ECONOMETRIC INSTITUTE

ON IDENTIFICATION OF LINEAR SYSTEMS AND THE ESTIMATION LIE-ALGEBRA OF THE ASSOCIATED NONLINEAR FILTERING PROBLEM

B. HANZON AND M. HAZEWINKEL

REPORT 8122/M

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

ON IDENTIFICATION OF LINEAR SYSTEMS AND THE ESTIMATION LIE-ALGEBRA OF THE ASSOCIATED NONLINEAR FILTERING PROBLEM.

Ben Hanzon Michiel Hazewinkel
Dept. Math., Erasmus Univ.Rotterdam
P.O. Box 1738, 3000 DR ROTTERDAM
The Netherlands

ABSTRACT. In this paper we are concerned with linear (stochastic) systems like $dx_t = (Ax_t+B_1u_t)dt + B_2dw_t$, $dy_t = Cx_tdt + dv_t$ or (more or less equivalent) ARMAX models and the problem of identifying A,B_1,B_2,C on the basis of observations of the inputs u_t and outputs y_t. In particular we are interested in the problem of whether there exists a machine (a system) driven by the instanteneous observations (u_t,y_t) which as output produces a "best" estimate of the unknown system (recursive estimation). And even more particularly we are interested on how big (in state space dimension) such a machine must be. Introducing additional state space parameters $a_{ij}, b^1_{k\ell}, b^2_{k\ell}, c_{r,s}$ and equations $da_{ij} = db^1_{k\ell} = db^2_{k\ell} = dc_{rs} = 0$ converts the original problem into a *nonlinear* filtering problem. For such problems the socalled estimation Lie algebra contains a good deal of information (on how hard the problem is), and this is what we try to exploit in this paper.

CONTENTS.

1. INTRODUCTION

Consider a continuous time linear state space model (system)

$$(1.1)\quad dx_t = Ax_t dt + B_1 u_t dt + B_2 dw_t,\ dy_t = Cx_t + Ddv_t,\ x_t \in \underline{\underline{R}}^n$$

or a discrete time ARMAX model

$$(1.2)\qquad \sum_{i=o}^{p} A_i y(t-i) = \sum_{i=1}^{m} D_i u(t-i) + \sum_{i=o}^{q} B_i w(t-i)$$

or the discrete time analogue of (1.1) or the continuous time analogue of (1.2) (Do not confuse the A,B's and D's in (1.1) with those in (1.2); they refer to rather different things). In this paper we are concerned with the problem of identifying optimally the various matrices in (1.1) (resp. 1.2) given observations of the deterministic inputs u_t and the outputs y_t. More precisely we are interested in finding a machine which does this in a recursive way (i.e. on line). Such a machine should proceed as follows: at time t - 1 there is available a model M(t-1) and perhaps an additional memory vector R(t-1) and on the basis of the state (M(t-1),R(t-1)) and the new data u(t), y(t) the new "best" model M(t) and new memory vector R(t) can be calculated by some formula Ψ. There are a number of rather obvious desiderata: e.g. Ψ must not depend on time and R(t) must (in dimension) remain bounded in time. In other words the identification machine we are looking for (in the discrete time case) is itself a system (most probably nonlinear) of the form

$$(1.3)\qquad \xi(t) = \Psi(\xi(t-1),u(t),y(t)),M(t) = \gamma(\xi(t)),\xi(t) \in \underline{\underline{R}}^N$$

and a continuous time identification machine could look like

$$(1.4)\qquad d\xi_t = \alpha(\xi_t)dt + \beta_1(\xi_t,u_t)dt + \beta_2(\xi_t)dy_t,$$

$$M(t) = \gamma(\xi_t),\ \xi(t) \in \underline{\underline{R}}^N$$

One particular question we would like to raise in this context is: "how big must N be"; i.e. we are interested in the minimal realization theory of the map

$$\begin{Bmatrix}\text{sequences or functions}\\ \text{of input/output data}\end{Bmatrix} \longmapsto \begin{Bmatrix}\text{best linear model of}\\ \text{given dimensions}\end{Bmatrix}$$

Of course the minimal model for this input/output map may involve more general spaces (manifolds) than the $\underline{\underline{R}}^N$.

One of the first issues is then "identifiability": can one distinguish between all models of type (1.1) (resp. (1.2)) on the basis of input/output data alone. In the case of the models (1.1) this is obviously not the case: there are superfluous parameters to be removed. The next question is finite identifiability how many data do we need to distinguish the various candidate models. This also provides a lower bound for N (provided we do not alow pathological (continuous) maps like the Peano curve (from the unit interval onto the unit square); it suffices to require Ψ and γ in (1.3) to be algebraic or differentiable to avoid this). This is the topic of section 2 below.

Section 3 then continues with some remarks and some precise (but open) suggestions concerning the possible structure of an identification machine (1.3) or (1.4).

