from $\{A,B_i\}_{LA}$ to $\{F,G_i\}_{LA}$. Can we determine the fundamental solution of (34) from the fundamental solution of (33)?

It is clear that since (33) and (34) are related by a Lie algebra isomorphism, we can pick an ordered basis for the Lie algebras such that the Wei-Norman equations of (34) are the same as those of (33). However, this only tells us that the fundamental solutions are related for g near to zero. The global picture may be radically different as becomes clear from an example.

Example:

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = u\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}$$

$$\begin{bmatrix}\dot{z}_1\\ \dot{z}_2\end{bmatrix} = u\begin{bmatrix}0 & 2\\ -2 & 0\end{bmatrix}\begin{bmatrix}z_1\\ z_2\end{bmatrix}$$

The Lie algebra's of these two equations are isomorphic, i.e.

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \simeq \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}$$

but the solution of the first is

$$\Phi_x = \begin{bmatrix} \cos \int udt & \sin \int udt \\ -\sin \int udt & \cos \int udt \end{bmatrix}.$$

while the solution of the second is

$$\Phi_z = \begin{bmatrix} \cos \int 2udt & \sin \int 2udt \\ -\sin \int 2udt & \cos \int 2udt \end{bmatrix}$$

and so only for $\int udt$ small can the fundamental solutions Φ_x and Φ_z be mapped onto each other. The z-system is a "double covering" of the x-system in the sense that when the x-system makes one rotation the z-system rotates two times.

In this example the group that we obtained by exponentiating the Lie algebra was S^1 which is not simply connected. To arrive at an x-system that reveals as much as possible about exponentiating the Lie algebra we must arrange the x-system in such a way that the corresponding group is simply connected. Such an x-system would cover all systems with the same Lie algebra.

If we assume that the set of matrices $\exp A_i X$ is a simply connected group and if

$$\phi : A_i \to F_i$$

is extendable to a Lie algebra homomorphism. Then there exists a group homomorphism

$$\psi : \exp\{A_i x\} \to \exp\{F_i z\}$$

such that for

$$\dot{x} = (\Sigma u_i A_i)x \text{ and } \dot{z} = (\Sigma u_i F_i)z$$

we have $\psi(X) = Z$. Thus X solves all systems generated this way.

In our example we can take as the covering system

$$\dot{x} = u, \qquad x \in \mathbb{R} \tag{35}$$

Exponentiating this Lie algebra gives $\mathbb{R}$, which is the simply connected covering group of S^1 (the circle), and clearly from (35) we can obtain the solution of (33) and (34). This idea has a limitation: there may be no matrix group which is "big enough", i.e. the simply connected covering group may not have a matrix representation. A well known example due to Birkhoff shows that this is the case for $S\ell(2)$ (the 2×2 matrix of det +1). The study of global equivalence is important for recursive filters. If we don't take care the resulting filter will work only for a finite time and questions of a steady-state behavior will be completely inaccessible.

A less ambitious approach to global equivalence is the following. Let us consider $x = (x_1, \ldots, x_n)$. Introduce the monomials homogeneous of degree p in x_i

$$x_\alpha^{[p]} := (x_1^p, {}_{p-1,1}x_1^{p-1}x_2, \ldots, x_n^p)$$

This is a vector with $\binom{n-p-1}{p}$ components. The constants α denote a certain normalization. In fact we can define a set of constants α^p such that for the ℓ_p-norm $||x|| = (\Sigma x_i^p)^{1/p}$ we have the equality: $||x||^{2p} = ||x^{[p]}||^2$. For each normalization we can define a mapping

$$\text{sub}[p]: \text{ n×n matrices } \binom{n-p-1}{p} \times \binom{n-p-1}{p} \text{ matrices}$$

such that if

$$\dot{x} = \left(\sum_{i=1}^{m} u_i A_i\right)x \tag{36}$$

then

$$\frac{d}{dt}x^{[p]} = \left(\sum_{i=1}^{m} u_i A_{i[p]}\right)x^{[p]} \tag{37}$$

and if we map $x(0) \longmapsto x^{[p]}(0)$ as above then solving (36) solves (37). Since we can find a linearly independent set in $\mathbb{R}^\beta$, $\beta = \binom{n-p-1}{p}$ consisting of vectors of the form $x^{[p]}$, this means that we can find the fundamental solution of (37) based on (36). Thus, again, (36) is a kind of universal simulator for all equations of the form (37). Note that (37) has the same Lie algebra as (36); this follows from the fact (37) can be regarded as being just (36) in a different coordinate system.

<u>Example</u>

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} u & v \\ w & -u \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad u, v, w : \mathbb{R} \to \mathbb{R}$$

then

$$\frac{d}{dt}\begin{bmatrix} x_1^2 \\ 2x_1x_2 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2u & v & 0 \\ 2w & 0 & 2v \\ 0 & w & -2u \end{bmatrix}\begin{bmatrix} x_1^2 \\ 2x_1x_2 \\ x_2^2 \end{bmatrix}$$

(We have chosen the normalization α corresponding to the ℓ^1-norm.) One might think that under the mapping $x(0) \to x^{[p]}(0)$ we only reach a linear subspace of $\mathbb{R}^\beta$, because the image of the map in $\mathbb{R}^\beta$ is very thin. It is a basic fact from tensor theory that the image of this map contains a basis. For instance in our example:

$$\begin{bmatrix}1\\0\end{bmatrix} \to \begin{bmatrix}1\\0\\0\end{bmatrix}, \qquad \begin{bmatrix}1\\1\end{bmatrix} \to \begin{bmatrix}1\\2\\1\end{bmatrix}, \qquad \begin{bmatrix}0\\1\end{bmatrix} \to \begin{bmatrix}0\\0\\1\end{bmatrix}$$

The advantage of the present approach over the Wei-Norman g-equations is that here we have a construction which works for all u and all t. The limitation is that there are isomorphisms between matrix Lie algebras which do not arise in this way. Thus this is not a completely general tool.

