LIE ALGEBRAIC METHOD IN FILTERING AND IDENTIFICATION

M. Hazewinkel
Centre for Mathematics and Computer Science
P.O.Box 4079, 1009 AB Amsterdam
The Netherlands

ABSTRACT. These lectures concern (nonlinear) filtering. Very roughly the art of obtaining best estimates for some stochastic time-varying variable x on the basis of observations of another process y. The more concrete object under consideration being a stochastic dynamical system $dx = f(x)dt + G(x)dw$, where w is Wiener noise, with observations $dy = h(x)dt + dv$, corrupted by further noise. The subject as presented here involves ideas and techniques from Lie algebra theory, stochastics, differential topology, approximation theory and partial differential equations and has relations with quantum theory and stochastic physics. The lectures are adressed to practitioners in any one of these areas assuming that as a rule they are not experts in the other ones.

1. INTRODUCTION

Filtering is concerned with making estimates of quantities associated with a stochastic process $\{x_t\}$ on the basis of information gleaned from a related process $\{y_t\}$. The process $\{x_t\}$ is called the *signal* or *state* process and $\{y_t\}$ is the *observation* process. In this paper the following more concrete realization will be considered

$$dx_t = f(x_t)dt + G(x_t)dw_t, \quad x_t \in \mathbb{R}^n, \ w_t \in \mathbb{R}^m \tag{1.1}$$

$$dy_t = h(x_t)dt + dv_t, \ y_t \in \mathbb{R}^p, \ v_t \in \mathbb{R}^p \tag{1.2}$$

Here f is a function $\mathbb{R}^n \to \mathbb{R}^n$; G is an $n \times m$ matrix valued function on $\mathbb{R}^n$, h is a function $\mathbb{R}^n \to \mathbb{R}^p$ and w_t and v_t are Wiener processes, assumed independent of each other and also independent of the initial random variable x_0. More precisely these equations can be written

$$x_t = x_0 + \int_0^t f(x_s)ds + \int_0^t G(x_s)dw_s \tag{1.3}$$

$$y_t = \int_0^t h(x_s)ds + v_t \tag{1.4}$$

where the last term of (1.3) is a stochastic integral in the sense of *Ito*.

Much more loosely one can look at equations (1.1) and (1.2) as

$$\dot{x} = f(x) + G(x)\dot{w} \tag{1.5}$$

$$\dot{y} = h(x) + \dot{v} \tag{1.6}$$

with $\dot{w}$ and $\dot{v}$ white noise. Thus we have a differential equation $\dot{x} = f(x)$ on $\mathbb{R}^n$ which is subject to continuous random shocks whose intensity and direction (distribution) is state dependant and as observations we have an integral of some function of x and these observations are corrupted by more noise.

The general filtering problem for the state process $\{x_t\}$ with observation process $\{y_t\}$ is now to calculate for (interesting) functions ϕ of the state the conditional expectation

$$E[\phi(x_t) \mid y_s, \ 0 \leqslant s \leqslant t] = \widehat{\phi(x_t)}, \tag{1.7}$$

i.e. the best (least squares) estimate of $\phi(x_t)$ given the observations y_s up to time t. That is we are interested in calculation procedures for $\widehat{\phi(x_t)}$. In many (engineering) applications the data come in sequentially and one does not really want a calculating procedure which needs all the data y_s, $0 \leqslant s \leqslant t$, every time t that it is desired to find $\widehat{\phi(x_t)}$; rather we would like to have a procedure which uses a statistic m_t which can be updated using only the new observations y_s, $t \leqslant s \leqslant t'$ to its value $m_{t'}$ i.e.

$$m_{t'} = a(m_t, t', t, \{y_s : t \leqslant s \leqslant t'\}) \tag{1.8}$$

and from which the desired conditional expectation can be calculated directly, i.e.

$$\widehat{\phi(x_t)} = E[\phi(x_t) | y_s, 0 \leqslant s \leqslant t] = b(t, y_t, m_t). \tag{1.9}$$

Finally to actually implement the filter it would be nice if m_t were a finite dimensional quantity. All this leads to the (ideal) notion of a *finite dimensional recursive filter*. By definition such a filter is a system

$$d\xi_t = \alpha(\xi_t)dt + \sum_{i=1}^{p} \beta_i(\xi_t)dy_{it} \tag{1.10}$$

driven by the observations y_{it}; y_{it} is the i-th component of y_t, $i = 1,...,p$; together with an output map

$$\widehat{\phi(x_t)} = \gamma(\xi_t) \tag{1.11}$$

More precisely formulated our problem is now the following: given a system (1.1)-(1.2) and a function ϕ on $\mathbb{R}^n$, how can we decide whether for these data there exists a finite dimensional recursive filter (1.10)-(1.11) which calculates $\widehat{\phi(x_t)}$, the best least squares estimate, and how do we find the functions (vectorfields) $\alpha, \beta_1, \cdots, \beta_p, \gamma$ of (1.10)-(1.11).

Now this may of course be a totally unreasonable question to ask. It could be that such nice filters virtually never exist. That is not the case though. In the case of linear systems

$$dx_t = Ax_t dt + Bdw_t \tag{1.12}$$

$$dy_t = Cx_t dt + dv_t \tag{1.13}$$

where now A, B, C are matrices of the appropriate sizes (which may be time varying), the well known Kalman-Bucy filter is precisely such a filter as (1.10)-(1.11). The equations are as follows. The statistic ξ_t is a pair (m_t, P_t) consisting of an n-vector and a symmetric $n \times n$ matrix P_t. These evolve according to

$$dP_t = (AP_t + P_t A^T + BB^T - P_t C^T CP_t)dt \tag{1.14}$$

$$dm_t = Am_t dt + P_t C^T(dy_t - Cm_t dt). \tag{1.15}$$

Here X^T denotes the transpose of a matrix X. This filter was discovered in 1961 and it is hard to overestimate its importance: whole books are devoted to its applications into single specialized fields and substantial companies can make a good living doing little more than Kalman-Bucy filtering. Naturally, efforts immediately started to find similar filters for more general systems than (1.12)-(1.13). This turned out to be unexpectedly difficult and this is still the case though there exists hosts of approximate filters of various kinds which (seem to) work well in a variety of situations; there is very little systematically known about how to construct approximate filters or about how to predict that a given one or class will work well when applied to a given collection of systems.

The approach based on Lie-algebraic considerations which I will try to discuss and explain below seems to hold great promise both in understanding the difficulties involved and in providing some kind of systematic foothold in the area of constructing approximate filters. For, as will become clear below, the existence of finite dimensional recursive filters for a nontrivial statistic will be a rare event.

Let me pause at this point to point out that identification problems can easily be construed as

filtering problems. By way of illustrating this point consider again a linear system

$$dx_t = Ax_t + Bdw_t, \quad dy_t = Cx_tdt + dv_t \tag{1.16}$$

where now the matrices A, B, C are (partially) unknown. By adding to (1.16) the stochastic equations

$$da_{ij} = 0, \quad db_{kl} = 0, \quad dc_{qx} = 0 \tag{1.17}$$

for all unknown a_{ij}, b_{kl}, c_{qr}, one obtains a system (1.16)-(1.17) (of much larger state space dimension). And solving the filtering problem for the functions which project the vector $(x, (a_{ij}), (b_{kl}), (c_{qr}))$ onto a suitable component means identifying that particular coefficient.

2. THE DMZ-EQUATION AND THE ESTIMATION ALGEBRA

Let $\{x_t\}$ be a diffusion process as in (1.1)-(1.2) above. Given sufficient regularity of f, G, h the conditional expectation $\hat{x}_t$ will have a density $\pi(x, t)$.

Theorem 2.1. Under appropriate regularity conditions there exists an unnormalized version $\rho(x, t)$ of $\pi(x, t)$ which satisfies an equation

$$d\rho = \mathcal{L}\rho dt + \sum_{i=1}^{p} h_i(x)\rho dy_{it} \tag{2.2}$$

where $\mathcal{L}$ is the second order differential operator given by

$$(\mathcal{L}\psi) = \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}((GG^T)_{ij}\psi) - \sum_{i=1}^{n} \frac{\partial}{\partial x_i}(f_i\psi) - \frac{1}{2}\sum_{j=1}^{p} h_j^2\psi \tag{2.3}$$

Here $(GG^T)_{ij}$ is the (i,j)-th component of the $n \times n$ matrix $G(x)G(x)^T$ and f_i, h_j are the i-th and j-th component respectively of f and h.

