

# On Lie Algebras and Finite Dimensional Filtering

MICHIEL HAZEWINDEL

*Erasmus University Rotterdam, Econometric Institute, Bur. Oudlaan 50, Rotterdam, The Netherlands*

and

STEVEN I. MARCUS

*Department of Electrical Engineering, The University of Texas at Austin, Austin, Texas 78712, U.S.A.*

*(Accepted for publication July 20, 1981)*

A Lie algebra  $L(\Sigma)$  can be associated with each nonlinear filtering problem, and the realizability or, better, the representability of  $L(\Sigma)$  or quotients of  $L(\Sigma)$  by means of vector fields on a finite dimensional manifold is related to the existence of finite dimensional recursive filters. In this paper, the structure and representability properties of  $L(\Sigma)$  are analyzed for several interesting and/or well known classes of problems. It is shown that, for certain nonlinear filtering problems,  $L(\Sigma)$  is given by the Weyl algebra

$$W_n = \mathbb{R}\langle x_1, \dots, x_n, \frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n} \rangle.$$

It is proved that neither  $W_n$  nor any quotient of  $W_n$  can be realized with  $C^\infty$  or analytic vector fields on a finite dimensional manifold, thus suggesting that for these problems, no statistic of the conditional density can be computed with a finite dimensional recursive filter. For another class of problems (including bilinear systems with linear observations), it is shown that  $L(\Sigma)$  is a certain type of filtered Lie algebra. The algebras of this class are of a type which suggest that “sufficiently many” statistics are exactly computable. Other examples are presented, and the structure of their Lie algebras is discussed.

## 1. INTRODUCTION

This paper is motivated by the problem of recursively filtering the state  $x_t$  of a nonlinear stochastic system, given the past observations  $z^t$

$= \{z_s, 0 \leq s \leq t\}$ . The systems we consider satisfy the Ito stochastic differential equations

$$\begin{aligned} dx_t &= f(x_t) dt + G(x_t) dw_t \\ dz_t &= h(x_t) dt + R_t^{1/2} dv_t \end{aligned} \quad (\Sigma)$$

where  $x \in \mathbb{R}^n$ ,  $w \in \mathbb{R}^m$ ,  $z \in \mathbb{R}^p$ ,  $w$  and  $v$  are independent unit variance Wiener processes, and  $R > 0$ . The optimal (minimum-variance) estimate of  $x_t$  is of course the conditional mean  $\hat{x}_t \triangleq E[x_t | z^t]$  (also denoted  $\hat{x}_{t|t}$  or  $E^t[x_t]$ );  $\hat{x}_t$  satisfies the (Ito) stochastic differential equation [1]–[3]

$$d\hat{x}_t = \hat{f}(x_t) - (\hat{x}_t \hat{h}^T - \hat{x}_t \hat{h}^T) R^{-1}(t) \hat{h} dt + (\hat{x}_t \hat{h}^T - \hat{x}_t \hat{h}^T) R^{-1}(t) dz_t \quad (1.1)$$

where  $\hat{\cdot}$  denotes conditional expectation given  $z^t$  and  $h$  denotes  $h(x_t)$ . The conditional probability density  $p(t, x)$  of  $x_t$  given  $z^t$  itself (we will assume that  $p(t, x)$  exists) satisfies the stochastic partial differential equation [3], [4]

$$dp(t, x) = \mathcal{L}^* p(t, x) dt + (h(x) - \hat{h}(x))^T R^{-1}(t) (dz_t - \hat{h}(x) dt) p(t, x) \quad (1.2)$$

where

$$\mathcal{L}^*(\cdot) = - \sum_{i=1}^n \frac{\partial(\cdot f_i)}{\partial x_i} + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2(\cdot (GG^T)_{ij})}{\partial x_i \partial x_j} \quad (1.3)$$

is the forward diffusion operator.

Notice that the differential Eq. (1.1) is in general both infinite dimensional and nonrecursive (because of the occurrence of the expectations  $\hat{f}$ ,  $\hat{x} \hat{h}^T$ , and  $\hat{h}$ ). Equation (1.2) is recursive but of course still infinite dimensional. Aside from the linear-Gaussian case in which the Kalman filter is optimal, there are very few known cases in which the conditional mean, or indeed *any* nonconstant statistic of the conditional distribution, can be computed with a *finite dimensional recursive* filter (a number of these are summarized in [5]). More precisely, a *finite dimensional recursive filter* is a stochastic differential equation driven by the observations of the form

$$d\eta_t = a(\eta_t) dt + \sum_{i=1}^p b_i(\eta_t) dz_{it}, \quad (1.4)$$

where  $\eta$  evolves on a *finite dimensional* manifold and  $a$  and  $b_i$  are sufficiently smooth to insure existence and uniqueness (these conditions

will be strengthened later). The conditional statistic  $E[c(x_t)|z^t]$  is said to be *finite dimensionally computable* (FDC) if it can be computed “pointwise” as a function of the state of a finite dimensional recursive filter:

$$\hat{c}(x_t) \triangleq E[c(x_t)|z^t] = \gamma(\eta_t). \quad (1.5)$$

As a practical matter, it is also useful to require that the combined estimator (1.4)–(1.5) yield a statistic  $\hat{c}(x_t)$  which is a continuous function of  $z$ ; we will comment on this later in this section.

Recently, Brockett and Clark [38] and Brockett [6], [7] have shown that Lie algebras and concepts of nonlinear system theory play an important role in nonlinear recursive estimation theory, and Mitter [8], [9] has emphasized the importance of functional integration and group representations and has shown the connection between certain Lie algebras arising in estimation and those arising in mathematical physics. The approach of Brockett [6] is the following. Instead of studying the Eq. (1.2) for the conditional density, we consider the Duncan–Mortensen–Zakai (D–M–Z) equation for an unnormalized conditional density  $\rho(t, x)$  [10], [45, Chapter 6]:

$$d\rho(t, x) = \mathcal{L}^* \rho(t, x) dt + \sum_{i=1}^p h_i(x) \rho(t, x) dz_{it} \quad (1.6)$$

where  $z_i$  and  $h_i$  are the  $i$ th components of  $z$  and  $h$ , and  $\rho(t, x)$  is related to  $p(t, x)$  by the normalization

$$p(t, x) = \rho(t, x) \cdot \left( \int \rho(t, x) dx \right)^{-1}. \quad (1.7)$$

The D–M–Z Eq. (1.6) looks much simpler than (1.2); indeed, (1.6) is an (infinite dimensional) bilinear differential equation [11] in  $\rho$ , with  $z$  considered as the input. This is the first indication (given work on the roles of Lie algebras in solving finite dimensional bilinear equations [32], [33]) that the Lie algebraic and differential geometric techniques developed for finite dimensional systems of this type may be brought to bear here. Modulo some conjectured infinite dimensional extensions of some known results in the finite dimensional case (to be discussed below) this can be made more precise as follows: *suppose* that, for some given initial density, some statistic of the conditional distribution of  $x_t$  given  $z^t$  can be calculated with a finite dimensional recursive estimator of the form (1.4)–(1.5), where  $a$ ,  $b_i$ , and  $\gamma$  are  $C^\infty$  or analytic. Of course, this statistic can also be obtained from  $\rho(t, x)$  by

$$\hat{c}(x_t) = \int c(x) \rho(t, x) dx \left( \int \rho(t, x) dx \right)^{-1}. \quad (1.8)$$

For the rest of the development, it is more convenient to write (1.4) and (1.6) in Fisk–Stratonovich form (so that they obey the ordinary rules of calculus and so that Lie algebraic calculations involving differential operators can be performed as usual):

$$d\eta_i = \tilde{a}(\eta_i) dt + \sum_{i=1}^p b_i(\eta_i) dz_{ii}$$

$$d\rho(t, x) = \left[ \mathcal{L}^* - \frac{1}{2} \sum_{i=1}^p h_i^2(x) \right] \rho(t, x) dt + \sum_{i=1}^p h_i(x) \rho(t, x) dz_{ii} \quad (1.10)$$

where the  $i$ th component

$$\tilde{a}_i(\eta) = a_i(\eta) - \frac{1}{2} \sum_{j,k} b_{jk}(\eta) \frac{\partial b_{ik}}{\partial \eta_j}(\eta)$$

(here  $b_{jk}$  is the  $k$ th component of  $b_j$ ).

The two systems (1.9), (1.5) and (1.10), (1.8) are thus two representations of the same mapping from “input” functions  $z$  to “outputs”  $\hat{c}(x_i)$ : (1.10), (1.8) via a bilinear infinite dimensional state equation, and (1.9), (1.5) via a nonlinear finite dimensional state equation. Motivated by the results of [12], [13] for finite dimensional state equations, the major thesis of [6] is that, under appropriate hypotheses, the Lie algebra  $F$  generated by  $\tilde{a}, b_1, \dots, b_p$  (under the commutator  $[a, b] = (\partial a / \partial \eta) b - (\partial b / \partial \eta) a$ ) should be a homomorphic image (quotient) of the Lie algebra  $L(\Sigma)$  generated by  $e_0 = \mathcal{L}^* - 1/2 \sum_{i=1}^p h_i^2(x)$  and  $e_i = h_i(x)$ ,  $i = 1, \dots, p$  (under the commutator  $[e_0, e_i] = e_0 e_i - e_i e_0$ ), with  $e_0 \rightarrow \tilde{a}$  and  $e_i \rightarrow b_i$ ,  $i = 1, \dots, p$ . On the other hand, if there is a homomorphism  $\phi$  of  $L(\Sigma)$  onto a Lie algebra generated by  $p+1$  complete vector fields  $\tilde{a}, b_1, \dots, b_p$ , on a finite dimensional manifold, then this is an indication (possibly via appropriate globalized and/or integrated infinite dimensional generalizations of some results of [34], [35]) that some conditional statistic may be computable by an estimator of the form (1.9), (1.5). It is not known in what generality such results are valid, especially for cases in which  $L(\Sigma)$  is infinite dimensional, and much work remains to be done (the fact that existence of a finite dimensional filter implies the existence of a Lie algebra homomorphism *has* been made rigorous for a class of estimation problems, including the cubic sensor discussed in Section II, in [36]). However, it is clear (in part, from a number of examples discussed below) that there is a strong relationship in general between the structure of  $L(\Sigma)$  and the existence of finite

dimensional filters. In this paper, we discuss the properties of  $L(\Sigma)$  for some interesting classes of examples. These Lie algebraic calculations give some new insights into certain nonlinear estimation problems and guidance in the search for finite dimensional estimators.

