LECTURES ON LINEAR AND NONLINEAR FILTERING

M. Hazewinkel
CWI, Amsterdam, The Netherlands

1. Introduction

Quite generally the *filtering problem* can be described as follows. Given a stochastic process $x(t)$, $t\in I\subset\mathbb{R}$, i.e. a sequence of random variables, and a (more or less related) second process $y(t)$, $t\in I$, it is desired to find the *best estimate* of x at time t, i.e. the best estimate of $x(t)$, given the (past observations) $y(s)$, $0\leqslant s\leqslant t$. Usually $I=\mathbb{Z}$ (discrete time) or $I=\mathbb{R}$ (continuous time).

Much related problems are *prediction:* calculate the best estimate of $x(t)$ given $y(s)$, $0\leqslant s\leqslant t-r$, and *smoothing:* calculate the best estimate of $x(t-r)$ given $y(s)$, $0\leqslant s\leqslant t$. In all these it may of course be the case that $y(t)$ and $x(t)$ are the same stochastic process.

In these lectures we shall be concerned with the (model) case that the continuous time processes $x(t)$ and $y(t)$ are related as follows

$$dx(t) = f(x(t))dt + G(x(t))dw(t)\,,\quad x(t)\in\mathbb{R}^n,\ w(t)\in\mathbb{R}^m, \tag{1.1}$$

$$dy(t) = h(x(t))dt + dv(t)\,,\ y(t)\in\mathbb{R}^p,\ v(t)\in\mathbb{R}^p \tag{1.2}$$

with initial conditions $x(0)\in\mathbb{R}^n$, $y(0)\in\mathbb{R}^p$. Here $w(t)$ and $v(t)$ are supposed to be independent Wiener noise processes also independent of the initial random variable $x(0)$, and $f(x)$, $G(x)$ and $h(x)$ are known vector and matrix valued functions. Thus $\dot{w}(t)$ and $\dot{v}(t)$ are white noise and (1.1) can be looked at as a dynamical system

$$\dot{x}(t) = f(x(t)) \tag{1.3}$$

subject to continuous random shocks whose direction and intensity is further modified (apart from being random) by $G(x(t))$. And equation (1.2), the observation equation, says that the observations at time t

$$y(t) = y(0)+\int_0^t h(x(s))ds\ ,\quad \dot{y}(t)=h(x(t)) \tag{1.4}$$

are corrupted by further (measurement) noise $v(t)$. Technically speaking, equations (1.1), (1.2) are to be regarded as Ito stochastic differential equations; cf section 5 below for more remarks.

The phrase 'best estimate' of $x(t)$, or, more generally, of an interesting function $\phi(x(t))$, is to be understood in the mathematical sense of conditional expectation $\hat{x}(t)=E[x(t)|y(s),\ 0\leqslant s\leqslant t]$, or, in the more general case, $E[\phi(x(t))|y(s),0\leqslant s\leqslant t]$. This is a mathematically well defined object.

Unfortunately the (mathematical) proof of this statement contains nothing in the way of methods of calculating these conditional expectations (effectively).

There are many techniques and approaches to filtering. It is definitely not the idea of these lectures to give a general survey of the field. Instead I shall try to give an account of one particular approach pioneered by Roger Brockett, Martin Clark and Sanjoy Mitter, [6,7,8,9,53,54] which is variously known as the Lie-algebra approach, the reference probability approach, or the unnormalized density approach. This is a rather recent set of ideas, which has several merits. First, it takes geometrical aspects of the situation into account. Second, it explains convincingly why it is easy to find exact recursive filters for linear dynamical systems while it is very hard to filter something like the cubic sensor - for over 20 years a notoriously hard case to handle. The notion of a recursive exact filter will be discussed below in section 2. Thus excitement about this approach was high in the very first years of the 1980's. The book [34] well reflects this. Since then interested and excitement have waned perceptibly. There are also several connected reasons for this. First the method itself indicates clearly - through this remains to be proved in one sense or another - that one can not expect many cases (beyond the case of linear systems) where finite dimensional exact recursive filters exist. 'Generally', it seems, such filters will not exist and though there remains the tantalizing possibility of whole new classes of useful models for which they do exist, there are at the moment no clear ideas as to how and where to look for them. All the same a number of new filters, both 'model cases' and filters of importance in practice, have been discovered using these Lie-algebra ideas [4,13,18,19,47-51,56]. Since exact finite dimensional filters can not exist in many cases it is natural to look for approximate ones. Here it is not immediately apparent how to proceed on the basis of the Lie-algebra approach, and little has been done.

Still there are a number of very promising (heuristic) ideas, which definitely work in some cases. It is the second purpose of these lectures to examine some of these ideas for obtaining approximate recursive filters. All seem to lead to far from trivial unsolved, and possibly quite difficult, mathematical questions, which invite major research efforts.

2. Recursive filters

The basic quantities we have available at time t are observations up to and including time t, i.e. the $y(s)$, $0 \leq s \leq t$. A priori an algorithm to calculate $\hat{x}(t)$, say, could involve all the $y(s)$. Now if the observations come in at a high rate and the algorithm really needs all the $y(s)$ each time an estimate is calculated, one is likely to run into (i) long computation times, and (ii) storage (memory) problems. In such a situation it would be much more practical and much nicer if it were possible to calculate $\hat{x}(t+dt)$ on the basic of $\hat{x}(t)$ and the new information $y(t+dt)$ which has come in. (It is easier to think of this situation in discrete time with $dt=1$.)

This turns out to be too optimistic. Such filters almost cannot exist (in nontrivial situations). The next best thing would be the existence of some other quantity $\xi(t)$ which does have the recursive update property "$\xi(t+dt)=\alpha(\xi(t),y(t+dt))$" and from which the desired quantity can be directly calculated. Of course then $\xi(t)$ must be a reasonable quantity and not some hard to handle infinite dimensional object like the time history function $\{y(s):0 \leq s \leq t\}$ itself. E.g. $\xi(t)$ could be a finite dimensional quantity, or something in an infinite dimensional space describable by a finite number of parameters or well approximatable in terms of a finite numbers of parameters.

Such filters do exist sometimes. E.g. in the case of linear time invariant systems

$$\begin{aligned} dx(t) &= Ax(t)dt + Bdw(t)\,, \quad x \in \mathbb{R}^n,\ w \in \mathbb{R}^m \\ dy(t) &= Cx(dt + dv(t)\,, \quad y \in \mathbb{R}^p,\ v \in \mathbb{R}^p \end{aligned} \tag{2.1}$$

where A,B,C are constant real matrices of dimensions $n \times n$, $n \times m$, $p \times n$ respectively, and where $w(t)$ and $v(t)$ are Wiener noise processes, independent of each other and also independent of the initial random vector $x(0) \in \mathbf{R}$. (One sets $y(0)=0$). In this case one has the well-known Kalman-Bucy filter for the conditional state

$$dP(t) = (AP(t)+P(t)A^T+BB^T-P(t)C^TCP(t))dt \tag{2.2}$$
$$dm(t) = Am(t)dt+P(t)C^T(dy(t)-Cm(t)dt)$$

Thus in this case the quantity $\xi(t)=(P(t),m(t))$ has the desired "recursive updating property", and the quantity we want to filter for, i.e. the quantity $\hat{x}(t)$ in this case, is obtained by a simple projection $\hat{x}(t)=m(t)$.

All this leads to the following initial definition of a *finite dimensional exact recursive filter* for a quantity (statistic) $\widehat{\phi(x(t))}$. By definition such a filter is a finite dimensional dynamical system of the form

$$d\xi(t) = \alpha(\xi(t))dt+\sum_{j=1}^{p}\beta_j(\xi(t))dy_j(t)\,,\quad \xi(t)\in M \tag{2.3}$$

where $y_j(t)$ is the j-th component of $y(t)$, together with an output map

$$\widehat{\phi(x(t))} = \gamma(\xi(t)) \tag{2.4}$$

Here M is supposed to be a finite dimensional manifold and the α and β_j are smooth vectorfields on M. (One can usually think of M as an $\mathbb{R}^n$ so that (2.3) becomes an ordinary stochastic differential equation).

Of course, more generally, one could let the α and β_j in (2.3) depend on the y_j as well. This does not bring very much more because we can, so to speak, add the $y_1,\dots,y_p$ to the state variables $x_1,\dots,x_n$. However, certainly more general potential filters could be considered then (2.3); in particular one can allow the output map γ at time t to depend explicitly on $y_1(t),\dots,y_p(t)$, and we shall have occasion to use this. Again this can be taken care of by extension. This time by extending the filter state vector $\xi(t)$ to the filter state vector $(\xi(t),y(t))$.

3. Robustness

The $y(t)$ are stochastic processes. As it stands (2.3) is a stochastic (partial) differential equation and as such its solutions are only defined apart from a set of measure zero. On the other hand the possible observations paths are piece wise differentiable and these constitute a set of measure zero (in the space of all paths under Wiener measure or therewith mutually continuous measures). Thus solutions of (2.3) may, so to speak, well be undefined precisely on the possible observation paths, [14].

More importantly - in my view - in actual situations we do not have available the stochastic process $y(t)$ but just one possible realization of it. Thus it would be nice if (2.3) made sense pathwise and if it could be replaced with something involving just functions of the $y_1(t),\dots,y_p(t)$, say polynomials, and not their derivatives.

For the filter (2.2) this can be done. The transformation $\overline{m}=m-PC^Ty$ yields

$$\begin{aligned} d\overline{m} &= dm-PC^Tdy-dPC^Ty=Amdt+PC^Tdy-PC^{T^T}Cmdt-PC^Tdy-dPC^Ty \\ &= (A-PC^TC)\overline{m}dt-(PA^TC^T+BB^TC^T)ydt \end{aligned}$$

or

$$\begin{aligned} \frac{d}{dt}\overline{m} &= (A-PC^TC)\overline{m}-(PA^TC^T+BB^TC^T)y \\ \frac{d}{dt}P &= AP+PA^T+BB^T-PC^TCP \qquad (3.1)\\ \hat{x}(t) &= \overline{m}(t)+P(t)C^Ty(t) \end{aligned}$$

a set of equations which makes perfect sense for an arbitrary single continuous but possibly almost everywhere non-differentiable observation path $y(t)$.

Such filters are called robust.

4. THE UNNORMALIZED DENSITY APPROACH

One obvious approach to try to find a filter for, say, the conditional state $\hat{x}(t)$ is to try to derive a differential equation for it. This can actually be done [44-46] and yields

$$d\hat{x} = \hat{f} - (\widehat{(xh_T)} - \hat{x}\hat{h}^T)\hat{h}dt + (\widehat{(xh_T)} - \hat{x}\hat{h}^T)dy \tag{4.1}$$

where f and h stand for $f(x(t))$ and $h(x(t))$ and a $\hat{}$ over a symbol means taking the conditional expectation. The trouble with this equation is that it involves the expectations $\hat{f}$, $\hat{h}$ and $\widehat{(xh_T)}$ (and these are for nonlinear f,h of course not equal to $f(\hat{x})$, $h(\hat{x})$ and $\hat{x}h^T(\hat{x})$). One can also write down equations for the conditional quantities $\hat{f}$, $\hat{h}$, $\widehat{xh^T}$, but these then involve conditional expectations of still more complicated expressions etc, etc. As a rule this process will not stop and there results an infinite system of equations.

The conditional density $p(x,t)$, that is the density of the stochastic variable $\hat{x}$ at time t, satisfies a nicer looking equation

$$dp = \bar{\mathfrak{L}}pdt + (h - \hat{h})^T(dy - \hat{h}dt)p \tag{4.2}$$

Given $p(t,x)$, $\hat{h}$ (at time t) is calculated by $\hat{h} = \int(hx)p(x,t)dx$ and inserting this gives in any case an integro partial differential equation of recursive type.

Still nicer is the equation satisfied by a certain unnormalized version $\rho(x,t)$ of the conditional density. And it is this equation, the so called Duncan-Mortensen-Zakai equation, or DMZ-equation, which is at the basis of the Lie algebraic approach. Here unnormalized means that $\rho(x,t) = r(t)p(x,t)$ for some (unknown) function $r(t)$ depending only on time (and not on x).

The DMZ equation for $\rho(x,t)$ reads ([20,33,34,55,65])

$$d\rho(x,t) = \bar{\mathfrak{L}}\rho(x,t)dt + \sum_{j=1}^{p} h_j(x)\rho(x,t)dy_j(t) \tag{4.3}$$

where $\bar{\mathfrak{L}}$ is the second-order semi-elliptic operator (in the $x_1,\ldots,x_n$) defined by

$$(\bar{\mathfrak{L}}\phi(x)) = \frac{1}{2}\sum_{j,k}\frac{\partial_2}{\partial x_j \partial x_k}((G(x)G(x)^T)_{jk}\phi) - \sum_{i=1}^{n}\frac{\partial}{\partial x_i}(f_i(x)\phi). \tag{4.4}$$

Here $(G(x)G^T(x))_{jk}$ is the (j,k)-entry of the $n \times n$ matrix GG^T and f_i is the i-th component of f.

Note that (4.2) is recursive, but being a *partial* stochastic differential equation, it is of course infinite dimensional. Note also that for the calculation of conditional expectations the unknown factor $r(t)$ does not matter much. Indeed $r(t) = \int\rho(x,t)dx$ and correspondingly, if $\phi(x)$ is some interesting function of the state, one has that

$$\widehat{\phi(x(t))} = \int_{\mathbf{R}^n}\phi(x)\rho(x,t)dx \Big/ \int_{\mathbf{R}^n}\rho(x,t)dx \tag{4.5}$$

5. ITO AND FISK-STRATONOVIČ STOCHASTIC DIFFERENTIAL EQUATIONS

Most of the equations written down so far, e.g. (1.1), (1.2), (2.1), (2.2), (2.3), (4.1), (4.2), (4.3) are *stochastic* differential equations. It is definitely not my intention to give extensive explanations of what this means, but a few words seem in order. The meaning of (1.1) e.g. is that there is a well defined notion of stochastic integral such that

$$x(t) = x(0) + \int_0^t f(x(t))dt + \int_0^t G(x(t))dw(t) \tag{5.1}$$

There are in fact several possible definitions. Two of these are the Ito integral and the Fisk-Stratonovič integral [1]. And they are definitely different in the sense that different stochastic processes result depending on whether the second integral (5.1) is interpreted in the Ito or Fisk-

Stratonovič sense. Both definitions have their advantages and disadvantages. The Fisk-Stratonovič integral has the major advantage that the usual rules of the differential-integral calculus still hold. This makes it the preferred interpretation on manifolds. Thus equation (2.3) is intended to be read as a Fisk-Stratonovič equation.

