Proceedings ICIAM 87, Paris-La Villette, June 29-July 3 1987

Introduction to Nilpotent Approximation Filtering

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

The socalled reference probability of unnormalized probability method for nonlinear filtering problems leads to a (robust) infinite dimensional filter of bilinear type. If the associated Lie algebra is topologically solvable or nilpotent an infinite dimensional version of Wei-Norman theory applies. If not then ideas of nilpotent approximation lead to (potential) approximation filters. This note is not so much a definite report on results as on outline of a research program.

1980 Mathematics subject classification: 93E11, 93B30, 93E10, 93D25, 60H15, 93B15, 17B65, 17B99, 57R25, 35H05.
Keywords and phrases: nonlinear filtering, estimation Lie algebra, topologically nilpotent Lie algebra, Duncan-Mortenson-Zakai equation, unnormalized conditional probability density, robustness, nilpotent approximation, Wei-Norman theory, stochastic differential equations.

1. Statement of the problem

In full generality *filtering* is concerned with obtaining estimates concerning a stochastic process $\{x_t\}$, the *signal process,* on the basis of another related process $\{y_t\}$, the *observation process.* In this paper we have the following realization of this situation in terms of stochastic differential equations.

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

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

where f, G, h are vector and matrix valued functions of the right dimensions and w_t and v_t are independent Wiener noise processes also independant of the initial state x_0. The problem is the following. For a given (interesting) function $\phi(x)$ of the state x, give a calculation procedure for the best estimation $\widehat{\phi(x_t)}$ at time t given the observations y_s, $0 \leqslant s \leqslant t$. More generally one also considers finding $\widehat{\phi(x_t)}$ given y_s, $0 \leqslant s \leqslant t_1$, $t_1 < t$ *(prediction)* and finding $\widehat{\phi(x_t)}$ given y_s, $0 \leqslant s \leqslant t_2$, $t < t_2$ *(smoothing).* Of particular importance is finding $\hat{x}_t$ *(state estimation).*

Ideally one would like the calculation procedure to be *finite dimensional, exact, recursive,* and *robust.* The first three adjectives here mean (more or less by definition) that the calculation procedure, the *filter,* should be of the form

$$dm_t = \tilde{\alpha}(m_t)dt + \sum_{j=1}^{r} \tilde{\beta}_j(m_t)d\tilde{\zeta}_j(y_t) \tag{1.3}$$

$$\widehat{\phi(x_t)} = \gamma(m_t, y_{1t}, \ldots, y_{pt}) \tag{1.4}$$

Here $\tilde{\alpha}, \tilde{\beta}_j, \tilde{\zeta}_j, \gamma$ are known functions and vectorfields and m_t evolves over a finite dimensional manifold (finite dimensionality); recursiveness is embodied by the fact that (1.3) is directly driven by the observations and that $\widehat{\phi(x_t)}$ only depends on the filter state m_t; and the current observations; (1.4) of course also reflects exactness. For robustness one requires that the filter equations be driven by y_t itself instead of also involving the dy_t. I.e. one requires (1.3) to be replaced by an equation

$$\frac{dm_t}{dt} = \alpha(m_t) + \sum_{j=1}^{r} \beta_j(m_t)\zeta_j(y_{1t}, \ldots, y_{pt}). \tag{1.5}$$

Thus while (1.3) is a stochastic differential equation its robust version (if it exists) (1.3) can be treated pathwise and makes sense as a family of differential equations, one for each possible observation path $\{y_t\}$.

The problem now is: given a system (1.1), (1.2) and a function ϕ how to find a filter (1.4), (1.5); i.e. how to determine the functions γ and ζ_j and vectorfields α and β_j occurring in (1.4), (1.5).

2. The DMZ filter

Under mild regularity assumptions on f,G,h and reachability and observability conditions on the system (1.1), (1.2) the conditional state $\hat{x}_t = E[x_t \mid y_s, 0 \leqslant s \leqslant t]$ has a density $\pi(x,t)$.