If  $L(\Sigma)$  is finite dimensional (this seems to occur only in very special cases [9], [37]), a finite dimensional estimator can in some cases be constructed by integrating the Lie algebra representation [9]. Indeed, if  $L(\Sigma)$  or any of its quotients is finite dimensional, then by Ado's Theorem [27, p. 202] this Lie algebra has a faithful finite dimensional representation; thus it can be realized with *linear* vector fields on a finite dimensional manifold, which may result in a *bilinear* filter computing some nonzero statistic (see, e.g., [16] and [26] for examples). However, actually computing the mapping from  $\rho(t, x)$  to  $\hat{c}(x_t)$  (i.e., deciding which statistic the filter computes) is a difficult problem from this point of view; at the moment at least, one must usually use other, more direct, methods, to actually construct this mapping or to derive the filter for a particular conditional statistic (see, e.g., [14]–[17]). Also, just a Lie algebra homomorphism from  $L(\Sigma)$  to a Lie algebra of vector fields is not enough. In addition to the homomorphism of Lie algebras, one needs compatibility conditions in terms of isotropy subalgebras [34], [35], or equivalently, in terms of the natural representations of the Lie algebras operating on the spaces of functions on the manifolds involved. Even if  $L(\Sigma)$  or its quotients are *infinite* dimensional, it is still possible that these Lie algebras can be realized by *nonlinear* vector fields on a finite dimensional manifold. Conditions under which this can be done is an unsolved problem in general; we prove in Section 2 that this is *not* possible for certain classes of Lie algebras. As an almost totally trivial example that two vector fields on a finite dimensional manifold *can* generate an infinite dimensional Lie algebra, consider the vector fields  $a = x^2 \partial / \partial x$  and  $b = x^3 \partial / \partial x$  on a one-dimensional manifold; it is easy to see that  $a$  and  $b$  generate the infinite dimensional Lie algebra of vector fields of the form  $x^2 p(x) \partial / \partial x$ , where  $p$  is a polynomial.

If a statistic  $\hat{c}(x_t)$  is finite dimensionally computable, the Lie algebraic approach also gives some insight into the continuity of the estimator. Since there is a Lie algebra homomorphism as discussed above, the vector fields  $b_1, \dots, b_p$  are homomorphic images of the operators  $e_1, \dots, e_p$  which all commute with each other (these are just multiplication operators). Thus  $b_1, \dots, b_p$  also commute, and the results of [18] imply that the filter (1.9) represents a continuous map (in the  $C^0$  and  $L_p$  topologies) from the space of "inputs"  $z$  to the solutions  $\eta$ . Hence, the estimator (1.9), (1.5) gives a continuous map from  $z$  to  $\hat{c}(x_t)$ ; this is a very useful property, indicating the "robustness" of the filter (see also [19], [20]).

Brockett and Clark [38] used this approach to study the estimation of a finite state Markov process observed in additive Brownian motion; the Lie algebraic approach led to the discovery of new low dimensional filters for the conditional distribution, even in some cases when the number of states was arbitrarily large. And even in the extremely well known case of linear systems (Kalman filter), the Lie algebraic approach gives an additional result in that it tells us how to propagate a non-Gaussian initial density.<sup>1)</sup> In this case the Lie algebra is finite dimensional; in fact, one finds higher dimensional relatives of the so called oscillator algebra of some fame in physics (incidentally, this is no accident [9]). In [21], a similar analysis is carried out for an example of the class of estimation problems considered in [14]–[16]; for this class of nonlinear stochastic systems, the conditional mean (and all conditional moments) of  $x_t$  given  $z^t$  are finite dimensionally computable. For this example, the Lie algebra  $L(\Sigma)$  is infinite dimensional but has many finite dimensional quotients corresponding to the Lie algebras of the finite dimensional filters; these are analyzed in detail in [21]. These last two examples, as well as the example of Beneš [17], are special cases of the class considered in Section 3.

In Section 2, we consider estimation problems for which  $L(\Sigma)$  is the Weyl algebra  $W_n$ . A number of examples are given and useful properties of the Weyl algebra are derived; some of these results have been obtained independently by Mitter [9]. The major results of Section 2 are proofs that neither  $W_n$  nor any quotient of  $W_n$  can be realized by vector fields with either  $C^\infty$  or formal power series coefficients on a finite dimensional manifold; this suggests that for these problems, no statistic of the conditional density can be computed with a finite dimensional recursive filter. This does not imply that there will not be appropriate approximation methods. Possibly partial homomorphisms of Lie algebras [39] of  $L(\Sigma)$  into Lie algebras of vector fields will play a role here. Also “deformations of algebras” techniques [40]–[42] suggest a possible approach to approximate methods. For example, the Lie algebra of  $dz_t = dw_t$ ,  $dz_t = (x + \varepsilon x^3) dt + dv_t$  is  $W_1$  for all  $\varepsilon \neq 0$ , but mod  $\varepsilon^n$  this algebra is finite dimensional for all  $n$  [43], [31]. Finally, in Section 4 we present another estimation problem with an interesting Lie algebraic structure and discuss the possible implications of this structure.

## 2. THE WEYL ALGEBRAS $W_n$

The *Weyl algebra*  $W_n$  [22], [23, Chapter 1] is the algebra of all polynomial differential operators; i.e.,  $W_n = \mathbb{R}\langle x_1, \dots, x_n; \partial/\partial x_1, \dots, \partial/\partial x_n \rangle$ . A basis for  $W_n$

consists of all monomial expressions

$$e_{\alpha, \beta} \triangleq x^\alpha \frac{\partial^\beta}{\partial x^\beta} \triangleq x_1^{\alpha_1} \cdots x_n^{\alpha_n} \frac{\partial^{\beta_1}}{\partial x_n^{\beta_1}} \cdots \frac{\partial^{\beta_n}}{\partial x_n^{\beta_n}} \quad (2.1)$$

where  $\alpha, \beta$  range over all multiindices  $\alpha = (\alpha_1, \dots, \alpha_n)$ ,  $\beta = (\beta_1, \dots, \beta_n)$ ,  $\alpha, \beta \in \mathbb{N} \cup \{0\}$  (the non-negative integers).  $W_n$  is a Lie algebra under the Lie bracket; as an example, we state the general formula for  $W_1$ :

$$\begin{aligned} \left[ x^i \frac{\partial^j}{\partial x^j}, x^k \frac{\partial^l}{\partial x^l} \right] &= \sum_{r=1}^j \binom{j}{r} \binom{k}{r} r! x^{i+k-r} \frac{\partial^{j+l-r}}{\partial x^{j+l-r}} \\ &\quad - \sum_{s=1}^l \binom{l}{s} \binom{i}{s} s! x^{i+k-s} \frac{\partial^{j+l-s}}{\partial x^{j+l-s}} \end{aligned} \quad (2.2)$$

where

$$\binom{j}{r} = \frac{j!}{(j-r)!r!}$$

is the binomial coefficient and we have used the convention that  $\binom{j}{r} = 0$  if  $r < 0$  or  $j < r$ . As is easily checked, the center of  $W_n$  (i.e., the ideal of all elements  $Z \in W_n$  such that  $[X, Z] = 0$  for all  $X \in W_n$ ) is the one-dimensional space  $\mathbb{R} \cdot 1$  with basis  $\{1\}$  [22, p. 148]. We next prove the simplicity of the Lie algebra  $W_n/\mathbb{R} \cdot 1$ ; this is of course stronger than showing that  $W_n$  is simple as an associative algebra [22, p. 148]. Our proof follows that of Avez and Heslot [24] for the Lie algebra  $P_n$  of polynomials under the Poisson bracket. A number of the following results are common to  $P_n$  and  $W_n$ , but these two Lie algebras are *not* isomorphic (this is basically because the expression in  $P_n$  corresponding to (2.2) would retain only the terms for  $r=1$  and  $s=1$ ). Hence, one must be careful in literally interpreting results proved for  $P_n$  in the context of  $W_n$  [30].

**THEOREM 2.1** *The Lie algebra  $W_n/\mathbb{R} \cdot 1$  is simple; i.e., it has no ideals other than  $\{0\}$  and  $W_n/\mathbb{R} \cdot 1$ . Equivalently, the only ideals of  $W_n$  are  $\{0\}$ ,  $\mathbb{R} \cdot 1$ , and  $W_n$ .*

*Proof* Suppose  $I$  is an ideal of  $W_n$  which contains a nonconstant element  $X = \sum c_{\alpha\beta} x^\alpha \partial^\beta / \partial x^\beta$ . Since commuting with  $x_i$  reduces  $\beta_i$  by 1 and commuting with  $\partial/\partial x_i$  reduces  $\alpha_i$  by 1, repeated commutation implies that an element of the form  $x_i$  or  $\partial/\partial x_i$  is in  $I$ . Since every element  $Y \in W_n$  can

be obtained by commutation of  $x_i$  (or  $\partial/\partial x_j$ ) with another element of  $W_n$ , this shows that  $I = W_n$ .

This theorem basically shows that if  $W_n$  occurs as the Lie algebra  $L(\Sigma)$  for some estimation problem, then either the unnormalized conditional density itself is finite dimensionally computable or no statistic at all is finite dimensionally computable. The next two theorems complete the argument by showing that in fact neither  $W_n$  nor its quotients can be realized by vector fields on a finite dimensional manifold.

Let  $\hat{V}_m$  be the Lie algebra of vector fields

$$\hat{V}_m \triangleq \left\{ \sum_{i=1}^m f_i(x_1, \dots, x_m) \frac{\partial}{\partial x_i} \right\}$$

with (formal) power series coefficients  $f_i \in \mathbb{R}[[x_1, \dots, x_m]]$ , and let  $V(M)$  be the Lie algebra of  $C^\infty$ -vector fields on a  $C^\infty$ -manifold  $M$ . The proofs of the following theorems are contained in Appendix A.

**THEOREM 2.2** *Fix  $n \neq 0$ . Then there are no non-zero homomorphisms from  $W_n$  to  $\hat{V}_m$  or from  $W_n/\mathbb{R} \cdot 1$  to  $\hat{V}_m$  for any  $m$ .*

**THEOREM 2.3** *Fix  $n \neq 0$ . Then there are no non-zero homomorphisms from  $W_n$  to  $V(M)$  or  $W_n/\mathbb{R} \cdot 1$  to  $V(M)$  for any finite dimensional  $C^\infty$ -manifold  $M$ .*

These results suggest (assuming the appropriate analogs of the results of [6], [12]) that if a system  $\Sigma$  has estimation algebra  $L(\Sigma) = W_n$  for some  $n$ , then neither the conditional density of  $x_i$  given  $z^t$  nor any nonzero statistic of the conditional density can be computed with a finite dimensional filter of the form (1.9) with  $a$  and  $b$   $C^\infty$  or analytic. This is indeed the case for the cubic sensor (Example 2.1) [36] (as was mentioned before). We will give several examples of such systems, but first we present a general method for showing that  $L(\Sigma) = W_n$ .