The original system (1.1), (1.2) however is a set of Ito equations and the DMZ equation (4.3) in also an Ito equation.

There is a fairly simple conversion rule from Ito equations to Fisk-Stratonovič equations and vice versa as follows.

Let

$$dx = f(x)dt + G(x)dw(t), \quad x \in \mathbb{R}^n, \ w \in \mathbb{R}^m \tag{5.2}$$

be an Ito stochastic differential equation. Then the equivalent Stratonovič equation is

$$dx = f(x)dt - \frac{1}{2}\sum_{j=1}^{m}\sum_{i=1}^{n}(\frac{\partial G}{\partial x_i})_j G_{ij}dt + G(x)dw(t) \tag{5.3}$$

Here $(\frac{\partial G}{\partial x_i})_j$ is the j-th column of the $n \times n$ matrix $\frac{\partial G}{\partial x_i}$, i.e. $(\frac{\partial G}{\partial x_i})_j = (\frac{\partial G_{1j}}{\partial x_i}, \ldots, \frac{\partial G_{nj}}{\partial x_i})^T$, and G_{ij} is the (i,j)-entry of the $n \times m$ matrix G. Here equivalent means that the same stochastic processes $x(t)$ occur as solutions of the Ito equation (5.2) and the Fisk-Stratonovič equation (5.3).

Note in particular that for an Ito equation of bilinear type

$$dx = Axdt + \sum_{k=1}^{m} B_k x dw_k(t) \tag{5.4}$$

the equivalent Fisk-Stratonovič equation is

$$dx = Axdt - \frac{1}{2}\sum_{k=1}^{m}(B_k^2 x)dw_k(t) + \sum_{k=1}^{m} B_k x dw_k(t) \tag{5.5}$$

The equivalent Fisk-Stratonovič equation to the DMZ filter equation (4.3) is

$$d\rho = £\rho dt + \sum_{j=1}^{p} h_i \rho dy_i(t) \tag{5.6}$$

where now £ is the operator given by

$$£(\phi) = \frac{1}{2}\sum \frac{\partial^2}{\partial x_i \partial x_j}((GG^T)_{ij}\phi) - \sum_i \frac{\partial}{\partial x_i}(f_i\phi) - \frac{1}{2}\sum_j h_j(x)^2\phi. \tag{5.7}$$

Though I have not given a discussion of the Ito-Stratonovič equivalence for partial stochastic differential equations this is easily understood by analogy from (5.5) if nothing else. Note that in (4.3) the unknown is ρ and that the $h_j(x)$ are (commuting) diagonal linear operators $\rho \mapsto h_j(x)\rho$.

As is turns out the stochastic aspects of the filtering problem in this approach largely disappear. This happens because there is an equivalent version of the Fisk-Stratonovič type DMZ equation (5.6) which is robust and can be interpreted pathwise, i.e. as a family of deterministic partial differential equations indexed by the possible observation paths, say, by the continuous functions $\mathbb{R}(\geqslant 0) \rightarrow \mathbb{R}^p$; cf below.

If from now on a stochastic differential equation appears then unless the contrary is explicitly stated, it will always be a Fisk-Stratonovič one.

6. THE ROBUST VERSION OF THE DMZ EQUATION

Consider the (Fisk-Stratonovič) equation (5.6) for an unnormalized conditional density $\rho(x,t)$. This involves the $dy_j(t)$. Now, as also mentioned in section 3 above, what we have available in terms of observations is one realization, one possible path, of a stochastic process $y(t)$. Hence, apart from smoothing effects introduced by the measurements process an almost surely nowhere differentiable function, which makes it more difficult to handle the integrals involved and to find numerical approximations.

Consider the time-dependent (gauge) transformation

$$\tilde{\rho}(x,t) = e^{-h_1(x)y_1(t)-\dots-h_p(x)y_p(t)}\rho(x,t) \tag{6.1}$$

As we are dealing with a Fisk-Stratonovič equation the ordinary rules of calculus apply and equation (5.6) transforms into an equation

$$\frac{\partial\tilde{\rho}}{\partial t} = \pounds\tilde{\rho} - \sum_{i=1}^{p} y_i(t)\pounds_i\tilde{\rho} + \sum_{i,j=1}^{p} y_i(t)y_j(t)\pounds_{ij}\tilde{\rho} \tag{6.2}$$

where the differential operators $\pounds_i$ and $\pounds_{ij}$ are given by

$$\pounds_i = [h_i,\pounds] = h_i\pounds - \pounds h_i\ , \quad \pounds_{ij} = \frac{1}{2}[h_i,[h_j,\pounds]]. \tag{6.3}$$

Cf. below for a derivation of (6.2). The terms $dy_j(t)$ cancel after this transformation and we are left with a family of partial differential equation indexed by the possible observations paths, i.e. with one equation from this family in a given filtering situation.

If $\tilde{\rho}(x,t)$ (for a particular path $y(t)$) has been found from (6.2), then $\rho(x,t)$ is given, as a function of x and t, by formula 6.1. Note in particular that the $h_i(x)$ in (6.2) should not be read as functions of the stochastic process $x(t)$; instead (6.1) is simply the exponential of the known multiplication operator $\rho \mapsto (\Sigma_{i=1}^{p} h_i(x)y_i(t))\rho$ on densities.

To obtain (6.2) observe that substituting (6.1) into the DMZ equation (5.6) gives

$$\frac{\partial\tilde{\rho}}{\partial t} = e^{-\sum h_j(x)y_j(t)}\pounds e^{\sum h_j(x)y_j(t)}\tilde{\rho} \tag{6.4}$$

Thus writing A for the operators of multiplication with $\sum h_j(x)y_j(t)$, we have to calculate $e^{-A}\pounds e^{A}$. By the adjoint action formula (cf the short tutorial on Lie algebras in this volume, [28]) this is equal to

$$e^{-A}\pounds e^{A} = \pounds - [A,\pounds] + \frac{A,[A,\pounds]]}{2!} - \frac{[A,[A,[A,\pounds]]]}{3!} + \dots \tag{6.5}$$

In our case $\pounds$ is a second order differential operator and A is multiplication with a function. Hence $[A,\pounds]$ is a first order differential operator, $[A,[A,\pounds]]$ is a zero-th order differential operator, i.e. (multication with) a function, and $[A,\dots,[A,\pounds]\dots]=0$ if three or more A's occur. Writing out $[A,\pounds]$ and $[A,[A,\pounds]]$ yields (6.2)-(6.3).

Even though now we can work with nonstochastic partial differential equations (6.2) the numerics of the situation are daunting, cf however, also [57]. Typically x is a large dimensional vector of, say, dimension 27, in certain practical problems involving helicopters. So we have a second order semi-elliptic PDE in 27 space dimensions and one time dimension. This rules out standard approximation schemes. Also of course we need a solution method which deals in principle not with one instance of equation (6.2) but with the whole family (6.2). I.e. the parameters $y_1(t),\dots,y_p(t)$ must enter into the calculation algorithm in a reasonable way. These remarks constitute some of the motivation for the approach via Wei-Norman equations discussed in the sections below.

7. WEI-NORMAN THEORY [64]

It is important to note that the filtering equation (6.2) (or (5.6)) is of the general form

$$\dot{x} = (A_1 x)u_1 + ... + (A_k x)u_k \tag{7.1}$$

where the A_i are linear operators and the u_i known functions of time. Of course in (6.2) the role of x is played by $\tilde{\rho}$, an infinite dimensional object. Here, for the moment, lets consider (7.1) 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 (an Ansatz)

$$x(t) = e^{g_1 A_1} e^{g_2 A_2} ... e^{g_k A_k} x(0) \tag{7.2}$$

where the $g_i(t)$ are still to be determined functions of t. Differentiating (7.2) gives

$$\dot{x} = \dot{g}_1 A_1 e^{g_1 A_1} e^{g_2 A_2} ... e^{g_k A_k} x(0) + e^{g_1 A_1} \dot{g}_2 A_2 e^{g_2 A_2} ... e^{g_k A_k} x(0) + ... \tag{7.3}$$

and inserting

$$e^{-g_{i-1} A_{i-1}} ... e^{-g_1 A_1} e^{g_1 A_1} ... e^{g_{i-1} A_{i-1}}$$

just after $\dot{g}_i A_i$ in the i-the term, equation (7.3) can be rewritten

$$\dot{x} = \sum_{i=1}^{k} \dot{g}_i (A_i + \sum_{\substack{j_1,...,j_{i-1} \\ j_1 + ... + j_i - 1 > 0}} \frac{g_1^{j_1} ... g_{i-1}^{j_{i-1}}}{j_1! ... j_{i-1}!} ad_{A_1}^{j_1} ... ad_{A_{i-1}}^{j_{i-1}} (A_i)) x \tag{7.4}$$

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

with $h_{ij}(0,...,0) = 0$, where, again, the adjoint action formula (6.5) has been used. Here $ad_A(B) = [A,B]$, $ad_A^i(B) = ad_A(ad_A^{i-1}(B))$. Thus it remains to solve (equating the coefficients of the basis elements A_i in (7.4) and (7.1))

$$\begin{aligned}
\dot{g}_1 + \dot{g}_1 h_{11}(g_1,...,g_k) + \dot{g}_2 h_{21}(g_1,...,g_k) + ... + \dot{g}_k h_{k1}(g_1,...,g_k) &= u_1 \\
\dot{g}_2 + \dot{g}_1 h_{12}(g_1,...,g_k) + \dot{g}_2 h_{22}(g_1,...,g_k) + ... + \dot{g}_k h_{k2}((g_1,...,g_k) &= u_2 \\
&... \\
\dot{g}_k + \dot{g}_1 h_{1k}(g_1,...,g_k) + \dot{g}_2 h_{2k}(g_1,...,g_k) + ... + \dot{g}_k h_{kk}(g_1,...,g_k) &= u_4
\end{aligned} \tag{7.5}$$

which can be done for small t and $g_1(0) = ... = g_k = (0) = 0$ because $h_{ij}(0,...,0) = 0$. These equations are called the Wei-Norman equations of (7.1). In general a representation (7.2) for the solution is only possible for small t. However things change if the Lie-algebra in question is solvable [64], then 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:

$$L \underset{\neq}{\supset} L^{(1)} = [L,L] \underset{\neq}{\supset} L^{(2)} = [L, L^{(1)}] \underset{\neq}{\supset} ... \underset{\neq}{\supset} L^{(m)} = [L, L^{(m-1)}] = 0 \tag{7.6}$$

Indeed let $A_1,...,A_{k_1}$, $A_{k_1+1},...,$ $A_{k_2},...,A_{k_{m-1}}+1,...,A_{k_m} = A_k$, $k_1 < k_2 < ... < k_m$ be a basis 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 (7.4) that $h_{ij} = 0$ for $j \leq i$ and the set of equations (7.5) 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 (7.4), so the h_{1j} in (7.5) are always all zero) and the resulting equations (7.5) for the nilpotent case are therefore of the form

$$\begin{aligned}
&\dot{g}_1 = u_1, \quad ... \quad ,\dot{q}_{k_1} = u_{k_1} \\
&\dot{g}_{k_1+1} = u_{k_1+1} + \alpha_{k_1+1}(u_1,...,u_{k_1}; g_1,...,g_{k_1}),..., \dot{g}_{k_2} = u_{k_2} + \alpha_{k_2}(u_1,...,u_{k_1}; g_1,...,g_{k_1}) \\
&\dot{g}_{k_2+1} = u_{k_2+1} + \alpha_{k_2+1}(u_1,...,u_{k_2}; g_1,...,g_{k_2}),..., \dot{g}_{k_3} = u_{k_3} + \alpha_{k_3}(u_1,...,u_{k_2}; g_1,...,g_{k_2}) \\
&...
\end{aligned} \tag{7.7}$$

where the α_j are known (universal) functions of the u's and g's. The initial conditions are $g_i(0)=0$, $i=1,...,k$.

It is quite important (for applications of Wei Norman theory) to note that the equations (7.5), i.e. the functions h_{ij} are universal and depend only on the abstract structure of the Lie algebra and the chosen basis. This means the following. That the matrices $A_1,...,A_n$ form the basis of a Lie algebra L of $n\times n$ matrices means that

$$[A_i,A_j] = A_iA_j-A_jA_i=\sum_r\gamma_{ij}^rA_r \tag{7.8}$$

for certain real numbers γ_{ij}^r, the structure constants of L with respect to this basis. Now let L' be a second Lie algebra, say of $m\times m$ matrices. Suppose that L and L' are isomorphic under $\phi:L\to L'$ and let $B_i=\phi(A_i)$. Then the $B_1,...,B_k$ are a basis for L' and because ϕ is an isomorphism

$$[B_i,B_j] = \sum_r\gamma_{ij}^rBr \tag{7.9}$$

with precisely the same γ_{ij}^r. As a result the Wei-Norman equations for the bilinear system of equations

$$\dot y = (B_1y)u_1+...+(B_ky)u_k \quad ,y\in\mathbb{R}^m \tag{7.10}$$

are exactly the same.