II. FINITE STATE ESTIMATION PROBLEMS

2.1 Stochastic Differential Equations

Consider the stochastic differential equation (in Itô form) on the manifold X

$$dx = f(x)dt + \sum_{i=1}^{m} g_i(x)dw_i \qquad x \in X \tag{1}$$

where w_i are independent standard Wiener processes. This class includes, for instance, linear equations:

$$dx = Axdt + \sum_{i=1}^{m} b_i dw_i \tag{2}$$

and bilinear equations:

$$dx = Axdt + \sum_{i=1}^{m} B_i x dw_i \tag{3}$$

We recall the <u>Itô rule</u> for differentiating a function

$$d\phi = \left\langle \frac{\partial\phi}{\partial x}, f(x)\right\rangle dt + \left\langle \frac{\partial\phi}{\partial x}, \sum_{i=1}^{m} g_i(x)dw_i\right\rangle + \frac{1}{2}\Sigma \frac{\partial^2\phi}{\partial x \partial x}(g_i, g_i)dt$$

The adjoint of the operator which appears on the right hand side is the <u>Fokker-Planck operator</u>.

One of the early results exemplifying the interplay between stochastic equations and geometry is the connection between the controllability of certain control systems related to the stochastic differential equation and the smoothness of the solutions of the Fokker-Planck equation. Consider an equation $L\phi = \psi$ where L is a differential operator and ψ and ϕ are distributions in sense of L. Schwartz. Both ϕ and ψ are defined on a manifold M and may have singular parts (e.g. delta functions).

Definition: L is called hypoelliptic if every solution of $L\phi = \psi$ is C^∞ off the support of the singular part of ψ after a suitable modification of ϕ on a set of measure zero. For instance elliptic operators with C^∞ coefficients are hypoelliptic.

Theorem: (Hörmander [11]) Let L be a differential operator of the form $L = L_o + \sum_i (L_i)^2$ where $L_o = \sum_j a_j(x) \frac{\partial}{\partial x_j}$ and $L_i = \sum_j b_{ij}(x)\frac{\partial}{\partial x_j}$ in which the functions $a_j(\cdot)$ and $b_{ij}(\cdot)$ are smooth functions on a manifold of M. If at every point $x \in M$ the Lie algebra $\{L_o, L_i\}_{LA}$ spans the tangent space $T_x M$ then L is hypoelliptic.

Example: Let's look at the Green's function of the diffusion equation:

$$- \frac{\partial p}{\partial t} + \frac{1}{2}\frac{\partial^2 p}{\partial t^2} = -\delta(t,x) \tag{4}$$

In this case the manifold is $M = \mathbb{R} \times \mathbb{R}$, $L_o = \frac{\partial}{\partial t}$ and $L_1 = \frac{\partial}{\partial x}$. Clearly $\frac{\partial}{\partial t}$ and $\frac{\partial}{\partial x}$ commute and $\{\frac{\partial}{\partial t}, \frac{\partial}{\partial x}\}_{LA}$ spans, at every point, the tangent space to $\mathbb{R} \times \mathbb{R}$ so that $(-\frac{\partial}{\partial t} + \frac{1}{2}\frac{\partial^2}{\partial x^2})$ is hypoelliptic. The impulse response is

$$p(x,t) = \frac{\theta(t)}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} \tag{5}$$

where $\theta(t)$ is Heaviside's step function.

Consider again the stochastic differential equation

$$dx = f(x)dt + \sum_{i=1}^{m} g_i(x)dw_i \tag{6}$$

The associated Fokker-Planck equation is of the form:

$$(-\frac{\partial}{\partial t} + L + \sum_i (L_i)^2)\rho = 0 \tag{7}$$

The differential operator at the left hand side is again of the form of a first order part plus a second order part which is a sum of squares. We have now the following result.

Theorem: The Fokker-Planck operator is hypoelliptic if and only if the associated system

$$\dot{x} = f(x) - \frac{1}{2}\sum_{i=1}^{m} \frac{\partial g_i}{\partial x} g_i + \sum_{i=1}^{m} u_i g_i(x) \tag{8}$$

is controllable in the sense that the Lie algebra $\mathscr{L}$ spans the tangent space of X at each point.

It we write the stochastic differential equation as a Stratonovich equation

$$dx = a(x)dt + \sum_{i=1}^{m} b_i(x)dw_i \tag{9}$$

then the Fokker-Planck operator is hypoelliptic if and only if the associated control system

$$\dot{x} = a(x) + \sum_{i=1}^{m} u_i b_i(x) \tag{10}$$

has the exact time reachability property.

2.2 The Conditional Probability Equation

The deterministic system

$$\dot{x} = Ax+uBx \tag{11}$$

may be thought of as being a Fisk-Stratonövich equation

$$\eth x = Axdt + Bx\eth w \tag{12}$$

($\eth$ denotes a Stratonovich differential as opposed to Itô notation) with the noise replaced by a control. It is sometimes useful to notice that the substitution

$$z = e^{\int uBdt}x$$

which implies

$$\dot{z}(t) = e^{-\int_0^t Bud\sigma} A e^{\int_0^t Bud\sigma} z(t)$$

gives an equivalent equation in which u only appears as an integrand. Thus only the integrated form of u enters the equation. The same is true of (12)

$$x = e^{-\int dwB}z \quad => \dot{z} = e^{-Bw}Ae^{Bw}$$

Stated differently, for the input-output system

$$\dot{x} = Ax+\dot{u}Bx, \qquad y = x, \qquad \dot{u} = \frac{du}{dt}$$

we can find z so that

$$\dot{z} = C(u)z, \quad y = f(z,u)$$

Moreover, the same is true for

$$\dot{x} = Ax+(\dot{u}_1B_1+\dot{u}_2B_2+\ldots+\dot{u}_mB_m)x$$

if the B_i's commute. These ideas have been explored by Friedmann and Willems [12], and by Doss [13] and Sussmann [14]. Under such a hypothesis the description of what the system does to continuous functions tells us what it will do to white noise " $\frac{dw}{dt}$ " even though differentiable functions are a set of (Wiener) measure zero. In a certain sense this class of systems is more robust than others.