In section 4 below we discuss the nonlinear filtering approach to identification. This amount to considering the entries of A, B_1, B_2, C, D in (1.1) as additional state variables and adding the equations

$$dA = dB_1 = dB_2 = dC = dD = 0 \tag{1.5}$$

(where if E is a matrix of variables $dE = 0$ stands for $de_{ij} = 0$ for all the entries e_{ij}). Adding (1.5) to (1.1) gives us a (rather large) nonlinear filtering problem, namely that of finding the best estimate of the state vector (x, A, B_1, B_2, C, D) given the observations(y_s, u_s), $0 \leqq s \leqq t$. To every nonlinear filtering problem there is associated a certain Lie-algebra called the estimation Lie algebra and there is a philosophy (an almost theorem), due to [Brockett-Clark 1978] algebra to Lie algebras of vectorfields correspond to exact filters for certain statistics of the system; cf. also [Hazewinkel-Marcus 1980], [Marcus-Mitter-Ocone 1978] and quite a few papers in [Hazewinkel-Willems, 1981] for more information on this. In our particular case of a filtering problem coming from an identification problem the estimation Lie algebra turns out to be pro-finite dimensional (cf. [Hazewinkel-Marcus 1980] for this notion and what it implies) which suggests that there will be "sufficiently many" statistics which can be computed recursively.

A priori the use of the identification Lie algebra seems restricted to finding out things about the existence or nonexistence of *exact* filters. This is probably not the case and the last two sections of this paper (section 5 on Gaussian approximation; section 6 on the Extended Kalman Filter) provide positive evidence that it also contains information (when considered not as a bare Lie algebra but as a Lie algebra with a given representation) on approximate filters.

2. FINITE IDENTIFIABILITY OF ARMA MODELS.

2.1 The set-up. The class of models we are interested in this section is the class of models (1.2) with zero inputs; i.e. we are interested in all models

$$\sum_{i=o}^{p} A_i y(t-i) = \sum_{i=o}^{q} B_i w(t-i), \tag{2.1}$$

where $y(t) \in \underline{\underline{C}}^m$, $t \in \underline{\underline{Z}}$ (outputs) and the $u(t) \in \underline{\underline{C}}^m$ are random vectors, independently and identically distributed with mean zero and covariance Σ (positive definite hermitian). The integers p,q are supposed given and in additive we assume that A_o is nonsingular and the causality (stability) condition

$$\det \sum_{k=o}^{p} A_k z^k \neq 0 \text{ for all } z \in \underline{\underline{C}} \text{ of norm} \leqq 1 \tag{2.2}$$

We are interested in the identifiability of this class of models, i.e. in the question of whether two different models of type (2.1) give different probability distributions on the to-be-observed outputs and given a identifiable subclass we are interested in finite identifiability which roughly means that we want to be able to decide this on the basis of the probability properties of a finite collection of outputs and how many are needed. In particular we want to know how many of the cross covariances (which by stationarity are independed of t)

$$\Gamma_k = E\ y_t y^*_{t-k} \tag{2.3}$$

must be known in order to determine all others.

Let $A(z) = \Sigma A_k z^k$, $B(z) = \Sigma B_j z^j$ be the generating functions of the sequences of matrices $(A_o,\ldots,A_p)$, $(B_o,\ldots,B_q)$ Form the (backwards) transfer function and expand it as a power series

$$T(z) = A(z)^{-1} B(z) = \sum_{k=1}^{\infty} T_k z^{k-1} \tag{2.4}$$

and in terms of the T's we are interested in how many of them we need to know (in terms of p and q) so that all others are determined.

The rank of the associated block H nkel matrix H of T(z)

$$H = \begin{pmatrix} T_1 & T_2 & T_3 & \cdots \\ T_2 & T_3 & T_4 & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix} \tag{2.5}$$

is of course finite and equal to the MacMillan degree n of the ARMA model (2.1).

2.6. Relations between Hankel matrix and ARMA model. Let $r = \max(p,q)$. Then the observability Kronecker indices of the system (F,G,H) are $\leqq r$. There is even, as is wellknown, a representation of the ARMA model for which the row degrees of the partitioned matrix $[A(z):B(z)]$ are equal to the observability Kronecker indices.

This gives $n \leqq mr$. Assume first $p \geqq q$ and consider the "shifted" transfer function $z^{p-q}T(z)$ which corresponds to an ARMA model

(2.7) $$\sum_{k=o}^{p} A_k \tilde{y}(t-k) = \sum_{j=o}^{q} B_j u(t-j-p-q)$$

with MacMillan degree $\tilde{n} \leqq mp$. We shall need a similar upper bound for the case $p < q$. In this case one considers the transfer functions

(2.8) $$\sum_{j=o}^{\infty} T_{j+k} z^j = A(z)^{-1} B^{(k)}(z)$$

where $B^{(o)}(z) = B(z)$ and $B^{(k)}(z) = z^{-1}\{B^{(k-1)}(z) - A(z)A(0)^{-1}B^{(k-1)}(0)\}$

One has that degree $B^{(k)}(z) \leqq q-k$ for $k \leqq q-p$ so that the MacMillan degree of $\Sigma\, T_{j+q-p} z^j$ is $\leqq mp$. Combining these two observations we see that the rank of the Hankel matrix

(2.11) $$\begin{pmatrix} T_{q-p+1} & T_{q-p+2} & \cdots \\ T_{q-p+2} & T_{q-p+3} & \cdots \\ \vdots & \vdots & \end{pmatrix}$$

with $T_i = 0$ if $i < 0$ has rank $\leqq mp$.

The next result we need is the following wellknown continuation lemma (due to Kalman) of partial realization theory. For a proof cf e.g. [Hazewinkel 1980].