This idea also extends to the case that we have a set $£_1,...,£_k$ of operators on some function space which form the basis of a Lie algebra L (so that for all i,j $[£_i,£_j]=\Sigma\gamma_{ij}^r£_r$ for certain γ_{ij}^r). Then again the Wei-Norman equations are identical to the ones of any finite dimensional copy L' of L (and by Ado's theorem, cf the short tutorial in Lie algebras in this volume, such a finite dimensional copy always exists). Of course in such a case of operators we still need to be able to calculate the $e^{g_i(t)L_i}$ for the individual operators L_i. Thus Wei-Norman theory can be seen as a method to integrate (solve) an equation of the form (7.1) in terms of the more elementary equations

$$\dot x = A_ix \tag{7.11}$$

Let me illustrate all this by means of an explicit example. Consider the four differential operators in one variable

$$£ = \frac{1}{2}\frac{d^2}{dx^2}-\frac{1}{2}x^2\ ,\ x,\ \frac{d}{dx},\ 1 \tag{7.12}$$

We then have

$$[£,x]\phi = £(x\phi)-x£\phi = \frac{1}{2}\frac{d^2}{dx^2}(x\phi)-\frac{1}{2}x^3\phi-\frac{1}{2}x\frac{d^2}{dx^2}\phi+\frac{1}{2}x^3\phi = \frac{1}{2}\frac{d}{dx}(\phi+x\frac{d\phi}{dx})-\frac{1}{2}x\frac{d^2\phi}{dx^2}$$
$$=\frac{1}{2}\frac{d\phi}{dx}+\frac{1}{2}\frac{d\phi}{dx}+\frac{1}{2}x\frac{d^2\phi}{dx^2}-\frac{1}{2}x\frac{d^2\phi}{dx^2} = \frac{d\phi}{dx} = \frac{d}{dx}(\phi).$$

Thus

$$[£,x] = \frac{d}{dx} \tag{7.13}$$

and similarly one finds

$$[£,\frac{d}{dx}] = x,\ [\frac{d}{dx},x]=1,\ \ [£,1]=[x,1]=[\frac{d}{dx},1]=0 \tag{7.14}$$

Thus the four operators (7.12) span a four dimensional Lie algebra. It is called the oscillator Lie algebra.

A finite dimensional copy in terms of 4×4 matrices of this algebra is given by the assignment

$$£\mapsto\begin{pmatrix}0&1&0&0\\1&0&0&0\\0&0&0&0\\0&0&0&0\end{pmatrix},\ x\mapsto\begin{pmatrix}0&0&1&0\\0&0&0&0\\0&0&0&0\\0&1&0&0\end{pmatrix},\ \frac{\partial}{\partial x}\mapsto\begin{pmatrix}0&0&0&0\\0&0&1&0\\0&0&0&0\\-1&0&0&0\end{pmatrix},\ 1\mapsto\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&-2&0\end{pmatrix}$$

(Exercise: check this; NB this does not contradict the Stone-von Neumann theorem that it is impossible to represent the communication relation $[\frac{\partial}{\partial x},x]=1$ in terms of finite dimensional operators *in such a way that 1 is represented by the unit operator.)*

Now, by way of example, let us explicitly calculate the Wei-Norman equations for the Lie algebra (7.12).

So the equation we want to solve is

$$\rho_t = \pounds\rho u_1 + x\rho u_2 + \frac{\partial}{\partial x}\rho u_3 + \rho u_4, \quad \rho(x,0)=\pi(x) \tag{7.15}$$

The 'Ansatz' is a solution of the form

$$\rho = e^{g_1(t)\pounds} e^{g_2(t)x} e^{g_3(t)\frac{d}{dx}} e^{g_4(t)}\pi(x) \tag{7.16}$$

Differentiating (7.16) gives

$$\begin{aligned}
\frac{d}{dt}\rho &= \pounds\dot{g}_1 e^{g_1\pounds} e^{g_2 x} e^{g_3\frac{d}{dx}} e^{g_4}\pi(x) \\
&+ e^{g_1\pounds}\dot{g}_2 x e^{g_2 x} e^{g_3\frac{d}{dx}} e^{g_4}\pi(x) \\
&+ e^{g_1\pounds} e^{g_2 x}\dot{g}_3 \frac{d}{dx} e^{g_3\frac{d}{dx}} e^{g_4}\pi(x) \\
&+ e^{g_1\pounds} e^{g_2 x} e^{g_3\frac{d}{dx}} \dot{g}_4 e^{g_4}\pi(x) \\
&= \pounds\dot{g}_1\rho + e^{g_1\pounds}\dot{g}_2 x e^{-g_1\pounds}\rho + e^{g_1\pounds}e^{g_2 x}\dot{g}_3\frac{d}{dx}e^{-g_2 x}e^{-g_1\pounds}\rho \\
&+ e^{g_1\pounds}e^{g_2 x}e^{g_3\frac{d}{dx}}\dot{g}_4 e^{-g_3\frac{d}{dx}}e^{-g_2 x}e^{-\pounds g_1}\rho
\end{aligned}$$

Now $\dot{g}_4$ commutes with the operators £, x, $\frac{\partial}{\partial x}$ (cf (7.14))). So the last term above simply gives $\dot{g}_4\rho$. To calculate the two middle terms we use the adjoint action formula again

$$\begin{aligned}
e^{g_1\pounds}xe^{-g_1\pounds} &= x + g_1[\pounds,x] + \frac{g_1^2}{2!}[\pounds,[\pounds,x]] + \frac{g_1^3}{3!}[[\pounds,[\pounds,[\pounds,x]]] + \dots \\
&= x + g_1\frac{d}{dx} + \frac{g_1^2}{2!}x + \frac{g_1^3}{3!}\frac{d}{dx} + \dots \\
&= \cosh(g_1)x + \sinh(g_1)\frac{d}{dx} \\
e^{g_2 x}\frac{d}{dx}e^{-g_2 x} &= \frac{d}{dx} + g_2[x,\frac{d}{dx}] + \frac{g_2^2}{2!}[x,[x,\frac{d}{dx}]] + \dots \\
&= \frac{d}{dx} - g_2 + 0 + 0 + 0 + \dots \\
e^{g_1\pounds}e^{g_2 x}\frac{d}{dx}e^{-g_2 x}e^{g_1\pounds} &= e^{g_1\pounds}(\frac{d}{dx} - g_2)e^{-g_1\pounds} = \\
&= \frac{d}{dx} - g_2 + g_1[\pounds,\frac{d}{dx}] + \frac{g_1^2}{2!}[\pounds,[\pounds,\frac{d}{dx}]] + \dots
\end{aligned}$$

$$= \frac{d}{dx} - g_2 + g_1 x + \frac{g_1^2}{2!}\frac{d}{dx} + \frac{g_1^3}{3!}x + \ldots$$

$$= -g_2 + \sinh(g_1)x\cosh(g_1)\frac{d}{dx}$$

Thus we find

$$\frac{\partial\rho}{\partial t} = \dot{g}_1 £_\rho + \dot{g}_2\cosh(g_1)x\rho + \dot{g}_2\sinh(g_1)\frac{d}{dx}\rho + \dot{g}_3\sinh(g_1)x\rho + \dot{g}_3\cosh(g_1)\frac{d}{dx}\rho - \dot{g}_3 g_2\rho + \dot{g}_4\rho$$

Comparing this with the original equation (7.15) we find the following ordinary differential equations for the $g_1,\ldots,g_4$

$$\begin{aligned} &\dot{g}_1 = u_1 &&, g_1(0)=0 \\ &\cosh(g_1)\dot{g}_2 + \sinh(g_1)\dot{g}_3 = u_2 &&, g_2(0)=0 \\ &\sinh(g_1)\dot{g}_2 + \cosh(g_2)\dot{g}_3 = u_3 &&, g_3(0)=0 \\ &\dot{g}_4 - \dot{g}_3 g_2 = u_4 &&, g_4(0)=0 \end{aligned} \tag{7.17}$$

or

$$\begin{aligned} &\dot{g}_1 = u_1 &&, g_1(0)=0 \\ &\dot{g}_2 = \cosh(g_1)u_2 - \sinh(g_1)(u_3) &&, g_2(0)=0 \\ &\dot{g}_3 = -\sinh(g_1)u_2 + \cosh(g_1)u_3 &&, g_3(0)=0 \\ &\dot{g}_4 = u_4 + g_2\dot{g}_3 &&, g_4(0)=0 \end{aligned} \tag{7.18}$$

which are of course trivial to solve. In order to find $\rho(t,x)$ itself it now remains to calculate the $e^{g_1 £}, e^{g_2 x}, e^{g_3 \frac{e}{dx}}, e^{g_4}$; or, in other words to solve the simpler initial value problems

$$\frac{\partial\sigma}{\partial t} = \dot{g}_1 £\sigma, \ \frac{\partial\sigma}{\partial t} = \dot{g}_2 x\sigma, \ \frac{\partial\sigma}{\partial t} = \dot{g}_3\frac{\partial}{\partial x}\sigma, \ \frac{\partial\sigma}{\partial t} = \dot{g}_4\sigma$$

The last three of these are trivial and the first one is the harmonic oscillator. Some more remarks on solving 'harmonic oscillator type' equations occur below in section 9.8.

The oscillators Lie algebra (7.12) is solvable, and not nilpotent. Hence the occurrence of the coupled equations block consisting of the second and third equation of (7.17). This is typical for the case of a solvable a Lie algebra. In the nilpotent case the equations can always be solved by quadratures only.

8. The estimation Lie - algebra

In section 4, 5 and 7 above two things have become clear. Firstly that the DMZ equation (5.6) or (6.2) is of bilinear type, i.e. of the general form

$$\frac{\partial\rho}{\partial t} = (£_1\rho)u_1(t) + \ldots + (£_n\rho)(u_n(t)) \tag{8.1}$$

where the $£_i$ are linear (differential) operators on some suitable space of unnormalized densities (functions), and secondly that for bilinear type equations the Lie algebra generated by the operators $£_1,\ldots,£_n$ is important. If this Lie algebra is finite dimensional we have at least small time solutions and if it is finite dimensional and solvable we have explicit methods to solve the initial value provided one can do the same for the simpler equations

$$\frac{\partial\rho}{\partial t} = v_i(t)£_i\rho \qquad i=1,\ldots,n \tag{8.2}$$

Incidentally the phrase 'bilinear' for equations of type (8.1) comes out of system and control theory. Analysts would call this simply a system of linear equations. In control theory however, a linear dynamical system is one of the form

$$\frac{\partial \rho}{\partial t} = A\rho + Bu(t) \tag{8.3}$$

where A is a linear operator on the space of ρ's, the state space, and B is a linear operator from some space of imputs to state space. The term bilinear is used to denote control systems of the form

$$\frac{\partial \rho}{\partial t} = A\rho + \sum_{j=1}^{m} B_i \rho u_i(t) \tag{8.4}$$

with A and $B_1,...,B_m$ operators on state space. In both cases the $u_i(t),...,u_m(t)$ are thought of as inputs or controls.

Thus it is clear that the Lie-algebra generated by the operators occurring in the DMZ equations (5.6) is important and has much to say about how difficult the filtering problem is. Indeed, as will be explained, it can serve to formulate a necessary criterion (the *BC* principle) for the existence of a recursive finite dimensional filter and this can be used to prove that for certain system nontrivial exact finite dimensional recursive filters cannot exist. (E.g. for the cubic sensor, cf section 12 below).

By definition the *estimation Lie algebra* of a system (1.1) -(1.2) is the Lie algebra spanned by the operators occurring in the DMZ equation (5.6); i.e. it is the Lie algebra generated by the second order differential operator £ given by (5.7) and the multiplication operators $h_1,...,h_p$. Notation: $ELie(\Sigma)$ if Σ denotes the system (1.1)-(1.2).

If one works with the robust version of the DMZ equation the natural object to study is the Lie-algebra:

$$E^s Lie(\Sigma) = \text{ Lie algebra generated by the operators } £,\ [£,h_i],\ [[£,h_i],h_j].$$

This is in any case a subalgebra; it is often equal to $ELie(\Sigma)$, and is in any case very similar to $ELie(\Sigma)$ as will be shown now. Indeed $E^sLie(\Sigma)$ is an ideal in $ELie(\Sigma)$. To see this it suffices to check that the generator of $ELie(\Sigma)$ when bracketed with the generators of $E^sLie(\Sigma)$ yield elements of $E^sLie(\Sigma)$. For the generators $£ \in ELie(\Sigma)$ this is trivial because £ is also in $E^sLie(\Sigma)$ and for the generators h_k of $ELie(\Sigma)$ we have that $[h_k,£]$ and $[h_k,[£,h_i]]$ are in $E^sLie(\Sigma)$ by definition, and that $[h_k,[[£,h_i],h_j]]=0$. This is the case because £ is second order and the h's are functions; thus $[£,h_i]$ is first order, $[[£,h_i],h_j]$ is a function and hence $[[h_k,[[£,h_i],h_j]]=0$. Now consider the quotient of $ELie(\Sigma)$ by $E^sLie(\Sigma)$

$$0 \to E^sLie(\Sigma) \to ELie(\Sigma) \to Q \to 0$$

Q is generated by the images of the commuting operators $h_1,...,h_p$ so that Q is abelian (= commutative; i.e. $[a,b]=0$ for all $a,b, \in Q$) of dimension $\leqslant p$. It follows that in particular that $ELie(\Sigma)$ is finite dimensional (resp. solvable) if and only if $E^sLie(\Sigma)$ is finite dimensional (resp. solvable). It also follows that doing Wei-Norman theory for $ELie(\Sigma)$ is practically the same as doing it for $E^sLie(\Sigma)$, the only difference being a number of initial quadratures, cf section 13 below.

9. Examples of Estimation Lie algebras

Let us look at some examples to see what kind of Lie algebras can arise as estimation Lie algebras.

9.1 Example *Wiener noise linearly observed.* This is the simplest non-zero linear system

$$\begin{aligned} dx(t) &= w(t), & x,w \in \mathbb{R} \\ dy(t) &= x(t)dt + dv(t), & y,v \in \mathbb{R} \end{aligned} \tag{9.2}$$

In this case $£ = \frac{1}{2}\frac{d^2}{dx^2} - \frac{1}{2}x^2$, $h = x$. The Lie-algebra generated by this is the four-dimensional Lie algebra with basis £, x, $\frac{d}{dx}$, 1. Cf. section 7 above for some of the calculations.