**THEOREM 2.4** *The Lie algebra  $W_n$  is generated by the elements*

$$x_i, \frac{\partial^2}{\partial x_i^2}, x_i^2 \frac{\partial}{\partial x_i}, i = 1, \dots, n; \quad \text{and} \quad x_i x_{i+1}, i = 1, \dots, n-1.$$

*Proof* (similar to that of [24] for Poisson brackets): Let  $L$  be the Lie algebra generated by these elements. Since

$$\left[ x_i^2 \frac{\partial}{\partial x_i}, x_i^k \right] = k x_i^{k+1},$$



$L$  contains  $x_i^k$ ,  $k \geq 1$ . Now,

$$\left[ \frac{\partial^2}{\partial x_i^2}, x_i \right] = 2 \frac{\partial}{\partial x_i} \quad \text{and} \quad \left[ \frac{\partial}{\partial x_i}, x_i \right] = 1.$$

Also,

$$\left[ \frac{\partial^2}{\partial x_i^2}, x_i^k \left( \frac{\partial}{\partial x_i} \right)^l \right] = 2k x_i^{k-1} \left( \frac{\partial}{\partial x_i} \right)^{l+1} + k(k-1) x_i^{k-2} \left( \frac{\partial}{\partial x_i} \right)^l, \quad k \geq 2; \quad (2.3)$$

with  $l=0$ , (2.3) implies that  $x_i^k (\partial/\partial x_i) \in L$ ,  $k \geq 0$ . Then by induction (2.3) implies that

$$x_i^k \left( \frac{\partial}{\partial x_i} \right)^l \in L \quad \text{for all } k, l \geq 0.$$

Notice that

$$\left[ x_i \frac{\partial^2}{\partial x_i^2}, x_i x_{i+1} \right] = 2 x_i x_{i+1} \frac{\partial}{\partial x_i},$$

and commuting this with  $x_i^k (\partial/\partial x_i)^l$  gives  $x_{i+1} \cdot \mathbb{R} \langle x_i, (\partial/\partial x_i) \rangle \in L$ . Repeated commutation with

$$x_{i+1}^2 \frac{\partial}{\partial x_{i+1}} \quad \text{and} \quad \left( \frac{\partial}{\partial x_{i+1}} \right)^2$$

yields (as above)

$$\mathbb{R} \left\langle x_i, x_{i+1}, \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_{i+1}} \right\rangle.$$

By induction, we have that  $L = W_n$ .

Theorem 2.4 provides a relatively systematic method for showing that  $L(\Sigma) = W_n$  for a particular estimation problem: one need only show that by taking repeated Lie brackets of  $\mathcal{L}^* - (1/2)h^2$  and  $h$ , the generating elements of  $W_n$  given in Theorem 2.4 are obtained. Notice that if  $n=1$ , the generating elements are  $x$ ,  $(\partial^2/\partial x^2)$ , and  $x^2(\partial/\partial x)$ . There is a ‘‘dual’’ result obtained by interchanging  $x_i$  and  $\partial/\partial x_i$  in Theorem 2.4. Some interesting examples are the following.

*Example 2.1 (the cubic sensor problem [9], [25])* Consider the system

$$dx_t = dw_t$$

$$dz_t = x_t^3 dt + dv_t$$

The Lie algebra  $L(\Sigma)$  is generated by the operators

$$e_0 = \frac{1}{2} \frac{\partial^2}{\partial x^2} - \frac{1}{2} x^6, \quad e_1 = x^3.$$

We can compute a sequence of Lie brackets to obtain a sequence of elements  $e_i \in L(\Sigma)$ , eventually obtaining the desired generators of  $W_n$ :

$$[e_0, e_1] = 3x^2 \frac{\partial}{\partial x} + 3x \Rightarrow e_2 = x^2 \frac{\partial}{\partial x} + x$$

$$ad_{e_2}^k e_1 = 3 \cdot 4 \dots (k+2) x^{k+3} \Rightarrow x^k \in L(\Sigma), k \geq 3$$

(where  $ad_{e_2}^0 e_1 = e_1$  and  $ad_{e_0}^{k+1} e_1 = [e_0, ad_{e_2}^k e_1]$ ). Combined with  $e_0, x^6 \in L(\Sigma)$  implies that  $e_3 = \partial^2 / \partial x^2 \in L(\Sigma)$ . Continuing,

$$[e_3, e_2] = 4x \frac{\partial^2}{\partial x^2} + 4 \frac{\partial}{\partial x} \Rightarrow e_4 = x \frac{\partial^2}{\partial x^2} + \frac{\partial}{\partial x}$$

$$[e_4, e_2] = 3x^2 \frac{\partial^2}{\partial x^2} + 6x \frac{\partial}{\partial x} + 1 \Rightarrow e_5 = 3x^2 \frac{\partial^2}{\partial x^2} + 6x \frac{\partial}{\partial x} + 1$$

$$[e_4, e_1] = 6x^3 \frac{\partial}{\partial x} + 9x^2 \Rightarrow e_6 = 2x^3 \frac{\partial}{\partial x} + 3x^2$$

$$[e_3, e_6] = 12x^2 \frac{\partial^2}{\partial x^2} + 24x \frac{\partial}{\partial x} + 6,$$

which combined with  $e_5$  implies that  $e_7 = 1$  and  $e_8 = x^2(\partial^2/\partial x^2) + 2x(\partial/\partial x)$

are in  $L(\Sigma)$ . A few more calculations will complete the demonstration:

$$[e_3, e_8] = 4x \frac{\partial^3}{\partial x^3} + 6 \frac{\partial^2}{\partial x^2} \Rightarrow e_9 = x \frac{\partial^3}{\partial x^3}$$

$$[e_1, e_8] = -6x^4 \frac{\partial}{\partial x} - 12x^3 \Rightarrow e_{10} = x^4 \frac{\partial}{\partial x}$$

$$[e_2, e_9] = -5x^2 \frac{\partial^3}{\partial x^3} - 9x \frac{\partial^2}{\partial x^2} \Rightarrow e_{11} = 5x^2 \frac{\partial^3}{\partial x^3} + 9x \frac{\partial^2}{\partial x^2}$$

$$[e_3, e_{10}] = 8x^3 \frac{\partial^2}{\partial x^2} + 12x^2 \frac{\partial}{\partial x} \Rightarrow e_{12} = 2x^3 \frac{\partial^2}{\partial x^2} + 3x^2 \frac{\partial}{\partial x}$$

$$[e_3, e_{12}] = 12x^2 \frac{\partial^3}{\partial x^3} + 24x \frac{\partial^2}{\partial x^2} + 6 \frac{\partial}{\partial x} \Rightarrow e_{13} = 2x^2 \frac{\partial^3}{\partial x^3} + 4x \frac{\partial^2}{\partial x^2} + \frac{\partial}{\partial x}$$

Now  $e_{13}$ ,  $e_{11}$ , and  $e_4$  are all linear combinations of the elements  $x^2(\partial^3/\partial x^3)$ ,  $x(\partial^2/\partial x^2)$ , and  $\partial/\partial x$ , and the coefficient matrix

$$\begin{bmatrix} 0 & 1 & 1 \\ 5 & 9 & 0 \\ 2 & 4 & 1 \end{bmatrix}$$

is nonsingular. It follows that  $L(\Sigma)$  contains  $e_{14} = \partial/\partial x$ ,  $e_{15} = x(\partial^2/\partial x^2)$ , and  $e_{16} = x^2(\partial^3/\partial x^3)$ . Finally,

$$[e_{14}, e_1] = 3x^2 \Rightarrow e_{17} = x^2$$

$$[e_{14}, e_{17}] = 2x \Rightarrow e_{18} = x$$

which combined with  $e_2$  gives  $x^2(\partial/\partial x) \in L$ ; thus by Theorem 2.4,  $L(\Sigma) = W_1$ .

Analogous computation of selected Lie brackets and the use of Theorem 2.4 yields similar results for the following examples.

*Example 2.2* For the system

$$dx_t = x_t^3 dt + dw_t$$

$$dz_t = x_t dt + dv_t,$$

$L(\Sigma)$  is generated by  $1/2(\partial^2/\partial x^2) - x^3(\partial/\partial x) - (7/2)x^2$  and  $x$ , and  $L(\Sigma) = \mathbb{R}$

*Example 2.3 (mixed linear-bilinear type)* Consider the system with state equations

$$dx_t = dw_{1t}$$

$$dy_t = x_t dt + x_t dw_{2t}$$

with observations

$$dz_t = y_t dt + dv_t.$$

$L(\Sigma)$  is generated by

$$\frac{1}{2} \frac{\partial^2}{\partial x^2} + \frac{1}{2} x^2 \frac{\partial^2}{\partial y^2} - x \frac{\partial}{\partial y} - \frac{1}{2} y^2 \quad \text{and} \quad y;$$

it is shown in Appendix B that  $L(\Sigma) = W_2$ . The same result is obtained if the  $x_t dt$  term is absent in the  $y$  equation; in that case we have a multi-dimensional Wiener integral of Brownian motion observed in Brownian motion noise.

*Example 2.4* Consider the system with state equations

$$dx_t = dw_t$$

$$dy_t = x_t^2 dt$$

and observations

$$dz_{1t} = x_t dt + dv_{1t}$$

$$dz_{2t} = y_t dt + dv_{2t}.$$

$L(\Sigma)$  is generated by

$$\frac{1}{2} \frac{\partial^2}{\partial x^2} - x^2 \frac{\partial}{\partial y} - \frac{1}{2} x^2 - \frac{1}{2} y^2, x, \quad \text{and} \quad y;$$

it is easily shown that  $L(\Sigma) = W_2$ . This is the example studied in [21], but here we have the additional observation  $z_2$ ; the relationship between these two examples will be examined in the next section.

### 3. PRO-FINITE DIMENSIONAL FILTERED LIE ALGEBRAS

A Lie algebra  $L$  is defined to be a *pro-finite dimensional filtered Lie algebra* if  $L$  has a decreasing sequence of *ideals*  $L = L_{-1} \supset L_0 \supset L_1 \supset \dots$  such that

$$a) \cap L_i = 0$$

b)  $L/L_i$  is a finite dimensional Lie algebra for all  $i$ .

The terminology is somewhat analogous to that of pro-finite groups [28]; no completeness assumptions are made, however. Notice that (a) implies that there is an injection from  $L$  to  $\bigoplus_i L/L_i$ . In the context of the estimation problem, this would correspond to  $L(\Sigma)$  having an infinite number of finite dimensional quotients; if each of these can be realized with a recursively filterable statistic (a plausible conjecture), then the injectivity of the map makes it reasonable to conjecture that these statistics represent some type of power series expansion of the conditional density. Of course, in addition to those discussed in Section 1, other difficult technical questions such as moment determinacy will also be relevant here, but the structure of the Lie algebra should provide some guidance as to possible successful approaches to the problem and some insight into the structure of the resulting approximations.