Several comments are in order. First of all equation (2.2) is in Fisk-Stratonovic form. The corresponding Ito equation looks the same with $\mathcal{L}$ changed by removing the $-\frac{1}{2}\sum h_j^2\psi$ term. The word "unnormalized" means that $\rho(x, t) = \sigma(t)\pi(x, t)$ where $\sigma(t)$ is an unknown function of time. Under appropriate reachability conditions on (1.1) $\rho(x, t)$ is a positive function. That ρ is unnormalized does not hurt much as $\rho(x, t)$ still suffices to calculate such things as $\hat{\phi(x_t)}$. Indeed

$$\hat{\phi(x_t)} = (\int \rho(x,t)dx)^{-1} \int \rho(x,t)\phi(x)dx \tag{2.4}$$

Theorem 2.1 was proved by Duncan [13], Mortensen [28] and Zakai [36] and the corresponding equation 2.1 is often refered to as the Duncan-Mortensen-Zakai or DMZ equation.

It is a stochastic partial differential equation being driven by the stochastic processes $y_1, ..., y_p$.

It is important to note (Brockett [5]), that equations (2.2), (2.4) together constitute in fact a recursive filter in the sense of (1.10)-(1.11). The role of ξ_t is played by $\rho(x, t)$ so that instead of a point ξ evolving on a finite dimensional M we have an evolving density, i.e. a point ρ in an infinite dimensional space of positive functions evolving with time.

The simplest nontrivial example of a system (1.1)-(1.2) is

$$dx = dw, \quad dy = xdt + dv \tag{2.5}$$

i.e. one dimensional Wiener noise linearly observed corrupted by further noise. In this case the DMZ-equation becomes

$$d\rho = (\frac{1}{2}\frac{\partial^2}{\partial x^2} - \frac{1}{2}x^2)\rho dt + x\rho dy_t, \quad (\frac{\partial\rho}{\partial t} = \frac{1}{2}\frac{\partial^2\rho}{\partial x^2} - \frac{1}{2}x^2\rho + x\rho\dot{y}) \tag{2.6}$$

i.e. we are dealing with the Euclidean Schrödinger equation with an extra forcing term. This is not an accident but part of a general pattern of which we shall see a further manifestation below in section 8,

cf. also Mitter [26,27] for other remarks on this theme. I do not know wether the use Bismut makes of the filtering equations when dealing with a stochastic approach to index theorems and the Dirac operator can also be fitted into this framework.

3. ROBUSTNESS AND NUMERICAL MATTERS

As it stands equation (2.2) is not a very useful object for applications. It is a stochastic partial differential equation (with as probability space a space of paths $\{y\}$) and as such a solution is in principle only defined apart from a set of measure zero. On the other hand actual observations will always consist of piecewise smooth $y(t)$ and the class of all such is of measure zero. Thus there arises the question whether there exist a version of (2.2) which can be interpreted pathwise for all $y(t)$ and for which the solutions of (2.2) for piecewise smooth $(y(t)$ carry (approximative) information, cf. Clark [9] and Davis [11]. Fortunatedly the time dependent gauge transformation

$$\tilde{\rho}(x,t) = \exp(-h_1(x)y_1(t) - \cdots - h_p(x)y_p(t))\rho(t,x) \tag{3.1}$$

transforms (2.2) into an equation

$$\frac{\partial \tilde{\rho}}{\partial t} = \mathcal{L}\tilde{\rho} - \sum_{i=1}^{p} y_i(t)\mathcal{L}_i\tilde{\rho} + \sum_{i,j=1}^{p} y_i(t)y_j(t)\mathcal{L}_{ij}\tilde{\rho} \tag{3.2}$$

where $\mathcal{L}_i = [h_i, \mathcal{L}] := h_i\mathcal{L} - \mathcal{L}h_i$ and $\mathcal{L}_{ij} = \mathcal{L}_{ji} = \frac{1}{2}[h_i, [h_j, \mathcal{L}]]$, and this equation, which does not anymore involve derivatives of y, can simply be interpreted as a family of partial differential equations parametrized by the possible observation paths $y(t)$.

Equation (3.2) can of course be verified directly (remembering that (2.2) is a Fisk-Stratonovic integral so that the ordinary rules of calculus apply; removing the term $-\frac{1}{2}\sum h_i^2$ from $\mathcal{L}$ gives the corresponding Ito equation and then Ito calculus of course also gives (3.2). An easier way of obtaining (3.2) is to observe that (3.1) in (2.2) gives $d\tilde{\rho} = \exp(-\sum h_i y_i)\mathcal{L} \exp(\sum h_i y_i)\tilde{\rho}$ and to use the version of the Baker-Campbell-Hausdorff formula which says

$$\exp(-rA)B\exp(rA) = \sum_{k=0}^{\infty}(-1)^k\frac{r^k}{k!}ad_A^k(B) \tag{3.3}$$

where $ad_A(B) = [A,B] = AB - BA$, $ad_A^k(B) = ad_A(ad_A^{k-1}(B))$ for linear operators A,B. In our case the contributions of (3.3) for $k \geq 2$ disappear because then A is a function, $B = \mathcal{L}$ is a second order differential operator, so $[A,B]$ is first order, $[A,[A,B]]$ is a function and $[A,[A,[A,B]]] = 0$.

Also of course there still remains the question of how to use equation (3.2) or (2.2) effectively to calculate certain desired conditional expectations. A direct numerical discretization approach is out of the question. Typically x is a fairly large dimensional object; for example around 27 for certain problems involving helicopters. Taking three data points per coordinate axis (which is ridiculous) then gives $3^{27} \approx 2.10^{14}$ space grid points! So other methods must be tried. It seems likely that the Lie-algebraic considerations to be discussed below will help. Other promising work into the numerics of the nonlinear filtering equations has been started by Pardoux-Talay [29].

4. WEI-NORMAN THEORY

It is important to note that the filtering equation (3.2) (or (2.2)) is of the general form

$$\dot{x} = (A_1x)u_1 + \cdots + (A_kx)u_k \tag{4.1}$$

where the A_i are linear operators and the u_i known functions of time. Of course in (3.2) the role of x is played by ρ, an infinite dimensional object. Here for the moment lets consider (4.1) as a finite dimensional object. Let us also assume that the $A_1,...,A_k$ who are now, say, $n \times n$ matrices, form the basis of a Lie algebra. (By adding a few more terms with corresponding u_i equal to zero this can of

course always be assured.) Let us look for solutions of the form (Wei-Norman [35]).

$$x(t) = e^{g_1 A_1 g_2 A_2}...e^{g_k A_k}x(0) \qquad (4.2)$$

Differentiating this gives

$$\dot{x} = \dot{g}_1 A_1 e^{g_1 A_1}e^{g_2 A_2}...e^{g_k A_k}x(0) + e^{g_1 A_1}\dot{g}_2 A_2 e^{g_2 A_2}...e^{g_k A_k}x(0) + ... \qquad (4.3)$$

and inserting

$$e^{-g_1 A_1}e^{-g_2 A_2}...e^{-g_i A_i}e^{g_i A_i}...e^{g_1 A_1}$$

just after $\dot{g}_i A_i$ in the i-th term equation (1.1) can be rewritten

$$\dot{x} = \sum_{i=1}^{k}\dot{g}_i(A_i + \sum_{r=1}^{i-1}\sum_{\substack{j_1,...,j_{i-1}\\ j_1+\cdots+j_{i-1}>0}}\frac{g_1^{j_1}...g_{i-1}^{j_{i-1}}}{j_1!\cdots j_{i-1}!}ad_{A_1}^{j_1}...ad_{A_{i-1}}^{j_{i-1}}(A_i))x \qquad (4.4)$$

$$= \sum_{i=1}^{k}\dot{g}_i(A_i + h_{ij}(g_1,...,g_k)A_j)$$

with $h_{ij}(0,...,0) = 0$, where, again, the Campbell-Baker-Hausdorff formula (3.3) has been used. Note that h_{ij} are universal functions which only depend on the Lie algebra and the chosen basis. Thus it remains to solve (equating the coefficients of the basic elements A_i in (4.4) and (4.1))

$$\dot{g}_1 + \dot{g}_2 h_{11}(g_1,...,g_k) + \dot{g}_2 h_{21}(g_1,...,g_k) + ... + \dot{g}_k h_{k1}(g_1,...,g_k) = u_1$$