This also means that starting from an arbitrary initial density $\phi(x)$ for x at time 0 we can solve the corresponding DMZ equation

$$\frac{\partial\rho}{\partial t} = (\frac{1}{2}\frac{\partial^2}{\partial x^2}-\frac{1}{2}x^2)\rho+x\rho\dot{y}(t) \tag{9.3}$$

by means of Wei-Norman theory. The explicit equation for g_1, g_2, g_3, g_4 occurring in the Ansatz

$$\rho(x,t) = e^{g_1(t)\pounds}e^{g_2(t)x}e^{g_3(t)\frac{d}{dx}}e^{g_4(t)}\phi(x) \tag{9.4}$$

are given by (7.18) above. They are

$$\dot{g}_1 = 1,\ \dot{g}_2=\cosh(t)\dot{y}(t),\ \dot{g}_3=-\sinh(t)\dot{y}(t),\ \dot{g}_4=-g_2\sinh(t)\dot{y}(t) \tag{9.5}$$

Not surprisingly the Kalman-Bucy filter for (9.2), given by

$$\begin{aligned}\dot{P} &= 1-P^2\\ \dot{m} &= P(\dot{y}-m)\end{aligned} \tag{9.6}$$

can be easily derived from (9.3). Indeed let us try for a solution of the form

$$\rho(x,t) = r(t)e^{-\frac{(x-m)^2}{2P}} \tag{9.7}$$

i.e. an unnormalized Gaussian density, where m and P are yet to be determined functions of t. One finds

$$\frac{\partial^2}{\partial x^2}\rho(x,t) = (\frac{(x-m)^2}{P^2}-\frac{1}{P})\rho(x,t)$$

$$\frac{\partial}{\partial t}\rho(x,t) = (\frac{(x-m)\dot{m}}{P}+\frac{(x-m)^2}{2P^2}\dot{P}+\dot{r})\rho(x,t)$$

Now substitute this in (9.3) and divide by $\rho(x,t)$. There results an expression of the form $ax^2+bx+c=0$ with a,b,c dependant on time alone. For this to hold we must have $a=0$, $b=0$, $c=0$. This gives

$$a=0:\ \frac{1}{2P^2}-\frac{1}{2}\frac{\dot{P}}{2P^2}=0,\qquad \text{i.e. } \dot{P}=1-P^2$$

$$b=0:\ \frac{2m}{2P^2}+\dot{y}(t)=\frac{\dot{m}}{P}-\frac{m\dot{P}}{P^2},\ \text{i.e., using } \dot{P}=1-P^2,\ \text{on finds } \dot{m}=-Pm+P\dot{y}$$

Finally $c=0$ gives some (complicated) expression for $\dot{r}$. This shows that the solutions are in fact of the form (9.7) provided the initial density is also an unnormalized Gaussian density. The precise result for $r(t)$ (and hence the precise expression for $\dot{r}$) is largely irrelevant because of formula (4.4) for the conditional expectation $E[\phi(x)|y(s),0\leq s\leq t]=\widehat{\phi(x(t))}$ of a function $\phi(x)$ of the state.

9.8. Example. Linear systems.
Now let us consider general linear systems

$$\begin{aligned}dx &= Axdt+Bdw \quad x\in\mathbf{R}^n, w\in\mathbf{R}^m\\ dy &= Cxdt+dv \quad y\in\mathbf{R}^p, v\in\mathbf{R}^p\end{aligned} \tag{9.9}$$

The system is said to be completely reachable if the $n\times(n+1)m$ matrix

$$R(A,B) = (B\ AB\ A^2B...A^nB)$$

Consisting the $(n+1)$ $n\times m$ blocks A^iB, $i=0,...,n$, has rank n. This means for the associated control system $\dot{x}=Ax+Bu$ every element $x(1)\in\mathbf{R}^n$ can be reached from $x(0)=0$ by a suitable choice of input functions $u_1(t),...,u_m(t)$, whence the terminology. Dually, the system (9.9) is completely observable if the $(n+1)p \times n$ matrix $Q(A,C)$

$$Q(A,C)^T = (C^T \ (CA)^T \ ... \ (CA^n)^T)$$

consisting of the blocks CA^i, $i=0,...,n$ has rank n. This means that (with zero inputs) one can see from $y(t)$ whether two initial states $x(0)$ and $x'(0)$ were different or not. Whence the name of the concept.

Let us assume in (9.9) that (A,B) and (A,C) are completely reachable and completely observable pairs of matrices. Then it is not difficult to show that the estimation Lie algebra L of (9.9) is the $2n+2$ dimensional Lie algebra with basis

$$£, \frac{\partial}{\partial x_1}, \ ...., \ \frac{\partial}{\partial x_n}, \ x_1,....,x_n, \ 1$$

and £ equal to

$$£ = \frac{1}{2}\sum_{ij}(BB^T)\frac{\partial^2}{\partial x_i \partial x_j} - \sum_{i,j} A_{ij}x_j\frac{\partial}{\partial x_i} - Tr(A) - \frac{1}{2}\sum_{i,j}(C^TC)_{ij}x_ix_j \tag{9.10}$$

The $(2n+1)$-dimensional Heisenberg algebra $\mathfrak{h}_n$ with basis

$$\frac{\partial}{\partial x_1},...,\frac{\partial}{\partial x_n},x_1,...,x_n, \ 1$$

is an ideal in L, i.e. $[£,\mathfrak{h}_n] \subset \mathfrak{h}_n$, cf the tutorial in Lie algebras [28] in this volume. Hence L is solvable and using Wei-Norman theory the DMZ equation can be integrated for arbitrary initial densities. Of course this requires that we be able to integrate the simpler equation

$$\frac{\partial \rho}{\partial t} = £\rho \quad , \quad \rho(0) = \pi(x) \tag{9.11}$$

All the others, i.e. $\partial\rho/\partial t = x_i\rho$ and $\partial\rho/\partial t = \partial\rho/\partial x_i$ are trivial, but £, cf (9.10), is itself a fairly complicated operator. One natural thing to try would be to try to do Wei-Norman theory once more, which leads tot the study of the Lie algebra generated by the various terms occurring in £, i.e. the $\frac{\partial^2}{\partial x_i \partial x_j}$, $x_j\frac{\partial}{\partial x_i}$, x_ix_j, 1. The constant $Tr(A)$ does not matter (as it commutes with everything). It turns out to be slightly more convenient to consider instead the operators

$$\begin{array}{ll} \frac{\partial^2}{\partial x_i \partial x_j} & , \ i,j=1,...,n; \ i \leqslant j \\ x_ix_j & , \ i,j=1,...,n; \ i \leqslant j \\ x_i\frac{\partial}{\partial x_j} & , \ i \neq j; \ i,j=1,....,n \\ x_i\frac{\partial}{\partial x_i} + \frac{1}{2} & , \ i=1,...,n \end{array} \tag{9.12}$$

It is a straightforward exercise to check these form in fact the basis of a $(2n^2+n)$ dimensional Lie algebra (of differential operators). As a matter of fact this Lie algebra is ismorphic to the Lie algebra of all $2n \times 2n$ sympletic matrices. i.e. the Lie algebra $sp_n(\mathbb{R})$

$$sp_n(\mathbb{R}) = \{ \begin{bmatrix} AB \\ CD \end{bmatrix} \in \mathbb{R}^{2n \times 2n}: \ A,B,C,D \in \mathbb{R}^{n \times n}, B = B^T, C = C^T, A = -D^T \} \tag{9.13}$$

This Lie algebra is simple and (thus) Wie-Norman theory only works for small time intervals. And indeed there are operators $M \in sp_n(\mathbb{R})$ such that $\frac{dx}{dt} = Mx$, $x \in \mathbb{R}^{2n}$, or, equivalently $\frac{\partial \rho}{\partial t} = \tilde{M}p$, where

$\tilde{M}$ is the differential operators corresponding to the $2n \times 2n$ matrix M, has finite escape time phenomena (for suitable initial conditions). However, it does turn out that this isomorphism of Lie algebras can be used effectively to integrate equations like (9.11). I shall not discuss this aspect further here, refering to [30] for all details. In the case of example 9.1. equation (9.11) is of course that of the harmonic oscillator which is well studied.

Again, there is of course a link of the DMZ equation with the Kalman-Bucy filter, and the latter can be deduced from the former [24]. As a matter of fact all Kalman-Bucy filters sort of fit together to define one large representation of the Lie algebra ls_n with basis

$$1;\ x_1,\ldots,x_n;\ \frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_n};\ \frac{\partial^2}{\partial x_i \partial x_j}, i,\ j=1,\ldots,n, i \leq j; \tag{9.14}$$

$$x_i x_j,\ i,j=1,\ldots,n, i \leq j;\ x_i \frac{\partial}{\partial x_j} i,j=1,\ldots,n$$

of dimension $2n^2+3n+1$, into the Lie-algebra of all vectorfields on $\mathbb{R}^N, N=\frac{1}{2}n^2+\frac{3}{2}n+1$. Cf. [24,29], for more details on this and the link with the so called oscillator or Segal-Shale-Weil representation of $sp_n(\mathbb{R})$. This is in fact the representation given by the operators (9.12) and it is precisely the fact that we know how to integrate this representation together with the availability of the matrix copy (9.13) of (9.12) which enables one to integrate equation like 9.11 [24].

9.15. Example. The cubic sensor
This is the system

$$dx = dw \qquad x,w \in \mathbb{R}$$
$$dy = x^3 dt + dv \quad y,v \in \mathbb{R}$$

In a certain sense this is the simplest nonlinear system. (The quadratic sensor where the observation equation is $dy = x^2 dt + dv$ instead is perhaps still simpler; on the other hand the noninjectivity of $x \mapsto x^2$ seems to be asking for additional trouble; as it turns out both are 'equally hard'). This example has a substantial literature devoted to it and has a reputation of being quite hard to handle [11]. A first indication of why this might be the case is the structure of its estimation Lie algebra.

Let $W_1 = \mathbb{R} < x, \frac{d}{dx} >$, i.e. the associative algebra generated by the symbols x and $\frac{d}{dx}$ subject to the relations suggested by the notation used, viz. $[\frac{d}{dx}, x] = (\frac{d}{dx})x - x(\frac{d}{dx}) = 1$. Consider W_1 as a Lie algebra under the commutator bracket $[A,B]=AB-BA$. In other words W_1 is the Lie algebra of all differential operators in x (any order) with polynomial coefficients:

$$W_1 = \{ \sum_{i,j}^{<\infty} c_{ij} x^i \frac{d^j}{dxj} : c_{i,j} \in \mathbb{R} \} \tag{9.16}$$

The estimation Lie algebra of (9.15) is the Lie-algebra generated by the two operators $\frac{1}{2}\frac{d^2}{dx^2} - \frac{1}{2}x^6$, x^3. It turns out that this is everything. I.e.

$$ELie(\text{cubic sensor}) = W_1 \tag{9.17}$$

This a very large infinite dimensional algebra and, as it turns out, cf below, a rather nasty one from the point of view of exact finite-dimensional filtering. For a proof of (9.17), cf [25].

9.18. Example. The quadratic sensor

$$dx = dw \qquad x, w \in \mathbb{R}$$
$$dy = x^2 dt + dv \quad y, v \in \mathbb{R}$$

Let W'_1 be the subalgebra of W_1 consisting of all differential operators of even total degree in x and $\frac{d}{dx}$ together:

$$W'_1 = \{ \sum_{i,j}^{<\infty} c_{ij} x^i \frac{d^j}{dx^j} \in W_1 \ : c_{ij} = 0 \ \text{ if } \ i \neq j \ \text{mod} 2 \} \tag{9.19}$$

The estimation Lie algebra of the quadratic sensor is generated by $\frac{d^2}{dx^2} - x^4$ and x^2. It turns out that

$$ELie(\text{quadratic sensor}) = W'_1 \tag{9.20}$$

9.12. Example. The weak cubic sensor

$$dx = dw \qquad x, \in \mathbb{R}$$
$$dy = (x + \epsilon x^3)dt + dv, \quad y, \ v \in \mathbb{R}$$

This time the generators of the estimation Lie algebra are $\frac{d^2}{dx^2} - (x + \epsilon x^3)^2$, $x + \epsilon x^3$. If $\epsilon = 0$ we have example (9.1) back. If $\epsilon \neq 0$ we have again [23], [25].

$$ELie \text{ (weak cubic sensor)} = W_1 \ , \ \text{ if } \ \epsilon \neq 0. \tag{9.22}$$

9.23. Example.

$$dx_1 = dw_1, \ dx_2 = x_1^2 dt \qquad , \ x_1, x_2, w \in \mathbb{R}$$
$$dy_1 = x_1 dt + dv_1, dy_2 = x_2 dt + dv_2 \ , \ y_i, v_i \in \mathbb{R}$$

Generalizing W_1, let W_n be the Lie algebra of all differential operators (any order) in n variables with polynomials coefficients, i.e.

$$W_n = \mathbb{R} < x_1, \ldots, x_n; \ \frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n} > \ = \{ \sum_{\alpha,\beta}^{<\infty} c_{\alpha,\beta} x^\alpha \frac{\partial^\beta}{\partial x^\beta} : c_{\alpha,\beta} \in \mathbb{R} \}$$

where $\alpha = (a_1, \ldots, a_n)$ and β are multiindices, $a_i \in \mathbb{N} \cup \{0\}$ and where x^α and $\frac{\partial^{|\beta|}}{\partial x^\beta}$ are short for $x_1^{a_1} \ldots x_n^{a_n}$ and $\frac{\partial^{b_1}}{\partial x_1^{b_1}} \ldots \frac{\partial^{b_n}}{\partial x_n^{b_n}}$ respectively. The generators of the estimation Lie algebra are in this case

$$\frac{1}{2} \frac{\partial^2}{\partial x_1^2} - x_1^2 \frac{\partial}{\partial x_2} \ , \ x_1, \ x_2$$

and we find (again) [25]

$$ELie(\text{example 9.23}) = W_2 \tag{9.24}$$

9.25. Example.

$$dx = x^3 dt + dw, \qquad x, w \in \mathbb{R}$$
$$dy = x dt + dv, \qquad y, v \in \mathbb{R}$$

This time the generators are $\frac{1}{2}\frac{d^2}{dx^2} - \frac{7}{2}x^2 - x^3\frac{d}{dx}$ and x and (again)

$$ELie(\text{example } 9.25) = W_1 \tag{9.26}$$

9.27. Example (mixed linear - bilinear type).

$$dx_1 = dw_1 \qquad x_1, w_1 \in \mathbb{R}$$
$$dx_2 = x_1 + x_1 dt + x_1 dw_2 \quad x_2, w_2 \in \mathbb{R}$$
$$dy = x_2 dt + dv \qquad y, v \in \mathbb{R}$$

The generators are $\frac{1}{2}\frac{\partial^2}{\partial x_2^2} + \frac{1}{2}x_1^2\frac{\partial^2}{\partial_2^2} - x_1\frac{\partial}{\partial x_2} - \frac{1}{2}x_2^2$ and x_2, and (again) [25]

$$ELie(\text{example } 9.27) = W_2 \tag{9.28}$$

Thus it would appear that the infinite dimensional Lie algebras W_n have a habit of appearing very often. This seems indeed to be the case. In fact I have conjectured

9.29. Conjecture.
Consider stochastic systems (1.1)-(1.2) with polynomial f, G and h. Then for almost all f, G, h the estimation Lie algebra of (1.1)-(1.2) will be equal to W_n.