*Example 3.1* [21] A simple example of the class considered in [14]–[16] is given by the state equations

$$dx_t = dw_t$$

$$dy_t = x_t^2 dt$$

and the observations

$$dz_t = x_t dt + dv_t$$

with  $x_0$  Gaussian. The computation of  $\hat{x}_t$  is of course straightforward by means of the Kalman filter; however, as shown in [14]–[16], all conditional moments of  $y_t$  can also be computed recursively with finite dimensional filters.  $L(\Sigma)$  is generated by

$$e_0 = -x^2 \frac{\partial}{\partial y} + \frac{1}{2} \frac{\partial^2}{\partial x^2} - \frac{1}{2} x^2 \quad \text{and} \quad e_1 = x;$$

as shown in [21], a basis for  $L(\Sigma)$  is given by  $e_0$  and

$$\left\{ x \frac{\partial^i}{\partial y^i}, \frac{\partial}{\partial x} \frac{\partial^i}{\partial y^i}, \frac{\partial^i}{\partial y^i}; i=0, 1, 2, \dots \right\}.$$

Defining  $L_i$  to be the ideal generated by  $x(\partial^i/\partial y^i)$ ,  $i=0, 1, 2, \dots$ , it is easy to see that  $L(\Sigma)$  is a pro-finite dimensional filtered Lie algebra, and realizations of the  $L(\Sigma)/L_i$  in terms of recursively filterable statistics are given in [21]. In addition,  $L(\Sigma)$  is solvable [21].

A similar analysis for systems of the form of Example 3.1, with  $x_t^2$  replaced by a general monomial  $x_t^p$  has also been done [31]; for  $p>2$ , a similar but more complex Lie algebraic structure is exhibited. It is interesting to compare Example 3.1 with Example 2.4, which is the same except for the additional observation  $dz_{2t} = y_t dt + dv_{2t}$ ; in that case  $L(\Sigma) = W_2$ , so that no conditional statistic can be computed *exactly* with a finite dimensional filter. However, it is probable that, due to the additional observation, a *suboptimal* approximate filter (such as the Extended Kalman Filter) for the conditional mean of  $y_t$  will result in lower mean-square error than the *optimal* filter which computes  $\hat{y}_t$  in Example 3.1. Thus some care must be taken in interpreting the Lie algebraic structure of a nonlinear estimation problem; this structure has direct implications on the *exact* computation of conditional statistics, but its implications on approximate filtering remains to be investigated (see [31]).

*Example 3.2 (degree increasing operators and bilinear systems)* Consider a system of the form  $(\Sigma)$ , and suppose that  $f$ ,  $G$ , and  $h$  are analytic with  $f(0)=0$  and  $G(0)=0$ , so that the power series expansions of  $f$  and  $G$  around zero are of the form

$$f(x) = \sum_{|\alpha| \geq 1} f_\alpha x^\alpha, \quad G(x) = \sum_{|\alpha| \geq 1} G_\alpha x^\alpha, \quad (3.1)$$

where  $|\alpha| = \alpha_1 + \dots + \alpha_n$ . It follows that

$$G(x)G'(x) = \sum_{|\alpha| \geq 2} \tilde{G}_\alpha(x) x^\alpha.$$

An example of such systems is the class of bilinear systems

$$\begin{aligned} dx_t &= Ax_t + \sum_{i=1}^p B_i x_t dw_t^i \\ dz_t &= Cx_t dt + dv_t \end{aligned} \quad (3.2)$$

Another example is

$$dx_t = x_t dt + \sin x_t dw_t$$

$$dz_t = h(x_t) dt + dv_t$$

with  $h$  analytic; in general, a wide variety of examples can be found.

Let  $M = \mathbb{R}[[x_1, \dots, x_n]]$  be the module of all (formal) power series in  $x_1, \dots, x_n$ , and define the submodules

$$M_i = \left\{ \sum a_\alpha x^\alpha \mid a_\alpha = 0 \text{ for } |\alpha| \leq i \right\}, \quad i = 0, 1, 2, \dots,$$

so that, e.g.,  $M_0$  consists of those power series with zero constant term. If  $\Sigma$  is a system satisfying the condition (3.1), it follows that for all  $i$ , the forward diffusion operator (1.3) satisfies

$$\mathcal{L}^* M_i \subset M_i;$$

hence

$$(\mathcal{L}^* - \frac{1}{2}h^2(x))M_i \subset M_i$$

and of course

$$h(x)M_i \subset M_i.$$

Since the two generators of  $L(\Sigma)$  thus leave  $M_i$  invariant, it is obvious that  $L(\Sigma)M_i \subset M_i$ ; thus, each element of  $L(\Sigma)$  can only increase (or leave the same) the degree of the first term in the power series expansion of an element of  $M$ . Let

$$L_i = \{X \in L(\Sigma) \mid XM \subset M_{i+1}\}, \quad i = -1, 0, 1, 2, \dots$$

Then  $L_i$  is an ideal in  $L(\Sigma)$  and we have an induced representation

$$\rho_i: L/L_i \rightarrow \text{End}(M/M_{i+1}).$$

Because  $M/M_{i+1}$  is finite dimensional, so is  $L/L_i$ , since  $\rho_i$  is injective (by definition of  $L_i$ ). It is obvious that  $\bigcap L_i = \{0\}$ ; thus  $L(\Sigma)$  is a pro-finite dimensional filtered Lie algebra, with filtration  $L_i$ . One additional

structural feature of this filtration is that  $L_0/L_i$  is a nilpotent Lie algebra for  $i=1,2,\dots$ ; also,  $L_i/L_{i+1}$  is abelian for all  $i \geq 0$ . The nilpotency of the  $L_0/L_i$  is a property also possessed by the filtration of Example 3.1.

The Lie algebraic structure of a scalar bilinear example of the form (3.2) has been derived independently, and in more detail, by Baras and Blankenship [44].

Since many systems can be well approximated by bilinear ones, these results may have important implications for approximate nonlinear filtering. We close this section with two interesting examples of this class; the first is a bilinear system of the form (3.2), but in which some elements of  $A$  are also unknown and must be estimated. The second is an angle modulation problem.

*Example 3.3 (Bilinear system with unknown parameter)* The simplest example of this type is

$$dx_t = \alpha_t x_t dt + x_t dw_t$$

$$d\alpha_t = 0$$

$$dz_t = x_t dt + dv_t$$

Here both the state  $x_t$  and parameter  $\alpha$  are to be estimated recursively. The Lie algebra  $L(\Sigma)$  is generated by

$$\frac{1}{2}x^2 \frac{\partial^2}{\partial x^2} + 2x \frac{\partial}{\partial x} + 1 - \alpha x \frac{\partial}{\partial x} - \alpha - \frac{1}{2}x^2$$

and  $x$ . Both of these operators are “degree increasing” when operating on  $\mathbb{R}[[x, \alpha]]$ , so  $L(\Sigma)$  is a pro-finite dimensional filtered Lie algebra.

*Example 3.4 (Angle modulation without process noise)* Consider the problem of observing

$$dz_{1t} = \sin(\omega t + \theta) dt + dv_{1t}$$

$$dz_{2t} = \cos(\omega t + \theta) dt + dv_{2t}$$

where  $\omega$  and  $\theta$  are constant random variables to be estimated. To place



this problem in the present framework, we have the three state equations

$$\dot{\omega} = 0$$

$$\dot{\theta} = 0$$

$$\dot{t} = 1$$

The Lie algebra  $L(\Sigma)$  is generated by

$$e_0 = \frac{\partial}{\partial t} - \frac{1}{2}, \quad e_1 = \sin(\omega t + \theta), \quad \text{and} \quad f_1 = \cos(\omega t + \theta).$$

It is easily shown that  $L(\Sigma)$  has basis elements  $e_0, e_i = \omega^i \sin(\omega t + \theta), f_i = \omega^i \cos(\omega t + \theta), i = 0, 1, 2, \dots$ . The nonzero commutation relations are  $[e_0, e_i] = f_{i+1}, [e_0, f_i] = -e_{i+1}$ . Hence  $L(\Sigma)$  is a pro-finite dimensional filtered Lie algebra, with filtration  $\{L_i\}$ , where  $L_i$  is the ideal generated by  $e_{i+1}$  and  $f_{i+1}, i = 0, 1, 2, \dots$ . Phase-lock loops are often used for filtering problems such as this, but the form of the *optimal* estimator is unknown. This calculation suggests that an infinite number of statistics of the conditional density may be finite dimensionally computable.

#### 4. A FINAL EXAMPLE

There are other filtering problems which do not fall into the above classes, but which have interesting Lie algebra structures with possible implications for finite dimensional filtering. One example is the following.

*Example 4.1* The system of this example is

$$dx_t = dw_t$$

$$dy_t = e^{x_t} dt$$

$$dz_t = x_t dt + dv_t$$

This does not quite fall into the class discussed in [14]–[16] (as does Example 3.1), since the  $y$  equation contains  $e^{x_t}$  rather than a polynomial in  $x_t$ . The conditional expectation  $\hat{x}_t$  is again computed by the Kalman filter, but the computation of  $\hat{y}_t$  is much more difficult. The Lie algebra

$L(\Sigma)$  is generated by

$$\frac{1}{2} \frac{\partial^2}{\partial x^2} - e^x \frac{\partial}{\partial y} - \frac{1}{2} x^2 \quad \text{and} \quad x;$$

the structure of  $L(\Sigma)$  is as follows. It has as basis the elements

$$x, \frac{\partial}{\partial x}, 1, \frac{\partial^2}{\partial x^2} - x^2; E_{ijk} = x^i e^{jx} \frac{\partial^j}{\partial y^j} \frac{\partial^k}{\partial x^k}, \quad j \geq 1, i, k \geq 0.$$

Let  $I_n (n \geq 1)$  be the subspace spanned by  $E_{ijk}$  with  $j \geq n$ , and let  $I'_n$  be the subspace spanned by 1 and  $E_{ijk}$  with  $j \geq n$ . Then the only ideals of  $L(\Sigma)$  are  $I_n, I'_n, \mathbb{R} \cdot 1$ , and  $I_1 \oplus \mathbb{R} \cdot 1 \oplus \mathbb{R}(\partial/\partial x) \oplus \mathbb{R} \cdot x$ . The quotients  $I_n/I_{n+1}$  are infinite dimensional and abelian, so that  $L(\Sigma)/I_n$  are successive extensions of the oscillator algebra  $L(\Sigma)/I_1$  (the algebra of the linear filtering problem [6]–[9]) by infinite dimensional abelian kernels. Also,

$$\bigcap_n I_n = \{0\}.$$

Due to this structure, it seems unlikely that there will be injections from  $L(\Sigma)$  itself into  $\hat{V}_m$ ; however, it does seem possible that the  $L(\Sigma)/I_n$  are realizable as (infinite dimensional) Lie algebras of vector fields on some finite dimensional manifold.