$$\dot{g}_2 + \dot{g}_1 h_{12}(g_1,...,g_k) + \dot{g}_2 h_{22}(g_1,...,g_k) + ... + \dot{g}_k h_{k2}(g_1,...,g_k) = u_2 \qquad (4.5)$$

$$...$$

$$\dot{g}_k + \dot{g}_1 h_{1k}(g_1,...,g_k) + \dot{g}_2 h_{2k}(g_1,...,g_k) + ... + \dot{g}_k h_{kk}(g_1,...,g_k) = u_k$$

which can be done for small t and $g_1(0) = ... = g_k(0) = 0$ because $h_{ij}(0,...,0) = 0$. In general a representation (4.2) for the solution is only possible for small t. However things change if the Lie-algebra in question is solvable, then ([35]) there is such a representation for all t. More precisely there is a suitable basis such that there is such a representation for all t. How this comes about is easy to see in the case that the Lie algebra L is nilpotent. Indeed let

$$L \supsetneq L^{(1)} = [L,L] \supsetneq L^{(2)} = [L,L^{(1)}] \supsetneq ... \supsetneq L^{(m)} = [L,L^{(m-1)}] = 0 \qquad (4.6)$$

be a basis such that $A_1,...,A_{k_1}$, $A_{k_1+1},...,$ $A_{k_2},...,A_{k_{m-1}+1},...,$ $A_{k_m} = A_k$, $k_1<k_2<...<k_m$ such that $A_{k_i+1},...,A_{k_m}$ is a basis for $L^{(i)}$, $i=0,...,m-1$ ($k_0=1,k_m=k$). Then it immediately follows from (4.4) that $h_{ij} = 0$ for $j\leq i$ and the set of equations (4.5) gets a nice triangular structure. Moreover $h_{ij}(g_1,...,g_k)$ involves only $g_1,...,g_{i-1}$ (this is always the case, cf.(4.5), so the h_{1j} in (4.5) are always all zero) and the resulting equations (4.5) for the nilpotent case are therefore of the form

$$\dot{g}_1 = u_1,...,\dot{g}_{k_1} = u_{k_1}$$

$$\dot{g}_{k_1+1} = u_{k_1+1}+\alpha_{k_1+1}(u_1,...,u_{k_1};g_1,...,g_{k_1}),...,\dot{g}_{k_2} = u_{k_2}+\alpha_{k_2}(u_1,...,u_{k_1};g_1,...,g_{k_1}) \qquad (4.7)$$

$$\dot{g}_{k_2+1} = u_{k_2+1}+\alpha_{k_2+1}(u_1,...,u_{k_2};g_1,...,g_{k_2}),...,\dot{g}_{k_3} = u_{k_3}+\alpha_{k_3}(u_1,...,u_{k_2};g_1,...,g_{k_2})$$

$$...$$

where the α_j are known (universal) functions of the u's and g's.

These considerations are not limited to Lie-algebras of matrices. Indeed the left hand sides of equations (4.5) only depend on the abstract structure of the Lie algebra in question and the choice of basis. Thus all this equally applies to Lie-algebras of say differential operators (given suitable definitions of exp(tA)), though in order to have a finite set of equations (4.5) one needs of course a

finite dimensional algebra. It also follows from (4.4) that the Wei-Norman equations are compatible with homomorphisms of Lie algebras, more precisely quotients. Indeed if $\mathfrak{A} \subset L$ is an ideal and $A_1,...,A_{k_1},A_{k_{i+1}},...,A_k$ is a basis of L such that $A_{k_1+1},...,A_k$ is a basis of $\mathfrak{A}$ then the h_{ij} are zero for $j \in \{1,...,k_1\}$ and $i \in \{k_1+1,...,k\}$. So in the case of a topologically nilpotent algebra L, or more generally one with a chain of ideals $\mathfrak{A}_1 \supset \mathfrak{A}_2 \supset \mathfrak{A}_3 \supset ...$ such that $\cap \mathfrak{A}_i = 0$ and $L / \mathfrak{A}_i$ finite dimensional for all i one can in principle still do Wei-Norman theory with now infinite ordered product expressions $x = e^{g_1 A_1} e^{g_2 A_2}...e^{g_k A_k}...x_0$ in the sense that the equations for the g_i belonging to a quotient $L / \mathfrak{A}_j$ involve only those same g_i. Of course now questions of convergence arise.

5. THE ESTIMATION LIE ALGEBRA

The considerations of the previous section already make it clear that the Lie algebra generated by the operators which occur in equation (2.2) or (3.2) contains important information concerning the filtering problem. One therefore defines the estimation Lie algebra $EL(\Sigma)$ of a system Σ given by (1.1)-(1.2) as the Lie algebra of differential operators generated by the 2-nd order differential operator $\mathfrak{L}$ and the multiplication operators $h_1,...,h_p$.

$$EL(\Sigma) = Lie(\mathfrak{L},h_1,...,h_p).\tag{5.1}$$

Note that the Lie algebra generated by the operators which occur in (3.2) is in any case a subalgebra of $EL(\Sigma)$. Often it is equal.

Example 5.2. Consider again the simplest nonzero linear system (2.5). Then $p=1$ and $\mathfrak{L}=\frac{1}{2}d^2 / dx^2 - \frac{1}{2}x^2$. So we have in this case the Lie algebra $Lie(\frac{1}{2}d^2 / dx^2 - \frac{1}{2}x^2, x)$. Now $[\frac{1}{2}d^2 / dx^2 - \frac{1}{2}x^2, x] = d / dx$ (as operators on functions), $[\frac{1}{2}d^2 / dx^2 - \frac{1}{2}x^2, \frac{d}{dx}] = x$, $[\frac{d}{dx}, x] = 1$ and $[?,1] = 0$. So in this case we obtain the well-known oscillator Lie algebra, which is four dimensional with basis $\frac{1}{2}d^2 / dx^2 - \frac{1}{2}x^2$, x, d / dx, 1. It is solvable (but not nilpotent) with as derived algebra the nilpotent Heisenberg algebra with basis x, d / dx, 1.

$EL(\Sigma)$ is (of course) an invariant of Σ meaning that a change of coordinates in Σ (a diffeomorphism $x \to x'$ taking Σ to Σ') will yield isomorphic estimation Lie algebras. The algebra also has a gauge transformation invariance. A gauge transformation $\rho(x,t) \to \psi(x)\rho(x,t)$, where $\psi(x) \neq 0$ for all x, transforms the DMZ-equation in such a way that the operators in the new equation generate an isomorphic Lie algebra.

The new equation may again have the form of a DMZ-equation, and in this way systems which are definitely not equivalent as systems may have equivalent filtering problems associated to them. An example are the 1-dimensional Benes systems (cf. various contributions in [19]).

In a way which will (hopefully) become clearer below the estimation Lie algebra $EL(\Sigma)$ encodes information about how difficult the filtering problem for Σ is. For example if it is finite dimensional (a very rare case) Wei-Norman theory does the job for small time; if it is also solvable one thus gets a filter. If it is infinite dimensional but solvable things become more difficult but asymptotic expansions are possible, cf. below; etc.

6. THE BC PRINCIPLE

Let me now describe a second reason why the Lie algebra $EL(\Sigma)$ of a system Σ is important for filtering problems. I like to call it the *BC principle*, not because it is very old, though it could have been maybe, nor is it named after Johny Hart's chartoon character; the BC stand for Brockett and Clark [6] who first enunciated it.

Suppose we have a filter (1.10)-(1.11) on a finite dimensional manifold M for a statistic $\widehat{\phi(x_t)}$. We may as well assume that it is minimal, i.e. has minimal $dim(M)$. The α and $\beta_1,...,\beta_p$ in (1.10) are vectorfields on M. Let $V(M)$ denote the Lie algebra of smooth vectorfields on M. Then the BC

principle states the following

6.1. BC Principle

If (1.10)-(1.11) is a minimal filter for a statistic then $\ell \mapsto \alpha$, $h_1 \mapsto \beta_1, \ldots, h_p \mapsto \beta_p$ defines an antihomomorphism of Lie algebras from $EL(\Sigma)$ into $V(M)$.

Here "anti" means the following: if $\phi: L_1 \to L_2$ is a map of vectorspaces from the Lie-algebra L_1 to the Lie-algebra L_2, it is called an antihomomorphism of Lie-algebras if $\phi([A,B]) = -[\phi(A),\phi(B)]$ for all $A,B \in L_1$.