2.12. <u>Lemma</u>. Let $T_o, T_1, \ldots, T_{\ell+j+1}$ be a sequence of $m \times m$ matrices. For all r,s with $r+s \leqq \ell+j+1$ write $H_{r,s}$ for the block Hankel matrix with the $r+1$ block rows $(T_o, \ldots, T_s)$, $(T_1, \ldots, T_{s+1})$, ..., $(T_r, \ldots, T_{r+s})$. Then if $\mathrm{rank}(H_{\ell,j}) = \mathrm{rank}(H_{\ell+1,j}) = \mathrm{rank}(H_{\ell,j+1})$ there is a unique continuation $T_{\ell+j+2}, \ldots$ such that $\mathrm{rank}(H_{\infty,\infty}) = \mathrm{rank}(H_{\ell,j})$.

There is also a partial converse. If $\mathrm{rank}(H_{\ell,j}) = \mathrm{rank}(H_{\ell,j+1}) < \mathrm{rank}(H_{\ell+1,j}) = n$ then $T_{\ell+j+2}$ such that $\mathrm{rank}(H_{\ell+1.j+1}) = n$ is not unique and there are even "free" parameters. Cf. [Kalman 1979] for more information concerning this in the scalar input, scalar output case.

2.12. <u>Finite Identifiability</u>. Now consider an ARMA(p,q)-model. Because the MacMillan degree of $(A(z), B^{(q-p)}(z))$

(resp. $A(z)$, $z^{p-q}B(z)$) is $\leqq pm$ if $q \geqq p$ (resp. $p \geqq q$) and all the observability Kronecker indices are $\leqq p$ it follows by the continuation lemma that the submatrix

$$(2.13)\qquad \begin{pmatrix} T_{q-p+1} & \cdots & T_{q+pm-p} \\ \vdots & & \vdots \\ T_{q+1} & \cdots & T_{q+pm} \end{pmatrix}$$

of (2.11) suffices to determine all of (2.11). It follows that the $T_o,\ldots,T_{q+pm}$ suffice to determine all further T's. If $q < p$ this can be sharpened to T_o, ..., T_{p+qm}, cf. [Hanzon 1981].

2.14. The Associated Covariance Systems. Let Γ_k be given by (2.3). These matrices for $k \in \underline{\underline{N}} \setminus \{0\}$ can be considered as impulse response matrices of some linear system which we call the covariance system corresponding to the ARMA-model. Using rather similar ideas as described above and using that the block Toeplitz matrix

$$\begin{pmatrix} \Gamma_o & \cdots & \Gamma_{q-p+1} \\ \vdots & \ddots & \vdots \\ \Gamma^*_{q-p+1} & \cdots & \Gamma_o \end{pmatrix}$$

is positive definite hermitian one obtains that the Γ_o, ..., Γ_e, $e = \min(pm+q, qm+p)$ suffice to determine the remaining ones so that (the statistics of) the first $e + 1$ outputs y_o, y_1, ..., y_e suffice for identifiability.

For $q \geqq p \geqq 1$ this sufficient condition can be proved to be necessary as well, cf. [Hanzon 1981].

3. IDENTIFICATION AND THE GEOMETRY OF THE MODULI SPACE.

3.1. The set-up. Let us consider the usual state space linear systems

$$(3.2)\qquad \dot{x} = Ax + Bu,\ y = Cx,\ x \in \underline{\underline{R}}^n,\ y \in \underline{\underline{R}}^p,\ u \in \underline{\underline{R}}^m$$

and consider at a fairly primitive level the problem of recursive "fitting" A, B, C in the "best" possible way to the available data u(t), y(t), $t \in \mathbb{R}$. In particular for the moment we are not going to worry about how to put in some stochastics so as to make sense of "best" in a probabilistic way. Here recursive should be interpreted as in the introduction. We shall also assume that (3.2) is completely observable and completely reachable and that we have available a (reasonable) guess for) n. The first remark is of course that y(t), u(t) for all t cannot determine (A, B, C) uniquely but only the orbit of (A, B, C) under state space equivalence, $(A,B,C)^S = (SAS^{-1}, SB, CS^{-1})$, $S \in GL_n(\mathbb{R})$. This leads to the quotient space $L^{co,cr}_{m,n,p}/GL_n = M^{co,cr}_{m,n,p}$ of the space of all cr and co systems of the indicated dimensions modulo state space equivalence. (And of course y(t), u(t) does distinguish between points of $M = M^{co,cr}_{m,n,p}$.) This space M is a nice smooth differentiable manifold which, perhaps unfortunately, is as a rule not diffeomorphic to an $\mathbb{R}^N$, ([Hazewinkel, 1977]).

Viewing identification as "walking around on M" makes the problem identifiable, i.e. it gets rid of the superfluous parameters. And it does so in a way which is much less ad hoc then the use of one or another canonical forms. Even when global continuous canonical forms do exist (which happens only when $m = 1$ or $p = 1$ [Hazewinkel 1977]) there are lots of them; and they are not equivalent e.g. in terms of the size of the gradient vectors of error functions and there is no especially favorable one.

So it seems much more natural to try to use the natural geometry of M and to do identification directly on M.