Here 'almost all' means an open dense set in the space of all triples (f, G, h) of vector and matrix valued functions of the right dimensions topologized by means of the natural topology on their sequences of coefficients. No proof of this conjecture appears to be in sight.

As we shall see the occurrence of W_n as the estimation Lie algebra of a stochastic system is bad news from the point of view of existence of exact recursive finite dimensional filters. So at least a few examples (besides the linear ones) where something else turns up would be welcome. One large class of such examples are the systems

$$dx = f(x)dt + G(x)dw, \qquad x \in \mathbb{R}^n, w \in \mathbb{R}^m$$
$$dy = h(x)dt + dv, \qquad y \in \mathbb{R}^p, v \in \mathbb{R}^p \tag{9.30}$$

with the extra conditions that f, G and h are real analytic and that $f(0) = 0, G(0) = 0$, cf [25]. Another example is

9.31. Example [25]

$$dx_1 = dw, \qquad x_1, w \in \mathbb{R}$$
$$dx_2 = x_1^2 dt, \qquad x_2 \in \mathbb{R}$$
$$dy = x_1 dt + dv, \qquad y, v \in \mathbb{R}$$

Here of course filtering for x_1, i.e. calculating $\hat{x}_1$, is straightforward by means of the Kalman filter. Finding $\hat{x}_2$ is a totally different matter. (NB: by the Ito formula $d(\frac{1}{3}x_1^3) = x_1^2 dx_1 + x_1 dt = x_1^2 dw + x_1 dt$ which does not have much to do with the equation for x_2). The generators in this case are

$$£ = \frac{1}{2}\frac{\partial^2}{\partial x_1^2} - x_1^2\frac{\partial}{\partial x_2} - \frac{1}{2}x_1^2 \ , \ x_1$$

and the Lie algebra generated by these two operators is the infinite dimensional Lie algebra with basis

$$£,\ x_1\frac{\partial^i}{\partial x_2^i},\ \frac{\partial}{\partial x_1}\frac{\partial^i}{\partial x_2^i},\ \frac{\partial^i}{\partial x_2^i},\ i=0,1,2,... \tag{9.32}$$

This is an infinite dimensional Lie algebra but still a solvable one in the appropriate sense, cf below.

9.33. Example. 'Polynomials' in n independent Brownian motions
The word 'polynomial' is here used in a very loose sense. What I mean are certain stochastic systems directly inspired by polynomials. E.g. the system corresponding to the 'polynomial' w^n where w is a one dimensional Wiener process is

$$\begin{aligned} dx_1 &= dw,\ dx_2=x_1dw,\ dx_3=x_2dw,...,\ dx_n=x_{n-1}dw \\ dy &= x_1dt+dv \end{aligned} \tag{9.34}$$

Here (9.34) are intended as Ito equations. Thus x_n does *not* really correspond to w^n. Indeed a system which has w^n as a second state variable, x_2, is

$$\begin{aligned} dx_1 &= dw \\ dx_2 &= nx_1^{n-1}dw+n(n-1)x_1^{n-2}dt \\ dy &= x_1dt+dv \end{aligned} \tag{9.35}$$

This one, obtained by adding the state variable, $x_2=x_1^n$, which is a function of the state variables already present, has an estimation Lie algebra which is isomorphic to the original system without the extra state variable. This is a general fact. Cf. section 10 below. It is a curious and somewhat remarkable fact that the estimation Lie algebra of (9.34) is also isomorphic to the oscillator Lie algebra. For this it is really necessary to 'unravel' w^n as in (9.34) by means of the intermediate states $x_2,...,x_{n-1}$: the system

$$\begin{aligned} dx_1 &= dw \\ dx_n &= nx_1^{n-1}dw \end{aligned}$$

does not have its estimation Lie algebra isomorphic to the oscillator Lie algebra. Instead, for $n\geqslant 4$, its estimation Lie algebra is a much more complicated infinite dimensional affair. (For $n=3$ the estimation Lie algebra is 5 dimensional with basis £, x_1, $\frac{\partial}{\partial x_1}+3x_1^2\frac{\partial}{\partial x_2}$, 1, $\frac{\partial}{\partial x_2}$.)

More generally, consider systems 'corresponding to a polynomial'

$$P = \sum c_\alpha w^\alpha \tag{9.36}$$

where $\alpha=(a_1,...,a_n)$ is a multiindex and w^α is short for $w_1^{a_1}...w_n^{a_n}$. These systems are defined as follows. Let $\mu=(m_1,...,m_n)$ be such that $c_\alpha=0$ unless $a_k\leqslant m_k$ for $k=1,...,n$. Consider now the system with state variables p; $x_{i,r}$, $i=1,...,n$; $r=1,...,m_i$ given by

$$\begin{array}{ccccc} dx_{1,1}=dw_1 & & dx_{2,1}=dw_2 & & dx_{n,1}=dw_n \\ dx_{1,2}=x_{1,i}dw_1 & & dx_{2,2}=x_{2,1}dw_2 & & dx_{n,2}=x_{n,1}dw_n \\ \cdot & & \cdot & \cdots & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ dx_{1,m_1}=x_{1,m_1-1}dw_1 & & dx_{2,m_2}=x_{2,m_2-1}dw_2 & & dx_{n,m_n}=x_{n,m_n-1}dw_n \end{array} \tag{9.37}$$

$$dp=\sum_\alpha\sum_{i=1}^n c_\alpha(x_{1,a_1}...x_{n,a_n})x_{i,a_i}^{-1}dw_i$$

with the observation equations

$$dy_1 = x_1 dt + dv_1, \ldots, dy_n = x_n dt + dv_n \tag{9.38}$$

Then the Estimation Lie algebra of (9.37)-(9.38) is generated by the 2nd order operator £ and $x_1, \ldots, x_n$. It is finite dimensional with as basis

$$£,\ £_1 = [£, x_1], \ldots, £_n = [£, x_n],\ x_1, \ldots, x_n,\ 1. \tag{9.39}$$

Besides the $£_i = [£, x_i]$ the nonzero brackets are $[£, £_i] = x_i$ and $[£_i, x_i] = 1$. For a (rather computational) proof cf [31]. There could well be, indeed should be, a conceptual proof of this, but so far the arguments I have in this direction are unconvincing. This is a solvable Lie algebra and so Wei-Norman theory is applicable. Indeed the estimation Lie algebra is the same one as the one of n-independent completely observed Wiener processes, so the Wei-Norman equations are the same as those in that case. The individual operators £ and $£_i$ though are very different and quite complicated, cf the particular example 9.41 below. So it remains to deal with the equations

$$\frac{\partial \rho}{\partial t} = £\rho \quad \text{and} \quad \frac{\partial \rho}{\partial t} = £_i \rho \tag{9.40}$$

It turns out that for both £ and $£_i$ the individual terms making up these operators themselves generate a solvable, albeit, as a rule, infinite dimensional Lie algebra. Here is a particular example

9.41. Example. System associated to the Brownian polynomial $w_1^2 + w_1 w_2 + w_2^3$

$dx_1 = dw_1$, $dx_2 = dw_2$, $dx_{2,2} = x_2 dw_2$, $dp = x_1 dw_1 + x_1 dw_2 + x_2 dw_1 + x_{22} dw_2$

$dy_1 = x_1 dt + dv_1$, $dy_2 = x_2 dt + dv_2$

In this (still quite simple) case the operators are equal to

$$\begin{aligned} £ = {} & \frac{1}{2}\frac{\partial^2}{\partial x_1^2} + \frac{1}{2}\frac{\partial^2}{\partial x_2^2} + \frac{1}{2}x_2^2\frac{\partial^2}{\partial x_{22}^2} + \frac{1}{2}\{(x_1 + x_{22})^2 + (x_1 + x_2)^2\}\frac{\partial^2}{\partial p^2} \\ & + (x_1 + x_2)\frac{\partial}{\partial x_1}\frac{\partial}{\partial p} + (x_1 + x_{22})\frac{\partial}{\partial x_2}\frac{\partial}{\partial p} + x_2\frac{\partial^2}{\partial x_2 \partial x_{22}} + (x_1 x_2 + x_2 x_{22})\frac{\partial^2}{\partial x_{22}\partial p} \\ & + x_2\frac{\partial}{\partial p} + \frac{\partial}{\partial x_{22}} + \frac{\partial}{\partial p} - \frac{1}{2}x_1^2 - \frac{1}{2}x_2^2 \end{aligned}$$

$$£_1 = [£, x_1] = \frac{\partial}{\partial x_1} + (x_1 + x_2)\frac{\partial}{\partial p}$$

$$£_2 = [£, x_2] = \frac{\partial}{\partial x_2} + (x_1 + x_{22})\frac{\partial}{\partial p} + x_2\frac{\partial}{\partial x_{22}}$$

Besides the cases where the W_n occur as estimation Lie algebras we thus have several classes of systems which yield infinite dimensional but solvable Lie algebras, viz. real analytic systems (9.30) such that $f(0)=0$, $G(0)=0$, and systems like example 9.31. A further class is furnished by the systems which arise when a identification problem for linear systems is considered as a filtering problem (cf section 15 below). This filtering problem is then nonlinear, reflecting the essential nonlinearity of identification, but its estimation Lie algebra is again solvable but infinite dimensional. Finally in tackling the equations (9.40) which form part of the filtering of Brownian polynomials again infinite dimensional solvable Lie algebras arise.

This then is ample motivation for investigating whether something like infinite dimensional Wei-Norman theory exists. This is a topic which we will take up below in section 13.

9.42. Example. Beneš systems [4].

$$dx = f(x)dt + dw$$
$$dy = xdt + dv$$
$$f_x + f^2 = ax^2+bx+c$$

I.e. we have Wiener noise with an extra nonlinear drift term given by $f(x)dt$; this drift term is required to be such that $\frac{df}{dx}+f^2$ is a quadratic polynomial in x. In this case also the estimation Lie algebra is the oscillator Lie algebra.

9.43. Some open questions.
All in all very little is known about estimation Lie algebras. It seems very difficult to find other (non-trivial and interesting) examples of finite dimensional estimation Lie algebras. Besides the linear and Beneš case and the (new) case of 'Brownian polynomials' (9.37) very few examples are known (Wing Wong, Marcus, Lie, Ocone, . . .). In particular it is unknown whether finite dimensional simple Lie algebras can ever arise as estimation Lie algebras.

10. Invariance properties of the estimation Lie algebra.
This section discusses some questions much related to the subjects discussed so far and what will still come. But these questions are not essential for the remainder of this paper, and this section is somewhat more abstract than the remainder of this paper. It can be skipped if desired.

The estimation Lie algebra $ELie(\Sigma)$ is clearly an invariant of a system Σ: (1.1)-(1.2) in the following sense. If Σ and Σ' are two system of the type (1.1)-(1.2) on $\mathbb{R}^n$ respectively, and $\phi:\mathbb{R}^n\to\mathbb{R}^n$ is an isomorphism of Σ and Σ' (a transformation of variables), then the estimation Lie algebras of Σ and Σ' are isomorphic. Here, as we are dealing with Ito differential equations, isomorphism means that under the change of variables $x'=\phi(x)$ the equation

$$dx = f(x)dt + G(x)dw \tag{10.1}$$

transforms into the equation

$$dx' = f'(x')dt + G'(x')dw \tag{10.2}$$

under the Ito formula (transformation rule) which says that $\phi(x)=x'$ satisfies the differential equations

$$dx'^k = d\phi^k(x)\sum = (\sum_i \frac{\partial\phi_k}{\partial x_i}f_i(x) + \frac{1}{2}\sum_{i,j}\frac{\partial^2\phi}{\partial x_i\partial x_j}(GG^T)_{ij})dt + \sum_i\frac{\partial\phi_k}{\partial x_i}G_i(x)dw \tag{10.3}$$

where $f_i(x)$ is the i-th component of $f(x)$ and $G_i(x)$ is the i-th row of $G(x)$. Substituting $x=\phi^{-1}(x')$ in (10.3) must then yield the equations (10.2). On a general manifold M the transformation rule (10.3) has no real meaning and then to talk about equivalent stochastic systems it is better to start with systems in Fisk-Stratonovič form.

In addition the DMZ equation

$$\frac{\partial\rho}{\partial t} = \pounds\rho + \sum_{i=1}^{p} h_i(x)\rho\dot{y}_i \tag{10.4}$$

can be gauge transformed, $\tilde\rho = e^{\phi(x)}\rho$ to give an equation

$$\frac{\partial\tilde\rho}{\partial t} = \tilde\pounds\tilde\rho + \sum_{i=1}^{p} h_i(x)\tilde\rho\dot{y}_i \tag{10.5}$$

with $\tilde\pounds = e^{\phi(x)}\pounds e^{-\phi(x)}$ Correspondingly there is an isomorphism of the Lie algebra L generated by $\pounds,h_1,...,h_p$ and the Lie algebra generated bij $\tilde\pounds$ and $h_1,...,h_p$. This isomorphism is given by

$A\in L\mapsto e^{\phi(x)}A\phi^{-\phi(x)}$ in $\tilde{L}$. Sometimes non-isomorphic dynamical systems are gauge equivalent in this sense. This happens e.g. for the Beneš systems 9.42 and corresponding 1-dimensional linear systems. Cf. [2] for material on 'invariants' in this context.