Example 6.2. Consider again the simplest nonzero linear system (2.5). It is linear so there is the Kalman-Bucy filter for the conditional state $\hat{x}$. This filter is

$$dP_t = (1-P_t^2)dt, \quad dm_t = P_t(dy_t - m_t dt). \tag{6.3}$$

So the two vectorfields α and β of the filter are respectively

$$\alpha = (1-P^2)\frac{\partial}{\partial P} - Pm\frac{\partial}{\partial m}, \quad \beta = P\frac{\partial}{\partial m}. \tag{6.4}$$

A simple calculation shows $[\alpha,\beta] = \frac{\partial}{\partial m}$, and it is now indeed a simple exercise to show that $\frac{1}{2}\frac{d^2}{dx^2} - \frac{1}{2}x^2 \mapsto \alpha$, $x \mapsto \beta$, induces an antimorphism of Lie-algebras. (It also induces a homomorphism, but that is an accident which happens for linear systems (1.12)-(1.13) if the drift term Ax is absent).

A feeling of why the BC principle should be true can be generated as follows. Think for the moment of two automata with given initial state and with outputs (Moore automata), which, when fed the same string of input data, produce exactly the same string of output data. Suppose the second automaton is minimal. Then it is wellknown (and easy to prove by constructing the minimal automaton from the input-output data) that there is a homomorphism of the subautomaton of the first consisting of the states reachable from the initial state to the second automaton; this homomorphism so to speak makes visible that the two machines do the same job. A similar theorem holds for initialized finite dimensional systems [Sussmann [34]), in particular for systems of the form

$$\dot{x} = \alpha(x) + \sum_{i=1}^{m}\beta_i(x)u_i, \quad y = \gamma(x) \tag{6.5}$$

Here the picture produced by theorem is the following commutative diagram
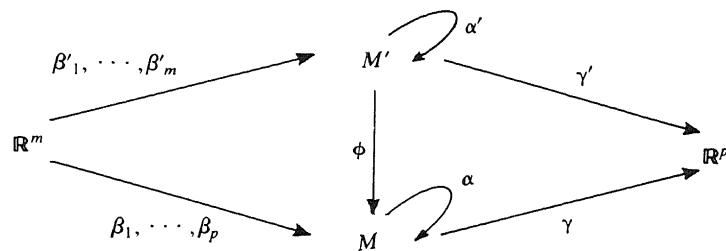
(The theorem asserts the existence of a differentiable map ϕ defined on the reachable from x'_0 subset of M' which makes the diagram commutative. This in particular implies that $d\phi$ takes the vectorfields

$\alpha', \beta'_1, ..., \beta'_m$ into $\alpha, \beta_1, ..., \beta_m$ respectively, and, ϕ being a differentiable map, $d\phi$ induces a homomorphism from the Lie algebra generated by $\alpha', \beta'_1, ..., \beta'_m$ to $V(M)$.

In the case of the BC principle we also have two "machines" which do the same job: one is the postulated minimal filter, the other is the infinite dimensional machinge given by the DMZ-equation (2.2) and the ouput map (2.4). So we are in a similar situation as above but with M' infinite dimensional. A proof in this case follows from considerations of Hijab [20].

The fact that in the case of the BC-principle we get an antihomomorphism arises from the following. Given a linear space V and an operator A on it we can define a (linear) vectorfield on V by assigning to $v \in V$ the tangent vector Av. (So we are considering the equation $\dot{v} = Av$.) This defines an anti-isomorphism of the Lie algebra of operators on V to the Lie algebra of linear vectorfields on V.

What about a converse to the BC principle? I.e. suppose that we have given an antihomomorphism of Lie-algebras $EL(\Sigma) \to V(M)$ into the vectorfields of some finite dimensional manifold. Does there correspond a filter for some statistic of Σ. Just having the homomorphism is clearly insufficient. There are also explicit counterexamples. This is understandable for in any case we completely ignored the output aspect when making the BC-principle plausible. This is not trivial contrary to what the diagram above may suggest. It is not true that given ϕ and any γ one can take $\gamma' = \gamma \circ \phi$. The problem is that γ' as a function on M' = space of unnormalized densities is of a very specific type cf. (2.4).

Even apart from that things are not guaranteed. What we need of course is a ϕ making the left half of the diagram above commutative. Then, if $m' \in M'$ is going to be mapped on $m \in M$, obviously the isotropy subalgebra of $EL(\Sigma)$ at m' will go into the isotropy subalgebra of $V(M)$ at m.

For the case of finite dynamical systems there are positive results of Krener [21] stating that in such a case this extra condition is also sufficient to guarantee the existence of ϕ locally.

The whole clearly relates to seeing to what extend a manifold can be recovered from its Lie algebra of vectorfields (via its maximal subalgebras of finite codimension) and whether differentiable maps can be recovered from the map between Lie-algebras they induce. This question has been examined by Pursell-Shanks [30].

A more representation theoretic way of looking at things is as follows. Both $EL(\Sigma)$ and $V(M)$ come with a natural representation on the space of functionals on M' and the space of functions on M respectively. If there were a ϕ as in the diagram above ϕ would also induce a map between these representation spaces compatible with the homomorphism of Lie algebras. That therefore is clearly a necessary condition. This way of looking at things contains the isotropy subalgebra condition and also contains output function aspects. Thus the total picture regarding a converse to the BC-principle is not unpromising but nothing is established.

Except for one quite positive aspect. If $EL(\Sigma)$ is finite dimensional, the Wei-Norman equations practically define the filter, for small time in the general case, for arbitrary time in the solvable case.

7. EXAMPLES OF ESTIMATION ALGEBRAS

7.1. The cubic sensor
This is the one dimensional system

$$dx_t = dw_t, \quad dy_t = x_t^3 dt + dv_t \tag{7.2}$$

and it is about the simplext nonlinear system imaginable. Its estimation Lie algebra is generated by $\frac{1}{2}\frac{d^2}{dx^2} - \frac{1}{2}x^6$, x^3.

Theorem 7.3. (Hazewinkel-Marcus [17]). EL(cubic sensor) $= W_1$, where $W_1 = \mathbb{R}<x, \frac{d}{dx}>$ is the Lie algebra of the differential operators (any order, zero included) with polynomial coefficients.

Example 7.4. $dx_1 = dw, dx_2 = x_1^2 dt$; $dy_1 = x_1 dt + dv_1, dy_2 = x_2 dt + dv_2$. In this case the estimation Lie algebra is $W_2 = \mathbb{R} < x_1, x_2, \dfrac{\partial}{\partial x_1}, \dfrac{\partial}{\partial x_2} >$, the Lie algebra of all differential operators in two variables with polynomial coefficients.

Example 7.5. $dx_1 = dw, dx_2 = x_1^2 dt, dy = x_1 dt + dv_1$. In this case the estimation Lie algebra has a basis A, B_i, C_i, D_i $i = 1, 2, ...$ with the commutation relations $[A, B_i] = C_i$, $[A, C_i] = B_{i+1} + 2B_{i+1}$, $[B_i, C_j] = -D_{i+j}$ and all other commutation relations between basis elements are zero. Note that in this case the Lie algebra is infinite dimensional but has many ideals $\mathfrak{A}_i$ such that $L / \mathfrak{A}_i$ is finite dimensional.

Example 7.6. $dx = dw_1 + \epsilon x dw_2, dy = xdt + dv$. Here again $EL = W_1$.

It has become clear that as a rule estimation algebras tend to be infinite dimensional (except in the linear case: then EL(linear system) has dimension $2n + 2$ if the linear system is completely reachable and observable); it has also become noticeable that the Weyl-Heisenberg algebras or Weyl algebras W_n have a tendency to appear very often.

Conjecture 7.7. Consider systems (1.1)-(1.2) with polynomial f, G, h. Then generically, i.e. for almost all f, G, h, the estimation algebra will be W_n.