We want to separate the <u>algebraic</u> complexity in nonlinear filtering theory from the analytical problems. For this purpose we will look at finite state processes. We follow the ideas in Brockett and Clark [15]. Let x(t) be a stochastic process taking on values in a finite set S $\mathbb{R}$, say

$$S = \{b_1,\ldots,b_n\}$$

Suppose $p_i(t)$ is the probability that x(t) is b_i and suppose

$$\dot{p} = Ap$$

with $p = (p_1,\ldots,p_n)^T$ and $A = (a_{ij})$ a generator (intensity matrix). Suppose we observe

$$dy = xdt+dw$$

with w a Brownian motion and want to propagate the conditional probability p(t) given y(s), $0 \leqslant s \leqslant t$. In the analogous discrete time situation the equation of the conditional density follows easily from Bayes' rule. Actually it's far more convenient to deal with an unnormalized version of the conditional probability which satisfies

$$\tilde{p}(k+1) = (B)^{y(k)}A\tilde{p}(k)$$

where B is the matrix

$$\begin{bmatrix} b_1 & & 0 \\ & \ddots & \\ 0 & & b_n \end{bmatrix}$$

and $p(k) = \tilde{p}/n$ where n is the normalization, $n = \tilde{p}_1+\tilde{p}_2+\ldots+\tilde{p}_n$. For the continuous time version Wonham [16] derived the conditional probability equation as an Itô equation

$$d\rho = A\rho dt + (B-\langle b,\rho\rangle I)\rho.(dy-\langle b,\rho\rangle dt)$$

where $b := (b_1,\ldots,b_n)^T$. By converting this to Stratonovich form we obtain

$$ \eth\rho = [A\rho - \frac{1}{2}(B-\langle b,\rho\rangle I)^2\rho + \frac{1}{2}\langle b,(B-\langle b,\rho\rangle I)\rho\rangle\rho]dt + (B-\langle b,\rho\rangle I)\rho(\eth y-\langle b,\rho\rangle dt) \quad (13) $$

In unnormalized form the equations take the much simpler form

$$ d\rho = A\rho dt + B\rho dy \quad (14) $$

and the Stratonovich version is

$$ \eth\rho = (A - \frac{1}{2}B^2)\rho dt + B\rho\eth y \quad (15) $$

where ρ now denotes the unnormalized conditional probability (so $\Sigma\rho_i$ need not be equal to 1).

We remark that for the Itô equation

$$ dx = Axdt + \Sigma B_i x dw_i $$

it is possible to study the moments by passing to the Stratonovich version

$$ \eth x = (A - \Sigma\frac{1}{2}B_i^2)xdt + \sum_{i=1}^{m} B_i x dw_i $$

and then using the map $x \to x^{[p]}$ to get a Stratonovich equation for $x^{[p]}$. This, in turn can be used to get a differential equation for the moments of x, e.g.

$$ \frac{d}{dt}\mathscr{E}x^{[p]} = ((A - \frac{1}{2}\sum_i B_i^2)_{[p]} + \sum_i (B_{i[p]})^2)\ x^{[p]} $$

where $\mathscr{E}$ denotes expectation. (see [17])

2.3 The UCP Equation as a Nonlinear System

Our point of view is that the conditional probability equation together with any functional of the conditional probability defines an input-output system. We may use the results from nonlinear system theory, i.e. the controllability, observability, and minimality properties, to study the unnormalized conditional probability equation. Specifically, consider the equation

$$ \dot\rho = (A - \frac{1}{2}B^2)\rho + uB\rho $$

with a scalar output map defined by the conditional mean

$$ y = \frac{\Sigma b_i\rho_i}{\Sigma\rho_i} $$

This defines an input-output system but as a realization of that input-output system it may not be controllable or observable. For efficient implementation of the filter we are interested in knowing if we can make the conditional density equation simpler. The questions of controllability and observability are very natural in this context.

The controllability algebra of the unnormalized conditional probability equation (UCP) is the matrix algebra

$$\{A - \frac{1}{2} B^2, B\}_{LA} \quad .$$

and the little observability algebra, for output $h(\rho)$, is

$$\{\frac{\partial}{\partial x_i} \hat{a}_{ij} x_j, h(\rho)\}_{LA}$$

with $(\hat{a}_{ij}) = A - \frac{1}{2}B^2$. We make the following remarks.

1. The UCP equation is never minimal, because the equation is unnormalized. The normalized version which evolves on a manifold of dimension n-1 also generates the same input-output map.

2. The controllability algebra of the unnormalized equation is closely related to the controllability algebra of the normalized equation.

3. One way to make sure the recursive filter is low dimensional is to force $\{A - \frac{1}{2} B^2, B\}_{LA}$ to be low dimensional. The dimension of this Lie algebra is an upper bound on the number of sufficient statistics.

We are lead by this approach to the following conceptual view of the nonlinear filtering problem. The UCP is to be thought of as an input-output system with, say, the conditional mean as the output

$$\xrightarrow{u} \boxed{\text{UCP}} \longrightarrow y$$

A recursive filter is a realization of this input-output map

$$u \longrightarrow \boxed{\begin{array}{c}\text{minimal}\\ \text{realization}\end{array}} \longrightarrow y$$

If we have a <u>finite</u> state process the question of <u>finite dimensionality</u> of the filter is not relevant because the UCP equation is automatically finite dimensional, however there can be a significant difference in the dimensionality of the UCP equation and its minimal realization.

We now focus on the following question. When does the UCP system have a simpler realization as an input-output system, and for what initial values of ρ is it the simplest? Recall that an intensity matrix (generator) A is called <u>irreducible</u> if there is no permutation matrix P such that

$$PAP^{-1} = \begin{bmatrix} A_1 & A_3 \\ 0 & A_2 \end{bmatrix} \quad \text{with } A_1,\ A_2 \text{ square}$$

A matrix representation of a Lie algebra is said to be <u>irreducible</u> if there is no choice of a basis such that all elements of the representation simultaneously appear as

$$\begin{bmatrix} A_1 & A_3 \\ 0 & A_2 \end{bmatrix} \quad \text{with } A_1,\ A_2 \text{ square}$$

<u>Fact:</u> If A is irreducible as an intensity matrix and the $\{b_i\}_{i=1}^{n}$ are distinct then $\{A-\frac{1}{2}B^2, B\}$ is irreducible as a representation of a Lie algebra.