More generally (than the case of, isomorphims, i.e. changes of variables), if $\Sigma\rightarrow\Sigma'$ is a morphism of stochastic dynamical systems then there is a corresponding homomorphism of their estimation Lie algebras. In particular consider a system (10.1) and let us add an additional state variable p which is a function of the original state: $p=\phi(x_1,\ldots,x_n)$. The resulting Ito differential equation for p is

$$dp = \sum_k \phi_{(k)}f_k dt + \sum_{k,l,j}\frac{1}{2}\phi_{(k,l)}G_{k,j}G_{l,j}dt + \sum_{k,l}\phi_{(k)}G_{k,l}dw_l \tag{10.6}$$

where

$$\phi_{(k)} = \frac{\partial\phi}{\partial x_k},\ \phi_{(k,l)} = \frac{\partial^2\phi}{\partial x_k\partial x_l}$$

Let L be the estimation Lie algebra of (10.1) with observations $dy_j=h_j(x)dt+dv_j$ and let $\tilde{L}$ be the estimation Lie algebra of (10.1) complemented with the p-equation (10.6) above and the same observations. Then the isomorphism $L\rightarrow\tilde{L}$ is induced by $\frac{\partial}{\partial x_i}\rightarrow\frac{\partial}{\partial x_i}+\phi_{(i)}\frac{\partial}{\partial p}$, $x_i\mapsto x_i$ and the inverse isomorphism $\tilde{L}\rightarrow L$ is induced by $x_i\mapsto x_i$, $\frac{\partial}{\partial x_i}\mapsto\frac{\partial}{\partial x_i}$, $\frac{\partial}{\partial p}\mapsto 0$. (These are, as is easily checked, homomorphisms of associative algebras $\mathbb{R}<x_1,\ldots,x_n;\ \frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_1},\frac{\partial}{\partial p}> \longrightarrow \mathbb{R}<x_1,\ldots,x_n,\frac{\partial}{\partial x_1},\ldots,\frac{\partial}{\partial x_n}>$.

Now, if

$$£ = \frac{1}{2}\sum_{i,j,k}\frac{\partial^2}{\partial x_i\partial x_j}G_{ik}G_{jk} - \sum_i\frac{\partial}{\partial x_i}f_i - \frac{1}{2}\sum_j h_j^2 \tag{10.7}$$

then the corresponding operator for the system extended with (10.6) is

$$\tilde{£} = £ + \frac{1}{2}\sum_l\Big(\sum_k\phi_{(k)}G_{k,l}\Big)^2 + \sum_{k,l,i}\frac{\partial^2}{\partial x_i\partial p}G_{il}\phi_{(k)}G_{k,l} - \frac{1}{2}\frac{\partial}{\partial p}\sum_{k,l,j}\phi_{(k,l)}G_{k,j}G_{l,j} - \sum_k\frac{\partial}{\partial p}\phi_{(k)}f_k \tag{10.8}$$

Replacing $\frac{\partial}{\partial x_i}$ with $\frac{\partial}{\partial x_i}+\phi_{(i)}\frac{\partial}{\partial p}$ in (10.7) yields the extra terms

$$\frac{1}{2}\sum_{i,j,k}\phi_{(i)}\frac{\partial}{\partial p}\frac{\partial}{\partial x_j}G_{ik}G_{jk} = \frac{1}{2}\sum_{i,j,k}\frac{\partial^2}{\partial x_j\partial p}\phi_{(i)}G_{ik}G_{jk} - \frac{1}{2}\sum_{i,j,k}\phi_{(i,j)}G_{ik}G_{jk}\frac{\partial}{\partial p} \tag{10.9}$$

$$\frac{1}{2}\sum_{i,j,k}\frac{\partial}{\partial x_i}\phi_{(j)}\frac{\partial}{\partial p}G_{ik}G_{jk} \tag{10.10}$$

$$\frac{1}{2}\sum_{i,j,k}\phi_{(i)}\phi_{(j)}G_{ik}G_{jk}\frac{\partial^2}{\partial p^2} \tag{10.11}$$

$$-\sum_i\phi_{(i)}\frac{\partial}{\partial p}f_i \tag{10.12}$$

Now the first term of the RHS of (10.9) and (10.10) combine to give the third term of the RHS of (10.8). The second term of the RHS of (10.9) is the fourth term of the RHS of (10.8); expression (10.11) is equal to the second term of the RHS of 10.8 and finally (10.12) is the last term of the RHS of 10.8. Thus $\frac{\partial}{\partial x_i}\mapsto\frac{\partial}{\partial x_i}+\phi_{(i)}\frac{\partial}{\partial p}$, $x_i\mapsto x_i$ does indeed take $£$ into $\tilde{£}$.

Inversely one can ask to what extent the estimation Lie algebra $ELie(\Sigma)$ determines the system Σ. Certainly nonisomorphic systems can have isomorphic estimation Lie algebras; e.g. the Beneš systems and one dimensional systems or the Brownian polynomial systems and the systems $dx_1 = dw_1,...,dx_n, = dw_n$, $dy_1 = x_1 dt + dv_1,...,dy_n = dt + dv_n$. But of course $ELie(\Sigma)$ is not just an abstract Lie algebra; it comes together with a natural (linear infinite dimensional) representation (on a suitable space of unnormalized densities). The more sensible question is therefore whether the pair ($ELie(\Sigma)$, corresponding representation) determines Σ up to isomorphism.

This is also not true as shown by the gauge equivalence of the Beneš systems with the corresponding linear ones. It would be nice to know more about just how much information is contained in the pair ($ELie(\Sigma)$, representation).

The question is akin to the following one for control systems of the form

$$\dot{x} = f(x)dt + \sum_{i=1}^{m} g_i(x)u_i, \quad x \in M, u_i \in \mathbb{R} \tag{10.13}$$

$$y = h(x), y \in \mathbb{R}^p$$

Associated to (10.13) we have the Lie algebra generated by f and the g_i; denote this one by $Lie(\Sigma)$. It also comes together with a natural representation. Indeed f and the g_i are vectorfields and hence are first order differential operators acting on functions on M, in particular the functions $h_1,...,h_p$. Let $V(\Sigma)$ be the smallest subspace of $\mathfrak{F}(M)$ containing $h_1,...,h_p$ and stable under D_f, D_{g_i}. Then $V(\Sigma)$ carries a linear representation of $Lie(\Sigma)$ and the question is to what extent the pair $(Lie(\Sigma), V(\Sigma))$ characterizes Σ up to isomorphism. A first problem here is to recover the manifold M from $(Lie(\Sigma), V(\Sigma))$. This is strongly related to the following question which has been studied in [59]. Given an n-dimensional manifold M let $V(M)$ be the Lie algebra of all vectorfields on M. Can one recover M from $V(M)$?

The reason I bring up these questions is the following. As we shall see in section 12 existence of an exact finite dimensional recursive filter implies the extence of a homomorphism of Lie algebras $ELie(\Sigma) \to V(M)$ where $V(M)$ is the Lie algebra of vectorfields on the manifold on which the filter for $\phi(x)$ exists (this filter is assumed to be of minimal dimension among all filters for $\phi(x)$.)

The questions briefly raised above relate to the inverse problem: given a homomorphism $ELie(\Sigma) \to V(M)$ for some M, plus suitable supplementary structure, does there exist a corresponding filter. Without additional hypotheses this is certainly not true cf e.g. the contributions by Krishnaprasad-Marcus and Hazewinkel-Marcus in [34].

11. The BC principle

We have already seen one set of reasons why $ELie(\Sigma)$ is important for filtering questions: If it is finite dimensional and solvable we can apply Wei-Norman theory; if it is at least finite dimensional we have in any case Wei-Norman theory for small time. If it is infinite dimensional but still solvable there are potential approximation schemes, cf below. Let me now describe a second reason why the estimation Lie algebra $ELie(\Sigma)$ 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 cartoon character; the BC stand for Brockett and Clark who first enunciated it, [9].

Suppose we have a filter (2.3)-(2.4) 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 (2.3) are vectorfields on M. Let $V(M)$ denote the Lie algebra of smooth vectorfields on M. Then the BC principle states the following:

11.1. BC Principle.
If (2.3)-(2.4) is a minimal filter for a statistic $\widehat{\phi(x_t)}$ of a system Σ then $£ \mapsto \beta_1,...,h_p \mapsto \beta_p$ defines an antihomomorphism of Lie algebras from $ELie(\Sigma)$ 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 $A,B \in L_1$.

11.2. Example.
Consider again the simplest nonzero linear system (9.2). It is linear so there is the Kalman-Bucy filter for the conditional state $\hat{x}$. This filter is (cf (9.6) and (2.2))

$$dP = (1-P^2)dt,\ dm = P(dy - m\ dt).$$

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

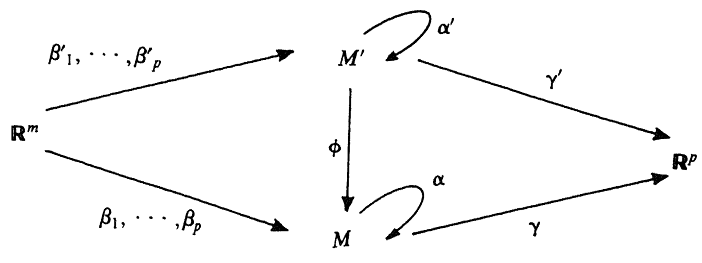
$$\alpha = (1-P^2)\frac{\partial}{\partial P} - Pm\frac{\partial}{\partial m},\ \beta = P\frac{\partial}{\partial m}$$

where we have used the '$\frac{\partial}{\partial x_i}$ notation'; cf the tutorial on differentiable manifolds [27] in this volume. 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 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 well known (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 [63], in particular for systems of the form

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

Here the picture produced by the theorem is the following commutative diagram



(The theorem assert 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 $V(M)$, cf [27].

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 machine given by the DMZ-equation (5.7) and the output map (4.4). So we are in a similar situation as above but with M' infinite dimensional. A proof in this case follows from considerations in [39].

The fact that in the case of the BC-principle we get an antihomomorphism instead of a homomorphism 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 antihomorphism of Lie-algebras $ELie(\Sigma) \rightarrow V(M)$ 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 [34]. 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 (4.4).

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 the mapped on $m \in M$, obviously the isotropy subalgebra of $ELie(\Sigma)$ at m' will go into the isotropy subalgebra of $V(M)$ at m. The isotropy subalgebra $V(M)_{(x)}$ of $V(M)$ at $x \in M$ consists of all vectorfields which are zero at x. The isotropy subalgebra of $ELie(\Sigma)$ at $x \in M$ is $ELie(\Sigma) \cap V(M)_{(x)}$. For the case of finite dynamical systems there are positive results, [41], 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 in [59].

A more representation theoretic way of looking at things, already touched upon in section 10 above, is as follows. Both $ELie(\Sigma)$ 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.

12. The cubic sensor

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. Apart from that it is simple; i.e. the quotient algebra $W_n/\mathbb{R}.\,1$ is simple.

12.1. Theorem ([25]).
Let $\alpha: W_n \rightarrow V(M)$ or $W_n/\mathbb{R}.1 \rightarrow V(M)$ be a homomorphism or antihomomorphism of Lie algebras, where M is a finite dimensional manifold. Then $\alpha = 0$.

The original 12 page proof of this result, [25], was long and computational. Another much shorter proof has more recently been given by Toby Stafford. Perhaps inevitably this more conceptual proof is based on the Stone- von Neuman result on the impossibility of representing the 3-dimensional Heisenberg Lie algebra $\mathfrak{h}_1$ with basis x, $\frac{d}{dx}$, 1, $[\frac{d}{dx}, x] = 1$ by means of finite dimensional matrices in

such a way that 1 is represented by the unit matrix.
Now consider again the cubic sensor, i.e. the one-dimensional system

$$dx = dw,\ dy = x^3 dt = dv \tag{12.2}$$

consisting of Wiener noise, cubically observed with further independent noise corrupting the observations. As noted before (example 9.14)

$$ELie(\text{cubic sensor}) = W_1. \tag{12.3}$$

Now suppose that we have a finite dimensional filter for some conditional statistic $\widehat{\phi(x_t)}$ of the cubic sensor. By the BC-principle 11.1 if follows that there is an antihomomorphism of Lie-algebras $W_1 = ELie$ (cubic sensor) $\xrightarrow{\alpha} V(M)$. By theorem 12.1 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 [26].

A similar statement holds for all other systems whose estimation Lie algebras are isomorphic to a W_n or $W_n/\mathbb{R}$, and in fact also for the quadratic sensor whose estimation Lie algebra is 'the even subalgebra' W'_1 of W_1. As we have seen W_n occurs often as an estimation algebra so often exact finite dimensional recursive filters will not exist. This makes approximate recursive filters doubly important, a point to which we will return several times below.

13. Infinite dimensional Wei-Norman theory

We have already seen a number of cases where estimation Lie algebras were infinite dimensional and were claimed to be solvable in a suitable infinite dimensional sense. The precise definition of this is as follows.

13.1. Definition

Let L be a (finite or infinite dimensional) Lie algebra (over a field k; take $\mathbb{R}$ for convience). Then L is solvable if there exists a sequence of ideals $I_1, I_2, \ldots, I_n, \ldots$ such that $\bigcap_n I_n = 0$ and such that each quotient algebra L/I_n is finite dimensional and solvable (as a finite dimensional Lie algebra).

This is a good concept in the context of Wei-Norman theory because as we shall see in a few moments the Wei-Norman equations are well behaved with respect to quotients (and not at all well with respect to subalgebras).

13.2. Example

Consider again the estimation Lie algebra L of example 9.31. Recall that it had a basis consisting of the operators

$$\pounds = \frac{1}{2}\frac{\partial^2}{\partial x_1^2} - x_1^2\frac{\partial}{\partial x_2} - \frac{1}{2}x_1^2,\ x_1\frac{\partial^i}{\partial x_2^i},\ \frac{\partial^i}{\partial x_2^i},\ \frac{\partial}{\partial x_1}\frac{\partial^i}{\partial x_2^i}, i = 0,1,2,\ldots$$

Let $\mathfrak{h}_i$ be the subspace spanned by the operators

$$x_1\frac{\partial^j}{\partial x_2^j},\ \frac{\partial^j}{\partial x_2^j},\ \frac{\partial}{\partial x_1}\frac{\partial^j}{\partial x_2^j},\quad j \geq i$$

It is easy to check that the $\mathfrak{h}_i$ are ideals, and not difficult to show that the quotients $L/\mathfrak{h}_i$ are finite dimensional and solvable.