3.3. A Riemannian metric on M. A first thing one needs for this is a Riemannian metric on M. A nice one is obtained as follows. Given a system (A, B, C) write, as usual

$$R(A,B) = (B; AB; \ldots; A^n B),$$

(3.4)

$$Q(A,C)^T = (C^T; A^T C^T; \ldots; (A^T)^n C^T)$$

(where the T denotes transposes). We shall use $dR(A,B)$ to denote the formal differential of $R(A,B)$, e.g. if $n = 2$ one has $dR(A,B) = (dB; (dA)B + A(dB); (dA)AB + A(dA)B + A^2(dB)$, where (dA,dB,dC) is a tangent vector to $L^{co,cr}_{m,n,p}$ at (A,B,C). Using this notation consider the following Riemannian metric on $L^{co,cr}_{m,n,p}$

$$(3.5) \qquad \|(dA,dB,dC)\| = Tr((dQ)RR^T(dQ)^T)+Tr((dR)^TQ^TQ(dR))$$

where Q is short for $Q(A,C)$ and R for $R(A,B)$. It is not difficult to check that this Riemannian metric is positive definite on $L^{co,cr}_{m,n,p}$ but that it degenerates on the boundary of $L^{co,cr}_{m,n,p}$, i.e. it becomes singular for systems of lower MacMillan degree.

One easily checks that this metric is invariant under $\underline{\underline{GL}}_n$ so that the metric descends to give us a Riemannian metric on M.

One interesting problem is to calculate the curvature of this metric also because of the connection between Gaussian curvature and "the Fisher second order efficiency" of a statistical estimation problem ([Efron 1975]).

3.6. Identifying systems. One could now imagine that an identification procedure would proceed as follows. At time t we have Σ_t and x_t. New data come in; assuming Σ_t, x_t evolves in a known way giving (given $u(t)$) a prediction for $y(t)$ which can be compared to the actual $y(t)$. Calculate the squared error (e.g.) function as a function of $\Sigma \in M$ and take the gradient. Now let Σ_t evolve along this gradient (possibly with a gain factor inserted). The question we would like to pose is does such an identification scheme exist and/or can any of the existing recursive identification schemes (cf. [Ljung 1981] for a very nice, up to date survey) be viewed in this way (perhaps in approximation)? Or more generally can there exist such a scheme? One would definitely conjecture yes. Given the fact that there are recursive identification schemes it is hard to see how they can avoid covering the moduli space M in some way.

4. THE NONLINEAR FILTERING APPROACH TO IDENTIFICATION.

4.1. The Estimation Lie algebra. Consider a general nonlinear stochastic system

$$(4.2) \qquad dx_t = f(x_t)dt + G(x_t)dw_t, \quad dy_t = h(x_t)dt + dv_t$$

where w_t, v_t are independent Wiener noise processes also independent of the initial random variable x_o. Here f,h,G are vector and matrix valued functions of the appropriate sizes. Assume sufficient regularity so that the conditional density p(x,t) exists of the state x_t given the past observations y_s, $0 \leqq s \leqq t$. Then an unnormalized version $\rho(x,t)$ satisfies the socalled Duncan-Mortenson-Zakai equation (Fisk-Stratonovic form)

$$(4.3) \qquad d\rho(x,t) = L_o\rho(x,t)dt + \sum_{i=1}^{p} h_i(x)\rho(x,t)dy_{it}$$

where h_i is the i-th component of h and L_o is the differential operator

$$(4.4) \qquad L_o\phi = \frac{1}{2}\sum_{i,j=1}^{n}\frac{\partial^2}{\partial x_i \partial x_j}((GG^T)_{ij}\phi) - \sum_{i=1}^{n}\frac{\partial}{\partial x_i}(f_i\phi) - \frac{1}{2}\sum_{j=1}^{p} h_j^2$$

The Lie algebra of differential operators generated by L_o and $h_1, \ldots, h_p$ is called the *estimation Lie algebra*. Cf the references cited in the introduction for more information on it.

4.5. The Estimation Algebra of an Identification Problem. Now consider the problem of identifying a system (1.1) where, for ease of notation mainly, we take $D = I_p$, $B_1 = 0$ (so that there are no deterministic inputs). Write it as a nonlinear filtering problem by adding the equations (1.5). This gives

$$(4.6) \qquad dx_t = Ax_t dt + Bdw_t, \quad dA = dB = dC = 0, \quad dy_t = Cx_t + dv_t$$

(Note that usually there is redundancy in A,B,C in spite of the fact that we have already normalized D). Writing out L_o and h_i in this case one notices that these operators are all sums of expressions of the form

$$c_{\alpha\beta} x^{\alpha} \frac{\partial^{\beta}}{\partial x^{\beta}} = c_{\alpha\beta} x^{\alpha} \partial_{\beta} \tag{4.7}$$

where $C_{\alpha\beta}$ is a polynomial in the entries of A, B, C and where $\alpha = (\alpha_1, \ldots, \alpha_n)$, $\beta = (\beta_1, \ldots, \beta_n)$ are multiindices such that $\|\alpha\|, \|\beta\| \leq 2$ where if γ is a multiindex $\|\gamma\|$ denotes $\gamma_1 + \ldots + \gamma_n$. Now the $x^{\alpha}\partial_{\beta}$ with $\|\alpha\|$, $\|\beta\| \leq 2$ form a $2n^2 + 3n + 1$ dimensional Lie algebra (under the commutator brackett) which we denote LS_n. It follows that the estimation Lie algebra of (4.6) is a sub-Lie-algebra of the Lie-algebra

$$LS_n \otimes \mathbb{R}[A, B, C] \tag{4.8}$$

where $\mathbb{R}[A,B,C]$ stands for the polynomial ring over $\mathbb{R}$ in the entries of A, B, C. In particular this implies that the estimation Lie algebra $L(\Sigma)$ of (4.6) is profinite dimensional. This means that there are ideals $I_1 \supset I_2 \supset I_3 \supset \ldots$ such that $L(\Sigma)/I_j$ is finite dimensional for all j and $\bigcap_j I_j = \{0\}$. And in turn this suggests that there are sufficiently many recursively computable statistics.