<u>Proof</u>: Since B is diagonal and has distinct eigenvalues its invariant subspaces are all spanned by collections of the basis vectors. A leaves no such space invariant by hypothesis.

2.4 A Class of Examples

Consider the Lie algebra $s\ell(2,\mathbb{R}) \simeq \text{span}\ \left\{\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right\}$ with the objective of finding A and B such that controllability algebra $\{A-\frac{1}{2}B^2, B\}_{LA}$ is equal to $s\ell(2,\mathbb{R})$. Now take the operators

$$I = x\frac{\partial}{\partial y}\ , \qquad II = y\frac{\partial}{\partial x}\ , \qquad III = x\frac{\partial}{\partial x} - y\frac{\partial}{\partial y}$$

An easy calculation gives

(a) $[I,\ II] = III$
(b) $[I,\ III] = -2I$
(c) $[II,\ III] = 2II$

These commutation relations are the same as for $\left\{\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right\}$, so the operators I, II, III form a representation of $s\ell(2,\mathbb{R})$. Notice that the vector space of forms $\phi(x,y) = \Sigma a_i x^i y^{n-i}$ homogeneous of degree n are mapped into itself by these operators. So we can construct a matrix representation of $s\ell(2,\mathbb{R})$ by matrices of <u>any</u> given dimension in the following way:

$$I \simeq \begin{bmatrix} 0 & n & 0 & & \\ 0 & 0 & n-1 & \ddots & \\ & 0 & \ddots & \ddots & \\ & & \ddots & 0 & 1 \\ & & & 0 & 0 \end{bmatrix}; \quad II \simeq \begin{bmatrix} 0 & 0 & & & \\ 1 & 0 & \ddots & & \\ & 2 & \ddots & \ddots & \\ & & \ddots & 0 & 0 \\ & & & n & 0 \end{bmatrix}$$

$$\mathrm{III} \simeq \begin{bmatrix} n & 0 & & & \\ 0 & n-2 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & -n+2 & 0 \\ & & & 0 & -n \end{bmatrix}$$

A second point of view is the following. Consider the mapping

$$\begin{bmatrix} x \\ y \end{bmatrix} \to \begin{bmatrix} x \\ y \end{bmatrix}^{[p]}$$

as applied to

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} a & b \\ c & -a \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \left(a \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right) \begin{bmatrix} x \\ y \end{bmatrix}$$

This gives

$$\frac{d}{dt}\begin{bmatrix} x^n y^o \\ \binom{n}{1} x^{n-1} y^1 \\ \vdots \\ x^o y^n \end{bmatrix} = \begin{bmatrix} na & b & & & \\ nc & (n-2)a & 2b & & \\ & (n-1)c & \ddots & \ddots & \\ & & \ddots & \ddots & nb \\ & & & c & -na \end{bmatrix}\begin{bmatrix} x^n y^o \\ \binom{n}{1} x^{n-1} y^1 \\ \vdots \\ x^o y^n \end{bmatrix}$$

where we have chosen the ℓ^1-normalization. If we want the map sub[p] to take intensity matrices into intensity matrices, then we have to take this normalization. Following [18] we show how to modify matrices of this form by a transformation R defined by a diagonal matrix with positive elements

$$A \longmapsto RAR^{-1}$$

such that $\{A-\frac{1}{2}B^2, B\}_{LA}$ is of a suitable form. Because B is diagonal and R is doagonal, $RBR^{-1} = B$. Now we want to select A in the Lie algebra together with R and β such that for given α

$$A = R\bar{A}R^{-1} + \frac{1}{2}\alpha B^2 + \beta I$$

is an <u>intensity matrix</u> (note that we always can add a multiple of the identity βI, because the identity commutes with everything in the algebra and does not change the Lie algebra except in a trivial way). We can enforce the condition that A be an intensity matrix by asking that $cA = 0$ where $c = (1,\ldots,1)$. Thus we require

$$cRAR^{-1} + \frac{1}{2}\alpha cB^2 + \beta cI = 0$$

or

$$cRA + \frac{1}{2}\alpha cB^2 R + \beta cIR = 0$$

which can be written as

$$cR(A + \frac{1}{2}\alpha B^2 + \beta I) = 0$$

It is a consequence of the Frobenious theorem on positive matrices that this equation always has a solution R. Notice that because R is diagonal and positive it induces only a change of measure.

To summarize, given a positive integer n, we can construct a finite state process with n states, which has a four-dimensional estimation algebra. (Four because we had to add the identity to the three dimensional Lie algebra $s\ell(2,\mathbb{R})$.) In fact we can write down explicitly a filter for this problem [18]. Moreover we can actually simulate the nonlinear filtering equation with a two-dimensional equation because this x-equation covers (globally!) the $x^{[p]}$-equation. What plays the role of the Gaussian initial conditions as in the Kalman filter? The analogue in our set-up comes from the binomial distributions transformed by the change of measure defined by R. Additional details are found in [18]. We see that for a binomial initial condition the conditional density is propagated in a manifold of dimension three.

III. ESTIMATION OF DIFFUSION PROCESSES

3.1 A Class of Models

Consider a stochastic process h(x) where x satisfies the Itô equation

$$dx = f(x)dt + \sum_{i=1}^{m} g_i(x)dw_i \qquad (1)$$

with $w_1, w_2, \ldots w_m$ independent, standard Wiener processes. If h(x) is observed in the presence of white noise, i.e. if there is available for processing

$$dy = h(x)dt + d\nu \qquad (2)$$

where ν is an additional independent Wiener process, then under suitable assumptions, the conditional density for x at time t given dy on the interval [0,t] satisfies a stochastic, nonlinear, partial differential equation, directly analogous to the situation we encountered in the finite state case. This is the class of problems which is of interest here.