13.3. Wei-Norman theory revisited

Now let us consider Wei-Norman theory again in the setting of a Lie-algebra L and an ideal $\mathfrak{a}$ in it. Let $\dim L = n$, $\dim \mathfrak{a} = n-k$, $0<k<n$. Choose a basis $A_1,...,A_k,A_{k+1},...,A_n$ for L in such a way that $A_{k+1},...,A_n$ are all in $\mathfrak{a}$ (i.e. they form a basis for the subspace $\mathfrak{a}$).

Recall from section 7, cf formulas (7.4) and 7.5), that the Wei-Norman equations are of the form

$$\dot{g}_r + \dot{g}_1 h_{1r}(g) + ... + \dot{g}_n h_{nr}(g) = u_r \tag{13.4}$$

where g is short for $(g_1,...,g_n)$ and where the h_{jr} are such that

$$\sum_{k_1,...,k_{j-1}} \text{function of}(g_1,...,g_{r-1})\ ad_{A_1}^{k_1}...ad_{A_{j-1}}^{k_{j-1}}(A_j) = \sum h_{jr}(g) A_r \tag{13.5}$$

It follows first of all from (13.5) that

$$h_{jr}(g) \text{ only depends on } \quad g_1,...,g_{r-1} \tag{13.6}$$

Second, let $j>k$, so that $A_j \in \mathfrak{a}$. Then $ad_{A_1}^{k_1}...ad_{A_{j-1}}^{k_{j-1}}(A_j) \in A$ and it follows that

$$\text{if } j>k,\ h_{jr}(g) = 0 \text{ for } r \leqslant k \tag{13.7}$$

Now take $r \leqslant k$ in (13.6). Taking account of (13.6) and (13.7) we see that (13.4) is of the form

$$\dot{g}_r + \dot{g}_1 h_{1r}(g_1,...,g_{r-1}) + ... + \dot{g}_k h_{kr}(g_1,...,g_{r-1}) = u_r \tag{13.4}$$

Thus in the situation under consideration of an ideal $\mathfrak{a}$ in L and a basis adapted to this situation the Wei-Norman equations for $g_1,...,g_k$ only involve the $g_1,...,g_k$, and can thus be written down, analysed and solved without any regard for the remainder of the Lie algebra.

As is now readily seen from what has been said, the Wei-Norman equations for $g_1,...,g_k$ are in fact the Wei-Norman equations belonging to the quotient algebra $L/\mathfrak{a}$ with respect to the basis $A_1 + \mathfrak{a}$, $A_2 + \mathfrak{a}$,...,$A_k + \mathfrak{a}$ of this quotient.

We are now ready to consider the infinite dimensional case. So suppose that L has ideals $I_1,I_2,...,I_n,...$ such that $\cap I_n = 0$ and each L/I_n is dimensional and solvable. There is than a basis of L of the form

$$A_1,...,A_{k_1},\ A_{k_1+1},...,A_{k_2}, \qquad k_1<k_2<... \tag{13.5}$$

such that

$$A_{k_n+1},\ A_{k_n+2},.. \cdots \tag{13.6}$$

is a basis for I_n. We are, as usual, interested in solving an equation

$$\frac{\partial \rho}{\partial t} = \sum A_i \rho u_i(t) \tag{13.7}$$

where, in our case at least, the sum on the right is a finite one and where we can assume that the A_i are part of the basis 13.5 (otherwise write out the operators in (13.7) in terms of that basis.) We can in effect assume that, say, the sum runs from $i=1$ to $i=k_n$. This does not mean, however, that the solution can be written in terms of the $e^{g_j(t)A_j}$, $1 \leqslant j \leqslant k_n$; the higher $A's$ will also tend to occur via higher brackets.

The idea now is to try an infinite product

$$e^{g_1 A_1}...e^{g_{k_1} A_{k_1}} e^{g_{k_1+1} A_{k_1+1}}...e^{g_{k_2} A_{k_2}}...e^{g_{k_n} A_{k_n}}...\pi(x) \tag{13.8}$$

as Ansatz. By the remarks made about quotients above, the infinite system of Wei-Norman for the g_i is such that

$$\dot{g}_1,...,\dot{g}_{k_1} \quad \text{only involve} \quad g_1,...,g_{k_1}\ ;\ u_1,...,u_{k_1}$$

$$\dot{g}_1,...,\dot{g}_{k_2} \quad \text{only involve} \quad g_1,...,g_{k_2} : u_1,...,u_{k_2}$$

$$...$$

So that in any case the infinite system of Wei-Norman equations makes sense. We can now calculate a sequence of densities

$$e^{g_1 A_1} ... e^{g_{k_1} A_{k_1}} \pi(x)$$

$$...$$

$$e^{g_1 A_1} ... e^{g_{k_1} A_{k_1}} ... e^{g_{k_n} A_{k_n}} \pi(x)$$

$$...$$

The question remains whether this sequence of densities converges in one sense or another. This is a largely uninvestigated mather. Scattered through the remainder of this article there are a number of comments on this.

There is more to be said about Wei-Norman type theory in infinite dimensional contexts. A number of estimation Lie algebras occur as solvable subalgebras of a Lie algebra of the form $\mathbb{R}[z_1,...,z_r]\otimes L$ where L is a finite dimensional Lie algebra of differential operators in $x_1,...,x_n$. The meaning of this symbolism is as follows. Let $A_1,...,A_s$ be a basis of L. Then a basis for $\mathbb{R}[z_1,...,z_r]\otimes L$ is formed by the differential operators

$$z^\alpha A_i, \quad \alpha=(a_1,...,a_r), \ a_j \in \mathbb{N}\cup\{0\} \text{ a multi index, } i=1,...,s$$

And the bracket between these basis elements is given by

$$[z^\alpha A_i, \ z^\beta A_j] = z^{\alpha+\beta} \ [A_i, A_j]$$

These are called current algebras and have been investigated in both the mathematics and the physics literature to a considerable extent [36-38,40,58]. The point here is that though $\mathbb{R}[z]\otimes L$ is infinite dimensional over $\mathbb{R}$ it is finite dimensional over the ring of functions $\mathbb{R}[z]$. Thus the natural object in which solutions of equations

$$\frac{\partial \rho}{\partial t} = \sum_{\alpha,i} u_{\alpha,i}(t) z^\alpha A_i$$

will live in something like the group of functions in $z_1,...,z_n$ to G, where G is the Lie group of L. In slightly more concrete terms this means that the Ansatz now becomes

$$\rho(t,x) = e^{g_1 A_1} ... e^{g_s A_s} \pi(x,z)$$

where now the g_i are supposed to be functions of both t and $z_1,...,z_r$, polynomial in z in this particular context.

The estimation Lie algebra of a linear system identification problem is of the 'subalgebra of current algebra' type, cf below in section 15. In [42] there are some more details on Wei-Norman theory and identification from this particular point of view.

14. The weak cubic sensor

Recall that this is the one dimensional system

$$dx = dw \tag{14.}$$

$$dy = (x + \epsilon x^3)dt + dv$$

with $\epsilon \neq 0$. Recall also that its estimation Lie algebra is equal to W_1 for $\epsilon \neq 0$ (and for $\epsilon = 0$ it reduc of course to the oscillator Lie algebra). Thus by the arguments of section 12 above it follows th there are no exact finite dimensional recursive filters for any statistic of the weak cubic sensor. C the other hand it is intuitively hard to believe that the filter for $\epsilon = 0$ will not give something of approximation. Also when one actually calculates $ELie$(weak cubic sensor) one notices that the higher order differential operators and higher order polynomials and products of these appear with high powers of ϵ in front of them, suggesting that neglecting these will (i) not matter too much, and (ii) give us something finite dimensional to work with. This can be made precise as follows. Consider ϵ in (14.1) as a parameter. Calculate $ELie$ (weak cubic sensor) in the usual way but with ϵ as a polynomial variable, i.e. calculate $ELie$ as a subalgebra of $\mathbb{R}<\epsilon, x, \frac{d}{dx}>$, $[\epsilon, \frac{d}{dx}]=0$, i.e. treat ϵ as a second variable (besides x). Now introduce the extra rule $\epsilon^m=0$ for $m \geqslant n$. Then the resulting algebra is finite dimensional and solvable. Let us call this the estimation Lie algebra modulo ϵ^n, $ELie\ mod(\epsilon^n)$. Technically speaking we are considering $ELie \otimes_{\mathbb{R}[\epsilon]} \mathbb{R}[\epsilon]/\epsilon^n$. That $ELie\ mod(\epsilon^n)$ is finite dimensional and solvable in this case is an instance of a much more general phenomenon.

14.2. Theorem ([23,29]).
Let Σ_ϵ be a stochastic system of the form

$$dx = (Ax+\epsilon P_A(x)dt+(B+\epsilon P_B(x)dw \quad x\in\mathbb{R}^n,\ w\in\mathbb{R}^m$$
$$dy = (Cx+\epsilon P_C(x)dt+dv \qquad g\in\mathbb{R}^p,\ v\in\mathbb{R}^p$$

where $P_A(x), P_B(x), P_C(x)$ are polynomial in x. Then $ELie(\Sigma_\epsilon)$ $mod(\epsilon^r)$ is finite dimensional and solvable for all r.

14.3. Example.
$ELie$ (weak cubic sensor) $mod(\epsilon^2)$ is 14 dimensional with basis

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

The next question is: do these finite dimensional solvable "quotients" of $ELie(\Sigma_\epsilon)$ calculate anything. Let us do the following. The solution of the DMZ equation will also depend on ϵ. Let us look for formal power series in ϵ solutions of the form

$$\rho(x,t,\epsilon) = \rho_0(x,t)+\epsilon\rho_1(x,t)+\epsilon^2\rho_2(x,t)+... \qquad (14.4)$$

Then the Wei-Norman equations for $ELie(\Sigma_\epsilon)$ $mod(\epsilon^r)$ precisely compute the first r coefficients of (14.4), i.e. $\rho_0(x,t),...,\rho_{r-1}(x,t)$. This is quite simple and is in fact related of the second group of ideas re infinite dimensional Wei-Norman theory as discussed in section 13 above. (It also has aspects of the first group of ideas and is in fact a sort of amalgam of both).

The next question is whether the formal series (14.4) will converge. This is again a matter which still needs a great amount of investigation. The series does converge for the weak cubic sensor and the weak quadratic one. It is also pleasing to note that the resulting $mod(\epsilon^2)$ filter for the weak quadratic sensor sensor already performs much better than the extended Kalman filter [35,48]. Further it can be shown that the series (14.4) is always an asymptotic series in the technical sense of the word. On the other hand there are arguments indicating that the series will not converge for the weak quartic sensor and higher. Thus there appears much to do and it may well be fruitful to take into account that $\rho(x,t,\epsilon)$ only matters up to a normalization factor $r(t,\epsilon)$, which can be chosen as one pleases.

15. 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

$$dx = Axdt+Bdw,\ dy = Cxdt+dv \qquad (15.1)$$

Now, however, the coefficients in A,B,C are unknown and also have to be estimated from the observations $y(t)$. That is the system (15.1) has to be identified. It is easy to turn this into a filtering problem by adding the (stochastic) equations

$$dA = 0, \ dB = 0, \ dC = 0 \tag{15.2}$$

(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.

15.3. Observation
The estimation Lie algebra of the system (15.1)-(15.2) is a sub-Lie-algebra of the current Lie algebra $ls_n \otimes \mathbb{R}[A,B,C]$ where $\mathbb{R}[A,BC]$ stands for the ring of polynomials in the indeterminates a_{ij},b_{kl},c_{rs}. Here ls_n is the $(2n^2+3n+1)$-dimensional Lie algebra with basis (9.14); i.e. the Lie algebra of all differential operators of total degree $\leqslant 2$ in x_i and $\frac{\partial}{\partial x_j}$, so that $ls_n = \{\Sigma c_{\alpha,\beta} x^\alpha \frac{\partial^{|\beta|}}{\partial x^\beta} : c_{\alpha,\beta}=0$ if $|\alpha|+|\beta|>2\}$. The Lie algebra ls_n contains a subalgebra isomorphic to sp_n (cf section 9 above), so this does not yet prove that $ELie$((15.1)-(15.2)) is solvable. But as a matter of fact it is. Thus the ideas and considerations of the previous two sections can be brought into play. Some initial results exploiting the current algebra based ideas briefly discussed in the second half of section 13 above are contained in [42]. In this rather special case it turns out that the higher approximations (the zero-th approximation is simply the family of Kalman-Bucy filters parameterized by A,B,C also discussed in section 9 above) have to do with sensitivity equations: sensitivities of the output $y(t)$ 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 (15.1) 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 $\mathbb{R}^n$ [32,33]. 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 more possible approach, which is in the spirit of the ideas of section 14 and the first half of section 13, rather than based on current algebra ideas. For the filters giving $\hat{x},\hat{A},\hat{B},\hat{C}$ for problem (15.1)-(15.2) one expects $\hat{x}$ to move fast relative $\hat{A},\hat{B},\hat{C}$. Thus it would make sense to consider a system

$$\begin{aligned} dx &= (A_0+\epsilon A_1)xdt+(B_0+\epsilon B_1)dw, \ dy = -(C_0+\epsilon C_1)dt+dv \\ dA_1 &= 0, \ dB_1 = 0, \ dC_1 = 0 \end{aligned} \tag{15.4}$$

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)).

16. Asymptotic expansions and approximate homomorphisms. The marking approach.

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 13 and 14 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 ([48]). for regularity results etc. cf e.g. [3,12,43,52] and references in these papers.

Let us consider a control system of the form

$$\dot{x} = f(x) + \Sigma u_i g_i(x) \tag{16.1}$$

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 (16.1) 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 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 may cases the behaviour of (16.1) 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 $f, g_1, ..., g_m$. More precisely brackets of order n of a nearby changed system differ by terms of the form $\Delta(n_1)^{i_1} ... \Delta(n_r)^{i_r}$ with $i_1 n_1 + ... + i_r n_r \geqslant n$ wher $\Delta(n_k)$ symbolizes an upper bound for the uncertainty, i.e., the changes, at level n_k in the Taylor expansions. Let me also stress that, in spite of the word 'Taylor expansion' in the previous sentences I am attempting to formulate global ideas of approximation and definitely not local ones around one point. If f is a function of one variable depicted as a graph in the plane, then a piecewise linear approximation of f with rounded corners would be a low order illustration of what is intended here. Spline approximation takes us up (at least) one order higher.