If M is a manifold let V(M) denote the Lie-algebra of vectorfields on M. For a fixed $\theta = (A, B, C)$ the Kalman-Bucy filter defines an anti-homomorphism of Lie-algebras of the estimation Lie algebra $L(\theta)$ to $V(\mathbb{R}^r)$, $r = \frac{1}{2}n^2 + \frac{3}{2}n$ (for a proof cf. [Brockett-Clark, 1978] in the simplest case and [Hazewinkel, 1981] in general). Letting θ vary these "combine" to define a Lie algebra homomorphism of the estimation Lie algebra of (4.6) to $V(\mathbb{R}^r \times \mathbb{R}^N)$, $N = n^2 + nm + np$ (the number of parameters in A, B, C). Or, better, one can use a certain representation of LS_n in $V(\mathbb{R}^r)$ which is essentially all possible Kalman filters combined, cf. [Hazewinkel 1981].

This Lie algebra anti-homomorphism (essentially a family of

Kalman-Filters) does calculate some statistic viz. the conditional density $p[x_t | y_s, 0 \leqq s \leqq t, \theta]$ as a function of θ. This and some related and/or derived entities which can be recursively computed can be used in a variety of ways, cf. e.g. [Krishnaprasad-Marcus, 1981], [Hanzon-Hazewinkel-Krishnaprasad 1981 a,b] and also below in section 5. But this filter does not give us an identification procedure of the type we want because it involves no equations which tell us how θ evolves.

5. GAUSSIAN APPROXIMATION.

The filter described above does give most useful information though, and combines nicely with Gaussian approximation ideas [Stratonovic 1960,1970]. Let us illustrate this by means of a most simple example

$$dx_t = \theta x_t dt + dw_t, \; d\theta = 0, \; dy_t = x_t dt + dv_t \tag{5.1}$$

The D-M-Z equation in this case looks like

$$d\rho = \left(\frac{1}{2}\frac{\partial^2}{\partial x^2} - \theta - \theta x \frac{\partial}{\partial x} - \frac{1}{2}x^2\right)\rho dt + x\rho dy_t \tag{5.2}$$

and the estimation Lie algebra is easy to calculate, cf. example 6.6 below. Write $\rho = e^{-S}$. Then S satisfies the equation

$$dS = \frac{1}{2}\frac{\partial^2 S}{\partial x^2} - \frac{1}{2}\left(\frac{\partial S}{\partial x}\right)^2 + \theta - \theta x \frac{\partial S}{\partial x} + \frac{1}{2}x^2 - x dy_t \tag{5.3}$$

which is a family of evolution equations for S parametrized by θ . Moreover if for a certain value Θ_0 of θ the initial distribution $\rho(x_0, \theta_0)$ is Gaussian so that $S(x, \theta, 0)$ is quadratic in x at Θ_0 ., then $\rho(x, \Theta_0, t)$ is Gaussian for all t (because given Θ_0 we are dealing with a linear system, i.e. $S(x, \Theta_0, t)$ is quadratic. This can also been seen from (5.3).

Assume $\rho(x_0, \Theta, 0)$ is Gaussian for all Θ. Write $S = ax^2+bx+c$, where a,b,c are functions of Θ and t. Then equation (5.3) gives us

$$(5.4)\qquad \begin{aligned} \dot{a} &= -2a^2 + \frac{1}{2} - 2a\theta , \\ \dot{b} &= -\ 2ab - b\theta - dy_t, \\ \dot{c} &= a + \theta - \frac{1}{2}\, b^2 \end{aligned}$$

which is simply another way of writing down (deriving) the family of Kalman filters alluded to before in the last part of section 4 above. In fact writing $a = -\frac{1}{2}\, p^{-1}$ the reader will recognize in the first equation of (5.4) the equation for the covariance p given θ.

Of course the Lie algebra of the filter (5.4) is a homomorphic image of the estimation Lie algebra. From the parametrized family of covariance equations $\dot{p} = 1 - p^2 + 2\theta p$ one obtains (also families of) equations for the (partial) derivative(s) $\frac{\partial p}{\partial \theta} = p$ (in this case $\dot{p}_\theta = -2pp_\theta + 2p + 2\theta p_\theta$) which is linear, given p.

This can be useful in view of a theorem of Nishimura (cf. [Jazwinsky 1970, page 254, Thm 7.8]) to the effect that $p(\Theta,t)$ assumes its minimum value at $p(\Theta_o,t)$ if Θ_o is the true value.