8. THE SEGAL-SHALE-WEIL REPRESENTATION AND ALL KALMAN-BUCY FILTERS

8.1. The linear systems Lie-algebra ls_n
Consider all differential operators in n indeterminates with polynomial coefficients

$$D = \Sigma c_{\alpha\beta} x^\alpha \frac{\partial^\beta}{\partial x^\beta} \tag{8.2}$$

where $\alpha = (\alpha_1, ..., \alpha_n)$, $\beta = (\beta_1, ..., \beta_n)$ are multiindices $\alpha_i, \beta_j \in \mathbb{N} \cup \{0\}$. Consider those D which are of total degree ≤ 2; i.e. such that $|\alpha| + |\beta| > 2 \Rightarrow c_{\alpha\beta} = 0$ where $|\alpha| = \alpha_1 + \cdots + \alpha_n$. As is readily verified these form a finite dimensional Lie algebra (under the commutator product $[D_1, D_2] = D_1 D_2 - D_2 D_1$) of dimension $2n^2 + 3n + 1$. A basis is

$$1; x_1, ..., x_n; \frac{\partial}{\partial x_1}, ..., \frac{\partial}{\partial x_n}; \frac{\partial}{\partial x_i \partial x_j}, i,j = 1, ..., n; x_i x_j, i,j = 1, ..., n; x_i \frac{\partial}{\partial x_j}, i,j = 1, ..., n. \tag{8.3}$$

The operators of total degree ≤ 1 form a subalgebra h_n (basis: $1; x_1, ..., x_n; \partial / \partial x_1, ..., \partial / \partial x_n$) which is in fact an ideal. The quotient is isomorphic to the symplectic algebra sp_n of all real $2n \times 2n$ matrices M such that

$$MJ + JM^T = 0, \quad J = \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}. \tag{8.4}$$

The isomorphism is given by

$$E_{i,n+j} + E_{j,n+i} \mapsto x_i x_j; \quad E_{n+i,j} - E_{n+j,i} \mapsto \frac{\partial^2}{\partial x_i \partial x_j}; \tag{8.5}$$

$$E_{i,j} - E_{n+j,n+i} \mapsto x_i \frac{\partial}{\partial x_j} + \frac{1}{2}\delta_{ij}; \quad i,j = 1, ..., n.$$

Here $E_{i,j}$ is the matrix with a 1 at spot (i,j) and 0 everywhere else; these linear combinations of the $E_{i,j}$ form a basis of sp_n; this isomorphism exhibits sp_n as a subalgebra complementary to h_n; i.e. as a Levi-factor for the short exact sequence $0 \rightarrow h_n \rightarrow ls_n \rightarrow sp_n \rightarrow 0$).

8.2. The oscillator representation

There is a famous representation of sp_n which occurs in the framework of symmetries of boson fields (Shale, Segal), in algebraic number theory (Weil), and a multitude of other places, known variously as the Segal-Shale-Weil representation or the oscillator representation. One way to obtain it is as follows. Let H_n denote the Heisenberg group, $H_n = \mathbb{R}^n \times \mathbb{R}^n \times S^1$ with multiplication

$$(x,y,z)(x',y',z') = (x+x', y+y', e^{-2\pi i <x,y'>} zz') \tag{8.7}$$

where $<\ ,\ >$ denotes the standard scalar product on $\mathbb{R}^n$. The Lie algebra of H_n is of course $h_n = \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$. And the Lie-bracket of h_n can be interpreted as giving (and given by) a bilinear form $\mathbb{R}^{2n} \times \mathbb{R}^{2n} \to \mathbb{R}$ defined by the matrix J, cf. (8.4) above. Thus the Lie group Sp_n of sp_n can be seen as a group of automorphism of h_n and H_n which is moreover the identity on the centre $S^1 \subset H_n$. Let ρ be the standard Schrödinger representation of H_n in $L^2(\mathbb{R}^n)$

$$(x,0,0) \to M_x, \quad (M_x f)(x') = e^{2\pi i <x,x'>} f(x'), \quad f \in L^2(\mathbb{R}^n)$$

$$(0,y,0) \to T_y, \quad (T_y f)(x') = f(x'-y), \quad f \in L^2(\mathbb{R}^n) \tag{8.8}$$

$$(0,0,z) \to S_z, \quad (S_z f)(x') = z f(x'), \quad f \in L^2(\mathbb{R}^n)$$

Now let $g \in Sp_n$ be seen as a group of automorphisms of H_n. Then $h \mapsto \rho(g(h))$ is another irreducible representation of H_n with the same central character. So by the Stone-von Neumann theorem there is an $\omega(g)$ intertwining them, i.e. such that

$$\omega(g)\rho(h)\omega(g)^{-1} = \rho(g(h)). \tag{8.9}$$

These $\omega(g)$ are unique up to scalar factors and therefore define a projective representation of Sp_n. The factors can be fixed up to define a representation of the two-fold covering $\widetilde{Sp}_n$ of Sp_n. This is the Segal-Shale-Weil representation.

8.3. All Kalman-Bucy filters

Now consider something apparently totally unrelated, namely the DMZ-filtering-equation (2.2) for a linear dynamical system

$$dx = Ax dt + B dw, \quad dy = Cx dt + dv, \quad x \in \mathbb{R}^n, \ w \in \mathbb{R}^m, \ y,v \in \mathbb{R}^p. \tag{8.11}$$

It is a trivial remark that the operators occurring in (2.2) are all in ls_n in this case. And in fact the Lie algebra generated by them will consist of the second order operator $\mathcal{L}$ and a subalgebra of h_n stable under $\mathcal{L}$. In most cases, to be precise in the case that the system (A,B,C) is completely reachable and completely observable, this will be all of h_n, giving us generically an estimation algebra of dimension $2n+2$ which is a subalgebra of ls_n, which has dimension $2n^2+3n+1$.

The Kalman-Bucy filter defines by BC principle an antihomomorphism of this Estimation Lie algebra $EL(A,B,C)$ into the vector Lie algebra of vector fields $V(\mathbb{R}^N)$, $N = n + \frac{1}{2}n(n+1)$.

Theorem [16] 8.12. For varying (A,B,C) these anti-homomorphisms fit together to define a antihomomorphism of all of ls_n into $V(\mathbb{R}^N)$ with as kernel the centre $\mathbb{R}1$. This representation can be lifted to one on $V(\mathbb{R}^{N+1})$ which is faithful.

The explicit formulas are as follows. Interpret a point $x \in \mathbb{R}^{N+1}$ as a triple $x = (c,m,P)$ consisting of a scalar c, an n-vector m, and a symmetric $n \times n$ matrix P. The antihomomorphism is then given by

$$1 \to \frac{\partial}{\partial c} \tag{8.13}$$

$$x \to m_i \frac{\partial}{\partial c} + \sum_i P_{ii} \frac{\partial}{\partial m_i} \tag{8.14}$$

$$\frac{\partial}{\partial x_i} \to -\frac{\partial}{\partial x_i} \tag{8.15}$$

$$x_i x_j \to (m_i m_j + P_{ij})\frac{\partial}{\partial c} + \sum_t (m_i P_{jt} + m_j P_{it})\frac{\partial}{\partial m_t}$$

$$+ \sum_{s,t} P_{is} P_{jt}\frac{\partial}{\partial P_{st}} + \sum_t P_{it} P_{jt}\frac{\partial}{\partial P_{tt}}$$ (8.16)

$$x_i \frac{\partial}{\partial x_j} \to -m_i \frac{\partial}{\partial m_j} - \delta_{ij}\frac{\partial}{\partial c} - P_{ij}\frac{\partial}{\partial P_{jj}} - \sum_t P_{it}\frac{\partial}{\partial P_{jt}}$$ (8.17)

$$\frac{\partial^2}{\partial x_i \partial x_j} \to \frac{\partial}{\partial P_{ij}} \text{ if } i \neq j, \quad \frac{\partial^2}{\partial x_i^2} \to 2\frac{\partial}{\partial P_{ii}}.$$ (8.18)

Inversely these formulas can be checked directly to give an antihomomorphism of Lie algebras and this thus verifies the BC principle for linear dynamical systems and also for families of such depending on a parameter.

Changing all the minus signs in (8.17) and (8.15) into plus signs gives a faithful homomorphism of ls_n into $V(\mathbb{R}^{N+1})$.

Restricting this homomorphism to sp_n as given by (8.5) then defines a homomorphism of sp_n into $V(\mathbb{R}^{N+1})$.

The final remark is that this realization of sp_n as a Lie algebra of vector fields on $\mathbb{R}^{N+1}$ has much to do with the Segal-Shale-Weil representation. The precise statement is that the mapping

$$(c,m,p) \to \exp(c + <2\pi i m, x> - 2\pi^2 P(x)) \in L^2(\mathbb{R}^n)$$ (8.19)

(where $P(x)$ is the quadratic form defined by the symmetric matrix P; note that apart from a scaling factor this is the normal distribution with mean m and covariance P) linearizes the vectorfields in the image of) ls_n in $V(\mathbb{R}^{N+1})$ and switching from a linear vectorfield to the operator which defines it then defines a representation of $ls_n \supset sp_n$. This is another real form of the Segal-Shale-Weil representation meaning that after tensoring with $\mathbb{C}$ (=extending scalars to the complexes), they become isomorphic.