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 introduced in the study of hypoellipticity [22,61], which are now also being investigated in control and system theoretic contexts [15,60].

Let me explain the context partly. Consider a homomorphism of a system of type (16.1) into an other one. That means a differentiable map $\phi: M \to M'$ where M and M' are the state space manifolds of (16.1) and (16.1)′ such that ϕ takes the vectorfields $f, g_1, ..., g_m$ into the vectorfields $f', g'_1, ..., g'_m$. (If there is also an output map $h: M \to \mathbb{R}^p$, then of course we must also have $h' \circ \phi = h$). Inversely if $Lie(\Sigma)$ is the Lie algebra generated by $f, g_1, ..., g_m$ and $\alpha: Lie(\Sigma) \to Lie(\Sigma')$ is a homomorphism of Lie algebras taking isotropy subalgebras into isotropy subalgebras, then, at least locally, there exists a ϕ such that '$d\phi = \alpha$'.

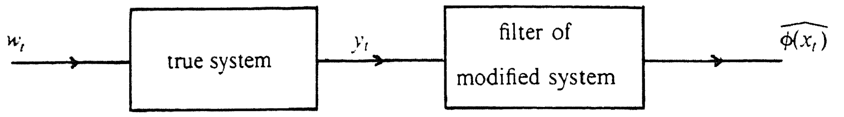
A first idea of an approximate homomorphism α of level m is that if σ resp. τ are elements in $Lie(\Sigma)$ which can be obtained by taking iterated brackets of the $f, g_1, ..., g_m$ at most a_σ resp. a_τ times and $a_\sigma + a_\tau \leqslant m$, then $\alpha[\sigma, \tau] = [\alpha(\sigma), \alpha(\tau)]$. This corresponds precisely to introducing markers, i.e. writing $zg_1, ..., zg_m$ and saying that a map $\alpha: Lie_z(\Sigma) \to Lie(\Sigma')$ is an approximate homomorphism of level m if it induces a homomorphism of Lie algebras $Lie_z(\Sigma) \, mod z^{m+1} \to Lie_z(\Sigma') \, mod z^{m+1}$

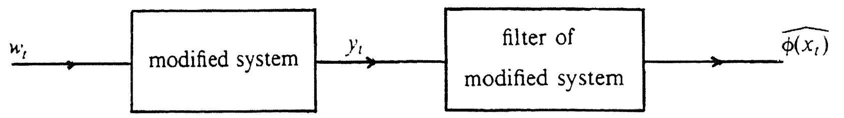
Thus in filtering theory, which can be seen as the theory of trying to find (approximate) homomorphism of the infinite dimensional system given by the DMZ equation (5.6) or (6.2) and the output map (4.4) to finite dimensional systems, it would seem natural to look at the Lie algebra of operators $ELie_z(\Sigma)$ generated by the operators

$$z_0 \pounds, \; z_1 h_1, ..., z_p h_p$$

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.

Let me conclude this section with an argument indicating that the approximation scheme indicated above should work. Let the true system be Σ and assume it is stable in the sense that modifying the $f, g_1, \ldots, g_m$ in the manner indicated does not change the input-output behaviour much. Models of real systems are expected to be like this simply because they must still do a reasonable job if some of the measured coefficients are slightly wrong, which is inevitable. Now suppose moreover that I can modify the higher derivatives of the $f, g_1, \ldots, g_n$ in such a way that there exists an exact finite dimensional filter for the modified system Σ'. Thus the situation is as depicted below





Now the filter of the modified system is also expected to be (input-output) stable. Indeed it will almost have to be that in order to do its job. Then the two composed systems shown above will be close in input-output sense, which means precisely that we have constructed an approximate filter for the true system.

Now, as far as I can see, for a given system Σ, there will as a rule not exist an approximation (in the given sense) which suddenly has a finite dimensional solvable estimation Lie algebra. Or even an infinite dimensional solvable one. In that case there certainly are lots of filters but it is less clear what quantities they filter for and it also remains to be investigated thoroughly whether they give usuable approximation to a $\rho(x,t)$, cf section 13, 14 above.

Thus it does not seem that the argument given above can be used to prove that the marking approach gives good approximate filters, but the argument certainly provides positive indications.

17. Removing outliers

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

$$dx = f(x)dt + G(x)dw, \quad dy = h(x)dt + dv. \tag{17.1}$$

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 the 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 (17.1) consider the modified system

$$dx = f(x)dt + Gdw, \quad y = e^{-a\|x\|^2}h(x)dt + dv \tag{17.2}$$

where $a > 0$ is a small parameter. Note that the only thing which (17.2) does with respect to (17.1) is to discount large y observations.

Now consider the estimation Lie algebra of the system (17.2).

17.3. Theorem

If $e^{-a\|x\|^2}$ *is differentially algebraically independent of* f, G, h *then the estimation Lie algebra of (17.2) is (infinite dimensional) solvable. To be more precise it is finite dimensional and solvable* $\text{mod}(a^i e^{-ja\|x\|^2}, i+j \geqslant n)$ *for all n.*

Thus the yoga of the previous sections can again be applied and the behaviour of the resulting 2-parameter family of filters as a goes to zero and n goes to infinity could be studied.

1. L. ARNOLD, *Stochastische Differentialgleichungen,* Oldenbourg Verlag, 1972. (Translation: Stochastic differential equations, Wiley, 1974).
2. J.S. BARAS, Group invariance methods in nonlinear filtering of diffusion processes, In: [34], 565-572.
3. 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.
4. V.E. BENES, Exact finite dimensional filters for certain diffusions with nonlinear drift, *Stochastics* **5** (1981), 65-92.
5. 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.
6. R.W. BROCKETT, Nonlinear systems and nonlinear extimation theory, In: [34], 441-478.
7. R.W. BROCKETT, Classification and equivalence in estimation theory, *Proc, 1979 IEEE Conf. on decision and Control,* Ft. Lauderdale, December 1979.
8. R.W. BROCKETT, Remarks on finite dimensional nonlinear estimation, In: C. Lobry (ed.), Analyse des systèmes, Astérisque **76** (1980), Soc. Math. de France.
9. 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.
10. R.S. BUCY, J.M.F. MOURA (eds.), Nonlinear stochastic problems, Reidel, 1983.
11. R.S. BUCY, J. PAGES, A priori error bounds for the cubic sensor problem, *IEEE Trans. Automatic Control* **AC-24** *(1979), 948-953.*
12. M. CHALEYAT-MAUREL, D. MICHE., Hypoellipticity theorems and conditional laws, *Z. Wahrsch. und verw. Geb.* **65** (1984), 573-597.
13. S.D. CHIKTE, J.T.-H. LO, Optimal filters for bilinear systems with nilpotent Lie algebras, *IEEE Trans, Automatic Control* **AC-24** (1979), 948-953.
14. 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.
15. P.E. CROUCH, Solvable approximations to control systems, *SIAM J. Control and Opt.* **32** (1984), 40-54.
16. F.E. DAUM, Exact nonlinear recursive filters and separation of variables, *Proc. 25-th CDC,* Athens, Dec. 1986.
17. M.H.A. DAVIS, Exact finite dimensional nonlinear filters, *Proc. 24-th CDC, Las Vegas December 1985.*
18. M.H.A. DAVIS, Pathwise nonlinear filtering, In: [34], 505-528.
19. M.H.A. DAVIS, S.I. MARCUS, An introduction to nonlinear filtering, In: [34], 565-572.
20. T.E. DUNCAN, Probability densities for diffusion processes with applications to nonlinear filtering theory, *Ph. D. thesis,* Stanford, 1967.
21. M. FUJISAKI, G. KALLIANPUR, H. KUNITA, Stochastic differential equations for the nonlinear filtering problem, *Osaka J. Math.* **1** (1972), 19-40.
22. R.W. GOODMAN, Nilpotent Lie groups: structure and applications to analysis, LNM **562,**

Springer, 1976.
23. M. HAZEWINKEL, On deformations, approximations and nonlinear filtering, *Systems Control Lett.* **1** (1982), 29-62.
24. 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.
25. M. HAZEWINKEL, S.I. MARCUS, On Lie algebras and finite dimensional filtering, *Stochastics* **7** (1982), 29-62.
26. 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.
27. M. HAZEWINKEL, A tutorial introduction to differentiable manifolds and calculus on differentiable manifolds, *This volume.*
28. M. HAZEWINKEL, A tutorial on Lie algebras, *This volume.*
29. M. HAZEWINKEL, Lie algebraic methods in filtering and identification, In: M. Mebkhout, R. Sénéor (eds.), *Proc. VIII-th Int. Cong. Math. Physics,* Luming 1986, World Scientific, 1987, 120-137 (also to appear in *Proc. 1-st World Cong. Bernouilli Society,* Tashkent, 1986, VNU, 1988).
30. M. HAZEWINKEL, Identification, filtering, the symplectic group, and Wei-Norman theory. To appear in *Proc. 2-nd workshop on the Road-vehicle system and related mathematics,* Torino, 1987, Reidel-Teubner, 1988.
31. M. HAZEWINKEL, On filtering and polynomials in *n* independent Brownian motions, *Acta Scientifica,* Caracas, to appear.
32. M. HAZEWINKEL, Moduli and canonical forms for linear dynamical systems II: the topological case, *Math. Systems Th.* **10** (1977), 363-385.
33. M. HAZEWINKEL, (Fine) moduli (spaces) for linear systems: what are they and what are they good for, In: C.I. Byrnes, C.F. Martin (eds.), *Geometric methods for linear systems theory* (Harvard, June 1979), Reidel, 1980, 125-193.
34. M. HAZEWINKEL, J.C. WILLEMS, (eds.), Stochastic systems: the mathematics of filtering and identification and applications, Reidel, 1981.
35. M. HAZEWINKEL, An introduction to nilpotent approximation filtering. To appear in *Proc. 2-nd European Symp. on Math. in Industry* (ECMI 2), Oberwolfach, 1987, Reidel-Teubner, 1988.
36. R.C. HERMANN, Lie algebras and quantum mechanics, Benjamin, 1970.
37. R.C. HERMANN, Infinite dimensional Lie algebras and current algebras, In: *Lect. Notes in Physics* **6,** Springer, 1969.
38. R.C. HERMANN, Current algebras, the Sugawara model and differential geometry, *J. Math. Phys.* **11:** 6, (1970), 1825-1829.
39. O.B. HIJAB, Finite dimensional causal functionals of brownian motion, In [6], 425-436.
40. A.A. KIRILOV, Local Lie algebras, *Usp. Mat. Mauk.* **81,** 57-76.
41. A.J. KRENER, On the equivalence of control systems and the linearization of nonlinear systems, *SIAM J. Control* **11,** (1973), 670-676.
42. P.S. KRISHNAPRASAD, S.I. MARCUS, M. HAZEWINKEL, Current algebras and the identification problem, *Stochastics* **11** (1983), 65-101.
43. 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.
44. H.J. KUSHNER, On the dynamical equations of conditional probability functions with application to optimal stochastic control theory, *J. Math. Anal. Appl.* **8** (1964), 332-344.
45. H.J. KUSHNER, Dynamical equations for optimal nonlinear filtering, *J. Diff. Equations* **3** (1967), 179-190.
46. R.S. LIPTSER, A.N. SHIRYAYEV, *Statistics of Random Processes I.* New York: Springer-Verlag, 1977.
47. 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), 277-279.

48. S.I. MARCUS, Algebraic and geometric methods in nonlinear filtering, *SIAM J. Control Opt.* **22** (1984), 817-844.
49. S.I. MARCUS, A.S. WILLSKY, Algebraic structure and finite dimensional nonlinear estimation, *SIAM J. Math. Anal.* **9** (1978), 312-327.
50. S.I. MARCUS, Optimal nonlinear estimation for a class of discrete-time stochastic systems, *IEEE Trans. Automatic Control* **AC-24** (1979), 297-302.
51. S.I. MARCUS, S.K. MITTER, D. OCONE, Finite dimensional nonlinear estimation for a class of systems in continuous and discrete time, In: *Analysis and Optimization of Stochastic Systems,* O.L.R. Jacobs et al. eds, Academic Press, 1980, 387-406.
52. D. MICHEL, Régularité des lois conditionelles en théorie du filtrage non-linéaire et calcul des variations stochastiques, *J. Funct. Anal.* **14** (1981), 8-36.
53. S.K. MITTER, On the analogy between mathematical problems of non-linear filtering and quantum physics, *Richerche di Automatica* **10** (1980), 163-216.
54. S.K. MITTER, Nonlinear filtering and stochastic machanics, In: [34], 479-504.
55. R.E. MORTENSEN, Optimal control of continuous time stochastic systems, *Ph. D thesis,* Berkeley, 1966.
56. D. OCONE, *Topics in non-linear filtering,* Thesis, M.I.T., June 1980.
57. E. PARDOUX, D. TALAY, Discretization and simulation of stochstic differential equations, *Acta Appl. Math.* **3** (1982), 182-203.
58. K.R. PARTHASARATHY, K. SCHMIDT, A new method for constructing factorizable representations for current group and current algebras, *Comm. Math. Phys.* **50** (1976), 167-175.
59. L.E. PURSELL, M.E. SHANKS, The Lie algebra of a smooth manifold, *Proc. Amer. Math. Soc.* **5** (1954), 468-472.
60. CH. ROCKLAND, Intrinsic nilpotent approximation, *Acta Appl. Math.* **8** (1987), 213-270.
61. L..P. ROTHSCHILD, E.M. STEIN, Hypoelliptic differential operators and nilpotent groups, *Acta Math.* **137** (1976), 247-320.
62. S. STEINBERG, Application of the Lie algebraic formulas of Baker, Campbell, Hausdorff and Zassenhaus to the calculation of explicit solutions of partial differential equations, *J. Diff. Equations* **26** (1977), 404-434.
63. H.J. SUSSMANN, Existence and uniqueness of minimal realizations of nonlinear systems, *Math. Syst. Theory* **10** (1977), 349-356.
64. 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.
65. M. ZAKAI, On the optimal filtering of diffusion processes, *Z. Wahrsch. verw. Geb.* **11** (1969), 230-243.