## References

- [1] H. J. Kushner, Dynamical equations for optimal nonlinear filtering, *J. Diff. Equations* **3** (1967), 179–190.
- [2] M. Fujisaki, G. Kallianpur, and H. Kunita, Stochastic differential equations for the nonlinear filtering problem, *Osaka J. Math.* **1** (1972), 19–40.
- [3] R. S. Liptser and A. N. Shiriyayev, *Statistics of Random Processes I*. New York: Springer-Verlag, 1977.
- [4] 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.
- [5] J. H. Van Schuppen, Stochastic filtering theory: A discussion of concepts, methods and results, in *Stochastic Control Theory and Stochastic Differential Systems*, M. Kohlmann and W. Vogel, eds., New York: Springer-Verlag, 1979, pp. 209–226.
- [6] R. W. Brockett, Remarks on finite dimensional nonlinear estimation, presented at the Conference on Algebraic and Geometric Methods in System Theory, Bordeaux, France, September 1978; also in *Analyse des Systèmes, Asterisque*, Vol. 75–76, 1980, pp. 47–55.
- [7] R. W. Brockett, Classification and equivalence in estimation theory, *Proc. 1979 IEEE Conf. on Decision and Control*, Ft. Lauderdale, December 1979.

- [8] S. K. Mitter, Filtering theory and quantum fields, presented at the Conference on Algebraic and Geometric Methods in System Theory, Bordeaux, France, September 1978; also in *Analyse des Systèmes, Asterisque*, Vol. 75–76, 1980.
- [9] S. K. Mitter, On the analogy between mathematical problems of non-linear filtering and quantum physics, *Recherche di Automatica* **10** (1980), 163–216.
- [10] M. Zakai, On the optimal filtering of diffusion processes, *Z. Wahr. Verw. Geb.* **11** (1969), 230–243.
- [11] R. W. Brockett, System theory on group manifolds and coset spaces, *SIAM J. Control* **10** (1972), 265–284.
- [12] H. J. Sussmann, Existence and uniqueness of minimal realizations of nonlinear systems, *Math. Systems Theory* **10** (1977), 263–284.
- [13] R. Hermann and A. J. Krener, Nonlinear controllability and observability, *IEEE Trans. Automatic Control* **AC-22** (1977), 728–740.
- [14] S. I. Marcus and A. S. Willsky, Algebraic structure and finite dimensional nonlinear estimation, *SIAM J. Math. Anal.* **9** (1978), 312–327.
- [15] S. I. Marcus, Optimal nonlinear estimation for a class of discrete-time stochastic systems, *IEEE Trans. Automatic Control* **AC-24** (1979), 297–302.
- [16] S. I. Marcus, S. K. Mitter, and 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., New York Academic Press, 1980, pp. 387–406.
- [17] V. E. Benes, Exact finite dimensional filters for certain diffusions with nonlinear drift, *Stochastics* **5** (1981), 65–92.
- [18] M. I. Freedman and J. C. Willems, Smooth representation of systems with differentiated inputs, *IEEE Trans. Automatic Control* **AC-23** (1978), 16–21.
- [19] J. M. C. Clark, The design of robust approximations to the stochastic differential equations of nonlinear filtering, in *Communication Systems and Random Process Theory*, ed. J. K. Skwirzynski, NATO Advanced Study Institute Series, Alphen aan den Rijn: Sijthoff and Noordhoff, 1978.
- [20] M. H. A. Davis, On a multiplicative functional transformation arising in nonlinear filtering theory, *Z. Wahrscheinlichkeitstheorie ver. Geb.* **54** (1980), 125–139.
- [21] C.-H. Liu and S. I. Marcus, The Lie algebraic structure of a class of finite dimensional nonlinear filters, in *Algebraic and Geometric Methods in Linear Systems Theory, Lectures in Applied Mathematics*, Vol. 18, C. I. Byrnes and C. F. Martin, eds., Providence: Amer. Math. Soc., 1980, pp. 277–297.
- [22] J. Dixmier, *Enveloping Algebras*. Amsterdam: North-Holland, 1977.
- [23] J.-E. Björk, *Rings of Differential Operators*. Amsterdam: North-Holland, 1979.
- [24] A. Avez and A. Heslot, “L’algèbre de Lie des polynômes en les coordonnées canoniques munie du crochet de Poisson, *C. R. Acad. Sci. Paris A*, **288**, 7 Mai 1979, 831–833.
- [25] R. S. Bucy and J. Pages, A priori error bounds for the cubic sensor problem, *IEEE Trans. Automatic Control* **AC-23** (1978), 88–91.
- [26] S. D. Chikte and J. T.-H. Lo, Optimal filters for bilinear systems with nilpotent Lie algebras, *IEEE Trans. Automatic Control* **AC-24** (1979), 948–953.
- [27] N. Jacobson, *Lie Algebras*, New York: Wiley-Interscience, 1962.
- [28] J.-P. Serre, *Cohomologie Galoisienne*, Lecture Notes in Math. 5. New York: Springer-Verlag, 1964.
- [29] M. Golubitsky and V. Guillemin, *Stable Mappings and Their Singularities*, New York: Springer-Verlag, 1973.
- [30] A. Joseph, Commuting polynomials in quantum canonical operators and realizations of Lie algebras, *J. Math. Phys.* **13** (1972), 351–357.

- [31] S. I. Marcus, C.-H. Liu, and G. L. Blankenship, Lie algebras and asymptotic expansions for some nonlinear filtering problems, *Proc. 1981 Joint Automatic Control Conf.*, Charlottesville, June 1981.
- [32] K. T. Chen, Decomposition of differential equations, *Math. Ann.* **146** (1962), 263–278.
- [33] J. Wei and E. Norman, On the global representation of the solutions of linear differential equations as a product of exponentials, *Proc. Am. Math. Soc.* **15** (1964), 327–334.
- [34] A. J. Krener, On the equivalence of control systems and the linearization of nonlinear systems, *SIAM J. Control* **11** (1973), 670–676.
- [35] H. J. Sussmann, An extension of a theorem of Nagano on transitive Lie algebras, *Proc. Am. Math. Soc.* **45** (1974), 349–356.
- [36] M. Hazewinkel, S. I. Marcus, and H. J. Sussman, Nonexistence of exact finite dimensional filters for the cubic sensor problem, in preparation.
- [37] D. Ocone, Nonlinear filtering problems with finite dimensional estimation algebras, *Proc. 1980 Joint Automatic Control Conf.*, San Francisco, August 1980.
- [38] R. W. Brockett and J. M. C. Clark, The geometry of the conditional density equation, *Analysis and Optimization of Stochastic Systems*, O. L. R. Jacobs, et al., eds., New York: Academic Press, 1980, pp. 299–309.
- [39] R. W. Goodman, *Nilpotent Lie Groups*, Lecture Notes in Math. 562. New York: Springer-Verlag, 1976.
- [40] M. Gestenhaber, On the deformations of rings and algebras, I-IV. *Ann. of Math.* **79** (1964), 59–103; **84** (1966), 1–19; **88** (1968), 1–34; **99**, (1974), 257–276.
- [41] A. Nijenhuis and R. Richardson, Cohomology and deformations in graded Lie algebras, *Bull. Amer. Math. Soc.* **72** (1966), 1–29.
- [42] R. Hermann, Analytic contribution of group representations II, *Comm. Math. Physics* **3** (1966), 53–74.
- [43] M. Hazewinkel, On deformations, approximations, and nonlinear filtering, *Systems and Control Letters* **1** (1981), 32–36.
- [44] J. S. Baras and G. L. Blankenship, Nonlinear filtering of diffusion processes: a generic example, to appear.
- [45] E. Wong, *Stochastic Processes in Information and Dynamical Systems*. New York: McGraw-Hill, 1971.

## Appendix A

### PROOF OF THEOREMS 2.2 AND 2.3

#### A.1 Filtrations and Preliminary Results

DEFINITION A.1 A Lie algebra  $L$  admits a *filtration* (or is a *filtered Lie algebra*) if there exists a sequence of subalgebras  $L = L_{-1} \supset L_0 \supset L_1 \supset \dots$  such that

$$\bigcap L_i = \{0\} \tag{A.1}$$

$$[L_i, L_j] \subset L_{i+j} \tag{A.2}$$

$$\dim(L_i/L_{i+1}) < \infty; \quad i = -1, 0, 1, \dots \tag{A.3}$$

*Example A.1* A prime example of filtered Lie algebras are the  $\hat{V}_n$ . The filtration is defined as follows:  $L_i$  consists of all vector fields  $\sum c_{\alpha,j} x^\alpha (\partial/\partial x_j)$  with  $c_{\alpha,j} = 0$  for all  $\alpha$  with  $|\alpha| \leq i$ , where the norm of the multiindex  $(\alpha_1, \dots, \alpha_n)$  is  $|\alpha| = \alpha_1 + \dots + \alpha_n$ .

Given a filtration  $L_{-1} \supset L_0 \supset L_1 \supset \dots$  on a Lie algebra  $L$ , we define a *valuation function*  $v: L \rightarrow \mathbb{N} \cup [0, -1] \cup \{\infty\}$  by

$$v(x) = \max \{j \mid x \in L_j\}.$$

Properties (A.1) and (A.2) of the filtration translate into

$$v(x) = \infty \Leftrightarrow x = 0 \quad (\text{A.4})$$

$$v([x, y]) \geq v(x) + v(y), \quad (\text{A.5})$$

and the fact that the  $L_i$  are vector spaces implies that

$$v(ax + by) \geq \min(v(x), v(y)); \quad x, y \in L, a, b \in \mathbb{R} \quad (\text{A.6})$$

and

$$\begin{aligned} v(x+y) &= v(x) \quad \text{if } v(x) < v(y) \\ v(ax) &= v(x) \quad \text{if } a \neq 0 \end{aligned} \quad (\text{A.7})$$

In addition, we will need the following results concerning  $W_1$ . First, we have the formula

$$\left[ \frac{\partial^n}{\partial x^n}, x^r \right] - \left[ \frac{\partial^{n-1}}{\partial x^{n-1}}, x^r \frac{\partial}{\partial x} \right] - r \left[ \frac{\partial^{n-1}}{\partial x^{n-1}}, x^{r-1} \right] = r x^{r-1} \frac{\partial^{n-1}}{\partial x^{n-1}}; \quad (\text{A.8})$$

this is easily proved by using (2.2) and formulas for the binomial coefficients. The following lemma, which also follows by a straightforward application of (2.2), shows that  $x^k (\partial^l / \partial x^l)$  is an ‘‘approximate eigenvector’’ of  $x^t (\partial^t / \partial x^t)$ .