As in the finite state case, it is more convenient to work with an unnormalized version of the conditional density equation. In the present context this equation, written as an Itô equation, takes the form

$$d\rho(t,x) = (L_o\rho(t,x)dt + h(x)\rho(t,x)dy \qquad (3)$$

where L_o is the Fokker-Planck operator. As was pointed out in [15] the Lie theoretic point of view suggests that this be rewritten as a Stratonovich equation

$$\eth\rho(t,x) = (L_o - \frac{1}{2}h^2)\rho(t,x)+h(x)\rho\eth y \tag{4}$$

thus bringing into prominence the Lie algebra generated by the operator $L_o - \frac{1}{2}h^2$, and h. Because this algebra will be important in determining the complexity of the estimation problem we call it the <u>estimation algebra</u>. We refer to (4) as the UCD equation.

We also point out that if one inserts a parameter to adjust the magnitude of the noise terms

$$dx = f(x)dt+\alpha\Sigma g_i(x)dw_i$$
$$dy = h(x)dt+d\nu$$

then the estimation algebra is

$$\{\alpha\hat{L}_o - F - \frac{1}{2}h^2, h\}_{LP}$$

Setting α to zero gives an algebra which is isomorphic to the little observability algebra.

3.2 Representation Theory for the Estimation Algebra

Writing the UCD equation in Stratonovich form makes it clear that our understanding of nonlinear filtering would be enhanced if we knew what kind of Lie algebras can be generated by $L_o - \frac{1}{2}h^2$ and h. This is a question about the representation of Lie algebras in terms of linear operators on an infinite dimensional space. However the sort of representation which is of direct concern has a number of special features which set it apart from the standard theory which was developed by physicists and mathematicians to fill the needs of quantum mechanics. Briefly, the situation is this:

(a) We cannot assume that the elements of the Lie algebras exponentiate to give a one parameter group. Typically there is a cone in the Lie algebra which exponentiates to generate a semigroup of bounded operators which cannot be extended to a group.

(b) The representations of interest act on real L_1 spaces not complex L_2 spaces.

(c) The Lie algebra contains a cone whose exponentials map the nonnegative functions in L_1 into themselves.

The most basic example consists of a four-dimensional Lie algebra represented by the operators

$$L_0 = \frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}x^2 \quad ,$$

$$L_1 = x, \quad L_2 = \frac{\partial}{\partial x}, \quad L_3 = 1$$

which are to be thought of as acting on $S \subset L_1(\mathbb{R})$ where S is a suitable dense subset. Of course, $\exp L_0 t$ makes sense only for $t \geq 0$. (One easily sees that L_0 does generate a semigroup of bounded operators on $L_1(\mathbb{R})$.) Likewise, L_2 is a generator, in fact it generates a group. However, e^{xt} is not bounded for t positive or negative. We define the cone

$$K = \{\alpha L_0 + \beta L_1 + \gamma L_2 + \delta L_3 | \alpha,\beta,\gamma,\delta \in \mathbb{R},\ \alpha > 0\}$$

Each element of K is a generator of a semigroup. Since the cone has nonempty interior in the four-dimensional Lie algebra and since the operators formally satisfy the desired commutation relations, we will call $\{L_0,L_1,L_2,L_3\}$ a representation of the given Lie algebra.

Because the operator L_0 is always second order the Lie algebra representation $\{L_0 - \frac{1}{2}h^2, h\}_{LA}$ must contain at least one 2nd order operator. With a view toward understanding as much as possible about representations involving 2nd order operators we begin by describing one such family. This family is of interest because it includes, as a subalgebra, the estimation algebra associated with linear systems and it suggests several possibilities for nonlinear filtering. There is also a great deal of interest in this representation from pure mathematicians [19].

The set of real 2n by 2n matrices of the form

$$M = \begin{bmatrix} A & B \\ C & -A' \end{bmatrix} ; \quad B = B'; \quad C = C'$$

form a n(2n+1)-dimensional vector space under ordinary matrix addition. It is easliy verified that if M_1 and M_2 are of this form, then $M_1M_2 - M_2M_1$ is also and thus they form a Lie algebra with respect to the standard commutator product. Matrices of this form are called <u>Hamiltonian</u> or <u>infinitesimally symplectic</u>. The Lie algebra is called the <u>symplectic Lie algebra</u> and is denoted here by Sp(n).

We are also interested in an extension of this algebra consisting of real matrices of dimension 2n+1 by 2n+1, and having the form

$$M = \begin{bmatrix} A & B & b \\ C & -A' & c \\ 0 & 0 & 0 \end{bmatrix}; \quad B = B', \quad C = C'$$

Again, it is easy to verify that this set of matrices forms a Lie algebra. We will call it the extended symplectic algebra.

Turning now to Lie algebras of operators, consider the n(2n+1)-dimensional vector space consisting of sums of the form

$$L = \sum_{i,j=1}^{n} b_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + a_{ij}\left(\frac{\partial}{\partial x_i} x_j - \frac{1}{2}\delta_{ij}\right) + c_{ij} x_i x_j$$

where δ_{ij} is one if $i = j$ and zero otherwise. For example, if $n = 1$, we are looking at real linear combinations of the type

$$L = b\frac{\partial^2}{\partial x^2} + a\left(\frac{\partial}{\partial x} x - \frac{1}{2}\right) + cx^2$$

The identities

$$\left[\frac{\partial^2}{\partial x^2}, x^2\right] = \frac{\partial^2}{\partial x^2} x^2 - x^2 \frac{\partial}{\partial x^2} = 4\frac{\partial}{\partial x} x - 2$$

$$\left[\frac{\partial^2}{\partial x^2}, \frac{\partial}{\partial x} x - \frac{1}{2}\right] = \frac{\partial^3}{\partial x^3} x - \frac{\partial}{\partial x} x \frac{\partial^2}{\partial x^2} = 2\frac{\partial^2}{\partial x^2}$$

$$\left[\frac{\partial}{\partial x} x - \frac{1}{2}, x^2\right] = \frac{\partial}{\partial x} x^3 - x^2 \frac{\partial}{\partial x} x = 2x^2$$

verify that in the case $n = 1$, the above set of operators are closed under commutation and hence form a Lie algebra. It is straightforward to verify that the same is true for $n > 1$. For example, for i,j and b distinct

$$\left[\frac{\partial^2}{\partial x_i \partial x_j}, x_k x_j\right] = \frac{\partial}{\partial x_i} x_k$$

$$\left[\frac{\partial^2}{\partial x_i \partial x_j}, \frac{\partial}{\partial x_k} x_j\right] = \frac{\partial}{\partial x_i}\frac{\partial}{\partial x_k}$$

$$\left[\frac{\partial}{\partial x_i} x_j, x_k x_i\right] = x_j x_k$$

and the other cases follow similarly.