Theorem 2.1. (Duncan [2], Mortensen [6], Zakai [9]). Under appropriate regularity conditions there exists an unnormalized version $\rho(x,t)$ of $\pi(x,t)$ (i.e. $\rho(x,t) = \sigma(t)\pi(x,t)$ for some unknown function $\sigma(t)$) which satisfies the stochastic partial differential equation

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

Here $\mathcal{L}$ is the second order partial differential operator defined by

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

Here G^T is the transpose of the matrix valued function G and $(GG^T)_{ij}$ is the (ij)-th entry of the matrix GG^T, f_i is the i-th component of the function f and h_j the j-th component of the function h.

The stochastic PDE (2.2) is to be regarded as a Fisk-Stratonovic stochastic PDE. To obtain the equivalent Ito version remove the term $-\frac{1}{2}\sum h_j^2\psi$ in (2.3).

Consider the time dependant gauge transformation

$$\tilde{\rho}(x,t) = \exp(-h_1(x)y_{1t} - \ldots - h_p(x)y_{pt})\rho(x,t). \tag{2.4}$$

Substituting this into (2.2) yields an equation

$$\frac{\partial\tilde{\rho}(x,t)}{\partial t} = \mathcal{L}\tilde{\rho} - \sum_{i=1}^{p} y_i(t)\mathcal{L}_i\tilde{\rho} - \sum_{i,j=1}^{p} y_i(t)y_j(t)\mathcal{L}_{ij}\tilde{\rho} \tag{2.5}$$

where

$$\mathcal{L}_i = [h_i,\mathcal{L}] := h_i\mathcal{L} - \mathcal{L}h_i, \quad \mathcal{L}_{ij} = \mathcal{L}_{ji} = \frac{1}{2}[h_i,[h_j,\mathcal{L}]]. \tag{2.6}$$

Given $\phi(x)$ and $\tilde{\rho}(x,t)$ the best estimate $\widehat{\phi(x_t)}$ can be calculated by

$$\rho(x,t) = \exp(h_1(x)y_{1t} + \ldots + h_p(x)y_{pt}) \tag{2.7}$$

$$\widehat{\phi(x_t)} = \left(\int \rho(x,t)dx\right)^{-1} \int \phi(x)\rho(x,t)dx. \tag{2.8}$$

Note that (2.5) together with the output map (2.7), (2.8) is a recursive, exact and robust filter. The only trouble with it (from the calculation point of view) is that it is infinite dimensional.

3. Wei-Norman theory [8].

For the moment let us consider control systems of the form

$$\dot{x} = u_1A_1x + \ldots + u_kA_kx, \quad x \in \mathbb{R}^n \tag{3.1}$$

where the A_i are $n \times n$ matrices and the u_i are inputs (known functions of time). Adding a few more terms (with $u_j = 0$, $j > k$) we may as well assume that $A_1,\ldots,A_k$ are a basis of a Lie algebra of $n \times n$ matrices (under the commutator difference product $[A,B] = AB - BA$). Let us look for solutions of the form

$$x(t) = e^{g_1 A_1} \cdots e^{g_k A_k} x(0) \tag{3.2}$$

ıe $g_i(t)$ are still to be determined functions of time. By differentiating (3.2), inserting $A_1) \cdots \exp(-g_i) \exp(g_i A) \cdots \exp(g_1 A)$ just after $\dot{g}_{i+1} A_{i+1}$ in the result, using the Baker-Hausdorff formula, using (3.1) and collecting terms, one finds a set of equations

$$\dot{g}_i + \sum_{j=1}^{k} \dot{g}_j h_{ji}(g_1,...,g_k) = u_i, \quad i = 1,...,k \tag{3.3}$$