The equations (5.3) also nicely show why Gaussian approximation might work well. Write $S = a_o + a_1x + a_2x^2 + a_3x^3 + a_1x^4 + \ldots$ Substituting this in (5.3) yields equations for the a_i, $i = 0,1,2,\ldots$, viz.

$$\begin{aligned} \dot{a}_o &= a_2 + \theta - \frac{1}{2}\, a_1^2 \\ \dot{a}_1 &= 3a_3 - 2a_1a_2 - \theta a_1 - dy_t \\ \dot{a}_2 &= -\ 2a_2^2 - 3a_1a_3 - 2a_2\theta + \frac{1}{2} + 6a_4 \\ \dot{a}_3 &= -\ 4a_1a_4 - 6a_2a_3 + 10a_5 - 3\theta a_3 \\ \dot{a}_4 &= -\ \frac{9}{2}\, a_3^2 - 8a_2a_4 - 5a_1a_5 + 15a_6 - 4\theta a_4 \\ &\ldots \end{aligned}$$

Taking the quadratic (Gaussian approximation) is stable in the sense that if $a_3 = a_4 = a_5 = \dots = 0$ at the starting time then they remain zero, but if e.g. $a_4 = a_5 = \dots = 0$ at the starting time then a_4, a_5, etc. do not remain zero. In fact Gaussian approximation is the only approximation which works in this sense.

The filter (5.4) calculates a,b,c as functions of θ, but is of course still an infinite-dimensional machine. Writing a,b,c as power series in $(\theta-\theta_o)$ (around a previous estimate θ_o e.g) we find from (5.4) differential equations for the coefficients $a_o, a_1, \dots; b_o, b_1, \dots; c_o, c_1, \dots$ of these power series and because the estimation Lie algebra is a subalgebra of the current algebra $LS_n \otimes \mathbb{R}[\theta]$ we have that $LS_n \otimes (\theta-\theta_o)^i \mathbb{R}[\theta]$ is an ideal so that these equations are such that a_k, b_k, c_k, $k \geq i$ remain zero for all time t if this is the case at time $t = 0$. This holds for all i, in particular for $i = 3$. Thus we can calculate the quadratic part (around θ_o) of S by an exact finite dimensional filter, and this quadratic part in turn contains all the data needed for the joint Gaussian approximation of the density $\rho(u,\theta)$ (up to a scalar factor) and from that an (approximate) estimate $\hat{\theta}$ results.

6. EXTENDED KALMAN FILTER AND ESTIMATION LIE ALGEBRA.

In this section we shall only consider two examples. The results, however, suggest a general theorem which remains to be established.

6.1. Example 1. Consider the identification type non linear filtering problem given by the equations

$$dx_t = b\,dw_t, db = 0, \; dy_t = x_t dt + dv_t, \; x_t \in \mathbb{R} \tag{6.2}$$

The estimation Lie algebra of this system is easily calculated. As a basis it has the operators

$$A = \frac{1}{2} a^2 \frac{\partial^2}{\partial x^2} - a \frac{\partial}{\partial x} - \frac{1}{2} x^2; \; B_i = a^{2i} x, \; i = 0,1,2\dots;$$

$C_i = a^{2i} \frac{\partial}{\partial x} - a^{2i-1}$, $i = 1,2,\dots$; $D_i = a^{2i}$, $i = 1,2,\dots$. The nonzero commutation relations are

$[A,B_i] = C_{i+1}$, $[A,C_i] = B_i$, $[B_i,C_j] = -D_{i+j}$.

Now consider the extended Kalman filter for (6.2) (cf.[Jazwinsky 1970, page 338]. This gives

$$dP_{11} = (2P_{12}+1-P_{11}^2)dt,\quad dP_{12} = (P_{22}-P_{11}P_{12})dt,$$

(6.3)

$$dP_{22} = -\ P_{12}^2 dt$$

$$d\hat{x} = P_{11}(dy_t-\hat{x}dt),\quad d\hat{b} = P_{12}(dy_t-\hat{x}dt)$$

Thus, writing x and b instead of $\hat{x}$, $\hat{b}$ for typographical convenience, the vectorfields defining the EKF (6.3) are (cf. also (1.4))

$$\alpha = (2P_{12}+1-P_{11}^2)\frac{\partial}{\partial P_{11}} + (P_{22}-P_{11}P_{12})\frac{\partial}{\partial P_{12}} -$$

$$- P_{12}^2 \frac{\partial}{\partial P_{22}} - xP_{11} \frac{\partial}{\partial x} - xP_{12} \frac{\partial}{\partial b}$$

$$\beta_o = P_{11} \frac{\partial}{\partial x} + P_{12} \frac{\partial}{\partial b}$$

Calculating interated bracketts in the α and β_o is a rather exhausting business. But a few are needed

$$\gamma_1 = [\beta_o,\alpha] = - \frac{\partial}{\partial x} - 2P_{12} \frac{\partial}{\partial x} - P_{22} \frac{\partial}{\partial a}$$

$$\beta_1 = [\gamma_1,\alpha] = P_{11} \frac{\partial}{\partial x} + P_{12} \frac{\partial}{\partial b} + P_{12}^2 \frac{\partial}{\partial b} + 2P_{22} \frac{\partial}{\partial x}$$

Now observe that if δ is a vectorfield of the general form

(6.4) $$\delta = r_1 \frac{\partial}{\partial x} + r_2 \frac{\partial}{\partial b},\ r_1,r_2 \text{ polymomials in } P_{11},P_{12},P_{22}$$

then $[\delta,\alpha]$ is of the same general form. Also observe that if δ,δ' are two (different) vectorfields of the form (6.4) then $[\delta,\delta'] = 0$. Finally observe that if γ is a vectorfield of the form (6.4) which can be written as