**LEMMA A.1** *Let  $l < t \leq k \neq 1$  be natural numbers. Then there are a nonzero  $c \in \mathbb{R}$  and  $d_1, \dots, d_{t-1} \in \mathbb{R}$  such that*

$$\left[ x^t \frac{\partial^t}{\partial x^t}, x^k \frac{\partial^l}{\partial x^l} \right] = c x^k \frac{\partial^l}{\partial x^l} + \sum_{i=1}^{t-1} d_i x^{k+i} \frac{\partial^{l+i}}{\partial x^{l+i}}$$

The proof of the next lemma is quite involved and is contained in Section A.3.

LEMMA A.2 *Suppose that  $W_1 = L_{-1} \supset L_0 \supset L_1 \supset \dots$  is a sequence of subalgebras of  $W_1$  satisfying (A.2), (A.3),  $\dim(W_1/L_2) < \infty$ , and either  $\cap L_i = \{0\}$  or  $\cap L_i = \mathbb{R} \cdot 1$ . Let  $v$  be the valuation function defined by the filtration. Then  $v(x^n) \rightarrow \infty$  as  $n \rightarrow \infty$ .*

## A.2 Proof of Theorem 2.2

The proof will be carried out for  $W_1$ ; the proof is virtually identical for  $W_1/\mathbb{R} \cdot 1$ , and the result is true a fortiori for  $W_m$  since  $W_1$  is clearly isomorphic to the subalgebra of  $W_m$  consisting of expressions in  $x_1$  and  $\partial/\partial x_1$  only. Suppose that there is a nonzero homomorphism  $\phi$  from  $W_1$  to  $\hat{V}_m$ . Then  $W_1$  has a filtration defined by the subalgebras  $M_i \triangleq \phi^{-1}(L_i)$ , where  $\{L_i\}$  is the filtration on  $\hat{V}_m$  defined in Example A.1; let  $v$  be the corresponding valuation function on  $W_1$ . Since  $\hat{V}_m/L_2$  is finite dimensional, so is  $W_1/M_2$ ; thus Lemma A.2 implies that  $v(x^i) \rightarrow \infty$  as  $i \rightarrow \infty$ . We claim it also follows that

$$v\left(x^{k+i} \frac{\partial^{l+i}}{\partial x^{l+i}}\right) \rightarrow \infty \quad \text{as } i \rightarrow \infty, \quad (\text{A.9})$$

and that this will lead to a contradiction.

First notice that

$$\left[ \frac{\partial^2}{\partial x^2}, x^{k+i+2} \right] = 2(k+i+2)x^{k+i+1} \frac{\partial}{\partial x} + (k+i+2)(k+i+1)x^{k+i},$$

so that from (A.5)–(A.7) and the fact that  $v(X) \geq -1$  for all  $X \in W_1$ ,

$$v\left(x^{k+i+1} \frac{\partial}{\partial x}\right) \geq \min \left\{ v(x^{k+i}), v\left[ \frac{\partial^2}{\partial x^2}, x^{k+i+2} \right] \right\} \geq \min \{v(x^{k+i}), v(x^{k+i+2}) - 1\}. \quad (\text{A.10})$$

Then taking  $r = k+i+1$  and  $n = l+i+1$  in formula (A.8) and using (A.10)

yields

$$\begin{aligned}
& v\left(x^{k+i} \frac{\partial^{l+i}}{\partial x^{l+i}}\right) \\
& \geq \min \left\{ v\left[\frac{\partial^{l+i+1}}{\partial x^{l+i+1}}, x^{k+i+1}\right], v\left[\frac{\partial^{l+i}}{\partial x^{l+i}}, x^{k+i+1} \frac{\partial}{\partial x}\right], v\left[\frac{\partial^{l+i}}{\partial x^{l+i}}, x^{k+i}\right] \right\} \\
& \geq \min \left\{ v(x^{k+i+1}) - 1, v\left(x^{k+i+1} \frac{\partial}{\partial x}\right) - 1, v(x^{k+i}) - 1 \right\} \\
& \geq \min \{v(x^{k+i+1}) - 1, v(x^{k+i+2}) - 1, v(x^{k+i}) - 1\}
\end{aligned}$$

which converges to  $\infty$  as  $i \rightarrow \infty$ , proving (A.9).

Now choose  $t_0 \in \mathbb{N}$  such that

$$v\left(x^t \frac{\partial^t}{\partial x^t}\right) \geq 1 \quad \text{for } t \geq t_0. \quad (\text{A.11})$$

Choose any  $k_0 \geq 1$  and consider the sequence  $\{v(x^{k_0+l}(\partial^l/\partial x^l)); l = 0, 1, 2, \dots\}$ . Then because by (A.9) this sequence converges to  $\infty$  there is for any  $l_0$  an  $l_1 \geq l_0$  such that

$$v\left(x^{k_0+l_1+i} \frac{\partial^{l_1+i}}{\partial x^{l_1+i}}\right) > v\left(x^{k_0+l_1} \frac{\partial^{l_1}}{\partial x^{l_1}}\right); \quad i \geq 1 \quad (\text{A.12})$$

Take  $l_0 = t_0 + 1$ , choose  $l_1$  such that (A.12) holds, and take  $t = l_1 + 1$ .

Then we can apply Lemma A.1 with  $t = l_1 + 1$ ,  $l = l_1$ , and  $k = k_0 + l_1$  (notice that the assumptions are satisfied). We find

$$\left[x^t \frac{\partial^t}{\partial x^t}, x^k \frac{\partial^l}{\partial x^l}\right] = cx^k \frac{\partial^l}{\partial x^l} + \sum_{i=1}^{t-1} d_i x^{k+i} \frac{\partial^{l+i}}{\partial x^{l+i}}$$

Because of (A.12), we have by (A.7) that

$$v\left(cx^k \frac{\partial^l}{\partial x^l} + \sum_{i=1}^{t-1} d_i x^{k+i} \frac{\partial^{l+i}}{\partial x^{l+i}}\right) = v\left(x^k \frac{\partial^l}{\partial x^l}\right). \quad (\text{A.13})$$

But because  $v(x^t(\partial^t/\partial x^t)) \geq 1$  (cf., (A.11)) we have by (A.5) that

$$v\left(\left[x^t \frac{\partial^t}{\partial x^t}, x^k \frac{\partial^l}{\partial x^l}\right]\right) \geq 1 + v\left(x^k \frac{\partial^l}{\partial x^l}\right).$$

Comparing this to (A.13) gives a contradiction, completing the proof of Theorem 2.2.

### A.3 PROOF OF LEMMA A.2

#### A.3.1 A preliminary reduction

LEMMA A.3 *Under the hypotheses of Lemma A.2, if there is an element  $x^n \in W_1$ ,  $n \geq 2$ , such that  $v(x^n) \geq 0$ , then  $v(x^m) \rightarrow \infty$  as  $m \rightarrow \infty$ .*

*Proof* Suppose we had such an element  $x^n$ . Because  $\dim(W_1/L_2) < \infty$ , there is an element

$$Y = \sum_{j=r}^s a_j \frac{\partial^j}{\partial x^j} \in W_1, a_s \neq 0, s \geq 2,$$

of valuation  $\geq 2$ . A simple computation shows that  $ad_{x^n}^s Y = n^s s! a_s x^{s(n-1)}$ , which has valuation  $\geq 2$  (by repeatedly using (A.5) and  $v(x^n) \geq 0$ ). Thus we now have an element  $x^k$ ,  $k \geq 2$ , with  $v(x^k) \geq 2$ . Now

$$Z = \left[ x^2 \frac{\partial^2}{\partial x^2}, x^k \right] = k(k-1)x^k + 2k x^{k+1} \frac{\partial}{\partial x}$$

has valuation  $\geq 1$ , and for any  $q$ ,  $ad_{x^k}^q Z = cx^{pk+q}$ ,  $c \neq 0$ . For any  $m \geq k$ , there exist nonnegative integers  $p, q$  such that  $m = pk + q$ , so we have for  $m$  large enough:

$$\begin{aligned} v(x^m) &= v(x^{pk+q}) = v(ad_{x^k}^p Z) \\ &\geq p \cdot v(Z) + v(x^q) \\ &\geq \left[ \frac{m}{k} \right] - 1 \geq \frac{m}{k} - 2, \end{aligned}$$

where  $[m/k]$  denotes the largest integer  $\leq m/k$ . Since  $k$  is fixed, this shows that  $v(x^m) \rightarrow \infty$  as  $m \rightarrow \infty$ .



### A.3.2 Some combinatorial lemmas

To prove that under the conditions of Lemma A.2 there is indeed an  $n \in \mathbb{N}$ ,  $n \geq 2$  such that  $v(x^n) \geq 0$ , we need some combinatorial lemmas.

LEMMA A.4 *Let  $r, s \in \mathbb{N}$  with  $r < s$ , and let  $a \in \mathbb{R}$ . Then*

$$\sum_{i=0}^s \binom{s}{i} (-1)^i (a+i+1)(a+i+2)\dots(a+i+r) = 0$$

*Proof*<sup>2)</sup> The proof is by induction on  $(r, s)$ ; in case  $s=2$  and  $r=1$ , we have

$$\begin{aligned} & \sum_{i=0}^2 \binom{2}{i} (-1)^i (a+i+1) \\ &= a \left[ \sum_{i=0}^2 \binom{2}{i} (-1)^i \right] + \binom{2}{0} - 2\binom{2}{1} + 3\binom{2}{2} = a \cdot 0 + 1 + 4 - 3 = 0. \end{aligned}$$

Now assume by induction that the lemma has been proved for  $(r-1, s-1)$ . Then

$$\begin{aligned} & \sum_{i=0}^s \binom{s}{i} (-1)^i (a+i+1)\dots(a+i+r) \\ &= a \left[ \binom{s}{0} (a+2)\dots(a+r) - \binom{s}{1} (a+3)\dots(a+r+1) + \dots \right] \\ & \quad + \binom{s}{0} (a+2)\dots(a+r) - 2\binom{s}{1} (a+3)\dots(a+r+1) \\ & \quad + 3\binom{s}{2} (a+4)\dots(a+r+2) - \dots \end{aligned} \tag{A.14}$$

Since each term in (A.14) has a product of  $r-1$  elements and

$$\binom{s}{i} = \binom{s-1}{i-1} + \binom{s-1}{i},$$

the induction hypothesis implies that the sum in the brackets is zero and

the other sum is equal to

$$\begin{aligned}
& -\binom{s}{1}(a+3)\dots(a+r+1)+2\binom{s}{2}(a+4)\dots(a+r+2) \\
& -3\binom{s}{3}(a+5)\dots(a+r+3)+\dots \\
& =-s\left[\binom{s-1}{0}(a+3)\dots(a+r+1)-\binom{s-1}{1}(a+4)\dots(a+r+2)\right. \\
& \left.+\binom{s-1}{2}(a+5)\dots(a+r+3)-\dots\right] \\
& =0
\end{aligned}$$

by the induction hypothesis, and the proof is complete.