9. W_n AND $V(M)$

We have seen that the Weyl-Heisenberg algebra $W_n = \mathbb{R}<x_1,...,x_n; \partial/\partial x_1,...,\partial/\partial x_n>$ of all differential operators with polynomial coefficients often occurs in filtering problems, i.e. as an Estimation Lie algebra. Given the BC-principle it is therefore of interest to know something about its relations with another class of infinite dimensional Lie algebras, viz the Lie algebras $V(M)$ of smooth vectorfields on a finite dimensional manifold. The algebra W_n has a one-dimensional centre $\mathbb{R}$. 1 consisting of the scalar multiples of the identity operator.

Theorem 9.1. (Hazewinkel-Marcus [17]). Let α: $W_n \to V(M)$ or $W_n / \mathbb{R}.1 \to V(M)$ be a homomorphism or antihomomorphism of Lie algebras, where M is a finite dimensional manifold. Then $\alpha = 0$.

The original proof of this result ([17]) was long and computational. Another much shorter proof based on the nonexistence of finite dimensional representations of h_n for which 1 gets mapped onto the unit operator has more recently been given by Toby Stafford.

10. THE CUBIC SENSOR

Consider again the cubic sensor, i.e. the one-dimensional system

$$dx = dw, \quad dy = x^3 dt + dv$$ (10.1)

consisting of Wiener noise, cubically observed with further independent noise corrupting the observations. As noted before (theorem 7.3)

$$EL(\text{cubic sensor}) = W_1.$$ (10.2)

Now suppose that we have a finite dimensional filter for some conditional statistic $\phi(\hat{x}_t)$ of the cubic sensor. By the BC-principle (6.1) if follows that there is an antihomomorphism of Lie-algebras $W_1 = EL$ (cubic sensor) $\xrightarrow{\alpha}$ $V(M)$. By theorem 9.1 it follows that $\alpha = 0$ and from this it is not hard to see that the only statistics of the cubic sensor for which there exists a finite dimensional exact recursive filter are the constants.

A direct proof of this, which sort of proves the BC-principle in this particular case along the way, is contained in Hazewinkel-Marcus-Sussmann [18].

11. PERTURBATIONS AND APPROXIMATIONS

Let us start with an example. Consider the weak cubic sensor

$$dx = dw, \quad dy = xdt + \epsilon x^3 dt + dv \quad (\Sigma_\epsilon) \tag{11.1}$$

If $\epsilon = 0$, this the simplest nontrivial linear system for which there is the Kalman filter. For $\epsilon \neq 0$ one can prove that $EL(\Sigma_\epsilon) = W_1$ again, [15]. So for all $\epsilon \neq 0$ there is no recursive exact filter for any nonconstant statistic. Yet it is hard to believe that for small ϵ the Kalman-Bucy filter would not do a good job of first approximation. A question thus arises whether the estimation Lie algebra also has things to say about approximate filters. In this section and the following ones I shall argue that it does.

The first observation is as follows. If one actually commutes the two generators $\frac{1}{2} d^2 / dx^2 - \frac{1}{2}(x + \epsilon x^3)^2$, $(x + \epsilon x^3)$ repeatedly of course eventually all the basis elements of W_1 appear. But they appear with higher and higher powers of ϵ and the ϵ-degree grows faster than the degree in (α) and (β) of the $x^\alpha \partial^\beta / \partial x^\beta$.

A precise version of this is as follows. Consider the two generators just listed as operators over the ring $\mathbb{R}[\epsilon]$ (or $\mathbb{R}[[\epsilon]]$), i.e. consider ϵ as an extra variable. Then it makes sense to consider

$$EL(\Sigma_\epsilon) \otimes_{\mathbb{R}[\epsilon]} \mathbb{R}[\epsilon] / \epsilon^n =: EL(\Sigma_\epsilon) \bmod(\epsilon^n). \tag{11.2}$$

This simply amounts to setting $\epsilon^m = 0$ for $m \geq n$ whenever it appears. The set of all $\epsilon^i x^j d^k / dx^k$ with $\epsilon \geq n$ form an ideal in $\mathbb{R}[\epsilon] < x, d / dx >$, so this makes sense. Now observe

Proposition 11.3. [15]. The Lie-algebras $EL(\Sigma_\epsilon)$ mod ϵ^n are finite-dimensional for all n.

As an example $EL(\Sigma_\epsilon)$ mod ϵ^2 turns out to be 14 dimensional with basis

$$\frac{1}{2} \frac{d^2}{dx^2} - \frac{1}{2} x^2 - \epsilon x^4, \ x, \ \epsilon x^3, \ \frac{d}{dx}, \ 1, \ \epsilon, \ \epsilon x^2 \frac{d}{dx}, \ \epsilon x,$$

$$\epsilon x \frac{d}{dx}, \ \epsilon \frac{d^2}{dx^2}, \ \epsilon \frac{d}{dx}, \ \epsilon \frac{d^3}{dx^3}, \ \epsilon x \frac{d^2}{dx^2}, \ \epsilon x^2.$$

This is a general phenomenon.

Theorem 11.4. [15]. Let Σ_ϵ be a system of the form

$$dx = (Ax + \epsilon P_A(x))dt + (B + \epsilon P_B(x))dw, \quad dy = (Cx + \epsilon P_C(x))dt + dv \tag{11.5}$$

where P_A, P_B, P_C are polynomial vector and matrix valued functions of the approximate dimensions. Then $EL(\Sigma_\epsilon)$ mod (ϵ^n) is finite dimensional for all n. It is also solvable.

In [15] this is proved for the case $P_B = P_A = 0$. The proof generalizes immediately (simply give ϵ a negative enough degree to make degree decreasing all terms in the generators of $EL(\Sigma_\epsilon)$ in which ϵ appears (both $x_i, \partial / \partial x_j$ are given degree 1 in this argument).)

The next obvious question is: do these "finite dimensional quotients of $EL(\Sigma_\epsilon)$" actually compute

anything, do they correspond to filters for some statistic? In the case of the weak cubic sensor this is (11.1) easy to answer. Consider the unnormalized conditional density $\rho(x,t,\epsilon)$ and (formally) expand it as a power series in ϵ

$$\rho(x,t,\epsilon) = \rho_0(x,t) + \epsilon\rho_1(x,t) + \epsilon^2\rho_2(x,t) + ... \tag{11.6}$$

Then $EL(\Sigma_\epsilon)$ mod (ϵ^n) corresponds to the first n coefficients $\rho_0(x,t),...,\rho_{n-1}(x,t)$, and via Wei-Norman theory actually computes them. This is generally true, also in the setting of theorem 11.4. In the case of the weak cubic sensor (11.6) actually converges (for small ϵ). That, it appears, is not generally true. But it is still true that (11.6) gives an asymptotic expansion (Blankenschip-Liu-Marcus [4]). The Lie algebras being solvable one can of course implement these approximate filters, using the Wei-Norman technique. This was done in [4] and also the results were compared with the extended Kalman filter (EKF). The zero-th order approximation (of course) performed worse than EKF but the first order approximation performed better!

These Lie algebras $EL(\Sigma_\epsilon)$ tend to become large rapidly and to actually produce the, say FOR-TRAN, code is a long, but mechanical, job, prone to errors. Even the simplest nontrivial case needs several pages of densely written code. It is thus natural to try to let the computer do the job itself and in this way these ideas and techniques are being implemented in an expert system which is a joint effort of INRIA and the Department of Electrical Engineering of the University of Maryland (cf. Blankenschip [3]; the system also contains many other facets of stochastic control, filtering and optimization).

From the point of view developed in section 4 above the fact that calculating $\rho_0(x,t),...,\rho_{n-1}(x,t)$ corresponds to $EL(\Sigma_\epsilon)$ mod (ϵ^n) can be understood as follows. Choosing a basis suitably the remarks made in section 4 about the compatibility of Wei-Norman theory with quotients say that $\rho(x,t,\epsilon)$ admits an "expansion"

$$\rho(x,t,\epsilon) = e^{g_1(t)A_1}e^{g_2(t)A_2}...e^{g_n(t)A_n}...\rho_0(x) \tag{11.7}$$

with $g_1,...,g_{m(n)}$ where $m(n) = \dim(EL(\Sigma_\epsilon)$ mod (ϵ^n)) depending only on $EL(\Sigma_\epsilon)$ mod (ϵ^n). The operators A_i in (11.7) involve higher and higher powers of ϵ. Writing out the exponentials one recovers (11.6). (And this point of view also strongly suggests (because also higher derivatives appear in the A_i) that the best one can hope for in general is an asymptotic expansion.)