In order to understand the structure of this Lie algebra we establish an isomorphism between it and the symplectic algebra introduced above. The isomorphism works as follows:

$$\Sigma a_{ij}\left(\frac{\partial}{\partial x_i} x_j - \frac{1}{2}\delta_{ij}\right) \longmapsto \begin{bmatrix} (a_{ij}) & 0 \\ 0 & -(a_{ji}) \end{bmatrix}$$

$$\frac{1}{2}\Sigma b_{ij}\frac{\partial^2}{\partial x_i \partial x_j} \longmapsto \begin{bmatrix} 0 & (b_{ij}) \\ 0 & 0 \end{bmatrix}$$

$$\frac{1}{2}\Sigma c_{ij} x_i x_j \longmapsto \begin{bmatrix} 0 & 0 \\ (c_{ij}) & 0 \end{bmatrix}$$

This then allows one to understand the symplectic algebra a different way, i.e., as the Lie algebra of 2nd order partial differential operations of the given form with commutation of operators being the Lie bracket.

We leave the verification of the fact that this is a Lie algebra isomorphism to the reader. However, the identities we have given do most of the work and a study of the case $n = 2$ involving

$$\frac{1}{2}\left(b_{11}\frac{\partial}{\partial x^2} + 2b_{12}\frac{\partial}{\partial x}\frac{\partial}{\partial y} + b_{22}\frac{\partial^2}{\partial y^2}\right) \longmapsto \begin{bmatrix} 0 & 0 & b_{11} & b_{12} \\ 0 & 0 & b_{12} & b_{22} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$a_{11}\frac{\partial}{\partial x} x + a_{12}\frac{\partial}{\partial x} y + a_{21}\frac{\partial}{\partial y} x + a_{22}\frac{\partial}{\partial y} y - \frac{1}{2}(a_{11} - a_{22})$$

$$\longmapsto \begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ 0 & 0 & -a_{11} & -a_{21} \\ 0 & 0 & -a_{12} & -a_{22} \end{bmatrix}$$

and

$$\frac{1}{2}\left(c_{11}x^2 + 2c_{12}x_y + c_{22}y^2\right) \longmapsto \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ c_{11} & c_{12} & 0 & 0 \\ c_{12} & c_{22} & 0 & 0 \end{bmatrix}$$

should be convincing.

With respect to the remarks on the generation of semigroups, if we regard the operators defined above as acting on a suitable dense subset of $L_1(\mathbb{R}^n)$, then the theory of diffusion processes tells us that

$$L = \Sigma b_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + a_{ij} \frac{\partial}{\partial x_i} x_j$$

generates a semigroup provided (b_{ij}) is a positive definite matrix. Moreover, if (c_{ij}) is positive definite along with (b_{ij}), then

$$\hat{L} = \sum_{i,j} b_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + a_{ij} \frac{\partial}{\partial x_i} x_j - c_{ij} x_i x_j$$

will generate a subgroup. Of course, without some kind of definiteness assumptions, no such conclusions are valid. Since the conditions on b_{ij} and c_{ij} give a cone in the Lie algebra having a nonempty interior, we have here a representation of the symplectic algebra which meets the conditions set forth at the beginning of this section.

There is a generalization of this construction which enables one to capture other important examples. Consider appending to the family of operators discussed above the first order constant coefficient operators $\partial/\partial x_i$, multiplication by linear functions x_i and multiplication by constants. We notice that

$$[\frac{\partial}{\partial x_k^2}, b\frac{\partial}{\partial x_i} + c_j x_i] = 2\frac{\partial}{\partial x_k} c_j \delta_{ij}$$

$$[\frac{\partial}{\partial x_i}, x_j] = \delta_{ij}$$

$$[\frac{\partial}{\partial x_i} x_j, x_i] = x_j \qquad i \neq j$$

$$[x_i x_j, \frac{\partial}{\partial x_i}] = x_j \qquad i \neq j$$

which together with other obvious identities show that $(2n^2+3n+1)$-dimensional vector space of operators of the form

$$L = \Sigma b_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + a_{ij} \frac{\partial}{\partial x_i} - \frac{1}{2} \delta_{ij} +$$

$$c_{ij}x_ix_j + b_i\frac{\partial}{\partial x_i} + c_ix_i + d$$

is a Lie algebra.

The mapping which sends this operator into the element of the extended symplectic algebra is given by

$$L \longmapsto 2\begin{bmatrix} A & B & b \\ C & -A' & c \\ 0 & 0 & 0 \end{bmatrix}.$$

defines a Lie algebra homomorphism. The constants ("d") lie in the kernel but otherwise this homomorphism is faithful. In Section 2 we discussed a 4-dimensional algebra which, under this homomorphism, goes into the set of all 3 by 3 matrices of the form

$$\begin{bmatrix} 0 & b & \alpha \\ -b & 0 & \beta \\ 0 & 0 & 0 \end{bmatrix}$$

There is one last comment on the structure of this Lie algebra. The subalgebra consisting of terms of the form

$$\hat{L} = \Sigma b_i\frac{\partial}{\partial x_i} + c_ix_i + d$$

forms an ideal in the whole algebra having dimension 2n+1. The terms of the form

$$\tilde{L} = \Sigma b_{ij}\frac{\partial^2}{\partial x_i\partial x_j} + a_{ij}\left(\frac{\partial}{\partial x_i}x_j - \frac{1}{2}\delta_{ij}\right) + c_{ij}x_ix_j$$

form a complementary subalgebra that is simple.