),...,0)=0 and the following properties of the $h_{ij}(g_1,...,g_k)$:

$$h_{ij} \text{ only involves } g_1,...,g_{i-1} \tag{3.4}$$

$_{+1},...,A_k$ are a basis of an ideal or $\mathfrak{a} \subset \mathfrak{g}$ (so that $[A_i, \mathfrak{a}] \subset \mathfrak{g}$ for all i) then

$$h_{ji} = 0 \text{ for } i = i,...,l;\ j = l+1,...,k \tag{3.5}$$

he equations for $g_1,...,g_l$ do not involve $g_{l+1},...,g_k$ at all. It is also important to note that the ıniversal functions depending only on the Lie algebra $\mathfrak{g}$ and the chosen basis and totally lent of the particular matrix realization (representation) we may be dealing with. In particu- s an ideal of $\mathfrak{g}$ and $A_1,...,A_k$ is a basis as above then

$$\text{equations for } g_1,...,g_l \text{ only depend on } \mathfrak{g}/\mathfrak{a}. \tag{3.6}$$

that $\mathfrak{g}$ is nilpotent (or more generally solvable) equations (3.3) therefore take a particularly triangular form which can be solved just using quadratures. Indeed if L is nilpotent, so that

$$L \supsetneq [L,L] = L_2 \supsetneq [L,L_2] = L_3 \supsetneq \cdots \supsetneq [L,L_r] = L_{r+1} = 0$$

e choose a basis

$$A_1,...,A_{k_1}, A_{k_1+1},...,A_{k_2},...,A_{k_{r-1}+1},...,A_{k_r}, \quad k_r = k$$

t

$$A_{k_{i-1}+1},...,A_{k_i}, \quad k_0 = 0$$

s for L_i, $i=1,...,r$, then the equations take the form

$$\begin{aligned}
&\dot{g}_1 = u \\
&... \\
&\dot{g}_{k_1} = uk_1 \\
&\dot{g}_{k_1+1} = u_{k_1+1} + \alpha_{k_1+1}(u_1,...,u_{k_1};g_1,\ldots,g_{k_1} \\
&... \\
&\dot{g}_{k_2} = u_{k_2} + \alpha_{k_2}(u_1,..,u_{k_1};g_1,\ldots,g_{k_1}) \\
&\dot{g}_{k_2+1} = u_{k_2+1} + \alpha_{k_2+1}(u_1,...,u_{k_2};g_1,...,g_{k_2}) \\
&... \\
&\dot{g}_{k_3} = u_{k_3} + \alpha_{k_3}(u_1,...,u_{k_3};g_1,...,g_{k_2}). \\
&...
\end{aligned} \tag{3.6}$$

ıte that the robust DMZ filter equation (2.5) is of the form (3.1) except that it takes place in a ı space. So in particular if the Lie algebra generated by the operators $\mathcal{L}, \mathcal{L}_i, \mathcal{L}_{ij}$ in (2.5) is nilpo- lvable) and finite dimensional with basis $A_1,...,A_k$ and we have given an initial density $\rho_0(x)$ ction ϕ then equations (3.6) together with the output equation

$$(g_1,\dots,g_k) \mapsto \tilde{\rho}(x,t) = \exp(g_1 A_1)\cdots\exp(g_k A_k)\rho_0(x)$$
$$\tilde{\rho}(x,t) \mapsto \rho(x,t) = \exp(h_1(x)u_1)\cdots\exp(h_p(x)y_p)\tilde{\rho}(x,t)$$
$$\hat{\phi(x)} = (\int \rho(x,t)dx)^{-1} \int \phi(x)\rho(x,t)dx$$

constitute a recursive exact robust filter for $\hat{\phi(x_t)}$. It is not really finite dimensional because the A_i here are operators and calculating $\exp(g_i A_i)$ (for known $g_i(t)$) amounts to solving $\frac{d}{dt}B_i = \dot{g}_i A_i B_i$, $B_0 = id$ which is again a partial differential equation.