12. THE PROFINITE DIMENSIONAL CASE

A Lie-algebra L is said to be profinite dimensional if there is a sequence of ideals $L \supset \mathfrak{A}_1 \supset \mathfrak{A}_2 \supset \mathfrak{A}_3 \supset ...$ such that

$$\dim L / \mathfrak{A}_i < \infty \text{ for all } i \tag{12.1}$$

$$\bigcap_i \mathfrak{A}_i = \{0\}. \tag{12.2}$$

Suppose the estimation Lie algebra $EL(\Sigma)$ has this property. Then again, as in the previous section, one can write an expansion

$$\rho(x,t) = e^{g_1A_1}e^{g_2A_2}...e^{g_nA_n}...\rho_0(x) \tag{12.3}$$

and consider the possible approximants

$$\rho^{(n)}(x,t) = e^{g_1A_1}e^{g_2A_2}...e^{g_nA_n}\rho_0(x). \tag{12.4}$$

Using, again, that the equations for $g_1,...,g_n$; $n = n(m) = \dim L / \mathfrak{A}_m$ do not depend on $g_{n+1},...$. Abstractly, there is no immediate reason to expect the higher $e^{g_nA_n}$ to be small, though one would expect this to the case in the majority of the interesting cases, even in more general cases than this, as I shall argue below in section 14.

Profinite dimensional estimation Lie algebras occur frequently. Consider systems

$$dx = f(x)dt + G(x)dw_t, \quad dy = h(x)dt + dv \tag{12.5}$$

with the additional assumptions that f, G and h are analytic (totally around zero) and that $f(0) = G(0) = 0$.

Theorem 12.6. [17]. Under the assumptions made immediately above $EL(\Sigma)$ is profinite dimensional.

If one adds the condition that $h(0) = 0$ (which surely does no harm; removing a known constant from the observation equation is a triviality) the resulting estimation Lie algebra is even solvable (meaning that all the quotients $L \, / \, \mathfrak{A}_i$ are solvable).

Another case of a profinite dimensional estimation Lie algebra (different from the class of theorem 12.6, the identification case to be treated below, and the perturbation case of section 11 above) is example 7.5. As a rule one should probably not expect that "the statistic calculated by $L \, / \, \mathfrak{A}_i$" of a system whose estimation Lie algebra happens to be profinite dimensional, is easily interpretable (recognizable) as the statistic of an interesting quantity. In the case of example 7.5 this is however the case, (Liu-Marcus [23]).

13. IDENTIFICATION OF LINEAR DYNAMICAL SYSTEMS

Suppose now that we are faced with a somewhat different problem. Namely suppose one has reason to believe, or simply does not know anything better to do, that a given phenomenon, say a time series, is modeled by a linear dynamical system

$$dx = Axdt + Bdw, \quad dy = Cxdt + dv \tag{13.1}$$

Now, however, the coefficients in A, B, C are unknown and also have to be estimated from the observation $y(t)$. That is the system (13.1) has to be identified. It is easy to turn this into a filtering problm by adding the (stochastic) equations

$$dA = 0, \quad dB = 0, \quad dC = 0 \tag{13.2}$$

(or just $dr_{ij} = 0$ whether the r_{ij} run through the coefficients which are unknown, if A, B, C are partly known; for example because of structural considerations). The resulting filtering problem is nonlinear.

13.1. Observation

The estimation Lie algebra of the system (13.1)-(13.2) is a sub-Lie-algebra of the current Lie algebra $ls_n \otimes \mathbb{R}[A, B, C]$ where $\mathbb{R}[A, B, C]$ stands for the ring of polynomials in the indeterminates a_{ij}, b_{kl}, c_{rs}.

A corollary is that these estimation algebras are profinite dimensional. And looking a bit more closely at them, they are solvable [37]. Thus the ideas and considerations of the previous two sections can be brought into play and one can try to do infinite dimensional Wei-Norman theory etc. This is attempted in Krishnaprasad-Marcus-Hazewinkel [37]. In this rather special case it turns out that the higher approximations (the zero-th approximation is simply the family of Kalman-Bucy filters parametrized by A, B, C also discussed in section 8 above) have to do with sensitivity equations: sensitivities of the ouput $y(t)$ with respect to changes in the parameters A, B, C.

As stated above, though, the problem is degenerate and likely to cause all kind of difficulties. The problem is that the conditional density $\rho(x, A, B, C, t)$ will be degenerate because the A, B, C are not uniquely determined by the observations. Indeed if S is an invertible $n \times n$ matrix then the system (13.1) given by the matrices SAS^{-1}, SB, CS^{-1} instead of A, B, C gives exactly the same input-output behaviour. Thus we should really be considering this problem on a suitable quotient space $\{(A, B, C)\} \, / \, GL_n$. These quotient spaces as a rule are not diffeomorphic to open sets in some $\mathbb{R}^n$. This is one way in which stochastic systems like (1.1)-(1.2) on nontrivial manifolds naturally arise and it leads to the necessity of finding a DMZ-equation in this more general context. Work in this direction has been done by Ji Dunmu and T.E. Duncan.

Let me add one observation. For the filters giving $\hat{x}, \hat{A}, \hat{B}, \hat{C}$ for problem (13.1)-(13.2) one expects $\hat{x}$

LIE ALGEBRAIC METHOD IN FILTERING AND IDENTIFICATION   173

to move fast relative $\hat{A}, \hat{B}, \hat{C}$. Thus it would make sense to consider a system

$$dx = (A_0 + \epsilon A_1)xdt + (B_0 + \epsilon B_1)dw, \quad dy = (C_0 + \epsilon C_1)dt + dv \qquad (13.4)$$

$$dA_1 = 0, \quad dB_1 = 0, \quad dC_1 = 0$$

(where A_0, B_0, C_0 are assumed known) and apply the ideas of section 11 above to find optimal directions of change (i.e. the A_1, B_1, C_1).

14. ASYMPTOTIC EXPANSIONS AND APPROXIMATE HOMOMORPHISMS

The ideas to be outlined below in this section are still speculative but there are quite a number of positive signs.

First however let me point out that the procedures based on Wei-Norman techniques as described in sections 11 and 12 above clearly indicate that existence, uniqueness and regularity results for solutions of the DMZ-equation have a lot to do with the existence of asymptotic expansions ([2,4]). For regularity results etc. cf. e.g. work of D. Michel, J.-M. Bismut, E. Pardoux, M. Chaleyat-Maurel, D. Ocone, Th. Kurtz, W.E. Hopkins Jr., H.J. Sussmann a.o. ([25,8,22,2] and references in these papers).

Let us consider a control system of the form

$$\dot{x} = f(x) + \Sigma u_i g_i(x) \qquad (14.1)$$

where the f and g_i are vectorfields. To make thinking easier assume that 0 is a stable and asymptotically stable equilibrium for the unforced equation. A system like (14.1) is intended as a model of something and as such one can argue that say the values of $f(x), g_i(x)$ are relatively well known, the values of their (partial) derivatives (w.r.t. the x_i) will be less well known, the second partial derivations are still less well determined etc..

Thus, intuitively, for systems which represent or model real (stable) things one would expect that in many cases the behaviour of (14.1) will depend primarily on the first few terms which appear in the Lie algebra generated by f and the g_i. The higher brackets should matter less and less.

That means that instead of looking at $Lie\{f, g_1, ..., g_m\}$, the Lie algebra generated by $f, g_1, ..., g_m$ as a Lie algebra without further structure, we should look at it as a Lie algebra with a given set of generators and sort of keep track of how often these generators are used to generate further elements of the algebra. (For each time a bracket is taken a differentiation is applied, and thus the higher brackets of the $f, g_1, ..., g_m$ depend only on the deeper parts of the Taylor expansions of $f, g_1, ..., g_m$.)

Personally I would also say that having noises rather than precise deterministic controls u_i would enhance this type of (structural?) stability.

A precise way to keep track of how often the generators are used is to introduce one extra counting indeterminate z and to consider instead of $L = Lie\{f, g_1, ..., g_m\}$ the Lie algebra generated by the vectorfields $\{zf, zg_1, ..., zg_m\}$. This Lie algebra L_z is topologically nilpotent, i.e. if $L_z^{(n)} = [L_z, L_z^{(n-1)}]$, $L_z^{(0)} = L_z$, then $\cap L_z^{(m)} = \{0\}$. And a homomorphism $L_z \to V(M)$ into the vectorfields on M with kernel $L_z^{(n)}$ precisely means "respecting the structure of the Lie algebra L up to brackets of order n". All this is very much related to the ideas of nilpotent approximation as introduced by Stein, Rothschild, Goodman and Rockland, [32,14,31] in the study of hypoellipticity and taken up by Crouch in system theory [10].