3.3 Linear Estimation Theory

A general linear model for recursive estimation theory is

$$dx = Axdt + \sum_{i=1}^{m} b_i dw_i$$

$$dy_i = \langle c_j, x\rangle dt + dv_j; \quad j = 1,2,\ldots,p$$

with $w_1,w_2,\ldots,w_m,v_1,v_2,\ldots v_p$ being independent brownian motions. The unnormalized conditional density equations for such systems take the form (as a Stratonovich equation)

$$\frac{\partial p}{\partial t} = (\Sigma\hat{b}_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + \frac{\partial}{\partial x_i} a_{ij}x_j - \frac{1}{2} <c_j,x>^2 + \dot{y}_i <c_i,x>)\rho$$

or

$$\frac{\partial p}{\partial t} = (L_o + \dot{y}L_1)\rho$$

Both the operator L_o and the operator L_1 belong to the $(2n^2+3n+1)$-dimensional Lie algebra defined in the previous section and so the Lie algebra which they generate is necessarily a subalgebra of the algebra discussed above. There is more to be said. Because L_1 belongs to the ideal L mentioned above, it is clear that the Lie algebra generated by L_o and L_1, which is the relevant algebra for studying the solution of the unnormalized conditional density equation, cannot have dimension higher than 2n+2. In fact, only L_o intersects the "symplectic part" of the algebra. Typically the Lie algebra generated by L_o and L_1 is of dimension 2n+2 in the case of single-input/single-output systems. In fact, if one assumes controllability and observability then only the presence of a degeneracy in the form of all-pass factors stands in the way of this conclusion (see [20]). Based on the homomorphism into the extended symplectic group given here, we see that a same general conclusion holds in the multivariable case.

We now consider a single-input/single-output situation and write down side-by-side the UCD realization

$$\partial\rho = L_o\rho dt + L_1\rho dy$$

$$\bar{y} = \int\rho<c,x>dx/\int\rho dx$$

and the usual Kalman-Bucy filter for the state estimation and the error variance

$$dz = (A-Pcc^T)2dt+Pcdy; \qquad \bar{y} = <c,z>$$

$$\dot{P} = AP + PA + bb^T - Pcc^TP \quad .$$

From the latter we obtain vector fields corresponding to the coefficient of dt and the coefficient of dy, respectively,

$$\begin{bmatrix} (A-Pcc^T)z \\ AP+PA^T+bb^T-Pcc^TP \end{bmatrix} , \quad \begin{bmatrix} Pc \\ 0 \end{bmatrix}$$

This pair is to be compared with the pair coming from UCD

$$(L_o - \frac{1}{2} L_1^2) , \quad L_1$$

Computing the successive Lie brackets of the vector fields which appear in the recursive filter gives vector fields of the form

$$\begin{bmatrix} P(A^T)^k c+d_k \\ 0 \end{bmatrix}$$

where d_k does not depend on P and $P(A^T)^k c$ is linear in P. There exist 2n vector fields of this form if c is cyclic for A^T. In fact there is a Lie algebra homomorphism which maps the unnormalized conditional density operators to the Lie algebra so obtained. The homomorphism has a one dimensional kernel corresponding to multiplication by a constant. (The reason for this kernel is that the UCD equation is not normalized.)

Finally, we show how to pass from the unnormalized conditional density equation to the recursive estimation equation in a logical way. Consider the case where x is a scalar and

$$dx = dw$$

$$dy = xdt+dv$$

The UCD equation is given by

$$d\rho = \frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2} x^2\rho+dyx\rho$$

The estimate is

$$\hat{x} = \frac{\int x\rho dx}{\int \rho dx}$$

Suppose x(0) has a gaussian distribution and look for a solution of the form

$$\rho(t,x) = e^{a(t)x^2+b(t)x+c(t)}$$

where a(t), b(t) and c(t) are not yet known. By differentiating this expression with respect to time and space we get three equations which make it possible to determine the unknowns a(t), b(t) and c(t).

$$\frac{\partial}{\partial t}(\rho(t,x)) = (\dot{a}x^2+\dot{b}x+\dot{c})\rho(t,x)$$

$$\frac{\partial}{\partial x}(\rho(t,x)) = [2ax+b]\rho(t,x)$$

$$\frac{\partial^2}{\partial x^2}(\rho(t,x)) = 2a\rho(t,x)+(2ax+b)^2\rho(t,x)$$

Substitution into the UCD equation and changing to Itô form gives

$$\dot{a}x^2+b\dot{x}+c = (ka-\tfrac{1}{2})x^2+(kab+\dot{y})x+2a+b$$

Equating the respective coefficients of x results in

$\dot{a} = ka^2(-\frac{1}{2})$ This plays the role of the Riccati equation

$\dot{b} = kab+\dot{y}$ This equation propagates the mean

$\dot{c} = 2a+b$ Normalization equation

We get three sufficient statistics, one of them is a function which does not depend on the sample path. (That is the reason why it is possible to compute the solution of the Riccati equation off-line.)

3.4 A Nonexistence Result

In section 3.2 we discussed a class of representations involving 2nd order differential operators, a simple example of which is the set spanned by

$$\{\frac{d^2}{dx^2}, \ x\frac{d}{dx}, \ x^2, \ 1\}$$

One direct way to generate different representations from a given one is to pick a nonvanishing function ψ and make the substitutions

$$\frac{d^2}{dx^2} \mapsto \psi^{-1}\frac{d^2}{dx^2}\psi \ ; \qquad x^2 \to x^2$$

$$x\frac{d}{dx} \mapsto \psi^{-1}x\frac{d}{dx}\psi \ ; \qquad 1 \to 1$$

This kind of substitution has been investigated by Mitter [21] in an estimation theory context and can be used to explain how the example of Beneš [22] can be obtained from a linear problem.