(6.5) $$\gamma = cP_{11}^s P_{12}^2 \frac{\partial}{\partial b} + \delta,\ s \in \underline{\underline{N}} \setminus \{0\},\ c \neq 0$$

with all polynomials in δ of degree $\leqq$ s+1. Then $[\gamma,\alpha]$ is of the form

$$c(s+2)P_{11}^{s+1}P_{12}^2 \frac{\partial}{\partial b} + \delta'$$

with all polynomials in δ' of degree $\leqq$ s+2. This proves that the $(ad\alpha)^n(\beta_o)$ are all linearly independent and that combined with the previous observations shows that $A \to \alpha$, $B_o \to \beta_o$ induces an anti-homomorphism of Lie algebras of the estimation Lie algebra of (6.2) onto the Lie algebra generated by the vectorfields of the corresponding extended Kalman Filter. The kernel of this anti-homomorphism is the centre of the estimation Lie algebra of (6.2).

6.6. Example 2. Now consider the identification type nonlinear filtering problem also considered in section 5 above given by the equations.

$$(6.7) \qquad dx_t = ax_t dt + dw_t, \; da = 0, \; dy_t = x_t dt + dv_t, x_t \in \mathbb{R}$$

Again it is not difficult to write down the estimation Lie algebra of (6.7). It has a basis consisting of

$$A = \frac{1}{2}\frac{\partial^2}{\partial x^2} - ax\frac{\partial}{\partial x} - a - \frac{1}{2}x^2; \; B_i = (a^2+1)^{2i}x,$$

$C_i = (a^2+1)^{2i}(\frac{\partial}{\partial x} - ax)$, $D_i = (a^2+1)^{2i}$, $i = 0,1,2,\ldots$ and the nonzero commutation relations are $[A,B_i] = C_i$, $[A,C_i] = B_{i+1}$, $[B_i,C_j] = -D_{i+j}$. So that modulo its center this Lie algebra is an infinite dimensional vectorspace with a shift operator exactly as in the previous example. In fact the Lie algebra modulo its centre of every identification type nonlinear filtering problem is a vectorspace with one endomorphism.

Now consider the extended Kalman filter of (6.7). The equations are

$$(6.8) \qquad \begin{aligned} dP_{11} &= (2aP_{11}+1-P_{11}^2+2xP_{12})dt, \\ dP_{12} &= (aP_{12}+xP_{22}-P_{11}P_{12})dt, dP_{22} = -P_{12}^2 dt, \end{aligned}$$

$$dx = axdt - P_{11}xdt + P_{11}dy_t,$$

$$da = P_{12}dy_t - P_{12}xdt$$

Thus the two vectorfields involved are

$$\alpha = (2aP_{11}+1-P_{11}^2+2xP_{12})\frac{\partial}{\partial P_{11}} +$$

$$(6.9) \qquad + (aP_{12}+xP_{22}-P_{11}P_{12})\frac{\partial}{\partial P_{12}} - P_{12}^2 \frac{\partial}{\partial P_{22}}$$

$$+ (ax-P_{11}x)\frac{\partial}{\partial x} - P_{12}x \frac{\partial}{\partial a}$$

$$\beta_o = P_{11} \frac{\partial}{\partial x} + P_{12} \frac{\partial}{\partial a}$$

Let L be the estimation Lie algebra of (6.7) and L' the Lie algebra of vectorfields generated by α and β_o. Note that if γ is a vectorfield of the form $P_{12}\delta + P_{22}\delta'$ then $[\alpha,\gamma]$ and $[\beta_o,\gamma]$ are of the same form and hence (Jacobi identity) all the vectorfields in L' of that form are an ideal I_1 of L'. Calculating L' mod I_1 is not difficult. Indeed α and β_o become

$$(6.10) \qquad \bar{\alpha} = (2aP_{11}+1-P_{11}^2)\ \frac{\partial}{\partial P_{11}} + (ax-P_{11}x)\ \frac{\partial}{\partial x},\ \bar{\beta}_o = P_{11}\ \frac{\partial}{\partial x}$$

Let $\bar{\gamma}$ be any vectorfield of the form $r(a,P_{11})\ \frac{\partial}{\partial x}$, then $[\bar{\alpha},\bar{\gamma}]$ is of the same form and the brackett of two such vectorfields is zero. Also the $(\mathrm{ad}\bar{\alpha})^n(\bar{\beta}_o)$ are all independent, this time because the unique highest degree term in $(\mathrm{ad}\bar{\alpha})^n(\bar{\beta}_o)$ is $a^nP_{11}\ \frac{\partial}{\partial x}$ as is easily checked. Thus $A \to \bar{\alpha}$, $B_o \to \bar{\beta}_o$ induces an antihomomorphism $L \to L'/I_1$ whose kernel is the centre of L.

Now consider all vectorfields in L' of the form $P_{12}^2\delta + P_{22}\delta' + P_{12}(1-P_{11}^2+2aP_{12})\delta''$. Check that bracketting with α or β_o gives again a vectorfield of the same form so that these vectorfields form an ideal I_2.