Thus in filtering theory it would seem natural to look at the Lie algebra of operators $EL_z(\Sigma)$ generated by the operators

$$z_0 \mathcal{L}, \ z_1 h_1, ..., z_p h_p$$

where the $z_0, z_1, ..., z_p$ are additional variables (so as to give, if desired, certain observations more weight than others and to be able to set certain of them, especially z_0, equal to 1). The idea would be then to study the filters produced by Wei-Norman type techniques for the various finite dimensional

quotients and to see whether this produces viable expansions.

15. REMOVING OUTLIERS

A final idea in much the same spirit as before is the following. Suppose we are again dealing with a system

$$dx = f(x)dt + G(x)dw, \quad dy = h(x)dt + dv. \tag{15.1}$$

Suppose also to make thinking easier that the thing is more or less stable so that x tends to remain in some bounded partion of $\mathbb{R}^n$ (f asymptotically stable) and maybe suppose also that h is proper, so that large y observations are exceedingly rare and should probably be discounted. Suppose that $e^{-\|x\|^2}$ is differential algebraically independent of f, G, h. This is for example this case if f, G, h are polynomial and also if they are of compact support. In other cases other functions with similar properties can presumably be found. Now instead of (15.1) consider the modified system

$$dx = f(x)dt + G(x)dw, \quad dy = e^{-a\|x\|^2} h(x)dt + dv \tag{15.2}$$

where $a > 0$ is a small parameter. Note that the only thing which (15.2) does with respect to (15.1) is to discount large y observations.

Now consider the estimation Lie algebra of the sytem (15.2).

Theorem 15.3. If $e^{-a\|x\|^2}$ is differentially algebraically independent of f, G, h then the estimation Lie algebra of (15.2) is pro-finite dimensional and solvable. To be more precise it is finite dimensional and solvable mod $(a^i e^{-ja\|x\|^2}, i + j \geq n)$ for all n.

Thus the yoga of the previous sections can again be applied and the behaviour of the resulting filters as a goes to zero could be studied.

REFERENCES

[1] J.S. BARAS, Group invariance methods in nonlinear filtering of diffusion processes, In: [19], 565-572.

[2] J.S. BARAS, G.L. BLANKENSCHIP, W.E. HOPKINS Jr., Existence, uniqueness and asymptotic behaviour of solutions to a class of Zakai equations with unbounded coefficients, IEEE Trans. AC-28 (1983), 203-214.

[3] G.L. BLANKENSCHIP, Lecture at MTNS'85, Stockholm, To appear North Holland Publ. Cy.

[4] G.L. BLANKENSCHIP, C.-H. LIU, S.I. MARCUS, Asymptotic expansions and Lie algebras for some nonlinear filter problems, IEEE Trans. AC-28 (1983), 787-797.

[5] R.W. BROCKETT, Remarks on finite dimensional nonlinear estimation, In: C. Lobry (ed.), Analyse des systèmes, Astérisque 76 (1980), Soc. Math. de France.

[6] R.W. BROCKETT, J.M.C. CLARK, The geometry of the conditional density equation In: O.L.R. Jacobs et al. (eds), Analysis and optimization of stochastic systems, New York, 1980, 299-309.

[7] R.S. BUCY, J.M.F. MOURA (eds), Nonlinear stochastic problems, Reidel, 1983.

[8] J.M.C. CLARK, The design of robust approximations to the stochastic differential equations of nonlinear filtering, In: J.K. Skwirzynski (ed.), Communication systems and random process theory, Sijthoff & Noordhoff, 1978.

[9] M. CHALEYAT-MAUREL, D. MICHEL, Hypoellipticity theorems and conditional laws, Z. Wahrsch. und verw. Geb. 65 (1984), 573-597.

[10] P.E. CROUCH, Solvable approximations to control systems, SIAM J. Control and Opt. 32 (1984), 40-54.

[11] M.H.A. DAVIS, Pathwise nonlinear filtering, In [19], 505-528.

[12] M.H.A. Dsavis, S.I. MARCUS, An introduction to nonlinear filtering, In [19], 565-572.

[13] T.E. DUNCAN, Probability densities for diffusion processes with applications to nonllinear filtering theory, Ph. D. thesis, Stanford, 1967.

[14] R.W. GOODMAN, Nilpotent Lie groups, structure and applications to analysis, LNM 562, Springer, 1976.

[15] M. HAZEWINKEL, On deformations, approximations and nonlinear filtering, Systems Control Lett. 1 (1982), 29-62.

[16] M. HAZEWINKEL, The linear systems Lie algebra, the Segal-Shale-Weil representation and all Kalman-Bucy filters, J. Syst. Sci. & Math. Sci. 5 (1985), 94-106.

[17] M. HAZEWINKEL, S.I. MARCUS, On Lie algebras and finite dimensional filtering, Stochastics 7 (1982), 29-62.

[18] M.HAZEWINKEL, S.I. MARCUS, H.J. SUSSMANN, Nonexistence of finite dimensional filters for conditional statistics of the cubic sensor problem, Systems Control Lett. 3 (1983), 331-340.

[19] M. HAZEWINKEL, J.C. WILLEMS (eds), Stochastic systems: the mathematics of filtering and identification and applications, Reidel, 1981.

[20] O.B. HIJAB, Finite dimensional causal functionals of brownian motion, In [6], 425-436.

[21] A.J. KRENER, On the equivalence of control systems and the linearization of nonlinear systems, SIAM J. Control 11 (1973), 670-676.

[22] Th.G. KURTZ, D. OCONE, A martingale problem for conditional distributions and uniqueness for the nonlinear filtering equations, Lect. Notes Control and Inf. Sci. 69 (1985), 224-235.

[23] C.-H. LIU, S.I. MARCUS, The Lie algebraic structure of a class of finite dimensional nonlinear filters, In: C.I. Byrnes, C.F. Martin (eds), Algebraic and geometric methods in linear systems theory, Amer. Math. Soc., 1980, 277-297.

[24] S.I. MARCUS, Algebraic and geometric methods in nonlinear filtering, SIAM J. Control Opt. 22 (1984), 817-844.

[25] D. MICHEL, Régularité des lois conditionelles en théorie du filtrage non-linéaire et calcul des variations stochastiques, J. Funct. Anal. 14 (1981), 8-36.

[26] S.K. MITTER, On the analogy between mathematical problems of non-linear filtering and quantum phisics, Ric. di Automatica 10 (1980), 163-216.

[27] S.K. MITTER, Nonlinear filtering and stochastic mechanics, In: [19], 479-504.

[28] R.E. MORTENSEN, Optimal control of continuous time stochastic systems, Ph.D. thesis, Berkeley, 1966.

[29] E. PARDOUX, D. TALAY, Discretization and simulation of stochastic differential equations, Acta Appl. Math. 3 (1982), 182-203.

[30] L.E. PURSELL, M.E. SHANKS, The Lie algebra of a smooth manifold, Proc. Amer. Math. Soc. 5 (1954), 468-472.

[31] Ch. ROCKLAND, Intrinsic nilpotent approximation, preprint MIT, LIDS-R-1482, 1985, to appear Acta Appl. Math.

[32] L.P. ROTHSCHILD, E.M. STEIN, Hypoelliptic differential operators and nilpotent groups, Acta Math. 137 (1976), 247-320.

[33] H.J. SUSSMANN, Approximate finite dimensional filters for some nonlinear problems, Stochastics 7 (1982), 183-203.

[34] H.J. SUSSMANN, Existence and uniqueness of minimal realizations of nonlinear systems, Math.

Syst. Theory 10 (1977), 349-356.

[35] J. WEI, E. NORMAN, On the global representation of the solutions of linear differential equations as a product of exponentials, Proc. Amer. Math. Soc. 15 (1964), 327-334.

[36] M. ZAKAI, On the optimal filtering of diffusion processes, Z. Wahrsch. verw. Geb. 11 (1969), 230-243.

[37] P.S. KRISHNAPRASAD, S.I. MARCUS, M. HAZEWINKEL, Current algebras and the identification problem, Stochastics 11 (1983), 65-101.