4. The identification case

The problem of identifying a linear system

$$dx_t = Ax_t dt + B dw_t, \quad dy_t = Cx_t + dv_t \tag{4.1}$$

i.e. the problem of determining the unknown matrices A,B,C on the basis of the observations, can be viewed as a nonlinear filtering problem for the system with state vector (x,A,B,C) obtained by adding the equations $dA = 0$, $dB = 0$, $dC = 0$ to (4.1). It can be proved that the Lie algebra generated by the $\mathcal{L},\mathcal{L}_i,\mathcal{L}_{ij}$ in this case is topologically solvable. I.e. there is a sequence of ideals $\mathfrak{a}_i$ such that $\mathfrak{g}/\mathfrak{a}_i$ is finite dimensional solvable for all i and $\bigcap_i \mathfrak{a}_i = \{0\}$. Because of (3.6) this yields a sequence of approximate filters via

$$e^{g_1 A_1}\cdots e^{g_{k_1} A_{k_1}}\rho_0, \quad e^{g_1 A_1}\cdots e^{g_{k_2} A_{k_2}}\rho_0, \cdots$$

where $A_1,\dots,A_{k_1},A_{k_1+1},\dots,A_{k_2},\cdots$ are such that the equivalence classes of $A_1,\dots,A_{k_r}$ mod $\mathfrak{a}_r$ are a basis for $\mathfrak{g}/\mathfrak{a}_r$. Cf [5] for more details.

5. Nilpotent and solvable approximations

However, in many cases, the Lie algebra generated by $\mathcal{L},\mathcal{L}_i,\mathcal{L}_{ij}$ will not be topologically solvable. For instance in the case of perturbed linear systems

$$dx = (Ax + \epsilon P_A(x))dt + (B + \epsilon P_B(x))dw_t, \quad dy = (C + \epsilon P_C(x))dt + dv_t \tag{5.1}$$

where the $P_A(x)$, $P_B(x)$, $P_C(x)$ are polynomial higher order disturbances. In this case the Lie algebra tends to be $W_n = \mathbb{R}<x_1,\dots,x_n; \frac{\partial}{\partial x_1},\dots,\frac{\partial}{\partial x_n}>$, the Lie algebra of all differential operators (any order) with polynomial coefficients. In this case the higher order operations come with higher powers of ϵ in the sense that

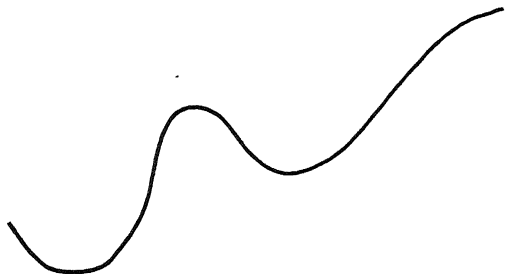
$$\text{Lie}(\mathcal{L},\mathcal{L}_i,\mathcal{L}_{ij}) \bmod \epsilon^n \text{ is finite dimensional for all } n \tag{5.2}$$

(and these algebras are solvable). Again there result approximate filters and they seem to perform well [3,4]. Still more generally there is no small parameter at all, but there still is a natural gradation structure on the Lie algebra. To see why this might be the case and why this will give us possibilities for constructing approximate filters observe that the operators $\mathcal{L},\mathcal{L}_i,\mathcal{L}_{ij}$ are of the general forms

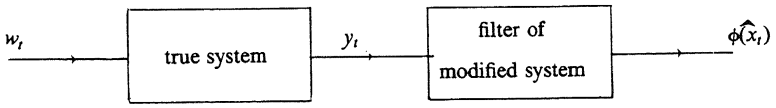
$$\mathcal{L} = \sum a_{ij}\frac{\partial^2}{\partial x_i \partial x_j} + \sum b_j \frac{\partial}{\partial x_j} + c$$
$$\mathcal{L}_i = \sum d_{ij}\frac{\partial}{\partial x_j} + e_i$$
$$\mathcal{L}_{ij} = f_{ij}$$

where the $a_{ij},b_{ij},f_{ij},e_i,c$ are explicit functions of the G_{ij},f_i,h_j and their derivatives. Commuting various $\mathcal{L}$'s brings at least one derivative of the G_{ij},f_i,h_j in each term, third order brackets bring second order derivatives or products of first order derivatives, ...