Another lemma from the same general family is the following.

LEMMA A.5 *Let  $s \in \mathbb{N}$ ,  $a \in \mathbb{R}$ ,  $k \in \mathbb{R}$ . Then*

$$\begin{aligned}
& \binom{s}{0}(a+s-1)\dots(a+1)a-\binom{s}{1}(a+s-2)\dots(a+1)a(a-k) \\
& +\binom{s}{2}(a+s-3)\dots(a+1)a(a-k)(a-k-1)-\dots \\
& +(-1)^{s-1}\binom{s}{s-1}a(a-k)\dots(a-k-s+2) \\
& +(-1)^s\binom{s}{s}(a-k)(a-k-1)\dots(a-k-s+1) \\
& =k(k+1)\dots(k+s-1)
\end{aligned}$$

*Proof* Using the fact that  $\binom{s}{i}=\binom{s-1}{i}+\binom{s-1}{i-1}$  and noticing that  $(a-k)$  is a factor of all terms except the first one and that  $a$  is a factor of all terms

except the last one, we rewrite the sum above as

$$\begin{aligned}
& a \left[ \binom{s-1}{0} (a+s-1) \dots (a+1) - \binom{s-1}{1} (a+s-2) \dots (a+1)(a-k) \right. \\
& \quad + \binom{s}{2} (a+s-3) \dots (a+1)(a-k)(a-k-1) \\
& \quad - \dots + (-1)^{s-1} \binom{s-1}{s-1} (a-k) \dots (a-k-s+2) \left. \right] \\
& - (a-k) \left[ \binom{s-1}{0} (a+s-2) \dots (a+1)a - \binom{s-1}{1} (a+s-3) \right. \\
& \quad \dots (a+1)a(a-k-1) + \dots \\
& \quad + \binom{s-1}{s-2} (-1)^{s-2} a(a-k-1) \dots (a-k-s+2) \\
& \quad \left. + (-1)^{s-1} \binom{s-1}{s-1} (a-k-1) \dots (a-k-s+1) \right] \tag{A.15}
\end{aligned}$$

The lemma obviously holds for  $s=1$ , since  $a-(a-k)=k$ . Assuming the lemma is true for  $s-1$ , we can by induction write the terms in (A.15) as

$$a(k+1) \dots (k+s-1)$$

( $s \rightarrow s-1$ ,  $a \rightarrow a+1$ ,  $k \rightarrow k+1$  with respect to the lemma as stated), and

$$(k-a)(k+1) \dots (k+s-1)$$

( $s \rightarrow s-1$ ,  $a \rightarrow a$ ,  $k \rightarrow k+1$  with respect to the lemma as stated). Summing these gives the desired result.

### A.3.3 Idea of the proof and more calculations

Because  $L/L_2$  is finite dimensional, there is some nonzero linear combination  $\sum a_m x^m$  of valuation  $\geq 2$ . Then

$$\left[ x \frac{\partial}{\partial x}, \sum a_m x^m \right] = \sum m a_m x^m$$

has valuation  $\geq 1$ . The idea is to produce enough elements of the form  $\sum m^i a_m x^m$  of valuation  $\geq 0$  to be able to conclude (via Vandermonde matrices) that the individual components  $a_m x^m$  have valuation  $\geq 0$ , and thus that the hypothesis of Lemma A.3 is satisfied. For example,

$$\left[ x^n \frac{\partial^n}{\partial x^n}, \sum a_m x^m \right] = \sum m(m-1) \dots (m-n+1) a_m x^m + \sum_{k=1}^{n-1} b_k x^{m+k} \frac{\partial^k}{\partial x^k}, \quad (\text{A.16})$$

and brackets of the form

$$\left[ x^{n+i} \frac{\partial^n}{\partial x^n}, \left[ x^{r-i} \frac{\partial^r}{\partial x^r}, \sum a_m x^m \right] \right]$$

produce similar terms. However, considerable effort is necessary (by another application of Vandermonde matrices) to eliminate unwanted terms (e.g., the final sum in (A.16)).

First, we perform some necessary calculations. For  $m \geq r+n$ , we shall need the sums

$$\sum_{i=0}^r (-1)^i \binom{r}{i} \left[ x^{n+i} \frac{\partial^n}{\partial x^n}, \left[ x^{r-i} \frac{\partial^r}{\partial x^r}, x^m \right] \right]. \quad (\text{A.17})$$

Now

$$\left[ x^{r-i} \frac{\partial^r}{\partial x^r}, x^m \right] = \sum_{j=0}^{r-1} \binom{r}{j} \frac{m!}{(m-r+j)!} x^{m-i+j} \frac{\partial^j}{\partial x^j},$$

so (A.17) becomes

$$\sum_{j=0}^{r-1} \left\{ \sum_{i=0}^r (-1)^i \binom{r}{i} \left[ x^{n+i} \frac{\partial^n}{\partial x^n}, \binom{r}{j} \frac{m!}{(m-r+j)!} x^{m-i+j} \frac{\partial^j}{\partial x^j} \right] \right\}. \quad (\text{A.18})$$

The terms of the inner sum in (A.18) which are obtained by the action of

$$\frac{\partial^s}{\partial x^s}, \quad 1 \leq s \leq j, \quad \text{on} \quad x^{n+i} \frac{\partial^n}{\partial x^n}$$

are of the form

$$-\binom{r}{j} \binom{j}{s} \frac{m!}{(m-r+j)!} x^{m+n+j-s} \frac{\partial^{n+j-s}}{\partial x^{n+j-s}} \left[ \sum_{i=0}^r (-1)^i \binom{r}{i} \frac{(n+i)!}{(n+i-j)!} \right];$$

this sum is zero by Lemma A.4, since  $s \leq j < r$ . The terms of the inner sum in (A.18) which are obtained by the action of  $\partial^s/\partial x^s$ ,  $i \leq s \leq j$ , on  $x^{m-i+j}(\partial^j/\partial x^j)$  are of the form

$$\binom{r}{j} \binom{n}{s} \frac{m!}{(m-r+j)!} x^{m+n+j-s} \frac{\partial^{n+j-s}}{\partial x^{n+j-s}} \left[ \sum_{i=0}^r (-1)^i \binom{r}{i} \frac{(m-i+j)!}{(m-i+j-s)!} \right];$$

this sum is also zero by Lemma A.4, since  $s \leq j < r$ . It follows that the only nonzero terms in (A.18) arise from the action  $\partial^s/\partial x^s$ ,  $j+1 \leq s \leq n$ , on  $x^{m-i+j}(\partial^j/\partial x^j)$ , so that (A.18) (and thus (A.17)) has the form

$$\sum_{k=1}^n b_k x^{m+n-k} \frac{\partial^{n-k}}{\partial x^{n-k}} \quad (\text{A.19})$$

The coefficients  $b_k$  remain to be calculated.

Fix a  $k$ ,  $1 \leq k \leq n$ ; the term in (A.18) which contributes to the  $k$ th term in (A.19) is

$$\begin{aligned} & \sum_{j=0}^{r-1} \left[ \sum_{i=0}^r (-1)^i \binom{r}{i} \binom{r}{j} \frac{m!}{(m-r+j)!} \binom{n}{j+k} \frac{(m-i+j)!}{(m-i-k)!} x^{m+n-k} \frac{\partial^{n-k}}{\partial x^{n-k}} \right] \\ &= \sum_{j=0}^{r-1} \binom{r}{j} \binom{n}{j+k} \frac{m!}{(m-k)!} \left[ \sum_{i=0}^r (-1)^i \binom{r}{i} \frac{(m+j-i)!}{(m-r+j)!} \frac{(m-k)!}{(m-k-i)!} \right] x^{m+n-k} \frac{\partial^{n-k}}{\partial x^{n-k}}. \end{aligned} \quad (\text{A.20})$$

According to Lemma A.5, with  $a \rightarrow m+j-r+1$ ,  $s \rightarrow r$ ,  $k \rightarrow k+j-r+1$ , the inner sum is equal to

$$(k+j)(k+j-1)\dots(k+j-r+1)$$

Thus (A.20) becomes

$$\begin{aligned} & \frac{m!}{(m-k)!} \left[ \sum_{j=0}^{r-1} \binom{r}{j} \binom{n}{j+k} (k+j)\dots(k+j-r+1) \right] x^{m+n-k} \frac{\partial^{n-k}}{\partial x^{n-k}} \\ &= \frac{m!}{(m-k)!} \binom{n}{k} \left[ \sum_{j=0}^{r-1} \binom{r}{j} \frac{(n-k)!}{(n-k-j)!} k(k-1)\dots(k-r+j+1) \right] x^{m+n-k} \frac{\partial^{n-k}}{\partial x^{n-k}} \end{aligned} \quad (\text{A.21})$$

The coefficient of  $k^r$  (the highest power of  $k$ ) in the inner sum of (A.21) is equal to

$$\sum_{j=0}^{r-1} \binom{r}{j} (-1)^j = (-1)^{r+1};$$

we will assume that  $r$  is odd, since the proof is the same for  $r$  even. It follows that the inner sum in (A.21) is of the form

$$k^r + c_{r-1}^{(r)}(n)k^{r-1} + \dots + c_1^{(r)}(n)k,$$

where the  $c_j(n)$  are polynomial functions of  $n$  and  $r$ . Hence (A.17) can be written as

$$\begin{aligned} & \sum_{i=0}^r (-1)^i \binom{r}{i} \left[ x^{n+i} \frac{\partial^n}{\partial x^n}, \left[ x^{r-i} \frac{\partial^r}{\partial x^r}, x^m \right] \right] \\ &= \sum_{k=1}^n \binom{n}{k} [k^r + c_{r-1}^{(r)}(n)k^{r-1} + \dots + c_1^{(r)}(n)k] \frac{m!}{(m-k)!} x^{m+n-k} \frac{\partial^{n-k}}{\partial x^{n-k}}. \end{aligned} \quad (\text{A.22})$$