There is a second group of transformations which can be used to generate new representations of Lie algebras from old ones in the present context. Consider the estimation algebra generated by

$$dx = f(x)dt+g(x)dw; \quad dy = h(x)dt+d\nu$$

For any given diffeomorphism of the real line into itself say $x \to \phi(x) = z$ we obtain from the Itô rule a new equation

$$dz = \tilde{f}(z)dt + \tilde{g}(z)dw; \quad dy = \tilde{h}(z)dt+d\nu$$

having an estimation algebra which is isomorphic to the original one. Thus this is a second way to generate new representations of Lie algebras in terms of 2nd order operations (see [18]).

It is obvious that as it stands $\{\frac{d^2}{dx^2}, \ x\frac{d}{dx}, \ x^2, \ 1\}$

does not contain operators of the form required in an estimation algebra. In view of the many alternative forms a representation of these same algebras may take, however, it makes sense to ask if there is any pair L_o, h such that L_o is a Fokker-Planck operator on $\mathbb{R}^1$, $h : \mathbb{R}^1 \to \mathbb{R}^1$, and $\{L_o, h\}_{LA}$ is isomorphic to this algebra. Correcting a claim to the contrary in [18] the answer is "no", provided the Fokker-Planck operator is sufficiently well behaved to avoid explosions.

The full proof is long and will not be given here. It can be put together out of the following remarks.

(a) If we have a representation of the Lie algebra in terms of 2nd order operators we can find a basis for the first derived algebra which takes the form

$$m \frac{d^2}{dx^2} + n \frac{d}{dx} + r, \quad p \frac{d}{dx} + q, \quad h$$

(b) The commutation relations then imply certain differential equations relating the coefficients m,n,r,p,q,n.

(c) There is an explicit criterion for explosions, see McKean [23] page 65, which is incompatible with the conditions on m etc. implied by the differential equations.

Some related work, but apparently not this particular result occurs in Occone's thesis [24].

REFERENCES

1. R. Hermann, "On the Accessibility Problem in Control Theory," in <u>Nonlinear Differential Equations and Nonlinear Mechanics</u>, (J.P. LaSalle and S. Lefschetz eds.) Academic Press, N.Y., 1963.

2. H. Hermes and G.W. Haynes, "On the Nonlinear Control Problem with Control Appearing Linearly," <u>SIAM J. on Control</u>, Vol. I, (1963) pp. 85-108.

3. C. Lobry, <u>Quelques Aspects Qualitatifs de la theories de la Commande</u>, L'Universite Scientifique et Medicak de Grenoble, pour obtenir le titre de Docteur es Sciences Mathematiques, May 19, 1972.

4. H. Sussmann and V. Jurdjevic, "Controllability of Nonlinear Systems," <u>J. of Diff. Eqs</u>. Vol. 12, (1972), pp. 95-116.

5. A. Krener, "A Generalization of Chow's Theorem and the Bang-Bang Theorem to Nonlinear Control Problems," <u>SIAM J. on Control</u>, Vol. 12, No. 1 (1974), pp. 43-52.

6. R.W. Brockett, "System Theory on Group Manifolds and Coset Spaces," SIAM J. on Control, Vol. 10, No. 2, (1972) pp. 265-284.

7. H.J. Sussmann, "Existence and Uniqueness of Minimal Realizations of Nonlinear Systems," Math. Systems Theory, Vol. 10, No. 3, (1975), pp. 1476-147

8. A.J. Krener and R. Hermann, "Nonlinear Controllability and Observability," IEEE Trans. on Automatic Control, Vol. 22, No. 5, (1977) pp. 728-740.

9. J. Wei and E. Norman, "On the Global Representation of the Solutions of Linear Differential Equations as a Product of Exponentials," Proc. Am. Math. Soc., April 1964.

10. K.T. Chen, "Decomposition of Differential Equations," Math. Ann., Vol. 146, pp. 263-278, 1962.

11. L. Hormander, "Hypoelliptic Second Order Differential Equations," Acta Math., Vol. 119, (1967), pp. 147-171.

12. M.I. Freedman and J.C. Willems, "Smooth Representation of Systems with Differential Inputs," IEEE Trans. on Aut. Control, Vol. 23, (1978), pp. 16-21.

13. H.J. Sussmann, "On the Gap Between Deterministic and Stochastic Ordinary Differential Equations," Ann. of Probability, Vol. 6, (1978), pp. 19-41.

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

15. R.W. Brockett and J.M.C. Clark, "The Geometry of the Conditional Density Equations," Proc. of the Oxford Conf. on Stochastic Systems, Oxford, England, 1978.

16. M.W. Wonham, "Some Applications of Stochastic Differential Equations to Optimal Nonlinear Filtering," SIAM J. on Control, Vol. 2, (1965).

17. R.W. Brockett, "Lie Theory and Control Systems Defined on Spheres," SIAM J. on Applied Mathematics, Vol. 25, No. 2, Sept. 1973, pp. 213-225.

18. R.W. Brockett, "Classification and Equivalence in Estimation Theory," 1979 Conf. on Decision and Control, IEEE, N.Y., 1979.

19. R. Howe, "On the Role of the Heisenberg Group in Harmonic Analysis," Bulletin of the AMS, Vol. 3, (1980), pp. 821-843.

20. R.W. Brockett, "Remarks on Finite Dimensional Nonlinear Estimation," in Analyse des Systèmes, Astérisque, Vol. 75-76 1980, pp. 47-55.

21. S.K. Mitter, "On the Analogy Between Mathematical Problems of Nonlinear Filtering and Quantum Physics," Ricerche di Automatica (to appear).

22. V. Benes, "Exact Finite Dimensional Filters for Certain Diffusions with Nonlinear Drift," Stochastics (to appear).

23. H.P. McKean Jr., Stochastic Integrals, Academic Press, N.Y. 1969.

24. D. Ocone, Topics in Nonlinear Filtering Theory, Sc.D. Thesis, Massachusetts Institute of Technology, 1980.