Consider any vectorfield of the form

$$(6.11) \qquad \gamma = r_1\ \frac{\partial}{\partial x} + P_{12}r_2x\ \frac{\partial}{\partial x} + P_{12}r_3\ \frac{\partial}{\partial P_{11}} + P_{12}r_4\ \frac{\partial}{\partial a}$$

where r_1,r_2,r_3,r_4 are polynomials in P_{11} and a. Observe that bracketing a vectorfield (6.11) with α yields a vectorfield of the same type modulo I_2. Also observe that bracketing two vectorfields of this type yields a vectorfield of the type (modulo I_2)

(6.12) $$\delta = sP_{12}\frac{\partial}{\partial x}$$

where s is a polynomial in a, P_{11}. Finally observe that if δ is of type (6.12) then $[\delta,\alpha] = [\delta,\beta_o] = 0 \bmod I_2$ so that these δ are in the centre of L'/I_2. Putting all this together we see that the natural projection $L'/I_2 \to L'/I_1$ induces an isomorphism of L'/I_1 with L'/I_2 mod its centre and that hence $A \to \alpha$, $B \to \beta_o$ induces an isomorphism of L and L'/I_2 modulo their centres.

6.13. <u>Remarks</u>. $L \to L'/I_1$ is in fact the "family of Kalman filters" antihomomorphism mentioned in section 4 above. This is a first order approximation to the filtering problem. L/Centre $\to (L'/I_2)$/centre is in the nature of a second order approximation. Indeed near the true parameters P_{22} is second order small, P_{12} is first order small and cf. equations (6.9), $(1-P_{11}^2+2aP_{11})$ is also first order small so that I_2 is a second order small ideal.

REFERENCES.

1. R.W. Brockett, J.M.C. Clark, 1978. The Geometry of the Conditional Density Equations, Proc. of the Oxford Conf. on Stochastic Systems, Oxford, 1978.
2. R.W. Brockett, 1981, Nonlinear Systems and Nonlinear Estimation Theory, In: M. Hazewinkel, J.C. Willems (eds), Stochastic Systems: the Mathematics of Filtering and Identification and Applications, Reidel Publ. Cy., 1981.
3. B. Efron, 1975, Defining the Curvature of a Statistical Problem (with applications to second order efficiency), Ann. of Stat. 3, 1189-1242.
4. G.D. Forney, 1975, Minimal Bases of Rational vectorspaces with Applications to Multivariable Linear Systems, SIAM J. Control 13, 493-520.
5. E.J. Hannan, 1971, The Identification Problem for Multiple Equation Systems with moving Average Errors, Econometrica 39 (1971), 751-765.
6. B. Hanzon, 1981, The Finite Identifiability Problem, Report 81.., Econometric Inst., Erasmus Univ. Rotterdam.
7. B. Hanzon, M. Hazewinkel, P.S. Krishnaprasad, 1981a, IEEE Conf. Dec. Control, San Diego Dec. 1981.
8. B. Hanzon, M. Hazewinkel, P.S. Krishnaprasad, 1981b, IEEE Conf. Dec. and Control, San Diego, Dec. 1981.
9. M. Hazewinkel, 1977, Moduli and Canonical Forms for Linear Dynamical Systems,II; the Topological Case, Math. System Theory 10 (1977), 363-385.
10. M. Hazewinkel, 1980, On the (Internal) Symmetry Groups of Linear Dynamical Systems, In: M. Dal Cin, P. Kramer (eds), Groups, Systems and many-body Physics, Vieweg 1980, 362-404.
11. M. Hazewinkel, 1981, The Linear Systems Lie Algebra, the Segal-Shale-Weil Representation and all Kalman Filters, Report 81.., Econometric Inst., Erasmus Univ. Rotterdam, 1981.
12. M. Hazewinkel, S.I. Marcus, 1980, On Lie Algebras and Nonlinear Filtering, Report 8019, Econometric Inst., Erasmus Univ. Rotterdam, to appear in Stochastics.

13. M. Hazewinkel, J.C. Willems (eds), 1981, Stochastic Systems: the Mathematics of Filtering and Identification and Applications, Reidel Publ. Cy., 1981.

14. A.H. Jazwinsky, 1970, Stochastic Processes and Filtering Theory, Acad. Pr., 1970.

15. R.E. Kalman, 1979, On Partial Realizations, Transfer Functions and Canonical Forms, Acta Polyt. Scan. Math. Comp. Sci Ser. no. 31 (1979), 9-32.

16. P.S. Krishnaprasad, S.I. Marcus, 1981, Identification and tracking: a class of Nonlinear Filtering Problems, JAAC, June 1981.

17. L. Ljung, 1981, Recursive Identification, In: M. Hazewinkel, J.C. Willems (eds), Stochastic Systems: The Mathematics of Filtering and Identification and Applications, Reidel Publ. Cy, 1981.

18. S.I. Marcus, S.K. Mitter, D. Ocone, 1978, Finite Dimensional Nonlinear Estimation for a Class of Systems in continuous and Discrete time, Proc. Conf. on Stochastic Systems, Oxford 1978.

19 R.L. Stratonovic, 1960, Application of the Theory of Markov Processes in Optimal Signal Determination, Radiotechn i Elekt. 5 (1960), 1751-1763.

20 R.L. Stratonovic, 1970, Detection and Estimation of Signals in noise when one or both are Nongaussian, Proc. IEEE 58 (1970), 670-679.

21. H.H. Tigelaar, 1978, The Identifiability Problem in Dynamic Simultaneous Equations with moving Average Errors, Disc. Paper 78.080, Cath. Univ. Tilburg, 1978.