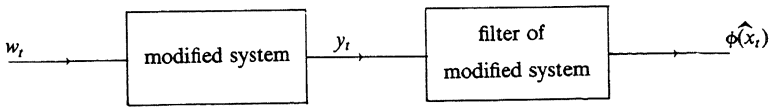
Now if the system described by the f_i, h_j, G_{ij} is supposed to model some real world phenomenon then we can not assume that we know these functions perfectly. In general one would expect that the values of the functions would be known very well, their derivatives less so, their second derivatives still less, etc., and by the time r-th derivatives come into play their values are almost totally unknown.



For $r=2$ the kind of approximation involved is somewhat like illustrated on the above, i.e. something like a piecewise linear approximation with rounded corners. One expects a system close to real one in this sense of diminishing importance of higher derivatives (globally) to behave much like the true one. The comulative effect of small inaccuracies in first derivatives, larger ones in second derivatives, ..., very large ones in r-th derivatives will be such that r order brackets are almost totally unknown. And thus a system approximation which just happened to have all these zero would perform much as the original one but that one would have a filter as in section 3 above and this filter should also give reasonable results for the true system by considering the stability properties of the composed system



which is close to the system with exact filter



Now such a modified system which just happens to have all terms in r-th order brackets of the $\mathcal{L}, \mathcal{L}_i, \mathcal{L}_{ij}$ equal to zero will probably not as a rule exist. But the corresponding filters can certainly be constructed. It suffices to introduce a counting mechanism and to consider the Lie algebra generated by the operator $z\mathcal{L}, z\mathcal{L}_i, z\mathcal{L}_{ij}$. This one is topologically nilpotent and so Wei-Norman theory can be applied to Lie $(z\mathcal{L}, z\mathcal{L}_i, z\mathcal{L}_{ij})$ mod z^n for all n (after which one sets $z=1$.) Here z is an extra parameter.

The argument above indicates that such a procedure could work well. Another not unrelated argument can be based on Volterra series expansions. These ideas have of course a good deal to do with nilpotent and solvable approximation ideas [1], [7].

REFERENCES

1. P.E. CROUCH, *Solvable approximation to control systems,* SIAM J. Control and Opt. 32 (1984), 40-54.
2. T.E. DUNCAN, *Probability densities for diffusion processes with applications to nonlinear filtering,* Ph.D. thesis, Standord, 1967.
3. M. HAZEWINKEL, *On deformations, approximations and nonlinear filtering,* Systems and Control Lett. 1 (1982), 32-36.
4. M. HAZEWINKEL, *Lie algebraic methods in filtering and identification,* Report PM-R86-6, November 1986, CWI, Amsterdam; to appear in Proc. 8-th Int. Symp. IAMP (Luminy, 1986), World Scientific and in Proc. 1-st World Congress Bernouilli Society (Taskent, 1986), VNU Science Press.
5. P.S. KRISHNAPRASAD, S.I. MARCUS, M. HAZEWINKEL, *Current algebras and the identification problem,* Stochastics 11 (1983), 65-101.
6. R.E. MORTENSEN, *Optimal control of continuous time stochastic differential equations,* Ph.D. thesis, Berkeley, 1966.
7. CH. ROCKLAND, *Intrinsic nilpotent approximation,* preprint MIT, LIDS-R-1482, 1985; to appear Acta Appl. Math., 1987.
8. J. WEI, E. NORMAN, *On the global representation of the solutions of linear differential equations as products of exponentials,* Proc. AMS 15 (1964), 327-334.
9. M. ZAKAI, *On the optimal filtering of diffusion processes,* Z. Wahrsch. und verw. Gebiete 11 (1969), 230-243.