For  $r=1$ , (A.22) becomes

$$\sum_{k=1}^n \binom{n}{k} k \frac{m!}{(m-k)!} x^{m+n-k} \frac{\partial^{n-k}}{\partial x^{n-k}}. \quad (\text{A.23})$$

Subtracting  $c_1^{(2)}(n)$  times (A.23) from (A.22) for  $r=2$  yields

$$\sum_{k=1}^n \binom{n}{k} k^2 \frac{m!}{(m-k)!} x^{m+n-k} \frac{\partial^{n-k}}{\partial x^{n-k}}.$$

Continuing by induction, we see that there are coefficients  $b(t, r, n)$  such that, for each  $t \in \mathbb{N}$ ,

$$\begin{aligned} & \sum_{r=1}^t b(t, r, n) \sum_{i=0}^r (-1)^i \binom{r}{i} \left[ x^{n+i} \frac{\partial^n}{\partial x^n}, \left[ x^{r-i} \frac{\partial^r}{\partial x^r}, x^m \right] \right] \\ &= \sum_{k=1}^n k^t \left[ \binom{n}{k} \frac{m!}{(m-k)!} x^{m+n-k} \frac{\partial^{n-k}}{\partial x^{n-k}} \right]. \end{aligned} \quad (\text{A.24})$$

### 3.4 Proof of Lemma A.2

According to Lemma A.3, we need only show that there is a  $p \in \mathbb{N}$ ,  $p \geq 2$ , such that  $v(x^p) \geq 0$ . By assumption  $W_1/L_2$  is finite dimensional; let  $r = \dim(W_1/L_2) + 1$ . Then there is for each  $u \in \mathbb{N}$  a nonzero sum of the form

$$X = \sum_{m=u}^{u+r-1} a_m x^m \quad (\text{A.25})$$

with valuation  $\geq 2$ . Take  $u \geq 2r$ , so that the calculations of the previous section are valid for all  $m$  in (A.25). Multiplying (A.24) by  $a_m$  and summing from  $m=u$  to  $m=u+r-1$  yields the expressions

$$\sum_{k=1}^n k^t X(k, n); t=0, \dots, r-1, n=1, \dots, r \quad (\text{A.26})$$

where

$$X(k, n) = \sum_{m=u}^{u+r-1} a_m \binom{n}{k} \frac{m!}{(m-k)!} x^{m+n-k} \frac{\partial^{n-k}}{\partial x^{n-k}}.$$

The elements (A.26) have thus been obtained from (A.25) by applying at most two brackets and taking linear combinations; therefore,

$$v\left(\sum_{k=1}^n k^t X(k, n)\right) \geq 0.$$

Using the nonsingularity of Vandermonde matrices, we can write the  $X(k, n)$  as linear combinations of the elements (A.26); thus

$$v(X(k, n)) \geq 0; k=1, \dots, n, n=1, \dots, r.$$

Taking  $k=n$  we obtain in particular the elements

$$Y(n) = \sum_{m=u}^{u+r-1} a_m \frac{m!}{(m-n)!} x^m, \quad n=1, \dots, r, \quad (\text{A.27})$$

with valuation  $\geq 0$ . It is easily shown that the coefficient matrix in (A.27) is nonsingular, implying that  $v(a_m x^m) \geq 0$ ,  $m=u, \dots, u+r-1$ , thus there is at least one  $m$  such that  $v(x^m) \geq 0$  (because not all  $a_m$  are zero). This includes the proof of Lemma A.2, thus proving Theorem 2.2.

#### A.4 PROOF OF THEOREM 2.3

Suppose that  $\phi: W_1 \rightarrow V(M)$  is a nonzero homomorphism, where  $M$  is an  $n$ -dimensional  $C^\infty$  manifold. Then there is a point  $m \in M$  such that the image of  $\phi$  contains an element which gives a nonzero tangent vector at  $m$ . Let  $G$  be the Lie algebra of germs of  $C^\infty$  vector fields around  $m$ ; i.e., in local coordinates centered at  $m$ ,  $G = \{\sum f_i(x) \partial / \partial x_i\}$ , where  $f_i$  are germs of  $C^\infty$  functions around  $m$ . Let  $A$  be the ideal in  $G$  consisting of all elements for which the  $f_i$  are flat functions in a neighborhood of  $m$  (a function germ in  $n$  variables  $x_1, \dots, x_n$  defined on a neighborhood  $N$  is flat on  $N$  if  $\partial^\alpha f / \partial x^\alpha(x) = 0$  for all  $x \in N$  and  $(\alpha)$ ).  $A$  is an ideal because derivatives of flat functions are flat. Restricting the vector fields of  $V(M)$  to their germs around  $m$ , we obtain a composed homomorphism of Lie algebras

$$W_1 \rightarrow V(M) \rightarrow G \rightarrow G/A \quad (\text{A.18})$$

which is nonzero because at least one vector field in  $\phi(W_1)$  was nonzero at  $m$ .

By Borel's extension lemma [29, p. 98],  $G/A$  is isomorphic to  $\hat{V}_n$ . Thus (A.18) gives a nonzero homomorphism from  $W_1$  to  $\hat{V}_n$ . However, since the only ideals of  $W_1$  are  $\{0\}$ ,  $W_1$ , and  $\mathbb{R} \cdot 1$ , this would yield a nonzero homomorphism from  $W_1$  or  $\mathbf{W}_1/\mathbb{R} \cdot 1$  to  $\hat{V}_n$ . This yields a contradiction by Theorem 2.2.

## Appendix B

### CALCULATIONS FOR EXAMPLE 2.3

The Lie algebra  $L(\Sigma)$  is generated by

$$e_0 = \frac{1}{2} \frac{\partial^2}{\partial x^2} + \frac{1}{2} x^2 \frac{\partial^2}{\partial y^2} - x \frac{\partial}{\partial y} - \frac{1}{2} y^2, \quad e_1 = y.$$

We proceed as in Example 2.1

$$[e_0, e_1] = -x + x^2 \frac{\partial}{\partial y} \triangleq e_2$$

$$[e_2, e_1] = x^2 \triangleq e_3$$

$$[e_0, e_3] = 2x \frac{\partial}{\partial x} + 1 \triangleq e_4$$



$[e_4, e_2] = -2x + 4x^2(\partial/\partial y)$ , which combined with  $e_2$  implies that  $e_5 = x$  and  $e_6 = x^2(\partial/\partial y)$  are in  $L(\Sigma)$ . Also,

$$[e_0, e_5] = \frac{\partial}{\partial x} \triangleq e_7$$

$[e_7, e_5] = 1 \triangleq e_8$ , which combined with  $e_4$  implies that  $e_9 = x(\partial/\partial x) \in L(\Sigma)$ .  
Now

$$[e_7, e_6] = 2x \frac{\partial}{\partial y} \Rightarrow e_{10} = x \frac{\partial}{\partial y}$$

$$[e_7, e_{10}] = \frac{\partial}{\partial y} \triangleq e_{11}$$

$$[e_7, e_0] = -\frac{\partial}{\partial y} + x \frac{\partial^2}{\partial y^2} \Rightarrow e_{12} = x \frac{\partial^2}{\partial y^2}$$

$$[e_7, e_{12}] = \frac{\partial^2}{\partial y^2} \triangleq e_{13}$$

$$[e_6, e_0] = -x^2 y - 2x \frac{\partial^2}{\partial x \partial y} - \frac{\partial}{\partial y} \Rightarrow e_{14} = 2x \frac{\partial^2}{\partial x \partial y} + x^2 y$$

$$[e_7, e_{14}] = \frac{\partial^2}{\partial x \partial y} + 2xy \triangleq e_{15}$$

$$[e_{10}, e_{15} + e_{11}] = x^3 - 2x \frac{\partial^2}{\partial y^2} \Rightarrow e_{16} = x^3$$

$$[e_{12}, e_{15}] = 4x^2 \frac{\partial}{\partial y} - \frac{\partial^3}{\partial y^3} \Rightarrow e_{17} = \frac{\partial^3}{\partial y^3}$$

$$[e_{13}, e_0] = -2y \frac{\partial}{\partial y} - 1 \Rightarrow e_{18} = y \frac{\partial}{\partial y}$$

$[e_{18}, e_{14}] = yx^2 - 2x(\partial^2/\partial x \partial y)$ , which combined with  $e_{14}$  implies that  $e_{19} = yx^2$  and  $e_{20} = x(\partial^2/\partial x \partial y)$  are in  $L(\Sigma)$ . Also,

$[e_{17}, e_{19}] = 3x^2(\partial^2/\partial y^2) \triangleq e_{21}$ , which combined with  $e_0$  and  $e_{10}$  implies that  $e_{22} = (\partial^2/\partial x^2) - y^2 \in L(\Sigma)$ . Continuing,

$[e_9, e_{22}] = -2(\partial^2/\partial x^2)$ , so that  $e_{23} = (\partial^2/\partial x^2)$  and  $e_{24} = y^2 \in L(\Sigma)$ . Now,

$$[e_{23}, e_{16}] = 3x^2 \frac{\partial}{\partial x} \Rightarrow e_{25} = x^2 \frac{\partial}{\partial x}$$

$$[e_{10}, e_{24}] = 2xy \Rightarrow e_{26} = xy$$

$$[e_{23}, e_{26}] = 2y \frac{\partial}{\partial x} \Rightarrow e_{27} = y \frac{\partial}{\partial x}$$

$$[e_{27}, e_{16}] = 3yx^2 \Rightarrow e_{28} = yx^2$$

$$[e_{27}, e_{28}] = 2y^2x \Rightarrow e_{29} = y^2x$$

$$[e_{27}, e_{29}] = y^3 \Rightarrow e_{30} = y^3$$

$$[e_{13}, e_{30}] = 3y^2 \frac{\partial}{\partial y} \Rightarrow e_{31} = y^2 \frac{\partial}{\partial y}$$

Noticing that the elements  $e_1, e_5, e_{13}, e_{23}, e_{25}, e_{26}$ , and  $e_{31}$  are precisely the generators of  $W_2$  given in Theorem 2.4, we conclude that  $L(\Sigma) = W_2$ .

#### Notes added in proof

- 1) See [6] for the simplest case of a one dimensional system.
- 2) A far better proof of lemma A4 is obtained by writing out  $n^{a+r}(1-n)^s$  in powers of  $n$ , calculating  $(d^r/dn^r) n^{a+r}(1-n)^s$  and then substituting  $n=1$ .

#### Acknowledgement

S. I. Marcus was supported in part by the National Science Foundation under grant ENG-76-11106 and in part by the Joint Services Electronics Program under contract F49620-77-C-0101. A portion of this research was conducted while this author was visiting the Econometric Institute, Erasmus University, Rotterdam, The Netherlands.